COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF DEVELOPMENTAL STAGES IN ISOLATED MAMMARY EPITHELIAL CELLS

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

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

by

Nicole Kristen Einfalt

December 2018

COMMITTEE MEMBERSHIP

TITLE:	Comparative Transcriptomic Analysis of Developmental Stages in Isolated Mammary Epithelial Cells
AUTHOR:	Nicole Kristen Einfalt
DATE SUBMITTED:	December 3, 2018
COMMITTEE CHAIR:	Dr. Daniel G. Peterson, PhD Professor, Molecular Physiology and Genomics Specialist
COMMITTEE MEMBER:	Dr. Mathew A. Burd, MS, DVM Professor, Veterinarian
COMMITTEE MEMBER	Dr. Juan F. Medrano, PhD Professor, Genomics

ABSTRACT

Comparative Transcriptomic Analysis of Developmental Stages in Isolated Mammary Epithelial Cells

Nicole Kristen Einfalt

The mammary gland is an organ common to all mammals; it is of value for neonatal nourishment, human nutrition through dairy consumption, and is a source of pathology in humans through the development of breast cancer. While transcriptomic analyses have been applied to cultured mammary epithelial cells (MEC) and to whole gland samples, few have studied purified MEC isolated directly from the gland *in vivo*. To identify the differentially expressed genes influencing MEC development during pregnancy and the differences between the nulliparous and primiparous quiescent states, primary MEC were isolated from virgin, pregnant, and primiparous quiescent sibling mice. Computational analysis was attempted using two differing platforms for the analysis of RNA sequencing data, the commercially-available CLC Genomics Workbench and the recently-launched, publicly-available Green Line Analysis. In the virgin-to-pregnant and virgin-to-post-lactational quiescent developmental comparisons, 31.02% and 26.97% of differentially detected genes, respectively, were dually detected by both platforms (p-value<0.05), with the remaining genes being detected in one platform but not the other. Expression was likewise compared for the dually differentially expressed genes detected with high (>500 RPKM), medium (10-500 RPKM), and low (0.02-9.99 RPKM) expression between the two developmental comparisons. In the virgin-to-pregnant and virgin-to-post-lactational quiescent developmental comparisons, 30.00% and 1.04% of differentially detected genes with high expression, respectively, were dually detected by both platforms (p-value<0.05); 30.51% and 7.60% of differentially detected genes with medium expression, respectively, were dually detected by both platforms (p-value<0.05); and 26.68% and 11.33% of differentially detected genes with high expression, respectively, were dually detected by both platforms (p-value<0.05). Although a small portion of differentially detected genes were dually detected between the two platforms, functional analysis for biological meaning revealed similar depictions of the underlying biological themes. The developmental comparison between the virgin and pregnant states suggests through enhanced mitochondrial processes, amino acid availability, cellular communication, and immune responses the lactational capacity is being established during the first half of pregnancy, when MEC are devoted to growth and proliferation and formation of the alveolus is not yet occurring. The developmental comparison between the virgin and primiparous quiescent states indicates an overall decrease in oncogenic pathways yet increase in ribosomal integrity may be associated with the parity-induced protection against breast cancer. Last, parallel analysis of the transcriptome and proteome from the same sample source allowed for the comparison of two differing means of analyzing the molecular phenotype and showed regulation of mRNA abundance may not necessarily reflect the expression pattern of the corresponding protein. A mathematical phenomenon was noted in the percent of dually detected transcripts relative to proteins, suggesting perhaps twenty percent of MEC genes are actively expressed at a given time.

ACKNOWLEDGMENTS

I would like to thank Dr. Dan Peterson not only for his guidance and scholarly influence, but also for all of the libations and life lessons that were had during my graduate career; Dr. Matthew Burd for his encouraging support, punctilious proofing, and amazing hugs; Drs. Medrano and Canovas for their analytical advice and input; both Mohammed Khalfan and Jason Williams for their patience and accommodations with the Green Line Analysis; my lab mates, Alyssa Conly and Joey Mancino, for making our time in the lab memorable; and the collective Cal Poly Animal Science faculty and staff for their assurance and encouragement.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1 – Literature Review

1.1 Introduction

Common to all female mammals is the milk-producing mammary gland that functions to nourish the neonate. Development of the mammary gland occurs in distinct ductal and secretory phases in response to endocrine signals; yet intriguingly, the majority of this development occurs postnatally. Upon the consummation of puberty and pregnancy, the mammary gland is fully differentiated and distinguished structurally by functional alveoli capable of synthesizing and secreting milk (Robinson et al., 1999). Lactation persists until weaning, at which time the gland involutes and is remodeled to the pre-pregnant state (Watson, 2006). Following involution, mammary epithelial cells (MEC) enter a state of reversible cell cycle arrest, remaining functionally quiescent until proliferative hormonal cues promote differentiation to reestablish the lactogenic alveoli in preparation for subsequent pregnancies (Harmes and DiRenzo, 2009).

While the hormonal regulation of mammary gland morphology is well understood, much remains to be learned about the molecular mechanisms governing development and differentiation. A combination of targeted disruption (gene knock-out) and tissue transplantation experiments has been utilized to explain the role of these molecular signals; however, further investigations to better understand these mechanisms may guide improvements not only in milk production but also in the prevention and treatment of breast cancer. This review describes the hormonal regulation of the developmental events and stages specific to mammary gland morphology and introduces the bioinformatics approach that was utilized for a whole system analysis. The following thesis addresses the comparisons between key developmental stages of isolated MEC to provide potential insights in lactation and cancer.

1.2 Mammary Gland Development and Functional Stages

Development of the mammary gland occurs in distinct phases in response to endocrine signals. Although anlage is established during fetal development, the majority of mammary gland morphogenesis is postnatal. Ductal elongation and branching are observed only after the onset of puberty, while functional differentiation is obtained only after pregnancy and parturition (Hennighausen and Robinson, 1998). To facilitate a more complete comprehension of the mechanisms cardinal to lactation and tumorigeneis, this section describes the anatomy and physiology of the mammary gland throughout a female's sexually reproductive lifespan.

1.2.1 Fetal Development

In all prenatal mammals the mammary gland originates from a localized thickening of the abdominal ectoderm, the outermost layer of embryonic epithelial cells (Hovey et al., 2002). As the epithelial cells constituting the thickening ectoderm proliferate, becoming columnar and multilayered, they thereby establish a protuberance that extends above and below the plane of the ectoderm (Hens and Wysolmerski, 2005). These protuberances form the mammary placodes, or designated pre-organ regions that eventually give rise to the mammary gland. In the mouse, five pairs of placodes form between embryonic days ten and eleven (Hens and Wysolmerski, 2005), while in humans, a single pair of placodes form between the seventh and eighth week of gestation

(Russo and Russo, 1987). A similar development takes place in ruminants, where the four placodes that develop into the four glands of the bovine udder form around embryonic day thirty-seven (Akers et al., 2000; Hovey et al., 2002). Moreover, it is well established that initial embryonic mammary gland development is dependent on intrinsic as opposed to systemic factors, as mammary placodes explanted *in vitro* follow the same proliferation and differentiation as their *in vivo* counterparts, even when cultured without supplemental hormones or growth factors (Levine and Stockdale, 1985; Robinson et al., 1999).

Little cell proliferation is observed by the embryonic MEC following placode formation; however, the migration and accretion of cells from the adjoining epidermis results in the formation of the rudimentary mammary epithelial bud. A study originally performed by Propper and Gomot (1973) demonstrated that embryonic mammary epithelium can induce non-mammary epithelium to form mammary buds when cultured *in vitro*. This study was later confirmed *in vivo* (Cunha et al., 1995) and complimented by tissue recombinant studies demonstrating how embryonic salivary epithelium, when combined with mammary epithelium, develops a ductal branching network resembling the salivary gland, yet when subsequently grafted into lactating hosts synthesizes milk proteins (Sakakura et al., 1976).

The morphology of the mammary placodes and buds is identical in males and females; however, the ensuing sexual dimorphism in mice introduces the importance of hormonal exposure during fetal development. Beginning on embryonic day twelve, transcription of the androgen, estrogen (E), and parathyroid hormone-related protein (PTHrP) receptors increases (Heuberger et al, 1982). By this stage of fetal development the mammary bud is surrounded by a primary layer of mesenchyme, the embryonic stroma with multipotent abilities to differentiate into supportive and connective tissues (Mele et al., 2013). Androgens produced by the fetal testes act on mesenchymal receptors to evoke condensation of the embryonic stroma around the male mammary bud, resulting in destruction of this rudimentary structure by embryonic day sixteen (Pamar and Cunha, 2004). In contrast, testes-lacking female mice to do not experience androgen-induced destruction of the mammary bud. Studies on the effects of exogenously-applied testosterone found that embryonic female mammary buds do regress when exposed to the androgen from embryonic days thirteen to sixteen, however testosterone exposure after embryonic day sixteen failed to result in the destruction of the female mammary bud. Thus, there is a window of embryonic endocrine sensitivity, after which androgens cannot evoke destruction of the female mammary gland (Kratochwil, 1977). Curiously, removal of endogenously-produced estrogens by X-irradiation of the ovaries at embryonic day thirteen does not result in mammary gland developmental effects until puberty, suggesting that the female morphology is the default state (Hovey et al., 2002; Raynaud, 1950). Sexual dimorphism of the mammary gland is not apparent in ruminants and humans until just prior to puberty and with the onset of puberty, respectively (Akers et al., 2000; Russo and Russo, 1987)

Upon the formation of the mammary bud, epithelial-stromal interactions play fundamental roles in mammary gland morphogenesis. This is not only apparent in the growth and preliminary branching observed during the remainder of gestation but also in the continued growth and development following parturition. In the final stages of fetal development, the MEC of the mammary bud proliferate, promoting not only the sprouting of the bud out through the dense mesenchyme and into the lower dermis in preparation for ductal branching, but also the concurrent differentiation of the overlying epidermal cells into the tissue of the nipple (Hens and Wysolmerski, 2005). As the mammary bud sprouts, its epithelial cells express PTHrP, the receptor for which is located in the mesenchyme. In both PTHrP and PTHrP-receptor knock-out mice, ductal outgrowth of the mammary bud, sexual dimorphisms, and nipple formation fail, thereby indicating PTHrP is necessary for the mesenchyme to support prenatal mammary development and to differentiate accordingly (Wysolmerski et al., 1998).

Prior to parturition, the mammary gland is characterized by primitive ducts that form a rudimentary branching network. The distal portion of these ducts resemble a bilayered structure and are composed of an outer layer of undifferentiated cap cells surrounding an inner layer of luminal epithelial cells (Hinck and Silberstein, 2005). Estrogen and growth hormone (GH) promote cell division and embryonic expansion of the primitive ducts through mesenchymal-derived growth factors, as opposed to exerting their effects directly on the ducts themselves (Silberstein, 2001). Separate knock-out (Cunha et al., 1997) and hypophysectomized (Wadden et al., 1998) studies in mice have shown that mesenchymal intermediaries, namely E-mediated epidermal growth factor (EGF) and GH-mediated insulin-like growth factor-1 (IGF-1), are necessary for ductal elongation. In contrast, progesterone (P) and prolactin (PRL) exert their effects on the mammary epithelium. Specific to the prenatal development of the mammary gland, PRL is essential for establishing the rudimentary branching network from the primitive ducts (Lamote et al., 2004). Knock-out experiments involving PRL and its receptor, PRLR, have demonstrated that mice deficient in the PRL gene possess a basic, immature branching network compared to their extensively-branched control group. Thus, PRL is inferred to be an obligate regulator of mammary gland development (Horseman, 1999).

At parturition, minor differences exist across species in the architectural arrangement of the mammary gland. In humans, a small group of lobules emerge from a terminal duct, giving them the appearance of a grape cluster found at the end of a stem. In mice, lobules appear more club-like; however, the stroma into which they invaginate is considerably higher adipose with only minor amounts of fibrous connective tissue (Parmar and Cunha, 2004). In the newborn calf, although the teats are discernable, the mammary fat pad and epithelium are barely palpable, existing in a negligible amount extending dorsally above each teat (Rowson et al., 2012).

1.2.2 Postnatal Development

From parturition to puberty, growth of the mammary gland is minimal and isometric, remaining proportional to that of the entire body (Lamote et al., 2004). Isometric growth persists until approximately age eight-to-twelve years in humans, four weeks in mice, and three months in cows, at which time the onset of puberty commences the allometric growth and functional differentiation that establishes the specialized lobulo-alveolar system prior to pregnancy (Hovey et al., 2002).

1.2.3 Pubertal Development

During isometric growth, an inhibition is maintained by the central nervous system, keeping the gonadotropin-releasing hormone (GnRH) neurons of the hypothalamus functionally quiescent. Upon removal of this inhibition, GnRH stimulates the onset of puberty through the production of reproductive hormones via the hypothalamic-pituitary-gonadal axis (Porterfield and White, 2007). Together, reproductive and metabolic hormones direct the cell proliferation, differentiation, and apoptosis that promote the further expansion and branching of the mammary ductal network into the surrounding stroma. Accordingly, while prenatal development of the mammary gland was influenced by interactions with the mammary mesenchyme, the development associated with puberty and pregnancy is dependent upon the stroma and its constituents (Hovey et al., 1999).

The club-like bulbs of the mouse mammary gland are collectively known as the terminal end buds (TEB) and represent the active site of ductal expansion and branching. Although the corresponding active sites of the human and ruminant mammary gland differ histologically and lack a bulb-like appearance, they can be considered a TEB-like structure as they undergo comparable proliferation and differentiation (Hovey et al., 2002). Similar to the primitive ductal establishment of the prenatal mammary gland, the TEB of the pubertal gland are a multilayered structure, composed of an inner layer of luminal epithelial cells surrounded by an outer layer of cap cells at the tip and myoepithelial cells along the neck and length of the duct. The concurrent apoptosis of the luminal epithelial cells with mitotically-proliferating cap cells extend the ducts through the stroma, which in turn is encompassed by the adipose cells of the mammary fat pad (MFP) (Klinowska et al., 1999). Lateral buds develop along the mature ducts, resulting in the open architecture of the post-pubertal branching network and ultimately decreasing the ratio of stroma to parenchyma (Hovey et al., 1999).

Estrogen is accepted as the primary ovarian hormone directing ductal expansion and branching during puberty. The E receptor (ER) exists in two subtypes, ER-α and ER-β. While ER-β knock-out mice display no developmental defects, ER-α knock-out mice are infertile and fail to develop as expected during the pubertal stage of mammary gland growth and differentiation (Couse and Korach, 1999). However, if the ER-α is knocked-out after puberty, alveolar differentiation is still able to occur, indicating the ER-α is necessary for and specific to pubertal mammary gland development (Briskin and Rajaram, 2006). Although the ER-α is solely restricted to the epithelial compartment in humans and ruminants, in mice it is located in both the stroma and the epithelium, the latter of which functions in a paracrine manner through the influential effects of growth factors (Lamote et al., 2004). Amphiregulin, another example of a stromally-derived, E-mediated growth factor, acts upon the epithelial cells of the TEB to direct ductal development. Accordingly, amphiregulin mRNA is most highly expressed during the pubertal growth phase, and its inactivation results in the failure to undergo ductal outgrowth and branching (Howlin et al., 2006).

Similarly, GH, mediating its effects through IGF-1, is also necessary for the TEB formation and ductal proliferation of pubertal development. Impaired pubertal ductal outgrowth is observed in both IGF-1 knock-out (IGF-1KO) and GH receptor knock-out (GHRKO) mouse models; however, the pubertal phenotype can be rescued in IGF-1KO mice upon the dual administration of IGF-1 and E. Growth hormone and IGF-1, therefore, function synergistically with E, as administration of IGF-1 alone is unable to promote ductal outgrowth (Howlin et al., 2006). This synergism has also been demonstrated in ruminants, where administration of both GH and E stimulates mammary gland development in peripubertal heifers; however, such a response is not observed upon the administration of GH alone (Hovey et al., 2002).

Although required for the alveolar development associated with pregnancy, the precise influence of P and its epithelial receptor, PR, specific to pubertal development of the mammary gland *in vivo* has yet to be determined (Howlin et al., 2006). Ductal branching is impaired in PR knock-out (PRKO); mice, however, the post-pubertal phenotype is not abolished, indicating P is not essential for ductal elongation by or in itself (Hovey et al., 2002).

Interestingly, the biologically active form of vitamin D3, 1,25-(OH)2D3, is a systemically-circulating hormone recently proposed to have a suppressive effect on mammary ductal development, potentially through the antagonism of proliferative signals (Howlin et al., 2006). The vitamin D3 receptor (VDR) is a nuclear steroid hormone receptor and is expressed in the epithelial cells of the TEB. In mice, mammary glands from VDR knock-outs are significantly heavier compared to wild-type glands and are characterized by a greater number and degree of TEB, branching, and ductal expansion. These findings indicate that the vitamin D3 signaling pathway participates in negative growth regulation of the pubertal mammary gland (Zinser et al., 2002).

The pubertal morphology of the mammary gland is not limited to influence by those hormones and locally-produced growth factors mentioned thus far. Mitogens such as epidermal growth factor (EGF), transforming growth factor (TGF), and hepatocyte growth factor (HGF) direct the mitotic cell division observable in ductal outgrowth. Stromally-derived matrix metalloproteases (MMP) are enzymes responsible for remodeling the extracellular matrix (ECM), and have been suggested as necessary in the invasion of the MFP by the ductal epithelium (Fata et al., 2004). Still other influential regulators may include transcription factors, cell cycle regulators, cytokines, and migrant cell types such as macrophages and eosinophils (Howlin et al., 2006; Gouon-Evans, et al., 2000). However, the focus of this review is limited to the hormonal regulation of mammary development. Consequently, while E and GH are recognized as the key regulators of pubertal development, it is crucial to understand proper morphology and function of the developing mammary gland are dependent upon the extensive interplay of numerous regulators.

1.2.4 The Virgin Adult

Upon the completion of puberty, ductal elongation and branching morphogenesis have established a mammary network extending throughout the stroma in the virgin adult, and the motile TEBs are no longer discernible. Alveolar buds, which develop into the functional terminal ductal lobular units in humans and ruminants and lobuloalveolar units in mice, remain rudimentary until dictated to differentiate further by the endocrine signals associated with pregnancy (Parmar and Cunha, 2004).

The mammary gland of the virgin adult is comprised of a heterogeneous mixture of cell types (Shackleton et al., 2006). The luminal and myoepithelial cells form a basal parenchymal layer that is separated from the stroma by the basement membrane. The stroma, in turn, consists of fibroblasts, adipocytes, inflammatory cells, vascular and lymphatic components, and the ECM. Cross-species differences exist in the cellular composition of the mammary stroma, most notable of which is the adipocyte-rich mouse stroma compared to the fibrous connective tissue-rich human and ruminant stroma.

Although postulated to affect the lactational composition and capacity of each species, the exact physiological significance of these histological abundance differences remains unclear (Parmar and Cunha, 2004; Hovey et al., 1999).

While the mammary gland of the virgin adult is relatively quiescent compared to the morphological changes associated with the previous developmental phases, minute morphological changes do occur in response to the estrous cycle. This cyclical remodeling is the result of hormonally-regulated cell proliferation, differentiation, and apoptosis that collectively promote rudimentary lateral branching and alveolar budding (Chua et al., 2010). The estrous cycles of the mouse and bovine last four-to-five days and twenty-one days, respectively, and are divided into proestrus, estrus, metestrus, and diestrus. In humans, the cycle lasts between twenty-five-to-thirty days and consists of follicular and luteal phases (Hovey et al., 2002). The greatest extent of alveolar budding is observed in diestrus in the mouse and in the luteal phase in humans. The morphological changes of the estrous cycle in ruminants, however, remain unexplored. Within all species, the transitory estrous cycle-associated appearance of alveolar buds is thought to indicate a developmentally prepared-for-pregnancy mammary gland (Chua et al., 2010; Hovey et al., 2002).

1.2.5 Development during Pregnancy

The hormonal influence of mammary gland development during pregnancy is impressive. Pregnancy begins with the implantation of the blastocyst into the uterus following conception, and the ensuing hormonal cues direct extensive proliferation and secretory differentiation to produce functional alveolar units capable of milk secretion.

The tissue remodeling associated with alveolar morphogenesis is dependent not only upon an initial synergy between P and PRL, but also on the influence of E, placental lactogen (PL), and GH. Still other potentially influential hormones include but are not limited to thyroid hormones, corticosteroids, insulin, leptin, and PYHrP. However, further investigations are needed to distinguish the developmental function of these hormones during pregnancy from the function they serve during lactation (Brisken and Rajaram, 2006; Neville et al., 2002; Tucker 1981).

Specific to the development of the mammary gland during pregnancy, MEC of the ductal network reorganize into polarized cells, forming a spherical layer of MEC that face an open lumen connected to the ductal network and are surrounded by contractile myoepithelial cells. Myoepithelial cells, in turn, are in direct contact with the basement membrane, a specialized structure of ECM that underlies the mammary epithelium. The cells of the ductal epithelium contain sparse cytoplasmic organelles and, remaining non-secretory, primarily function as a channel for conveying milk upon the initiation of lactation (Barcellos-Hoff et al., 1989). The luminal alveolar cells accomplish the synthesis of milk while the myoepithelial cells, in response to oxytocin, contract to expel milk out of the alveolus and into the ducts to nourish the young (Richert et al., 2000). However, while differentiation of the MEC into alveolar structures has been heavily investigated, little is currently known regarding the function and significance of coordinated changes required for alveolar formation within the stroma (Brisken and Rajaram, 2006).

Together, P and PRL promote the initial MEC cell proliferation associated with early pregnancy and are required for the polarization of the luminal alveolar cells.

Characteristic to humans and ruminants, PRL levels remain elevated throughout pregnancy (Anderson et al., 2007; Tucker, 1981). Characteristic to the mouse, however, this synergy is expunged in the later stages of pregnancy. While P supports the continuation of gestation, increased signaling by PRL is required for the initiation of lactation and the expression of most milk protein genes (Neville et al., 2002).

Molecular modulators and their mechanisms within MEC are of pivotal importance for understanding the signaling pathways by which hormonal regulation is able to induce morphological changes. The PR, a steroid hormone nuclear receptor, is not found in all MEC, so the proliferation of these cells in response to P is partly mediated by paracrine factors (Oakes et al., 2006). Similar to that mentioned in previous developmental stages, transplantation studies have demonstrated how wild-type mammary epithelium can promote alveolar differentiation in adjacent PRKO epithelium (Brisken et al., 1998). Studies by Conneely and colleagues have identified two different isoforms of the MEC PR, PR-A and PR-B. Their analysis on mammary glands of PRA-KO and PR-BKO mice has shown that alveolar morphogenesis is drastically diminished in mice lacking the PR-B isoform, while ablation of the PR-A does not affect the ability of PR-B to elicit normal alveolar development (Conneely et al., 2003; Mulac-Jericevic et al., 2000).

The wingless-related NMTV integration site 4 (Wnt4) and receptor activator of nuclear factor (NF)-κB ligand (RankL) are two proposed mediators of the P signaling pathway. When P binds to PR-B, it is speculated to achieve its developmental effects through one of these downstream signaling mediators (Brisken et al., 2000; Oakes et al., 2006). Wnt4 is the only *Wnt* gene directly induced by P, and within murine MEC in the early stages of pregnancy, P has been shown to induce Wnt4 expression. Wnt4 is thus thought to mediate the maturation of the ductal side branching leading to alveolar morphogenesis, as transplantation of mammary epithelia from Wnt4-knock out (Wnt4KO) mice has demonstrated that Wnt4 is key to this process. However, ductal side-branching and alveolar morphogenesis do develop later in pregnancy in Wnt4KO mice, indicating other P-induced factors might play a role in mediating proliferation and alveolar morphogenesis (Brisken et al., 2000). The other candidate, RankL, is likewise speculated as a P-signaling mediator. When RankL binds to its receptor, Rank, the ensuing IκB kinase-α (IKK-α) signaling pathway activates the downstream NF-κB transcription factor, the phosphatidylinsoitol 3-kinase (PI3K) Akt pathway, and the CAAT/enhancer binding protein-β (C/EBP-β) signaling pathway (Fernandez-Valdivia and Lydon, 2012), all of which are known to be influential in alveolar morphogenesis during pregnancy in mice (Oakes et al., 2006). Remarkably, in wild-type mouse hosts transplanted with PRKO mammary epithelium, ductal side branching and alveolar budding is observed upon the administration of exogenous RankL, supporting the notion that this paracrine mediator is fundamental to the P-signaling pathway (Fernandez-Valdivia and Lydon, 2012). Whether RankL has the same functional role in humans remains to be determined. Considering the profound proliferation that is triggered by Wnt4 and RankL in P-promoted mammary gland development, these mediators and their inhibitors are of clinical interest in future applications specific to the prevention and treatment of breast cancer (Tanos et al., 2013).

Another essential molecular modulator associated with pregnancy is the signal transducer and activator of transcription protein 5a (STAT5A). In response to PRL binding to PRLR, the receptor dimerizes, inducing the phosphorylation and activation of Janus kinase 2 (JAK2). The activated kinase consequentially phosphorylates STAT5A, which in turn translocates to the nucleus of MEC and serves as a mandatory transcription factor in the expression of specific genes related to alveolar morphogenesis (Liu et al., 1996). Examples of those genes induced by STAT5A include claudins and connexins, which are necessary for cell-cell interactions, collagens and laminins, which are necessary for stromal-epithelial interactions, the suppressor of cytokine signaling 2 (Socs2) protein, which functions as a negative regulator of the PRL-signaling pathway, and the E74-like factor 5 (Elf5) transcription factor, which is necessary for the structural and functional development of mammary alveoli (Oakes et al., 2006). Socs2 and Elf5 function as the most influential molecular mediators of the PRL-induced development of the mammary gland, and Harris and colleagues have used PRLKO mice to demonstrate that alveoli are capable of milk production following either the additional genetic ablation of Socs2 or the retrovial re-expression of Elf5 (Harris et al., 2006). While Socs2 knock-out mice exhibit an overall loss of growth control and disproportionately large organs, further analysis of the physiological significance of Elf5 is unfortunately hindered by the early embryonic lethality of Elf5-knock out mice (Zhou et al., 2005).

The precise significance of E on mammary gland development after ductal morphogenesis remains to be fully investigated. While P is directly essential for alveolar morphology during pregnancy, E, interacting with ER-α, is presumed not only to be influential in ductal growth but also to indirectly stimulate alveolar development through its subsequent induction of PR and PRLR in the mammary epithelium (Neville et al., 2002).

Likewise, the role of GH in MEC differentiation is unclear. While GH appears to signal through its stromal receptor, GHR, it is not necessary for alveolar development. Although ductal outgrowth and branching are diminished in GHRKO mice, lactation is still able to occur following parturition (Kelly et al., 2002). Furthermore, and specific to humans, female dwarves lacking GH are capable of lactating, maintaining the supportive yet non-essential role of GH in alveolar development (Neville et al., 2002).

The syncytiotrophoblast cells of the placenta, a transient organ that develops only during gestation, are able to secrete several hormones that function either to maintain the pregnant state of the maternal uterus or to promote alveolar formation within the mammary gland (Porterfield and White, 2007). Placental lactogen is a polypeptide hormone secreted by the placenta of humans, rodents, and ruminants, and is structurally similar to GH and PRL. Accordingly, while there currently is no known specific PL receptor (Neville et al., 2002), studies by Herman and colleagues have shown that isolated ovine PL is able to bind to bovine GHR and PRLR (Herman et al., 2000). Whether similar binding occurs in humans and mice has yet to be determined (Neville et al., 2002). However, considering PL is far more abundant in maternal circulation compared to fetal circulation, its functional role remains pivitol to the former through the support of GH- and PRL-mediated alveolar formation (Porterfield and White, 2007).

Equally interesting to note are the equally dramatic changes in other tissue types in accordance with pregnancy. For example, to function metabolically under the increased energy requirements of pregnancy and lactation, the intestines and liver enlarge. To provide the mammary gland with the increased quantities of energy, sugars, and amino acids required for milk production, there is a parallel increase in the vasculature of the stroma (Oakes et al., 2006). Maternal behavior is stimulated by PRL at the end of pregnancy and maintained by oxytocin following parturition (Uvnas-Moberg and Eriksson, 1996).

Proper morphogenesis of the functional mammary gland is thus dependent upon the coordination of endocrine induction, signaling pathways, and their corresponding molecular mediators to direct the formation of alveolar units from the ductal epithelium. The following sections will reference these developmental pathways as many of the same processes necessary for proper morphogenesis are reflected in lactation and manifest in metastatic tumorogenesis (Colletta et al., 2004). Consequently, the regulation of MEC proliferation and differentiation is significant in the application to milk production within the dairy industry and in the prevention and treatment of breast cancer.

1.2.6 Lactation

Milk production, which is blocked by P during pregnancy, is stimulated by PRL and the increased transcription of milk protein genes around parturition (Anderson et al., 2007). While the physical process of milk ejection is similar among species, differences do exist in the coordination of increased PRL signaling and P withdrawal. Although PRL levels remain elevated throughout pregnancy in humans, P does not fall until the placenta is removed following parturition. In contrast, in mice and ruminants, PRL rapidly spikes as P decreases just prior to parturition. Thus, while full lactation is slightly delayed in humans, milk is readily available for the newborn pups and calves in mice and ruminants (Neville at al., 2002).

Various changes are observed within the mammary gland upon the transition from pregnancy to lactation. Histologically, at the onset of lactation, the rudimentary alveolar buds have fully developed into functional terminal ductal lobular units in humans and ruminants and lobuloalveolar units in mice. Additionally, the ductal network has branched extensively throughout the MFP, and tight junctions between the alveolar cells have closed (Anderson et al., 2007). The most prominent histological change is the increase in size and abundance of lipid droplets and casein micelles within the alveolar cells, and the movement of these particles into the alveolar luminal space (Neville et al., 2002). Typical of any cell specialized for secretion, the endoplasmic reticulum (ER) of the lactating MEC is extensive and in contact with numerous mitochondria (Boisgard et al., 2001). On a more systemic level, the onset of lactation is also accompanied by increases in blood volume and cardiac output. The resulting increase in blood flow to and from the mammary gland is correlated to milk yield and provides the mammary gland with the nutrients required for synthesis of the various milk components (Svennersten-Sjaunja and Olsson, 2005).

Milk ejection from the alveoli is stimulated by suckling, regulated by a neuroendocrine reflex, and required for the continuation of lactation (Anderson et al., 2007). Specifically, the suckling stimulus depolarizes the somatosensory afferent neurons at the tip of the nipple, triggering the magnocellular and parvicellular neurons of the hypothalamus to promote the respective release of oxytocin and PRL from the pituitary gland (Porterfield and White, 2007). While oxcytocin binds to its receptors on the myoepithelial cells of the ductal network, thereby causing them to contract and transport milk through the ducts to nourish the young, PRL binds to its receptors on the secretory MEC of the alveolus, thereby promoting their continued production of milk components (Neville el at., 2002; Uvnas-Moberg and Eriksson, 1996).

Although PRL is required for the continuation of lactation, other metabolic hormones such as insulin, glucocorticoids, thyroid hormones, and GH have also been proposed to be potentially influential in milk production and yield. Insulin levels during lactation are relatively low, and a decreased responsiveness of adipose and skeletal tissues to insulin serves to increase the availability of glucose for the mammary gland (Svennersten-Sjaunja and Olsson, 2005). Adrenal steroids, such as glucocorticoids, are known to maintain blood glucose levels during periods of starvation, and are thus pertinent in the negative energy balance that exists in ruminants and mice during early lactation. Humans rarely enter a negative energy balance during lactation, and thus glucocorticoid levels remain relatively lower (Neville et al., 2002). Thyroid hormones are known to increase membrane Na^+-K^+ adenosine triphosphate (ATPase) concentration and activity, consequently increasing the metabolic pathways and overall energy expenditure of a cell (Porterfield and White, 2007). Triiodothyronine (T_3) is a tyrosine-based thyroid hormone that is formed from the enzymatic 5'-deiodination of thyroxine (T_4) within the thyroid and peripheral tissues such as the mammary gland. Interestingly, despite the decreased amount of deiodination in the liver and kidneys during lactation, there is an increased amount in the mammary gland, and the resulting hypothyroidism of the peripheral tissues decreases their metabolism yet enhances that of the mammary gland (Neville et al., 2002; Tucker 1981). The precise role of GH during lactation remains debatable. Exogenously administered GH is known to mobilize energy reserves and has been shown to enhance the blood flow, uptake of nutrients, milk synthesis, and activity of secretory cells (Bauman, 1999). The resulting increase milk yield has thus led to its commercial application in the dairy industry. However, it is unclear whether the effects of GH are restricted to the luminal alveolar cells of the mammary gland or are more relevant to the overall nutrient availability of the lactating female (Svennersten-Sjauja and Olsson, 2005).

Fascinatingly, in addition to the various nutritional components synthesized and secreted into the milk, the mammary gland has recently been shown capable of synthesizing several hormones and growth factors including PRL, leptin, PTHrP, and GH (Neville et al., 2002). Hence, the lactating mammary gland is not only functioning as directed by specific systemic cues but is also itself a site of hormone production. Prolactin serves as the primary reproductive hormone governing lactation, while other metabolic hormones have been speculated to be indirectly influential. These metabolic hormones are not essential to the alveolar functioning of the mammary gland, but rather may affect the synthesis of the various milk components by altering the nutrient availability of the lactating gland (Neville at al., 2002). Milk secreted by the mammary gland consists of water, proteins, lipids, carbohydrates, vitamins, and minerals. Specific synthesis of the protein, lactose, and lipid components will be discussed in the following subsections.

1.2.6.1 Protein Synthesis

The synthesis of proteins is important not only for the generation of those secreted in the milk but also for the generation of those necessary and responsible for proper cell function and survival. Protein synthesis is an energetically expensive process, yet the increase in efficiency of milk protein synthesis remains a profitable aspiration for the dairy industry (Bionaz and Loor, 2011). When stimulated by a PRL-mediated lactogenic environment, MEC selectively regulate the production of milk proteins. Caseins and whey are the main proteins produced by the mammary gland and comprise 95.6% of the total proteins secreted in milk, with the remaining 4.4% of total secreted proteins originating in the blood as immunoglobulins and lactoferrins (Maas et al., 1997). Protein synthesis is accomplished through the coordinated steps of selective transcription of deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) by RNA polymerase II within the nucleus, exportation of the RNA through nuclear pore complexes, translation of the RNA to a sequence of amino acids within a ribosome, and finally post-translational modifications such as removal of the signal peptide, phosphorylation, and glycosylation of the protein just prior to secretion (Alberts et al., 2008). The availability of amino acids for the translational process within MEC is generally regarded as the limiting factor in the synthesis and secretion of milk proteins (Boisgard et al., 2001).

Significant to this process is the function of the ribosome, a catalytic complex that uses the genetic information carried by RNA molecules to guide the synthesis of proteins. Eukaryotic ribosomes are assembled from a small 40S subunit and a 60S large subunit each consisting of thirty-three and forty-nine unique proteins, respectively. While the small ribosomal subunit provides a framework on which transfer RNA (tRNA) molecules match the RNA nucleotide sequences to specific amino acids, the large ribosomal subunit, links the peptide bonds between individual amino acids within the sequence of the protein (Alberts et al., 2008). Interestingly, individual ribosomal proteins have recently been highlighted as having extra-ribosomal functions such as DNA repair, regulation of apoptosis, and autoregulation of ribosomal protein synthesis. Furthermore, ribosomopathies, disorders resulting from impaired ribosome biogenesis and function, have recently been shown to be oncogenic and consequently detrimental to cellular homeostais (Shenoy et al., 2012; Warner and McIntosh, 2009).

The newly synthesized proteins found in milk are aqueous solutes and secreted through an exocytotic pathway, meaning they are packaged into secretory vesicles within the Golgi apparatus and transported to the apical region of MEC. The membrane of the transport vesicles fuse with the plasma membrane, ultimately resulting in the discharge of the synthesized protein contents into the luminal alveolar space (McManaman and Neville, 2003). The regulation of this exocytotic secretory pathway remains to be explored. While PRL and the resulting JAK-STAT signaling pathway are generally regarded as essential regulators of protein expression in non-ruminant mammary glands, recent studies in mice and ruminants have highlighted a role of the mammalian target of rapamycin (mTOR) signaling pathway in milk protein synthesis (Bionaz and Loor, 2011).

1.2.6.2 Lactose Synthesis

Lactose is the main carbohydrate found in milk, serving as a vital source of energy for the newborn offspring. Both ruminant and nonruminant offspring are able to digest this disaccharide, breaking it down into glucose and galactose. Synthesis of lactose is unique to mammary alveolar cells and occurs within the lumen of the Golgi apparatus by the enzyme lactose synthase (Anderson et al., 2007; Shennan and Peaker, 2000). Lactose synthase is composed of two protein units, α-lactalbumin and galactotransferase. While P represses the expression of α-lactalbumin within the mammary gland throughout pregnancy, PRL induces the expression of both of these protein components. Lactose synthase catalyzes the formation of lactose and uridine diphosphate- (UDP-) galactose from glucose and UDP-galactose (Turkington and Hill, 1969). Beginning with suckling by the offspring, the dam experiences a gradual increase in milk and lactose production, and the volume of milk secreted is closely related to the rate of lactose synthesis (Shennan and Peaker, 2000; Uvnas-Moberg and Eriksson, 1996)

Alveolar cells of the lactating mammary gland are characteristically high in cytoplasmic glucose concentration, a phenomenon that results from the presence of the non-insulin dependent glucose transporter, GLUT1, on the basolateral membrane (Shennan and Peaker, 2000). The GLUT1 transporter is also located on the membrane of the Golgi apparatus, allowing for the uptake of glucose into the Golgi apparatus and its subsequent interaction with UDP-galactose and lactose synthase (Anderson et al., 2007). However, while the membranes of the Golgi apparatus and apical alveolar cell are freely permeable to water, neither are permeable to lactose. As a result, the newly synthesized lactose osmotically draws water into the Golgi apparatus, thereby significantly contributing to the overall milk volume yield (Shennan and Peaker, 2000). Similar to protein secretion, lactose is packaged into vesicles within the Golgi apparatus and secreted into the luminal alveolar space through an exocytotic pathway (McManaman and Neville, 2003).

Significant to note are the other fates of glucose utilization by the alveolar cells of the lactating mammary gland. While glucose is required for the synthesis of lactose, it can also be converted into glucose-6-phosphate (G-6-PO_4) for adenosine triphosphate (ATP) production within the mitochondria, for glycerol production in the synthesis of triacylglycerol (TAG), or for nicotinamide adenine dinucleotide phosphate (NADPH) production through the pentose phosphate pathway (Anderson et al., 2007).

1.2.6.3 Lipid Synthesis

In addition to lactose, lipids also serve as a vital source of energy for the newborn offspring. The amount of fat in milk can range from less than 1% to greater than 50%, depending not only on the species but also on the breed within that species (Shennan and Peaker, 2000). While the percentage of fat is typically around 4% in humans and ruminants and 20% in mice, the exact percentage is highly influenced by diet intake and stage of lactation (Neville and Picciano, 1997; Gors et al., 2009). TAG, formed from three fatty acid tails connected to a glycerol backbone through an ester linkage, is the major component of milk fat, typically accounting for 98% of the fat found in milk (Anderson et al., 2007). Fatty acids, in turn, can either be taken up from the circulating blood or synthesized through liopogenesis by the lactating MEC (Shennan and Peaker, 2000).

Fatty acids within the blood are considered an exogenous lipid source and are derived either from the diet or adipose tissue. Transport to the mammary gland is facilitated through the formation of chylomicrons or through the binding to transport proteins such as albumin or very low-density lipoprotein (VLDL) (Anderson et al., 2007; Shennan and Peaker, 2000). TAG itself cannot enter mammary tissue; upon reaching the mammary gland, lipoprotein lipase breaks TAG down into its constituents. Whether fatty acids cross the MEC plasma membrane via diffusion or a transport system is currently unknown (Neville and Piacciano, 1997). Glycerol and fatty acids are then taken up from the blood and into the alveolar cells where they are subsequently used for TAG synthesis (Anderson et al., 2007).

In addition to hepatic and adipose tissues, lipogenesis can also occur in mammary tissue. Depending on the species, these tissues utilize different precursor molecules to convert acetyl-coenzyme A (acetyl CoA) into fatty acids for TAG synthesis and secretion. In non-ruminants such as humans and mice, glucose serves as the precursor for fatty acid synthesis, undergoing glycolysis within the cytosol for conversion to pyruvate, transport into the mitochondria for conversion to citrate via the tricarboxylic acid (TCA) cycle, and lastly transport out of the mitochondria for conversion into acetyl CoA by the enzyme ATP citrate lyase (Neville and Picciano, 1997). In ruminants, however, volatile fatty acids (VFA) such as acetate and butryate, *not* glucose, are the primary energy source, resulting from the ruminal fermentation of ingested carbohydrates. Specific to ruminant lipogenesis, acetate, as well as β-hydroxybutyrate, serve as the precursors for acetyl CoA synthesis, the conversion of which is performed by the enzyme acetyl CoA synthetase (Bernard et al., 2008).

Once generated in the ruminant or non-ruminant, acetyl CoA in the cytoplasm is converted into malonyl CoA by acetyl CoA carboxylase, after which fatty acid synthase catalyzes the formation of the growing fatty acid chain, a process that requires NADPH as an electron-donating reducing agent (Porterfield and White, 2007; Neville and Picciano, 1997). The production of this reducing agent varies between species. While ruminant MEC primarily produce NADPH from the conversion of isocitrate to α-ketoglutarate using the enzyme isocitrate dehydrogenase, non-ruminant MEC can produce NADPH from the pentose phosphate cycle and from the conversion of malate to pyruvate using the enzyme malate dehydrogenase, in addition to also utilizing isocitrate dehydrogenase (Anderson et al., 2007; Bernard et al., 2008; Neville and Picciano, 1997). Ultimately, the newly formed fatty acid chains are esterified to a glycerol-3-phosphate backbone by the actions of glycerol-3-phosphate acyl transferase and diacylglycerol acyltransferase enzymes located on the endoplasmic reticulum, thus completing the synthesis of the TAG molecule (Bernard et al., 2008)

The secretion of TAG into the milk is unique to MEC. Individual TAG molecules combine and incorporate themselves into cytoplasmic lipid droplets that are transported to the apical plasma membrane. Upon reaching the plasma membrane, the lipid droplets fuse with it, become embedded within it, and eventually are pitched off from it in a unique budding secretory process. Together, the membrane-enveloped lipid particle secreted into the alveolar luminal space is known as the milk fat globule (MFG) (McManaman and Neville, 2003; Neville and Picciano, 1997).

1.2.7 Involution

Upon the completion of lactation, involution is an essential process that returns the mammary gland to its pre-pregnant state in preparation for subsequent pregnancies. Accordingly, the cessation of suckling and milk removal by the young deems the lactating MEC redundant and initiates their removal. The resulting morphology of the post-involutional mammary gland is similar to that of the virgin mammary gland, characterized by a rudimentary ductal branching network. Of all the phases specific to mammary gland development discussed thus far, the process of involution is the least well understood (Pai and Horseman, 2011). Yet considering that the inability of mammary tissue to regress is associated with increased tumorogenesis, involution of the mammary gland is an essential developmental activity (Strange et al., 1992). The process can be divided into two main events, a reversible apoptotic phase followed by an irreversible remodeling phase (Watson, 2006).

The apoptotic phase of involution is reversible, meaning milk production and secretion can be rescued. In mice this phase is known to last for the first two days following forced weaning with pup removal. Provided the pups are returned and allowed to nurse within that time frame, apoptosis is halted and lactation resumes (Watson, 2006). The initial accumulation of milk within the alveolar luminal space results in a volume- and pressure-induced swelling that flattens the surrounding epithelial cells. This accumulation initiates distinct alterations that collectively promote apoptosis and shedding of the secretory epithelium (Richert et al., 2000). Apoptosis is distinct from necrosis. While the former is a programmed event involving coordinated cellular condensation of a tissue structure, the latter progresses from a stressed cellular environment, resulting in the loss of structure and the random destruction of protein and nucleic acids (Strange et al., 1992). Thus, and in accordance with the programming of cell death, the alterations induced by milk accumulation include the increased expression of leukemia inhibitory factor (LIF) and transforming growth factor β3 (TGF-β3). The binding of these factors to their receptors on the MEC luminal membrane promotes receptor dimerization, inducing the phosphorylation and activation of JAK2, which subsequently phosphorylates and activates signal transducer and activator of transcription protein 3 (STAT3). While activation of STAT5A was previously described as necessary in the development of the alveolar structures during pregnancy, STAT3 is likewise necessary for the initiation of involution following the cessation of milk removal. Accordingly, STAT3 translocates to the nucleus of MEC where it serves as a mandatory transcription factor in the expression of specific genes related to apoptosis (Watson, 2006; Liu et al., 1996). These include the genes for insulin-like growth factor binding protein 5 (IGFBP-5) and CCAAT-enhancer binding protein δ (C/EBPδ), as well as the genes for the negative regulatory subunits of the phosphatidylinsolitol-4,5-bisphosphate 3-kinase (PI3K) Akt signaling pathway (Watson, 2006). Akt, also known as protein kinase B (PKB), is a general mediator of cell survival that, when activated through phosphorylation, can itself phosphorylate and consequently deactivate pro-apoptotic proteins such as B-cell lymphoma-extra large (BCL-X), BCL-2-associated death promoter (BAD), BCL-2-like protein 4 (BAX), and BCL-2 homologous antagonist killer (BAK) (Alberts et al., 2008; Datta et al., 1997). Thus, the inhibition of Akt activation is key to the apoptosis associated with involution. Perturbation of this coordinated kinase signaling cascade can result either in excessive cell death or unnecessary survival, manifesting in tissue necrosis or cancer, respectively (Datta et al., 1997).

Following the apoptotic phase, the irreversible remodeling phase is characterized by decreased milk protein gene expression, disruption of the basement membrane, collapse of the ECM surrounding the alveolar structures, adipose and vascular remodeling, and clearance of cellular debris (Li et al., 1996; Pai and Horseman, 2011). In a study using casein hydrolysates that disturb the integrity of tight junctions between MEC in dairy cattle, Shamay and colleagues have proposed that leakage of milk components into the surrounding interstitial space, caused by disruption of the MEC tight junctions, triggers the remodeling phase of involution (Shamay et al., 2003).

Components of the basement membrane, such as laminin, collagen, and fibronectin, normally serve to engage integrins located on the basal MEC membranes, thereby biochemically-anchoring MEC to the ECM (Pai and Horseman, 2011). During involution, however, stromally-derived MMP break down the ECM surrounding each alveolus, disrupting the integrin signaling and thereby resulting in MEC detachment and collapse. While the epithelium becomes increasingly more disorganized as the alveolar structures collapse, the stroma increases in density (Richert et al., 2000). Concomitant with alveolar collapse is collapse of the vasculature enveloping each secretory unit. Invading macrophages phagocytize the accumulating cellular debris, a unique immune response that curiously lacks a vigorous inflammatory reaction. Activation of this involution-associated immune response is crucial, as failure to clear away cellular debris can result in ductal ectasia, mastitis, and inflammation (Pai and Horseman, 2011).

Cellular quiescence is a physiological state distinct from senescence, as the proliferative arrest is reversible in the former yet irreversible in the latter (Harmes and DiRenzo, 2009). Upon the finalization of involution the mammary gland is considered functionally quiescent, remaining dormant until hormonal cues promote differentiation to reestablish the lactogenic alveoli in preparation for subsequent pregnancies (Harmes and DiRenzo, 2009). Involution is complete within ten to fifteen days in mice (Lascelles and Lee, 1978). In ruminants, involution is complete within thirty to sixty days; however, this process does not include significant tissue remodeling or regression (Capuco and Akers, 1999; Pai and Horseman, 2011). Morphologically, although milk stasis does initiate apoptosis of ruminant MEC, detachment of the basement membrane and total alveolar collapse are less pronounced during the remodeling phase of involution than they are in mice (Capuco and Akers, 1999). Unfortunately, although the governing mechanisms of involution are similar between differing species, the general scarcity of human mammary tissue samples during involution limits the precise characterization of this post-lactational developmental phase (Faupel-Badger et al., 2012).

1.3 Lactational Capacity

Lactational capacity, defined here as the efficiency of milk production, is of specific interest to the dairy industry considering the application of novel management approaches may increase the profitability of milk production. In a basic fundamental sense, milk production is not only a function of the number of secretory MEC but also of the secretory activity per cell (Capuco and Akers, 1999). Various techniques have been adopted in an attempt to maximize the function of and production by the mammary gland. For example, the frequent milking of ruminants earlier in lactation and the use of bovine GH (somatotropin) each result in a tenacious increase of milk yield (Wall et al., 2006; Akers 2006). Similarly, the length of the dry period, the non-lactating state in between parturitions, can be managed and affects milk production and persistency in subsequent lactations just as much as does the nutritional status of the female (Capuco and Akers, 1999). However, the underlying mechanisms to milk production, that is the hormonally-mediated genetic expression of receptors, signaling proteins, transcription factors, and cell death/survival signals, also stand as a guideline for selection or intervention strategies that best support the lactational capacity. A complete understanding of these intracellular signaling mechanisms and their economical significance to the dairy industry for enhanced MEC form and function are innovative implementations of agriculturally-applied molecular biology (Akers, 2006).

1.4 Incidence of Breast Cancer

As discussed in prior sections of this review, the hormonal milieu that promotes mammary gland differentiation is the result of an intricate interplay of ovarian, pituitary, and placental hormones acting upon not only the mammary epithelium but also the surrounding stroma. The processes through which these hormones promote secretory differentiation are of critical interest to further advances in breast cancer prevention, diagnosis, and management. While there are an estimated 1.38 million new cases of breast cancer each year, of which greater than ninety percent are ductal in origin, those factors and mechanisms that initiate cancer progression remain largely ambiguous (Russo et al., 2001; Hinck and Silberstein, 2005; Eccles et al., 2013). An estimated 458,000 women die each year from breast cancer, making it not only the most frequently diagnosed cancer in the female population, but also the most common cause of cancer death (Eccles et al., 2013). Curiously, there does exist a parity-induced protection or risk dependent on the age of a female at first parity. An early full-term pregnancy is associated with a reduced risk of breast cancer development, whereas either nulliparity or late parity is associated with a greater risk of breast cancer development (Russo et al., 2001). Specific to women, a full-term pregnancy completed before age twenty-four is protective against breast cancer development yet, a full-term pregnancy completed after age thirty is precarious (Neville et al., 2002; Russo et al., 2001). Inarguably, there exists an urgency to improve the current body of knowledge relating to this parity-associated mechanism which may provide insight to guide further efforts in prevention and treatment. Significant to these efforts is an understanding the fundamental hallmarks of cancer as well as the current limitations in breast cancer research. These topics will be discussed in the following subsections.

1.4.1 Hallmarks of Cancer

Cancer may be thought of as a disease involving dynamic changes in the genome, where the signaling processes that once supported normal cell proliferation and homeostais become defective, consequently facilitating cancer cell proliferation and tissue invasion (Hanahan and Weinberg, 2000; Radisky and Hartmenn, 2009). These genomic changes are not limited solely to the parenchyma since alterations in the stroma likewise influence mammary tumorogenesis. Regulatory defects in cell growth, differentiation, and migration therefore impact not only cell-cell or cell-matrix interactions, but also epithelial-stromal interactions (Imagawa et al., 2002).

The progressive transformation of a normal cell or tissue into a defective derivative can be classified according to any one or combination of cancerous characteristics. These hallmarks of cancer are acquired capabilities, including but not limited to self-sufficiency in growth signals, insensitivity to anti-proliferative signals, avoidance of apoptosis, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg, 2000).

Growth signaling, primarily through extracellular mitogenic growth factors, stimulates a cell to grow and proliferate. While cells other than the targeted cell produce these extracellular signaling molecules, cancer cells are capable of synthesizing and responding to their own growth factors (Alberts et al, 2008; Hanahan and Weinberg, 2000). Under normal conditions, growth factors bind to receptors on the cell surface and initiate intracellular growth-promoting cascades. Interestingly, the receptors for insulin, EGF, and vascular endothelial growth factor (VEGF) all require integrin association for optimal activation. In turn, integrins bind to ECM proteins such as laminins, collagens, and fibronectins. Cells unable to maintain proper integrin-mediated adhesion to the ECM experience impaired proloferation and survival since many of the kinases activated for progression through the cell cycle are likewise regulated by integrin (Giancotti and Ruslahti, 1999). Yet cancer cells are able to adjust which integrins they express, thereby selectively promoting their continued survival (Hanahan and Weinberg, 2000).

Conversely, cancer cells can also acquire an insensitivity to anti-proliferative signals. Crucial to the cell cycle are those regulatory components governing progression through G_1, G_2/M, and metaphase-to-anaphase, the checkpoints immediately prior to chromosomal duplication, division of the nucleus during mitosis, and division of the cytoplasm during cytokinesis, respectively. The activation through phosphorylation of numerous cyclins by cyclin-dependent kinase (CDK) guides the progression through the G_1 and G_2/M checkpoints, while both protein phosphorylation and protein destruction guide progression through the metaphase-to-anaphase checkpoint (Alberts et al., 2008). Specifically, regulatory factors known as E2F proteins promote the transcription of genes required for chromosomal duplication. The E2F factors are typically bound to a retinoblastoma (Rb) protein, thus rendering them inactive. However, if Rb is phosphorylated (pRb), E2F is released for the expression of genes necessary for cell cycle progression and proliferation (Giacinti and Giordano, 2006). Disruption of this regulatory mechanism often liberates E2F, leaving cells insensitive to anti-proliferative signals that normally prevent progression through the cell cycle (Hanahan and Weinberg, 2000). Protein degradation likewise guides cell cycle progression, operating through ubiquitinating ligases that mark targets for destruction by proteasomes. These ligases, such as the anaphase-promoting complex (APC) and the Skp1/Cullin/F-box complex (SCF), allow for the completion of mitosis and the destruction of inhibitory CDK, respectively (Alberts et al., 2008). Disruption of these mechanisms is likewise associated with proliferation abnormalities. Thus, while normal cells monitor their external and internal environments during growth and proliferation, cancer cells circumvent the corresponding checkpoints, proliferating uncontrollably with an infinite potential to replicate (Nakayama and Nakayama, 2006).

Apoptosis is a control mechanism that eliminates abnormal, nonfunctional, unnecessary, or potentially dangerous cells. An acquired avoidance of apoptosis, in conjunction with abnormal proliferation, is typical of perhaps all types of cancer (Hanahan and Weinberg, 2000). Damages and cellular stress are managed by the regulatory transcription factor p53. When activated through phosphorylation, p53 translocates from the cytoplasm into the nucleus and stimulates the transcription of components for CDK inhibitors, thereby halting progression through the cell cycle. If the sensed damage cannot be repaired during cell cycle arrest, apoptosis can also be initiated through a p53-mediated pathway. This pathway includes both the increased expression of pro-apoptotic BAX proteins and the decreased expression of anti-apoptocic BCL-2 proteins. An apoptosome complex is then formed from the combination of chytochrome c proteins that were released from the mitochondria and the activation of the cytosolic apoptotic protease activity factor 1 (APAF-1) (Alberts et al., 2008). In turn, the apoptosome complex initiates a caspase cascade that ultimately leads to the coordinated cellular condensation, cytoskeletal collapse, and nuclear envelope disassembly associated with apoptosis (Strange et al., 1992). Thus, p53 is a key suppressor of tumor formation and has come to be referred to as the “guardian of the genome” (Alberts et al., 2008; Bose and Ghosh, 2007). Under normal conditions, the levels of p53 are kept low by mouse double minute 2 homolog (MDM2) proteins, which function as yet another ubiquitinating ligase that target p53 for destruction by proteasomes. p53 is mutated in approximately fifty percent of all cancers, while over expression of MDM2 contributes to the remaining prevalence (Bose and Ghosh, 2007). Consequentially, cancerous cells are able to avoid apoptosis, aiding their survival and proliferation despite the genetic abnormalities.

All cells require contact with the circulatory system for delivery of oxygen and nutrients, removal of waste products, and hormonal signaling. This obligates any cell to reside within 100 μm from a capillary (Hanahan and Weinberg, 2000). Sustained angiogenesis, the formation of new blood vessels, is yet another acquired capability that supports the ever-increasing metabolic needs of cancerous cells. As these cells proliferate, promoting tumor expansion, they become hypoxic and initiate an angiogenic switch favoring vascular development. The primary regulator of this hypoxia-induced angiogenesis is hypoxia inducible factor 1α (HIF-1α), a protein whose increased expression promotes the transcription of pro-angiogenic factors such as vascular endothelial growth factor (VEGF). These growth factors interact with their tyrosine kinase receptors on the surface of endothelial cells to promote new blood vessel formation and growth (Alberts et al., 2008; Liao and Johnson, 2007; Hanahan and Weinberg, 2000). Vascular endothelial growth actor is an obligate regulator of angiogenesis because knock-out experiments in mice possessing only one allele of VEGF result in embryonic fatality (Liao and Johnson, 2007). Normally adult blood vessels are relatively quiescent, with angiogenesis being tightly regulated unless activated during tissue renewal and wound healing. Its enhanced activation in cancer not only promotes tumor expansion but also provides a route through which cancerous cells may metastasize (Alberts et al., 2008; Hanahan and Weinberg, 2000).

Tissue invasion and metastasis, the translocation of cancerous cells by the circulatory system to distant and foreign environments for the establishment of new colonies, are the cause of ninety percent of all cancer-associated deaths (Alberts et al., 2008; Hanahan and Weinberg, 2000). Normally cells are tethered to the ECM and to other cells by integrins and cell-cell adhesion molecules (CAMs), respectively. However, when cancer cells alter these interactions, they may acquire invasive and metastatic capabilities. Additionally, cancerous cells often exhibit an increased transcription of extracellular protease genes concomitant with a decreased transcription of protease inhibitor genes. These proteases degrade the surrounding matrix, thereby facilitating the invasion of cancerous cells into the stroma or blood supply (Hanahan and Weinberg, 2000). Unfortunately, the mechanism by which cells acquire invasive and metastatic capabilities is the least understood of all the hallmarks of cancer (Alberts et al., 2008; Hanahan and Weinberg, 2000).

1.4.2 Current Limitations of Breast Cancer Research

The signaling pathways of cancer development are complex, yet these pathways remain a common focus of breast cancer research. Any cancer-promoting gene processing a tumorogenic phenotype may be termed an "oncogene," the identification of which helps guide numerous aspects associated with cancer prevention and treatment. For example, knock-out experiments in mice pertaining solely to post-lactational involution have identified over fifty different regulatory oncogenes (Radisky and Hartmann, 2009). Additionally, more than ninety different human breast cancer cell lines have been established, each representing unique characteristic of the malignancies described earlier (Ronnov-Jessen et al., 1996). These numerous identifications and establishments emphasize how easily defective signaling pathways influence cancer development and progression (Radisky and Hartmann, 2009).

While many limitations exist in cancer research as a whole, several specific limitations currently affect that which pertains to breast cancer. For example, it is not known if lactational differentiation of the mammary gland as a whole induces the parity-associated protection against breast cancer or if the protection is a result of a unique temporal hormonal combination. Here, time point studies involving specific hormonal isolations and applications are necessary to determine whether a parity-induced hormonal protection profile exists (Britt et al., 2007). Although multiple oncogenes such as BRCA1, BRCA2, CHEK2, ATM, PALB2, BRIP1, TP53, PTEN, CDH1, and STK11 have been identified as genetically predisposing a female to the development of breast cancer, there still lacks a detailed understanding of the associated epigenetic factors, point mutations, and psychosocial considerations (Britt et al., 2007; Thompson et al., 2008). Additionally, while animal models and cell culture applications have greatly facilitated studies centered on the molecular pathways involved in breast cancer development and progression, the cell lines utilized display few of the cellular properties characteristic of normal MEC since cell lines are often derived from late-stage tumors. Here, there exists a need to improve the current models of the cellular microenvironment and their influence in aiding breast cancer development (Thompson et al., 2008). While recent genomic studies have highlighted the molecular profiles of different cancer types, comprehending and applying the vast amount of information thus obtained towards improved clinical care is in its infancy (Eccles et al., 2013). Fortunately, bioinformatics can provide a global illustration of the complex molecular mechanisms specific to breast cancer prevention and treatment.

1.5 Global Profiling through Transcriptomic Analyses

Conventional scientific research, which typically adopts a series of sequential approaches, limits itself in its ability to understand the interacting biological processes and signaling pathways occurring at the molecular level. However, within the past decade, the development of -omic technologies has enabled the analyses of thousands of biomolecules simultaneously, providing a unique approach to better understanding the biology of the organism of interest (Klopfleisch and Gruber, 2012). Such systems-level research represents a more global approach aimed at comprehensively illustrating the complex molecular mechanisms underlying cell physiology and pathology. As further technological advances ameliorate the financial and experimental aspects of -omic technologies, there also exists an increasing requirement for the ability of researchers to interpret the information thus generated (Kitano, 2002).

Bioinformatics, the scientific discipline and computational study of biological data, has only recently emerged in tandem with the advances and breakthroughs in -omic technologies. Following the development of sequencing technologies, for example, computer applications became necessary to store, organize, and analyze the immense amount of information obtained from those outputs (Pop and Salzberg, 2007). This interdisciplinary science has revolutionized biological research by integrating quantitative experimental data with the available software infrastructure to allow for a computational system analysis that illustrates the underlying molecular dynamics (Kitano, 2002). The primary focus of bioinformatics applications includes the identification, quantification, and analysis of the genome, transcriptome, or proteome, the complete set of genes, RNA, or proteins within a tissue, respectively. Still other biochemical elements such as the complete set of metabolic intermediates comprising the metabolome and the complete set of cellular sugars comprising the glycome have likewise recently emerged as research emphases (Klopfleisch and Gruber, 2012). However, comparison of the diversity and illustrations made from the analyses on these varying biochemical elements remains relatively unexplored (Pop and Salzberg, 2008).

Bioinformatics and global profiling enable the explorative identification of expression patterns specific to a particular phenotype. Such “data mining” functions to identify and characterize these patterns and profiles for the inference of hypotheses that drive subsequent studies (Klopfleisch and Gruber, 2012; Kitano, 2002). While advances in all applications of “-omic” technologies have been made, those methods specific to analyzing the transcriptome, especially that of RNA sequencing, have made the greatest progress to date (Klopfleisch and Gruber, 2012; Garber et al., 2011; Ozsolak and Milos, 2011; Costa et al., 2010). RNA sequencing (RNA-seq), the generation of complimentary DNA (cDNA) fragments derived from RNA molecules for the sequencing, characterization, and quantification of the entire transcriptome, has provided a means for a more complete understanding of the molecular mechanisms underlying cell biology (Ozsolak and Milos, 2011). Not only does RNA-seq allow for the absolute quantification of transcript abundance compared to the relative quantification of microarray technologies, but it also permits transcript sequencing independent of transcript size or prior knowledge of the genome from which it originates (Mortazavi et al., 2008; Marguerat and Bahler, 2010). Although only first utilized in 2008, RNA-seq experiments have since given insight into novel regulatory mechanisms, differential splicing, single nucleotide polymorphisms (SNP), and allele-specific transcript expression, leading some researchers to suggest it may supersede all other established transcriptomic technologies (Marguerat and Bahler, 2010; Costa et al., 2010). However, the advantages of RNA-seq are not without their own complexities because the unprecedented amount of information thus generated consequently relies heavily upon bioinformatics for the interpretation of the underlying molecular dynamics (Costa et al., 2010).

1.6 Summary

The mammary gland is a dynamic organ. Development occurs across several yet distinct stages, from the gestational development of a rudimentary bud to the priming of the gland during pregnancy to the functional differentiation of the lactating phenotype to the regression and tissue remodeling of the post-involutional gland. Regulation through these developmental stages results from an intricate endocrine and epithelial interplay according to sexual maturity and reproductive requirements. However, for all that is known about the fundamental factors affecting mammary gland morphology, much remains to be analyzed intracellularly at the molecular level. Understanding the impact and influence of the mechanism behind this regulation is significant for future advances and implications in dairy production and breast cancer research.

To provide a more complete characterization of the developmental cycle, gene expression comparisons will be made to identify the intracellular changes taking place during transitional stages within murine primary MEC. Specifically, RNA-seq and differential analysis will enable an explorative whole system approach, with the resulting bioinformatics and global profiling providing a comprehensive illustration of the complex molecular mechanisms influencing MEC physiology and pathology. Comparison of the virgin to pregnant expression profiles will allow for the analysis of the developmental stage that establishes the lactational capacity. This comparison may provide insight into selection or intervention strategies that best support initial mammary gland development and subsequent milk production. Comparison of the virgin to post-involutional quiescent expression profiles may lead to insight into the molecular mechanism underlying early parity-induced protection against the development of breast cancer. Last, parallel analysis of the transcriptome and proteome from the same sample source will allow for the comparison of two differing means of analyzing the molecular phenotype. This is a novel joint approach unique to mammary gland development that has not yet been previously reported.

CHAPTER 2 – Transcriptomic Analysis to Identify Differentially Expressed Genes Associated with the Developmental Stages of Mammary Epithelial Cells

2.1 Introduction

Various methods exist through which the transcriptome may be analyzed, including quantitative polymerase chain reaction (qPCR), hybridization-based microarrays, and Sanger sequencing-based technologies. While these technologies were developed to characterize and quantify a set of transcripts within a cell, they lack the sensitivity and resolution obtained from RNA-seq (Nagalakshmi, et al., 2010). The platforms that conduct high-throughput next generation sequencing technologies can detect hundreds of millions of raw bases in a single run by directly sequencing cDNA produced from the RNA of interest (Nagalakshmi, et al., 2010; Pareek et al., 2011). In brief, the extracted RNA is fragmented, ligated to adaptors, and retrotranscribed by complementary primers to produce fragmented, double stranded cDNA. This cDNA library is allowed to hybridize to the surface of a flow cell, where it undergoes cluster generation through isothermal bridge amplification, producing up to 200 million spatially separated template clusters. Sequencing primers are hybridized to the templates, which are then sequenced base-by-base, in parallel, using fluorescently labeled nucleotides. Through the process of cyclic reversible termination (CRT), the clusters are excited by a laser to emit a light that identifies each newly incorporated base within the sequencing reaction. Thus, the basic output of RNA-seq is a list of short sequences along with their detected quantities that may then be assessed for quality control, aligned to a reference genome, and analyzed for differential gene expression. A further detailed discussion of the technical and methodological aspects to RNA-seq is beyond the scope of this experiment and can be found elsewhere (Mardis 2008; Ansorge, 2009; Costa et al., 2010; Mortazavi et al., 2008; Ozsolak and Milos, 2011; Pareek et al., 2011; Garber et al., 2011).

RNA-seq thus offers the ability to accurately measure transcript expression in a single assay, however the resulting output must be analyzed with equally accurate and robust mathematical and statistical algorithms. In practice, the focus of RNA-seq has shifted from the generation of experimental data to its biological interpretation (Costa et al., 2010). Compared to other biomolecules, RNA itself is relatively fragile and prone to degradation by ribonucleases. The success of an RNA-seq experiment depends heavily on the quality of the extracted RNA and the generation of equally high quality, full-length, cDNA (Alberts et al., 2008; Nagalakshmi, et al., 2010). Quality control and statistical analysis of the sequenced fragments are likewise critical to the data interpretation process. For example, the sequence alignment process, mapping short sequence reads to their corresponding location along the reference genome, is essential for all subsequent analytical applications and interpretations (Li and Homer, 2010). Numerous commercially available and open-source software packages have been developed to facilitate in these analytical procedures, each with its own application of mathematical and statistical algorithms. To date, more than 80 individual tools are available for the mapping of high-throughput sequencing data to a reference sequence (Fonseca et al., 2012). Indeed, variations exist in the application of read-alignments, transcriptome reconstruction, quantification, and differential expression (Trapnell et al., 2012). Theoretically, the algorithms specific to each software program are similar. Although not novel, a comparison between two differing programs using the same sample source would be of interest to the bioinformatics community (Costa et al., 2012).

While transcriptomic analyses have been applied to cultured MEC and to whole mammary gland samples, few have studied purified MEC isolated directly from the gland *in vivo*. Furthermore, and prior to 2012, all published transcriptomic studies conducted on mammary tissue primarily utilized microarray technologies (Wickramasinghe et al., 2012). Thus, the application of RNA-seq to mammary tissue is a novel assessment of mammary gland development, and this is the first known examination of RNA-seq analysis on isolated murine MEC. Considering the variation that exists in the software programs currently available, the computational analysis of the RNA-seq output will be performed twice, using both the commercially-available CLC Genomics Workbench and the recently-launched, publicly-available Green Line Analysis, a line within the DNA Subway provided by CyVerse (formerly the iPlant Collaborative). The characteristics and approaches specific to these two programs will be discussed later; however, a further detailed discussion specific to the mathematical and computational aspects to sequence alignment and differential analysis is beyond the scope of this experiment and can be found elsewhere (Li and Homer, 2010; Fonesca et al., 2012; Li and Durbin, 2010; Pepke et al., 2009; Trapnell et al., 2012; Trapnell et al., 2013).

2.2 Methods

2.2.1 Animal Management for Material Extraction

ICR mice (Taconic, Hudson, NY) were selected for the study and were housed in the Cal Poly Rodent Colony with a 12 h light schedule and *ad libitum* access to food and water. Samples were taken from virgin, pregnant, and post-involutional quiescent mice following euthanasia using CO_2 asphyxiation and cervical dislocation. Virgin and pregnant mice were between 10 and 11 weeks of age, with pregnant mice on day 10 of pregnancy. Post-involutional quiescent mice were approximately 23 weeks of age, with pups having been weaned at day 21 of lactation and samples collected 18 days post-weaning. The estrous cycle was not taken into account for the either the virgin or the post-involutional quiescent samples. The Cal Poly Institutional Animal Care and Use Committee (IACUC) approved all animal procedures.

2.2.2 Primary Mammary Epithelial Cell Isolation

Immediately following euthanasia, mammary tissue was removed from the left and right cervical, thoracic, abdominal, and inguinal glands and rinsed in 1X Hank's Balanced Salt Solution. The excised tissue was transferred to a digestion media containing collagenase, trypsin, and EDTA in Dulbecco's Modified Eagle Medium and minced to approximately 3 mm^3 particles. The digestion media and minced tissue were incubated with constant swirling at 37 °C for 90 minutes, with disruption by pipet every 30 minutes. Cells were pelleted via centrifugation and then removed of red blood cells following a water bath incubation in red blood cell lysis buffer (8.3 g/L ammonium chloride in 0.01 M Tris-HCl) at 37 °C for 5 minutes. Cells were again pelleted via centrifugation and then removed of fibroblasts following incubation in T-75 flasks at 37°C for 1 hour. Cell suspensions were washed with centrifugations using EDTA and DNase solutions and then filtered through a 100 μm filter. The isolated epithelial cells were resuspended in DMSO cell freezing media, brought down to -80 °C at approximately -1 °C/minute, and stored in liquid nitrogen until processing.

2.2.3 RNA Extraction and Sequencing Library Preparation

Total RNA was extracted from 3 samples per developmental stage (n=3) using an RNeasy® Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. RNA was quantified by a BioTek Synergy 2 microplate spectrophotometer (BioTek Instruments, Winooski, VT) and the quality and integrity were assessed with an Experion bio-analyzer (Bio-Rad, Hercules, CA) according to the manufacturer's instructions. Messenger RNA (mRNA) was isolated and purified using a TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) by the Medrano Lab at UC Davis. The mRNA was fragmented to approximately 200 bp fragments, synthesized to cDNA, and ligated with adapters and sequencing indexes according to the manufacturer's instructions.

2.2.4 RNA Sequencing

The cDNA libraries were sequenced by the Vincent J. Coates Genomics Sequencing Laboratory in the California Institute for Quantitative Biosciences (QB3) Facility at UC Berkeley. Sequencing was performed in one lane, multiplexing all samples within that lane, using a HiSeq 2000 Sequence Analyzer (Illumina, San Diego, CA). Single-end sequence read files were made available for download from the host to a local server, and were accessed through file transfer protocol (FTP) FileZilla software.

2.2.5 Quality Control Analysis

Quality control was performed twice, using both the commercially-available CLC Genomics Workbench version 6.5 (CLC Bio, Aarhus, Denmark) and the recently-launched, publicly-available Green Line Analysis (iPlant Collaborative, Tucson, AZ).

2.2.5.1 CLC Genomics Workbench

Single read sequences of 100 bp from each sample were assessed for quality control as directed through the Toolbox: NGS Core Tools expression analysis application. Graphical reports were created for all virgin, pregnant, and post-involutional quiescent sequenced samples to depict the quality control analysis according to per-sequence parameters such as length distribution, GC-content, ambiguous base content, and quality distribution, and according to per-base parameters such as coverage, nucleotide contributions, GC-content, ambiguous base-content, and quality distribution. Any sample that did not meet the quality control parameter requirements was eliminated from further analyses.

2.2.5.2 Green Line Analysis

Single read sequences of 100 bp from each sample were assessed for quality control as directed through the Manage Data QC application. Graphical reports were created for all virgin, pregnant, and post-involutional quiescent sequenced samples to depict the quality control analysis according to per-sequence parameters such as length distribution, GC-content, and quality distribution, and according to per-base parameters such as nucleotide contributions, GC-content, ambiguous base-content, and quality distribution. Any sample that did not meet the quality control parameter requirements was eliminated from further analyses.

2.2.6 Differential Gene Expression Analysis

Differential analysis was performed twice, likewise using both CLC Genomics and Green Line Analysis.

2.2.6.1 CLC Genomics

Those samples indicative of good quality were assembled on the annotated GRCm38.71 *Mus musculus* reference genome (http://www.ncbi.nlm.nih.gov/genome) as directed through the Toolbox: Transcriptomic: RNA-seq Analysis application. Map settings were dictated as follows: a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and a maximum number of 10 hits for a read. Exon discovery settings were dictated as follows: a required relative expression level of 0.20, a minimum number of 10 reads, and a minimum read length of 50 bp. The data were normalized by calculating the reads per kilo base per million mapped reads (RPKM = total exon reads/mapped reads in millions x exon length in kb) for each gene and annotated with *Mus musculus* genome assembly (38,124 genes). For the statistical analysis, a t-test was performed on $\log_{10}$-transformed data to identify the genes with significant changes in expression ($p < 0.05$) between virgin and pregnant samples and between virgin and post-involutional quiescent samples. Overall expression data was reported as the number of genes identified as expressed, the number of genes differentially expressed, and the number of genes detected but not differentially expressed. Expression data was further reported as the number of differentially expressed genes detected with high expression (>500 RPKM), medium expression (10-500 RPKM), and low expression (0.02-9.99 RPKM). Samples were sequenced on one lane, negating the normalization against any confounding factors due to lane or batch.

2.2.6.2 Green Line Analysis

Those samples indicative of good quality were aligned to the annotated GRCm38.74 *Mus musculus* reference genome (http://www.ncbi.nlm.nih.gov/genome) (39,174 genes) as directed through the TopHat application of the Analyze Transcriptome. Default parameter settings were used, except advanced parameter settings were dictated to match those of the CLC Genomics Workbench analysis as follows: a maximum number of 10 hits for a read, a minimum isoform fraction of 0.20, a segment length of 80 and the enabling of the "no novel junctions" option. Following alignment and annotation, the RPKM for each gene were accessed and downloaded using the cummerbund R package within the CyVerse Discovery Environment. Replicate values were fetched using the command gene.rep.matrix<-repFpkmMatrix(genes(cuff)). For the statistical analysis, a t-test was performed on $\log_{10}$-transformed data to identify the genes with significant changes in expression ($p < 0.05$) between virgin and pregnant samples and between virgin and post-involutional quiescent samples. Overall expression data was reported as the number of genes identified as expressed, the number of genes differentially expressed, and the number of genes detected but not differentially expressed. Expression data was further reported as the number of differentially expressed genes detected with high expression (>500 RPKM), medium expression (10-

500 RPKM), and low expression (0.02-9.99 RPKM). Samples were sequenced on one lane, negating the normalization against any confounding factors due to lane or batch.

2.2.7 Platform Comparison of Expressed and Dually Detected Genes between CLC Genomics and Green Line Analysis

Platform comparisons were performed using those genes identified as being dually differentially expressed by both CLC Genomics Workbench and Green Line Analysis. Overall expression was compared for the dually differentially expressed genes between the virgin and pregnant developmental stages and between virgin and post-involutional quiescent developmental stages. Expression was also compared for the dually differentially expressed genes detected with high expression (>500 RPKM), medium expression (10-500 RPKM), and low expression (0.02-9.99 RPKM) between the virgin and pregnant developmental stages and between virgin and post-involutional quiescent developmental stages.

Simultaneously considering both the virgin-to-pregnant and the virgin-to-post-lactational quiescent developmental comparisons, a general linear model (GLM) was used to compare the observed fold change in expression for those genes identified as being dually differentially expressed by both CLC Genomics Workbench and Green Line Analysis. The corresponding R^2 values were calculated using Statistical Analysis System (SAS) JMP Pro version 12.2.0. (SAS, Cary, NC).

2.3 Results

2.3.1 CLC Genomics Workbench

As mentioned in Chapter 2.2.5.1, graphical reports were created for all virgin, pregnant, and post-involutional quiescent sequenced samples to depict the quality control analysis (see Appendix A). All samples met the quality control parameter requirements, with the exception of Virgin Mouse 2. This sample yielded very low reads and was therefore eliminated from further analyses. Of the 8 remaining samples analyzed, a total of approximately 151.6 million sequence reads were obtained, with an average of 19 million reads for each sample, of which greater than 90 percent were categorized as mapped reads to the *Mus musculus* genome. The number of expressed genes for each sample assembled on the annotated GRCm38.71 *Mus musculus* genome were determined and calculated as a percentage out of 38,124 genes (Table 1).

Accordingly, three biological replicates were analyzed for the pregnant and post-involutional quiescent stages of MEC development, while only 2 replicates could be analyzed for the virginal stage. Comparing the virginal state to the pregnant state, 2,681 (7.03%) genes were detected as differentially expressed ($p < 0.05$), 15,582 (40.87%) genes were detected but not differentially expressed ($p> 0.05$), and 19,861 (52.10%) genes were not expressed. Of the 2,681 genes differentially expressed between the virgin and pregnant samples, 1,037 genes were down-regulated while the remaining 1,644 genes were up-regulated ($p< 0.05$). Comparing the virginal state to the post-involutional quiescent state, 2,341 (6.14%) genes were detected as differentially expressed ($p< 0.05$), 15,980 (41.92%) genes were detected but not differentially expressed ($p> 0.05$), and 19,803 (51.94%) genes were not expressed. Of the 2,341 genes differentially expressed between the virgin and post-involutional quiescent samples, 1,820 genes were down-regulated while the remaining 521 genes were up-regulated ($p < 0.05$) (Table 2).

Table 1: Total and Mapped Reads Detected by CLC Genomics Workbench Analysis

MEC were isolated from virgin, pregnant, and post-lactational quiescent mouse mammary glands. RNA was extracted, sequenced, and aligned to the annotated GRCm38.71 *Mus musculus* genome (38,124 genes). Graphical reports were created to depict the quality control analysis according to per-sequence parameters such as length, distribution, GC-content, ambiguous base-content, and quality distribution. The number of reads detected per sample analyzed, the number of those reads that were mapped to the reference genome, the percentage of those mapped reads, the number of genes expressed, and the percentage of those genes expressed have been listed for each mouse sample analyzed through the CLC Genomics Workbench. With the exception of the limited number reads derived from Virgin Mouse 2, an average of 19 million reads were obtained for each sample, with greater than 93 percent being mapped to the reference genome.

Sample Description	**Reads (n)**	**Mapped Reads (n)**	**% Mapped Reads**	**Genes Expressed (n)**	**% Genes Expressed**
Virgin Mouse 1	31,241,022	29,428,183	94.20	17,087	44.82
Virgin Mouse 2	6,176	N/A	N/A	N/A	N/A
Virgin Mouse 3	16,864,550	15,881,378	94.17	16,657	43.69
Pregnant Mouse 1	16,666,247	15,318,291	91.91	16,484	43.24
Pregnant Mouse 2	11,861,826	11,170,110	94.17	16,170	42.41
Pregnant Mouse 3	15,686,433	14,777,213	94.20	16,374	42.95
Quiescent Mouse 1	18,222,830	17,173,154	94.24	16,552	43.42
Quiescent Mouse 2	20,998,909	19,693,821	93.78	16,433	43.10
Quiescent Mouse 3	20,109,186	18,741,292	93.20	16,518	43.33
Total Reads	**151,657,179**	**112,755,259**	**93.73**	**-**	**-**

Table 2: Differentially Expressed Genes Detected by CLC Genomics Workbench Analysis

Differential analysis was performed for the detected genes successfully mapped to the *Mus musculus* genome. Developmental comparisons were made for the pregnant state relative to the virgin state, and for the post-lactational quiescent state relative to the virgin state. The numbers and percentages of those genes differentially expressed, detected but not differentially expressed, and not expressed are given for each developmental comparison. A total of 2,681 (7.03%) and 2,341 (6.14%) genes were differentially expressed ($p < 0.05$), with the corresponding number of down-regulated genes indicated in red and the corresponding number of up-regulated genes indicated in green.

<table>
<tr><th rowspan="2">Developmental Comparison</th><th colspan="4">Genes Differentially Expressed</th><th colspan="2">Genes Detected but not Differentially Expressed</th><th colspan="2">Genes not Detected</th></tr>
<tr><th colspan="2">n</th><th colspan="2">%</th><th>n</th><th>%</th><th>n</th><th>%</th></tr>
<tr><td><b>Virgin vs Pregnant</b></td><td colspan="2">2,681</td><td colspan="2">7.03</td><td>15,582</td><td>40.87</td><td>19,861</td><td>52.10</td></tr>
<tr><td><b>Downregulated/Upregulated</b></td><td>1,037</td><td>1,644</td><td>2.72</td><td>4.31</td><td colspan="4">N/A</td></tr>
<tr><td><b>Virgin vs Quiescent</b></td><td colspan="2">2,341</td><td colspan="2">6.14</td><td>15,980</td><td>41.92</td><td>19,803</td><td>51.94</td></tr>
<tr><td><b>Downregulated/Upregulated</b></td><td>1,820</td><td>521</td><td>4.77</td><td>1.37</td><td colspan="4">N/A</td></tr>
</table>

Of the 2,681 genes detected as differentially expressed in the comparison of the virginal state to the pregnant state ($p <0.05$), 23 genes were detected with high expression (>500 RPKM), 1,201 genes were detected with medium expression (10-500 RPKM), and 1,767 genes were detected with low expression (0.02-9.99 RPKM). Of the 2,341 genes detected as differentially expressed in the comparison of the virginal state to the post-involutional quiescent state ($p <0.05$), 70 genes were detected with high expression (>500 RPKM), 969 genes were detected with medium expression (10-500 RPKM), and 1,655 genes were detected with low expression (0.02-9.99 RPKM) (Table 3).

Table 3: High, Medium, and Low Differentially Expressed Genes Detected by CLC Genomics Workbench Analysis

Differential analysis was performed for the detected genes successfully mapped to the *Mus musculus* genome. The data were normalized by calculating RPKM for each gene, where RPKM = total exon reads/mapped reads in millions x exon length in kb. Developmental comparisons were made for the pregnant state relative to the virgin state, and for the post-lactational quiescent state relative to the virgin state. Of the 2,681 and 2,341 genes differentially expressed within each developmental comparison (p <0.05), the number of genes detected with high, medium, and low expression are given, with the corresponding number of down-regulated genes indicated in red and the corresponding number of up-regulated genes indicated in green. The number of high, medium, and low genes for each developmental comparison sums to more than the number of differentially expressed genes considering a given gene may be expressed with a specific strength in one developmental state and a different strength in the other developmental state.

Expression	**Virgin vs Pregnant**		**Virgin vs Quiescent**	
	n		**n**	
High **>500 RPKM**	23		70	
Downregulated/Upregulated	10	13	5	65
Medium **10-500 RPKM**	1,201		969	
Downregulated/Upregulated	457	744	654	315
Low **0.02-9.99 RPKM**	1,767		1,655	
Downregulated/Upregulated	694	1,073	1,409	246

2.3.2 Green Line Analysis

As mentioned in Chapter 2.2.5.2, graphical reports were created for all virgin, pregnant, and post-involutional quiescent sequenced samples to depict the quality control analysis (see Appendix B). All samples met the quality control parameter requirements, with the exception of Virgin Mouse 2. Again, this sample yielded very low reads and was therefore eliminated from further analyses. Of the 8 remaining samples analyzed, a total of approximately 154.1 million sequence reads were obtained, with an average of 19 million reads for each sample, of which greater than 90 percent were categorized as mapped reads to the *Mus musculus* genome. The number of expressed genes for each sample assembled on the annotated GRCm38.75 *Mus musculus* genome were determined and calculated as a percentage out of 39,174 genes (Table 4).

Accordingly, 3 biological replicates were analyzed for the pregnant and post-involutional quiescent stages of MEC development, while only 2 replicates could be analyzed for the virginal stage. Comparing the virginal state to the pregnant state, 1,470 (3.75%) genes were detected as differentially expressed ($p < 0.05$), 14,959 (38.19%) genes were detected but not differentially expressed ($p > 0.05$), and 22,745 (58.06%) genes were not expressed. Of the 1,470 genes differentially expressed between the virgin and pregnant samples, 756 genes were down-regulated while the remaining 715 genes were up-regulated ($p< 0.05$). Comparing the virginal state to the post-involutional quiescent state, 1,392 (3.56%) genes were detected as differentially expressed ($p< 0.05$), 15,201 (38.80%) genes were detected but not differentially expressed ($p> 0.05$), and 22,581 (57.64%) genes were not expressed. Of the 1,392 genes differentially expressed between the virgin and post-involutional quiescent samples, 725 genes were down-regulated while the remaining 667 genes were up-regulated ($p< 0.05$) (Table 5).

Table 4: Total and Mapped Reads Detected by Green Line Analysis

MEC were isolated from virgin, pregnant, and post-lactational quiescent mouse mammary glands. RNA was extracted, sequenced, and aligned to the GRCm38.75 *Mus musculus* genome (39,174). Graphical reports were created for all virgin, pregnant, and post-involutional quiescent sequenced samples to depict the quality control analysis according to per-sequence parameters such as length distribution, GC-content, and quality distribution, and according to per-base parameters such as nucleotide contributions, GC-content, ambiguous base-content, and quality distribution. The number of reads detected per sample analyzed, the number of those reads that were mapped to the reference genome, the percentage of those mapped reads, the number of genes expressed, and the percentage of those genes expressed have been listed for each mouse sample analyzed through the Green Line Analysis. With the exception of the limited number reads derived from Virgin Mouse 2, an average of 19 million reads were obtained for each sample, with greater than 91 percent being mapped to the reference genome.

Sample Description	**Reads (n)**	**Mapped Reads (n)**	**% Mapped Reads**	**Genes Expressed (n)**	**% Genes Expressed**
Virgin Mouse 1	31,772,780	29,209,123	91.93	20,127	51.38
Virgin Mouse 2	6,424	N/A	N/A	N/A	N/A
Virgin Mouse 3	17,146,830	15,780,379	92.03	19,269	49.19
Pregnant Mouse 1	16,944,508	15,393,785	90.84	18,988	48.47
Pregnant Mouse 2	12,063,133	11,166,600	92.56	18,473	47.16
Pregnant Mouse 3	15,952,008	14,737,176	92.38	18,879	48.19
Quiescent Mouse 1	18,503,205	17,184,240	92.87	19,224	49.07
Quiescent Mouse 2	21,352,039	19,558,433	91.59	18,952	48.38
Quiescent Mouse 3	20,441,670	18,535,917	90.67	19,062	48.66
Total Reads	**154,182,597**	**141,565,653**	**91.81**	**-**	**-**

Table 5: Differentially Expressed Genes Detected by Green Line Analysis

Differential analysis was performed for the detected genes successfully mapped to the *Mus musculus* genome. Developmental comparisons were made for the pregnant state relative to the virgin state, and for the post-lactational quiescent state relative to the virgin state. The numbers and percentages of those genes differentially expressed, detected but not differentially expressed, and not expressed are given for each developmental comparison. A total of 1,470 (3.75%) and 1,392 (3.56%) genes were differentially expressed ($p < 0.05$), with the corresponding number of down-regulated genes indicated in red and the corresponding number of up-regulated genes indicated in green.

<table>
<tr><th rowspan="2">Developmental Comparison</th><th colspan="4">Genes Differentially Expressed</th><th colspan="2">Genes Detected but not Differentially Expressed</th><th colspan="2">Genes not Detected</th></tr>
<tr><th colspan="2">n</th><th colspan="2">%</th><th>n</th><th>%</th><th>n</th><th>%</th></tr>
<tr><td>**Virgin vs Pregnant**</td><td colspan="2">1,470</td><td colspan="2">3.75</td><td>14,959</td><td>38.19</td><td>22,745</td><td>58.06</td></tr>
<tr><td>**Downregulated/Upregulated**</td><td>756</td><td>715</td><td>1.92</td><td>1.83</td><td colspan="4">N/A</td></tr>
<tr><td>**Virgin vs Quiescent**</td><td colspan="2">1,392</td><td colspan="2">3.56</td><td>15,201</td><td>38.80</td><td>22,581</td><td>57.64</td></tr>
<tr><td>**Downregulated/Upregulated**</td><td>725</td><td>667</td><td>1.85</td><td>1.71</td><td colspan="4">N/A</td></tr>
</table>

Of the 1,470 genes detected as differentially expressed in the comparison of the virginal state to the pregnant state ($p < 0.05$), 16 genes were detected with high expression (>500 RPKM), 741 genes were detected with medium expression (10-500 RPKM), and 887 genes were detected with low expression (0.02-9.99 RPKM). Of the 1,392 genes detected as differentially expressed in the comparison of the virginal state to the post-involutional quiescent state ($p < 0.05$), 27 genes were detected with high expression (>500 RPKM), 672 genes were detected with medium expression (10-500 RPKM), and 564 genes were detected with low expression (0.02-9.99 RPKM) (Table 6).

Table 6: High, Medium, and Low Differentially Expressed Genes Detected by Green Line Analysis Differential analysis was performed for the detected genes successfully mapped to the *Mus musculus* genome. The data were normalized by calculating RPKM for each gene, where RPKM = total exon reads/mapped reads in millions x exon length in kb. Developmental comparisons were made for the pregnant state relative to the virgin state, and for the post-lactational quiescent state relative to the virgin state. Of the 1,470 and 1,392 genes differentially expressed within each developmental comparison (p <0.05), the number of genes detected with high, medium, and low expression are given, with the corresponding number of down-regulated genes indicated in red and the corresponding number of up-regulated genes indicated in green. The number of high, medium, and low genes for each developmental comparison sums to more than the number of differentially expressed genes considering a given gene may be expressed with a specific strength in one developmental state and a different strength in the other developmental state.

Expression	Virgin vs Pregnant		Virgin vs Quiescent	
	n		n	
High >500 RPKM	16		27	
Downregulated/Upregulated	8	8	13	14
Medium 10-500 RPKM	741		672	
Downregulated/Upregulated	372	369	340	332
Low 0.02-9.99 RPKM	887		564	
Downregulated/Upregulated	694	422	315	249

2.3.3 Platform Comparison of Differential Analysis between CLC Genomics and Green Line Analysis

For the virgin-to-pregnant comparison, of the 4,151 total number of genes differentially detected by the CLC Genomics Workbench and the Green Line Analysis (p < 0.05), 983 (31.02%) of those genes were dually detected by both platforms. Of the remaining genes, 1,698 were detected only by the CLC Genomics Workbench platform while 487 were detected only by the Green Line Analysis (Figure 1). For the 39 total number of genes detected with high expression within the virgin-to-pregnant comparison (>500 RPKM, p <0.05), 9 (30.00%) of those genes were dually detected by both platforms. Of the remaining genes, 14 were detected only by the CLC Genomics Workbench platform while 7 were detected only by the Green Line Analysis. For the 1,942 total number of genes detected with medium expression within the virgin-to-pregnant comparison (10-500 RPKM, p <0.05), 454 (30.51%) of those genes were dually detected by both platforms. Of the remaining genes, 747 were detected only by the CLC Genomics Workbench platform while 287 were detected only by the Green Line Analysis. For the 2,654 total number of genes detected with low expression within the virgin-to-pregnant comparison (0.20-9.99 RPKM, p <0.05), 599 (26.68%) of those genes were dually detected by both platforms. Of the remaining genes, 1,208 were detected only by the CLC Genomics Workbench platform while 328 were detected only by the Green Line Analysis (Figure 2). For the virgin-to-post-lactational quiescent comparison, of the 3,733 total number of genes differentially detected by the CLC Genomics Workbench and the Green Line Analysis (p < 0.05), 793 (26.97%) of those genes were dually detected by both platforms. Of the remaining genes, 1,548 were detected only by the CLC Genomics Workbench platform while 599 were detected only by the Green Line Analysis (Figure 3). For the 97 total number of genes detected with high expression within the virgin-to-post-lactational quiescent comparison (>500 RPKM, p <0.05), 1 (1.04%) of those genes were dually detected by both platforms. Of the remaining genes, 69 were detected only by the CLC Genomics Workbench platform while 26 were detected only by the Green Line Analysis. For the 1,641 total number of genes detected with medium expression within the virgin-to-post-lactational quiescent comparison (10-500 RPKM, p <0.05), 116 (7.60%) of those genes were dually detected by both platforms. Of the remaining genes, 853 were detected only by the CLC Genomics Workbench platform while 556 were detected only by the Green Line Analysis. For the 2,219 total number of genes detected with low expression within the virgin-to-post-lactational quiescent comparison (0.20-9.99 RPKM, p <0.05), 226 (11.33%) of those genes were dually detected by both platforms. Of the remaining genes, 1,429 were detected only by the CLC Genomics Workbench platform while 338 were detected only by the Green Line Analysis (Figure 4).

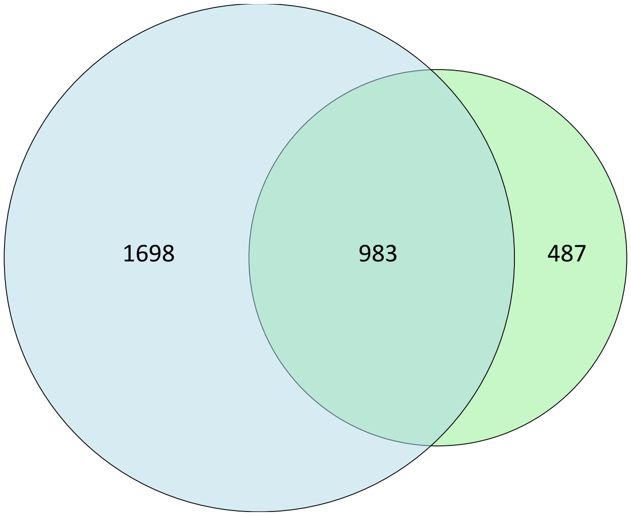


Figure 1: Platform Comparison of Differentially Detected Genes from the Analysis of Virgin vs. Pregnant MEC

An area-proportional Venn diagram was created to depict the number of dually differentially detected genes between the two transcriptomic platforms. While 983 genes were dually detected by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-pregnant developmental comparison (p-value < 0.05), 2,185 genes were detected in one platform but not the other. Those genes specific to the CLC Genomics Workbench are represented in blue while those genes specific to the Green Line Analysis are represented in green, with the dually differentially detected genes represented in the enclosed section.

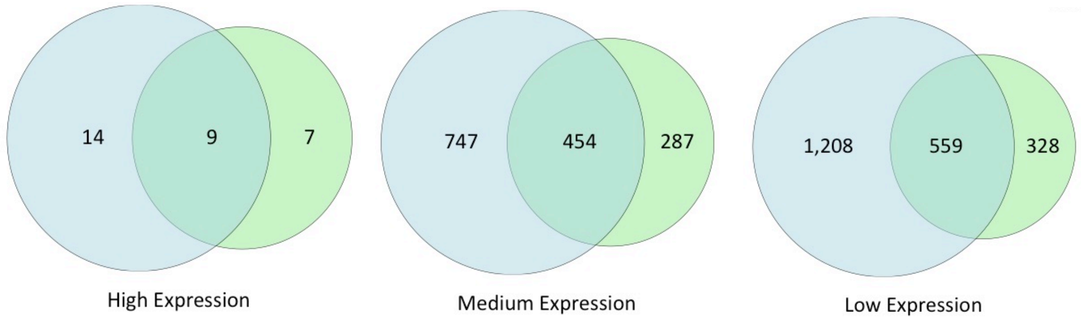


Figure 2: Platform Comparison of Genes from the Analysis of Virgin vs. Pregnant MEC Differentially Detected with High, Medium, and Low Expression

Area-proportional Venn diagrams were created to depict not only the number of dually differentially detected genes between the two transcriptomic platforms, but also the number of genes with high, medium, and low expression. While 9 genes were dually detected with high expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-pregnant developmental comparison (>500 RPKM p-value < 0.05), 21 genes were detected in one platform but not the other. While 454 genes were dually detected with medium expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-pregnant developmental comparison (10-500 RPKM, p <0.05), 1,034 genes were detected in one platform but not the other. While 559 genes were dually detected with low expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-pregnant developmental comparison (0.20-9.99 RPKM, p <0.05), 1,536 genes were detected in one platform but not the other. For all three comparisons, those genes specific to the CLC Genomics Workbench are represented in blue while those genes specific to the Green Line Analysis are represented in green, with the dually differentially detected genes represented in the enclosed sections.

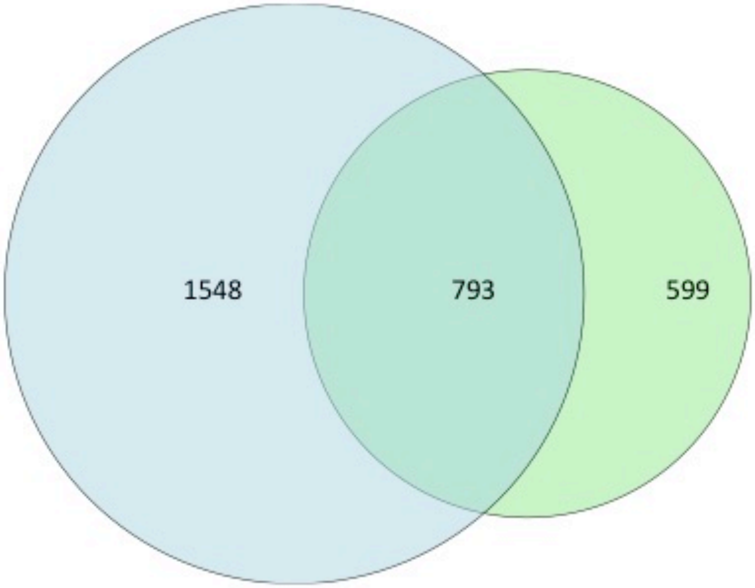


Figure 3: Platform Comparison of Differentially Detected Genes from the Analysis of Virgin vs. Quiescent MEC

An area-proportional Venn diagram was created to depict the number of dually differentially detected genes between the two transcriptomic platforms. While 793 genes were dually detected by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-post-lactational quiescent developmental comparison (p-value < 0.05), 2,147 genes were detected in one platform but not the other. Those genes specific to the CLC Genomics Workbench are represented in blue while those genes specific to the Green Line Analysis are represented in green, with the dually differentially detected genes represented in the enclosed section.

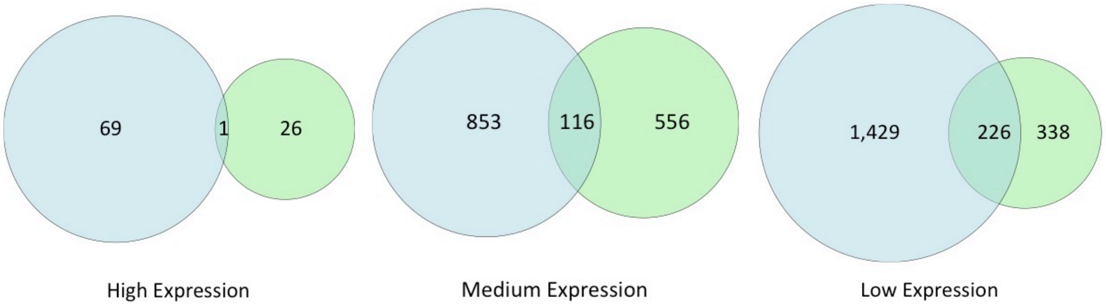


Figure 4: Platform Comparison of Genes from the Analysis of Virgin vs. Quiescent MEC Differentially Detected with High, Medium, and Low Expression

Area-proportional Venn diagrams were created to depict not only the number of dually differentially detected genes between the two transcriptomic platforms, but also the number of genes with high, medium, and low expression. While 1 gene was dually detected with high expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-post-lactational quiescent developmental comparison (>500 RPKM p-value < 0.05), the remaining 95 genes were detected in one platform but not the other. While 116 genes were dually detected with medium expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-post-lactational developmental comparison (10-500 RPKM, p <0.05), 1,409 genes were detected in one platform but not the other. While 226 genes were dually detected with low expression by both the CLC Genomics Workbench and the Green Line Analysis for the virgin-to-post-lactational quiescent developmental comparison (0.20-9.99 RPKM, p <0.05), 1,767 genes were detected in one platform but not the other. For all three comparisons, those genes specific to the CLC Genomics Workbench are represented in blue while those genes specific to the Green Line Analysis are represented in green, with the dually differentially detected genes represented in the enclosed section.

Considering only those dually detected genes as determined by either platform, 983 genes from the virgin-to-pregnant comparison and 793 genes from the virgin-to-post-lactational quiescent comparison were plotted against each other in a regression analysis, using the detected fold change values from CLC Genomics Workbench along the horizontal axis and the detected fold change values from the Green Line Analysis along the vertical axis. The corresponding coefficient of determination values were calculated as $R^2 = 0.70$ for the virgin-to-pregnant comparison and $R^2 = 0.78$ for the virgin-to-post-lactational quiescent comparison (Figure 5). A line of best fit was matched for each developmental comparison, with the corresponding slopes calculated as 0.6355x for the virgin-to-pregnant comparison and 0.7824x for the virgin-to-post-lactational quiescent comparison.

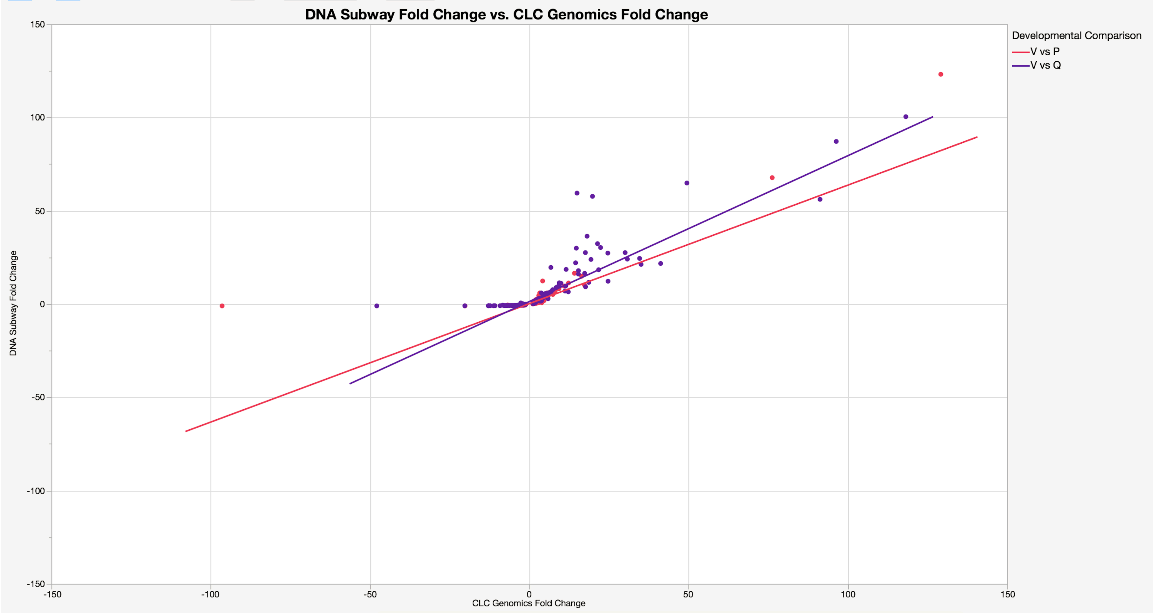


Figure 5: Platform Comparison Regression Analysis of Differentially Expressed Genes Dually Detected by CLC Genomics Workbench and Green Line Analysis

Considering only those dually detected genes as determined by either platform, 983 genes from the virgin-to-pregnant comparison and 793 genes from the virgin-to-post-lactational quiescent comparison were plotted against each other in a regression analysis, using the detected fold change values from CLC Genomics Workbench along the horizontal axis and the detected fold change values from the Green Line Analysis along the vertical axis. Those dually detected genes from the virgin-to-pregnant developmental comparison are shown in red, while those dually detected genes from the virgin-to-post-lactational quiescent developmental comparison are shown in purple. A line of best fit was drawn for each developmental comparison, with the corresponding as R^2 values calculated as an estimate of the amount of variation explained by the results obtained.

2.4 Discussion

MEC are responsible for the synthesis and secretion of milk components; however, the molecular mechanisms taking place within MEC remain relatively unexplored. To provide a more complete characterization of the developmental cycle, single-ended RNA-seq was performed on MEC isolated from virgin, pregnant, and post-involutional quiescent sibling mice. The objective of this experiment is to identify the differentially expressed genes influencing MEC development during pregnancy and the differences between the nulliparous and primiparous states. Although sequencing reads in pairs can help detect alignment errors and improve the sensitivity and specificity compared to that of single-end reads, such an experimental approach is necessary only when isoform annotation and exploration of the genetic architecture are the primary goals (Li and Homer, 2010).

Numerous commercially available and open-source software packages have been developed to facilitate in the assembly and differential analysis of RNA-seq data. However, while these computer programs all share a common goal to RNA-seq analysis, variations exist in their application of mathematical and statistical algorithms and computer science programming (Trapnell et al., 2012). For this reason the computational analysis of the RNA-seq output was performed twice, using both the CLC Genomics Workbench and Green Line Analysis platforms.

The CLC Genomics Workbench is a commercially-available software program for the analysis of RNA-seq data generated through Illumina sequencing technologies. The alignment algorithms applied by this program are based on hash tables, data structures that are used to store and sort information. Through a process known as "seeding," algorithms based on hash tables use the short sequenced reads as queries against a reference genome, finding areas of local alignment along the reference genome, and then using a q-gram calculation to filter out poor matches. A vectorized adaptation of Smith-Waterman programming accelerates this process and avoids repeated alignment of identical subsequences. The output is thus statistically significant alignments of the sequenced reads along the reference genome (Li and Homer, 2010; Li and Durbin, 2010; Fonesca et al., 2012; Mortazavi et al., 2008). The subsequent differential analysis is based on the methods by Mortazavi et al., 2008. In brief, through the application of an Enhanced Read Analysis of Gene Expression (ERANGE) program, the prevalence of transcripts against the reference genome is calculated and then normalized to a RPKM expression measure. Each RPKM is generated with a corresponding p-value, the probability of obtaining that value—or a value more extreme—given that the null hypothesis is true (Grafen and Hails, 2002). This normalization is a necessary process, considering that the length of each transcript must be taken into account when comparing the detected expression of transcripts since longer transcripts would naturally yield greater detection reads (Mortazai et al., 2008; Trapnell et al., 2012).

In contrast, the Green Line Analysis is an open-source program for the same analysis of Illumina-generated RNA-seq data. This line, which is currently in Beta testing for user acceptability and feedback, was developed as part of the DNA Subway, a publicly accessible analytical platform managed by the iPlant Collaborative. Although it is in the Green Line Analysis roadmap to add an assembly workflow as part of their analytical processes, unfortunately this feature is unavailable at present since the platform accepts only one FASTQ file for upload per replicate. As such, the assembled sequence files were generated by the bioinformatics directors of the Cold Springs Harbor Laboratory using a simple concatenation and then uploaded onto the Green Line Analysis for alignment to the mouse genome. TopHat, Cufflinks, and Cuffdiff are software tools composing what is known as the Tuxedo Protocol and which the Green Line Analysis utilizes for assessing RNA-seq data. Specifically, TopHat aligns sequenced reads to a reference genome using a Bowtie program, which in turn uses an FM index data structure to store and sort the sequence information (Trapnell et al., 2012). Distinct from the hash table algorithms discussed above, the FM index algorithms are based on suffice/prefix tries. Whereas non-vectorized Smith-Waterman hash table alignments are performed for each identical copy of a substring sequence within the reference genome, algorithms built upon suffice/prefix tries align multiple identical copies of the reference substring (Li and Homer, 2010). The resulting alignment files are then used to calculate differential expression levels and test the statistical significance of the observed changes (Trapnell et al., 2012). To normalize the expression data for transcript length, TopHat calculates the RPKM, essentially analogous to the RPKM. The two remaining tools within the Tuxedo protocol were not utilized in this experiment considering CuffLinks is only needed for the detection of novel isoform discovery and CuffDiff does not allow transformation of the raw data. Additionally, in CuffDiff each RPKM is generated with a corresponding q-value, a similar measure of the statistical significance of a p-value except it also considers conditionality and takes into account the fact that thousands of features are being tested simultaneously. In theory q-values, which are an extension of the FDR, are more intuitive for transcriptomic studies where a much higher FDR can be tolerated than with a p-value (Storey and Tibshirani, 2003). To avoid the compounding comparison of differential genes as determined by both p-values and q-values, the raw FKMP values were accessed and downloaded by the External Collaborations Directors of the Cold Springs Harbor Laboratory using the cummerbund R package within the CyVerse Discovery Environment. This accession then allowed the same statistical analysis to be performed on the expression values detected by the Green Line Analysis as was performed on those expression values detected by the CLC Genomics Workbench.

Here, we have dually demonstrated that high-quality cDNA was generated and successfully sequenced from eight of the nine samples analyzed. Both the CLC Genomics Workbench and the Green Line Analysis programs assessed several quality control conditions according to per-sequence and per-base parameters such as base content, ambiguous base content, quality scores, and length distribution. Base content considers the proportion of each nucleotide base, where no biases or overrepresentation should be detected within the sequence libraries. If the sequence analyzer is unable to identify a specific nucleotide during the sequencing processes, that base is considered ambiguous. Ideally, the proportion of ambiguous base content should be minimal, indicating successful base calling and data interpretation by the sequence analyzer when the library clusters are being fluorescently excited. Quality scores are calculated to assess the error probability of base detection. By measuring the accuracy of the sequencing process, systematic errors can be detected if a significant proportion of the sequences analyzed yield a low-quality score. Length distribution considers the fragment size for the libraries sequenced, where all fragments should be of a uniform length.

Sequence reads indicative of poor quality provide less biological information, hinder the assembly and alignment processes, and should therefore be eliminated from subsequent analyses. Although all nine samples were indicative of good quality, both software programs detected only approximately 6,000 reads for Virgin Mouse 2. Such low read detection is most likely due to degradation of the extracted RNA. Although it cannot be stated definitively where the degradation might have occurred, possible sources include the library construction and amplification processes, transportation between UC Davis and UC Berkeley, or any cleaning procedure performed by the QB3 Facility at UC Berkeley. How accurately RNA-seq reflects the original RNA population is dependent upon the quality of the extracted RNA throughout the sequencing process (Costa et al., 2010). Considering the eight remaining samples averaged approximately 19 million reads each, Virgin Mouse 2 was eliminated from the subsequent differential analysis.

Differential analysis individually performed by the CLC Genomics Workbench and the Green Line Analysis each produced relatively large sets of differentially expressed genes for the two developmental comparisons being considered. Figures 1 through 4 have illustrated the detection comparison between the two platforms. It was surprising to find that while some similarities are shared between the two platforms, the majority of those differentially expressed genes were found in one platform but not the other. A portion of this striking amount of dissimilarity may be attributed to the differing versions of the *Mus musculus* genome to which the sequenced reads were aligned. While feasibly impossible to interpret manually, regression analysis provides a method to interpret how similar the two gene lists are to each other. A GLM is a method of regression analysis that estimates how well a dataset is described by explanatory parameters, also known as independent variables. As part of the output from a GLM analysis, the coefficient of determination (R^2) is calculated for measuring the proportion of variance that is being explained. The greater R^2 is, the greater the fraction of variance or "goodness of fit" being explained by the model (Grafen at Hails, 2002). The comparison of those dually differentially detected genes was weakly similar, with unremarkable R^2 values. As shown in Figure 5, only 70% and 78% of the variation in the data can be explained by the fold changes detected by the two platforms within the virgin-to-pregnant and virgin-to-quiescent developmental comparison, respectively. Important to note, no gene can be found in quadrants II or IV, meaning no gene was differentially detected in a direction of regulation by one platform but detected in a differing direction of regulation by the other platform.

Considering the same input data were used in the analysis performed by the two RNA-seq platforms, these resulting differences and similarities question the validity of transcriptomic data. Several topics concerning technical and biological reproducibility will now be discussed. Reproducibility of biological data is of concern within every scientific discipline. Since the introduction of RNA-seq techniques in 2008, the reproducibility of transcriptomic expression data, particularly that of microarrays, has come under scrutiny (Mortazavi et al., 2008; Ioannidis et al., 2009). In a key evaluation of the repeatability of gene expression profiling from eighteen differing and independent microarray-based studies published between 2005-2006, Ioannidis and colleagues failed to reproduce the reported analyses in principle in sixteen of those studies (Ioannidis et al., 2009). In contrast to this apparent limited repeatability of microarrays, RNA-seq has consistently been demonstrated more repeatable, with few systematic differences among replicates. In a comparison of two technical replicates from isolated mouse brain samples, Mortazavi and colleagues observed the sequenced transcript abundances as being highly reproducible, with a correlation of $R^2 = 0.96$ (Mortazavi et al., 2008). Additionally, to further assess technical variance, Marioni and colleagues compared the RNA-seq results from liver and kidney samples, sequencing each sample seven times, to that of the microarray results from the same samples. Again, while sequenced transcript abundances across the technical replicates were observed as being highly reproducible, with a correlation of $R^2 = 0.96$, comparison of the differential results from two technologies was observed as being less similar, with a correlation of $R^2 = 0.73$. Specific to that study, despite that 6,534 genes were identified as being differentially expressed by both technologies, an additional 4,949 genes were identified through RNA-seq that were not identified through microarray analysis (Marioni et al., 2008). Thus, the strength and repeatability of RNA-seq across technical replicates is far superior to that of microarray hybridization technologies. While such technical replication is significant, the implementation of biological replicates and proper sample size is inarguably essential to the statistical power of any scientific study (Li et al., 2013).

Biological replication and proper sample size are crucial design considerations for the accuracy of any RNA-seq experiment, however their applications can be difficult considering possible financial or technical restrictions (Auer and Doerge, 2010). While analyses on unreplicated data that consider only a single subject per treatment group are not uncommon in the RNA-seq literature (Marioni et al., 2008; Brawand et al., 2001; Graveley et al., 2011; Hah et al., 2011; Soneson and Delorenzi, 2013), they provide no estimation of within-treatment-group variability (Auer and Doerge, 2010). Accordingly, inclusion of biological replicates is desirable as it allows for the estimation of within-treatment-group variability for the comparison to between-group variability and the ultimate generalization to the population of interest (Grafen and Hails, 2002). Researchers have recently begun to address the statistical principle of replication for proper sample size selection as applied to RNA-seq experimental design and analysis. While many equations have been proposed for the determination of the optimal number of biological replicates necessary to achieve a desired statistical power, all agree that replicated data are superior to unreplicated data, where each additional biological replicate increasingly improves the analytical accuracy and power of differential gene detection (Auer and Doerge, 2010; Li et al., 2013; Liu et al., 2014; Hart et al., 2013). Thus, it is possible to consider experiments with smaller numbers of replicates per condition, such as the transcriptional analysis of developmental stages of isolated MEC being considered here, for interpretation of biological insight (Anders and Huber, 2010).

Taken together, these data and findings are in support of successful RNA extraction, library generation, and sequencing application. Quality control analysis of the RNA-seq output through both the CLC Genomics Workbench and the Green Line Analysis further support the individual sequencing processes. Although unfortunate that the sequencing of RNA from Virgin Mouse 2 had to be eliminated from further analyses, the resulting experimental design and biological replicates themselves are sufficient for subsequent gene ontological and pathway analyses. However, there does exist a surprising dissimilarity in the differential expression of the two platforms, as shown by the strikingly small number of dually detected genes. While the detection was similar in direction of regulation, the parity of those dually-detected differentially-expressed genes from each developmental comparison was interestingly found to be weakly similar, as supported by relatively unremarkable R^2 values. The effect these global profiling differences and similarities have on the underlying molecular mechanisms influencing MEC physiology and pathology will be considered in the following experiment, which may also be thought of as the subsequent analytical effort to that just described.

CHAPTER 3 – Identification of Key Regulator Genes Affecting Developmental Stages in Mice using Functional Analysis

3.1 Introduction

The basic output from an RNA-seq experiment is a list of short sequences along with their detected quantities that may then be assessed for quality control, aligned to a reference genome, and analyzed for differential gene expression. Accordingly, differential analysis yields a set of genes showing different average expression levels across two populations. However, the interpretation and extraction of biological insight from such information poses a challenge to researchers as these sets often contain thousands of genes. The recent advances in high-throughput next generation sequencing technologies and the data thus generated have consequently made manual investigations in the literature for analysis and interpretation dauntingly exhaustive (Jiline et al., 2011). Differentially expressed genes can be ordered in a ranked list according to their change in magnitude of expression; yet, individual gene-by-gene analysis often fails to recognize the underlying themes of molecular biology such as cellular processes, metabolic pathways, and transcriptional programs (Subramanian et al., 2005).

To facilitate the secondary analysis of RNA-seq experiments, thereby providing insight into the relevant biological themes, structured vocabularies known as ontologies have been developed for the management of information in biological databases. By providing a centralized collection of known relationships between biological terms and all genes related to those terms, Gene Ontology (GO) databases automate the process of assigning attributes to experimentally-derived, differentially expressed gene sets for biological interpretation (Harris et al., 2004). Ontologies in and of themselves are not a novel concept to scientific applications since historical examples include the anatomical classification of body parts by Aristotle and the chemical classification of the periodic table of elements by Mendeleev (Splendiani et al., 2014). Similar to the numerous commercially available and open-source software packages developed for the initial RNA-seq data analysis, each with its own application of mathematical and statistical algorithms and computer science programming, numerous databases likewise exist for the functional analysis of differential gene sets. Since the launch of the first GO database in 2002, to date more than 70 additional databases have been developed as researchers increasingly depend upon them for the validation of large gene sets with more manageable and well-established sources of knowledge (Kitano, 2002; Huang et al., 2009a). Regardless of their distinctions or differences, all GO databases aim to represent the current knowledge of biological entities and their relationships and to statistically examine the enrichment of ontologies by relevant genes (Hill et al., 2002; Huang et al., 2009a).

Three domains of GO have been proposed to describe the molecular biology concepts universal to all living systems: cellular component, molecular function, and biological process. Cellular component, such as plasma membrane or ubiquitin ligase complex, refers to the area within a cell where a specific gene product is active. Molecular function, such as kinase activity or regulation of transcription, refers to the biochemical activity of a specific gene product. Biological process, such as cell death or oxidation reduction, refers to the biological objective, often a chemical or physical transformation, to which a specific gene product contributes. GO terms are interconnected in a dynamic network of relationships since any biological process is accomplished by one or more assemblies of molecular functions that take place within a designated cellular component (Harris et al., 2004). Furthermore, how closely a set of differentially expressed genes matches an ontology can be quantitatively assessed by determining that which best describes the experimental data. Specifically, the probability that an identified GO term meaningfully relates to the data set is calculated given that the input data set is a random list of genes. The lower the probability, the more likely the GO terms describe the underlying themes of molecular biology for the experimental data being analyzed (Splendiani et al., 2012). Such an analysis is more exploratory in nature, aiming to systematically extract biological meaning from large gene lists as opposed to confirming or refuting theories specific to a particular biological phenomenon (Huang et al., 2009b). To assimilate the down- and up-regulated genes detected by RNA-seq differential analysis between the virgin, pregnant, and post-involutional quiescent samples, GO and pathway analysis will provide a comprehensive illustration of the complex molecular mechanisms influencing MEC physiology and pathology.

3.2 Methods

GO and pathway analysis were performed twice, first on the differentially up- and down-regulated genes detected with high, medium, and low expression by the CLC Genomics Workbench, and second on the differentially up- and down-regulated genes detected with high, medium, and low expression by Green Line Analysis.

3.2.1 Gene Ontology Analysis

Detection of over-represented themes and classification into GO terms was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2009b). The complete mouse transcriptome was used as background to calculate expected frequencies of over-represented themes as directed by the default parameters and singular enrichment analysis (SEA) computations intrinsic to DAVID (Huang et al., 2009a). Specific to the DAVID database, observed frequencies and their associated p-values are calculated to that expected by chance using Benjamini statistics (Huang et al., 2009a). Adopting an exploratory approach to extracting biological meaning, over-represented GO terms were determined among the up- and down-regulated differentially detected genes with high, medium, and low expression between the virgin and pregnant developmental comparison and among the up- and down-regulated differentially detected genes with high, medium, and low expression between the virgin and post-involutional quiescent developmental comparison for both the CLC Genomics Workbench and Green Line Analysis.

3.2.2 Pathway Analysis

To identify the biological pathways significantly enriched in the data sets of the up- and down-regulated differentially detected genes with high, medium, and low expression between the virgin and pregnant developmental comparison and between the virgin and post-involutional quiescent developmental comparison, the corresponding gene data sets were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000).

3.3 Results

3.3.1 CLC Genomics Workbench

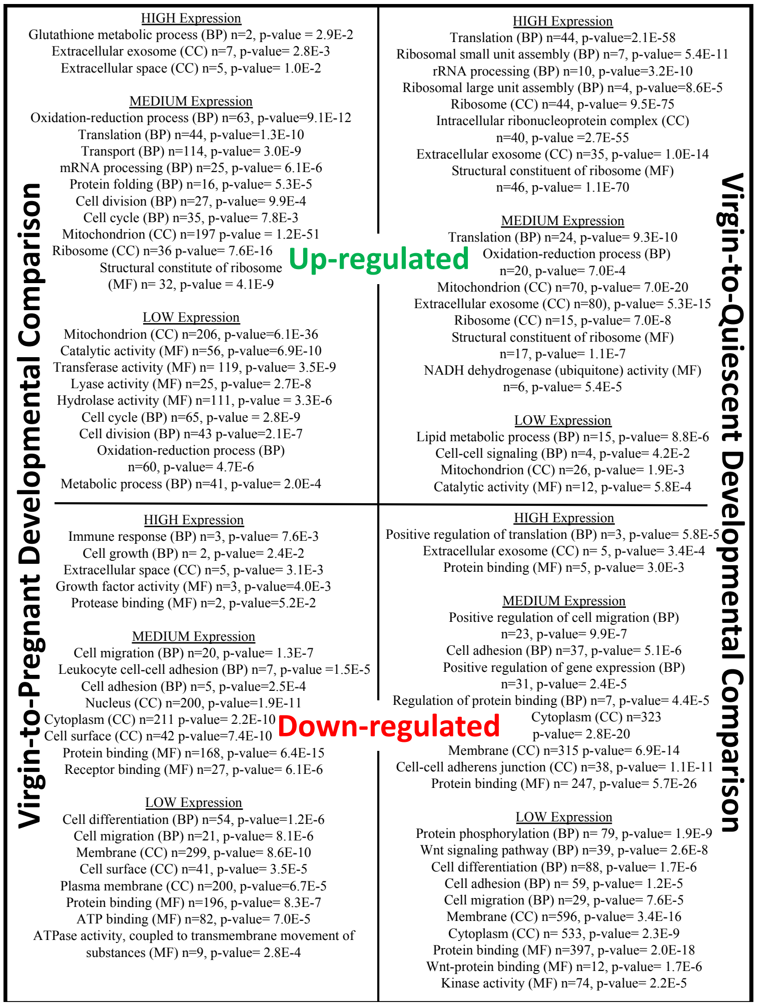
All up- and down-regulated genes differentially detected with high, medium, and low expression between the virgin and pregnant developmental comparison and all up- and down-regulated genes differentially detected with high, medium, and low expression between the virgin and post-involutional quiescent developmental comparison were individually uploaded into DAVID for GO term identification and pathway analysis. The number of high, medium, and low genes for each developmental comparison summed to more than the number of differentially detected genes considering a given gene may be expressed with a specific strength in one developmental state and a different strength in the other developmental state. GO term identification was utilized for a general description of the underlying themes influencing the MEC phenotype. Export from the DAVID database into KEGG was utilized for pathway analysis to provide a comprehensive illustration of the molecular mechanisms influencing the MEC phenotype.

3.3.1.1 Gene Ontology Analysis

For the virgin-to-pregnant developmental comparison, 1,037 differentially down-regulated and 1,644 differentially up-regulated genes were detected by the CLC Genomics Workbench ($p<0.05$). Of the 10 down-regulated genes detected with high expression 3, 4, and 14 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 457 down-regulated genes detected with medium expression, 57, 90, and 293 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 694 down-regulated genes detected with low expression, 64, 79, and 252 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 13 up-regulated genes detected with high expression, 4, 3, and 6 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 744 up-regulated genes detected with medium expression, 76, 62, and 111 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 1,073 up-regulated genes detected with low expression 32, 51, and 86 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively.

For the virgin-to-post-involutional quiescent developmental comparison, 1,820 differentially down-regulated genes and 521 up-regulated genes were detected by the CLC Genomics Workbench ($p<0.05$). Of the 70 down-regulated genes detected with high expression, 5, 3, and 3 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 654 down-regulated genes detected with medium, expression, 118, 102, and 289 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 1,409 down-regulated genes detected with low expression, 93, 94, and 294 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 65 up-regulated genes detected with high expression, 22, 11, and 31 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 315 up-regulated genes detected with medium expression, 55, 33, and 77 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 246 up-regulated genes detected with low expression, 5, 4, and 8 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively.

For each comparison made, considering the biological relevance and associated p-value for that GO term as determined using Benjamini statistics, an exploratory approach was taken to extract biological meaning and consideration for mapping to pathway analysis. A snapshot of applicable and compelling GO terms within each annotation category have been depicted (Figure 6), along with the corresponding number of genes pertaining to that GO term and the calculated p-value.



Virgin-to-Pregnant Developmental Comparison

Up-regulated

HIGH Expression
Glutathione metabolic process (BP) n=2, p-value = 2.9E-2
Extracellular exosome (CC) n=7, p-value= 2.8E-3
Extracellular space (CC) n=5, p-value= 1.0E-2

MEDIUM Expression
Oxidation-reduction process (BP) n=63, p-value=9.1E-12
Translation (BP) n=44, p-value=1.3E-10
Transport (BP) n=114, p-value= 3.0E-9
mRNA processing (BP) n=25, p-value= 6.1E-6
Protein folding (BP) n=16, p-value= 5.3E-5
Cell division (BP) n=27, p-value= 9.9E-4
Cell cycle (BP) n=35, p-value= 7.8E-3
Mitochondrion (CC) n=197 p-value = 1.2E-51
Ribosome (CC) n=36 p-value= 7.6E-16
Structural constitute of ribosome (MF) n= 32, p-value = 4.1E-9

LOW Expression
Mitochondrion (CC) n=206, p-value=6.1E-36
Catalytic activity (MF) n=56, p-value=6.9E-10
Transferase activity (MF) n= 119, p-value= 3.5E-9
Lyase activity (MF) n=25, p-value= 2.7E-8
Hydrolase activity (MF) n=111, p-value = 3.3E-6
Cell cycle (BP) n=65, p-value = 2.8E-9
Cell division (BP) n=43 p-value=2.1E-7
Oxidation-reduction process (BP) n=60, p-value= 4.7E-6
Metabolic process (BP) n=41, p-value= 2.0E-4

Down-regulated

HIGH Expression
Immune response (BP) n=3, p-value= 7.6E-3
Cell growth (BP) n= 2, p-value= 2.4E-2
Extracellular space (CC) n=5, p-value= 3.1E-3
Growth factor activity (MF) n=3, p-value=4.0E-3
Protease binding (MF) n=2, p-value=5.2E-2

MEDIUM Expression
Cell migration (BP) n=20, p-value= 1.3E-7
Leukocyte cell-cell adhesion (BP) n=7, p-value =1.5E-5
Cell adhesion (BP) n=5, p-value=2.5E-4
Nucleus (CC) n=200, p-value=1.9E-11
Cytoplasm (CC) n=211 p-value= 2.2E-10
Cell surface (CC) n=42 p-value=7.4E-10
Protein binding (MF) n=168, p-value= 6.4E-15
Receptor binding (MF) n=27, p-value= 6.1E-6

LOW Expression
Cell differentiation (BP) n=54, p-value=1.2E-6
Cell migration (BP) n=21, p-value= 8.1E-6
Membrane (CC) n=299, p-value= 8.6E-10
Cell surface (CC) n=41, p-value= 3.5E-5
Plasma membrane (CC) n=200, p-value=6.7E-5
Protein binding (MF) n=196, p-value= 8.3E-7
ATP binding (MF) n=82, p-value= 7.0E-5
ATPase activity, coupled to transmembrane movement of substances (MF) n=9, p-value= 2.8E-4

Virgin-to-Quiescent Developmental Comparison

Up-regulated

HIGH Expression
Translation (BP) n=44, p-value=2.1E-58
Ribosomal small unit assembly (BP) n=7, p-value= 5.4E-11
rRNA processing (BP) n=10, p-value=3.2E-10
Ribosomal large unit assembly (BP) n=4, p-value=8.6E-5
Ribosome (CC) n=44, p-value= 9.5E-75
Intracellular ribonucleoprotein complex (CC) n=40, p-value =2.7E-55
Extracellular exosome (CC) n=35, p-value= 1.0E-14
Structural constituent of ribosome (MF) n=46, p-value= 1.1E-70

MEDIUM Expression
Translation (BP) n=24, p-value= 9.3E-10
Oxidation-reduction process (BP) n=20, p-value= 7.0E-4
Mitochondrion (CC) n=70, p-value= 7.0E-20
Extracellular exosome (CC) n=80), p-value= 5.3E-15
Ribosome (CC) n=15, p-value= 7.0E-8
Structural constituent of ribosome (MF) n=17, p-value= 1.1E-7
NADH dehydrogenase (ubiquitone) activity (MF) n=6, p-value= 5.4E-5

LOW Expression
Lipid metabolic process (BP) n=15, p-value= 8.8E-6
Cell-cell signaling (BP) n=4, p-value= 4.2E-2
Mitochondrion (CC) n=26, p-value= 1.9E-3
Catalytic activity (MF) n=12, p-value= 5.8E-4

Down-regulated

HIGH Expression
Positive regulation of translation (BP) n=3, p-value= 5.8E-5
Extracellular exosome (CC) n= 5, p-value= 3.4E-4
Protein binding (MF) n=5, p-value= 3.0E-3

MEDIUM Expression
Positive regulation of cell migration (BP) n=23, p-value= 9.9E-7
Cell adhesion (BP) n=37, p-value= 5.1E-6
Positive regulation of gene expression (BP) n=31, p-value= 2.4E-5
Regulation of protein binding (BP) n=7, p-value= 4.4E-5
Cytoplasm (CC) n=323 p-value= 2.8E-20
Membrane (CC) n=315 p-value= 6.9E-14
Cell-cell adherens junction (CC) n=38, p-value= 1.1E-11
Protein binding (MF) n= 247, p-value= 5.7E-26

LOW Expression
Protein phosphorylation (BP) n= 79, p-value= 1.9E-9
Wnt signaling pathway (BP) n=39, p-value= 2.6E-8
Cell differentiation (BP) n=88, p-value= 1.7E-6
Cell adhesion (BP) n= 59, p-value= 1.2E-5
Cell migration (BP) n=29, p-value= 7.6E-5
Membrane (CC) n=596, p-value= 3.4E-16
Cytoplasm (CC) n= 533, p-value= 2.3E-9
Protein binding (MF) n=397, p-value= 2.0E-18
Wnt-protein binding (MF) n=12, p-value= 1.7E-6
Kinase activity (MF) n=74, p-value= 2.2E-5

FIGURE 6: DAVID GO Terms Originating from CLC Genomics Workbench Differential Analysis of Genes Detected with High, Medium, and Low Expression

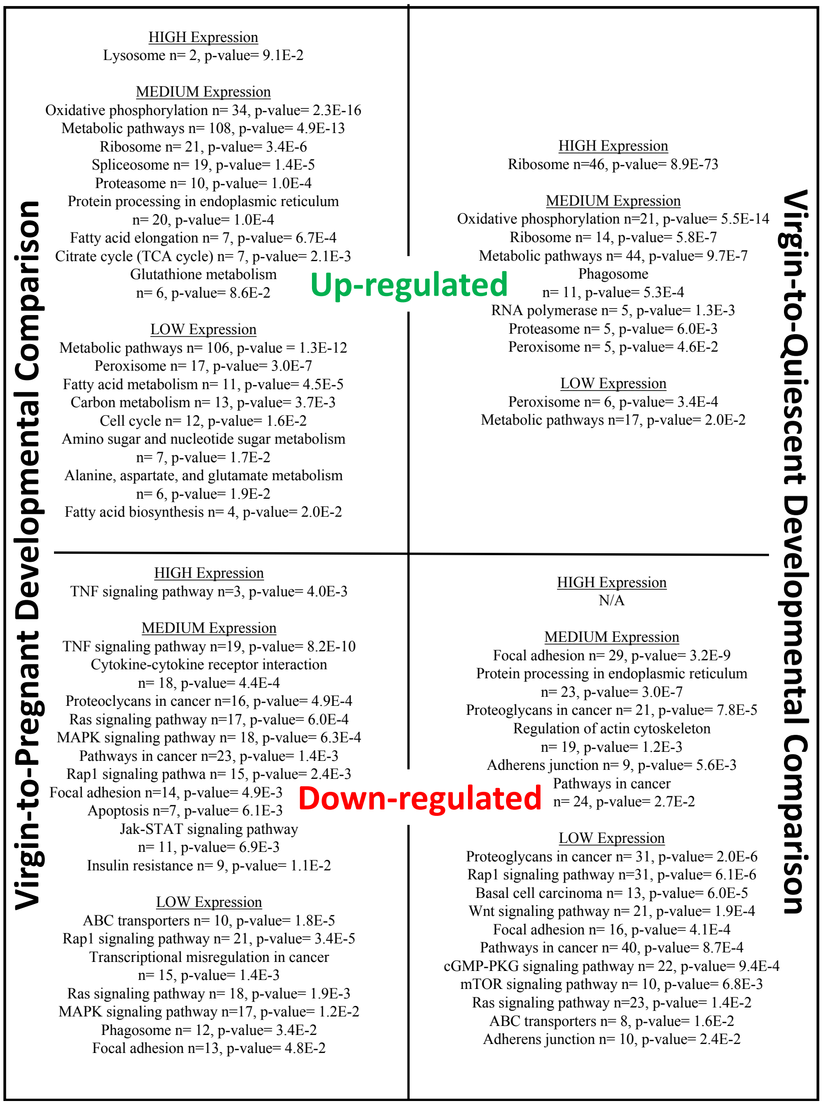
Detection of over-represented themes and classification into GO terms was performed using DAVID. The complete mouse transcriptome was used as background to calculate expected frequencies of over-represented themes as directed by the default parameters. GO terms were determined among genes detected with high, medium, and low differential expression for the virgin-to-pregnant developmental comparison and for the virgin-to-post-involutional quiescent developmental comparison. For each GO term depicted in the snapshot, the number of genes matching that term and the associated p-values are calculated to that expected by chance using Benjamini statistics. The corresponding GO domains representing the molecular biology concepts universal to all living systems are also indicated as either cellular component (CC), molecular function (MF), or biological process (BP).

3.3.1.2 Pathway Analysis

GO term identification was utilized for a general description of the underlying themes influencing the MEC phenotype. Export from the DAVID database into KEGG was utilized for pathway analysis to provide a comprehensive illustration of the molecular mechanisms significantly enriched and influencing the MEC phenotype. Following export into KEGG, for the virgin-to-pregnant developmental comparison, of the 10 down-regulated genes detected with high expression, 6 chart records were identified as compatible within a KEGG pathway. Of the 457 down-regulated genes detected with medium expression, 51 chart records were identified as compatible within a KEGG pathway. Of the 694 genes detected with low expression, 27 chart records were identified as compatible within a KEGG pathway. Of the 13 up-regulated genes detected with high expression, 1 chart record was identified as compatible within a KEGG pathway. Of the 744 up-regulated genes detected with medium expression, 32 chart records were identified as compatible within a KEGG pathway. Of the 1,073 genes detected with low expression, 33 chart records were identified as compatible within a KEGG pathway.

For the virgin-to-post-involutional quiescent developmental comparison, of the 5 down-regulated genes detected with high expression, no chart records were identified as compatible within a KEGG pathway. Of the 654 down-regulated genes detected with medium expression, 23 chart records were identified as compatible within a KEGG pathway. Of the 1,409 genes detected with low expression, 44 chart records were identified as compatible within a KEGG pathway. Of the 65 up-regulated genes detected with high expression, 2 chart records were identified as compatible within a KEGG pathway. Of the 315 up-regulated genes detected with medium expression, 18 chart records were identified as compatible within a KEGG pathway. Of the 246 genes detected with low expression, 4 chart records were identified as compatible within a KEGG pathway.

For each comparison made, considering the biological relevance and associated p-value identified for each pathway as determined using Benjamini statistics, an exploratory approach was taken to identify biological pathways significantly enriched within the data sets. A snapshot of applicable and compelling KEGG pathways have been depicted (Figure 7), along with the corresponding number of genes pertaining to that pathway and the calculated p-value.



Virgin-to-Pregnant Developmental Comparison

Virgin-to-Quiescent Developmental Comparison

Up-regulated

HIGH Expression
Lysosome n= 2, p-value= 9.1E-2

MEDIUM Expression
Oxidative phosphorylation n= 34, p-value= 2.3E-16
Metabolic pathways n= 108, p-value= 4.9E-13
Ribosome n= 21, p-value= 3.4E-6
Spliceosome n= 19, p-value= 1.4E-5
Proteasome n= 10, p-value= 1.0E-4
Protein processing in endoplasmic reticulum
n= 20, p-value= 1.0E-4
Fatty acid elongation n= 7, p-value= 6.7E-4
Citrate cycle (TCA cycle) n= 7, p-value= 2.1E-3
Glutathione metabolism
n= 6, p-value= 8.6E-2

LOW Expression
Metabolic pathways n= 106, p-value = 1.3E-12
Peroxisome n= 17, p-value= 3.0E-7
Fatty acid metabolism n= 11, p-value= 4.5E-5
Carbon metabolism n= 13, p-value= 3.7E-3
Cell cycle n= 12, p-value= 1.6E-2
Amino sugar and nucleotide sugar metabolism
n= 7, p-value= 1.7E-2
Alanine, aspartate, and glutamate metabolism
n= 6, p-value= 1.9E-2
Fatty acid biosynthesis n= 4, p-value= 2.0E-2

HIGH Expression
Ribosome n=46, p-value= 8.9E-73

MEDIUM Expression
Oxidative phosphorylation n=21, p-value= 5.5E-14
Ribosome n= 14, p-value= 5.8E-7
Metabolic pathways n= 44, p-value= 9.7E-7
Phagosome
n= 11, p-value= 5.3E-4
RNA polymerase n= 5, p-value= 1.3E-3
Proteasome n= 5, p-value= 6.0E-3
Peroxisome n= 5, p-value= 4.6E-2

LOW Expression
Peroxisome n= 6, p-value= 3.4E-4
Metabolic pathways n=17, p-value= 2.0E-2

Down-regulated

HIGH Expression
TNF signaling pathway n=3, p-value= 4.0E-3

MEDIUM Expression
TNF signaling pathway n=19, p-value= 8.2E-10
Cytokine-cytokine receptor interaction
n= 18, p-value= 4.4E-4
Proteoclycans in cancer n=16, p-value= 4.9E-4
Ras signaling pathway n=17, p-value= 6.0E-4
MAPK signaling pathway n= 18, p-value= 6.3E-4
Pathways in cancer n=23, p-value= 1.4E-3
Rap1 signaling pathwa n= 15, p-value= 2.4E-3
Focal adhesion n=14, p-value= 4.9E-3
Apoptosis n=7, p-value= 6.1E-3
Jak-STAT signaling pathway
n= 11, p-value= 6.9E-3
Insulin resistance n= 9, p-value= 1.1E-2

LOW Expression
ABC transporters n= 10, p-value= 1.8E-5
Rap1 signaling pathway n= 21, p-value= 3.4E-5
Transcriptional misregulation in cancer
n= 15, p-value= 1.4E-3
Ras signaling pathway n= 18, p-value= 1.9E-3
MAPK signaling pathway n=17, p-value= 1.2E-2
Phagosome n= 12, p-value= 3.4E-2
Focal adhesion n=13, p-value= 4.8E-2

HIGH Expression
N/A

MEDIUM Expression
Focal adhesion n= 29, p-value= 3.2E-9
Protein processing in endoplasmic reticulum
n= 23, p-value= 3.0E-7
Proteoglycans in cancer n= 21, p-value= 7.8E-5
Regulation of actin cytoskeleton
n= 19, p-value= 1.2E-3
Adherens junction n= 9, p-value= 5.6E-3
Pathways in cancer
n= 24, p-value= 2.7E-2

LOW Expression
Proteoglycans in cancer n= 31, p-value= 2.0E-6
Rap1 signaling pathway n=31, p-value= 6.1E-6
Basal cell carcinoma n= 13, p-value= 6.0E-5
Wnt signaling pathway n= 21, p-value= 1.9E-4
Focal adhesion n= 16, p-value= 4.1E-4
Pathways in cancer n= 40, p-value= 8.7E-4
cGMP-PKG signaling pathway n= 22, p-value= 9.4E-4
mTOR signaling pathway n= 10, p-value= 6.8E-3
Ras signaling pathway n=23, p-value= 1.4E-2
ABC transporters n= 8, p-value= 1.6E-2
Adherens junction n= 10, p-value= 2.4E-2

FIGURE 7: KEGG Pathway Records Originating from CLC Genomics Workbench Differential Analysis of Genes Detected with High, Medium, and Low Expression

To identify the biological pathways significantly enriched in the data sets of differentially down- and up-regulated genes between virgin and pregnant samples and between virgin and post-involutional quiescent samples as detected by the CLC Genomics Workbench, genes were mapped to KEGG directly from DAVID. For each pathway identified, the number of genes matching that pathway and the percentage of those genes matching that pathway as determined from the number of differentially detected input genes have been listed. The associated p-values are calculated to that expected by chance using Benjamini statistics.

3.3.2 Green Line Analysis

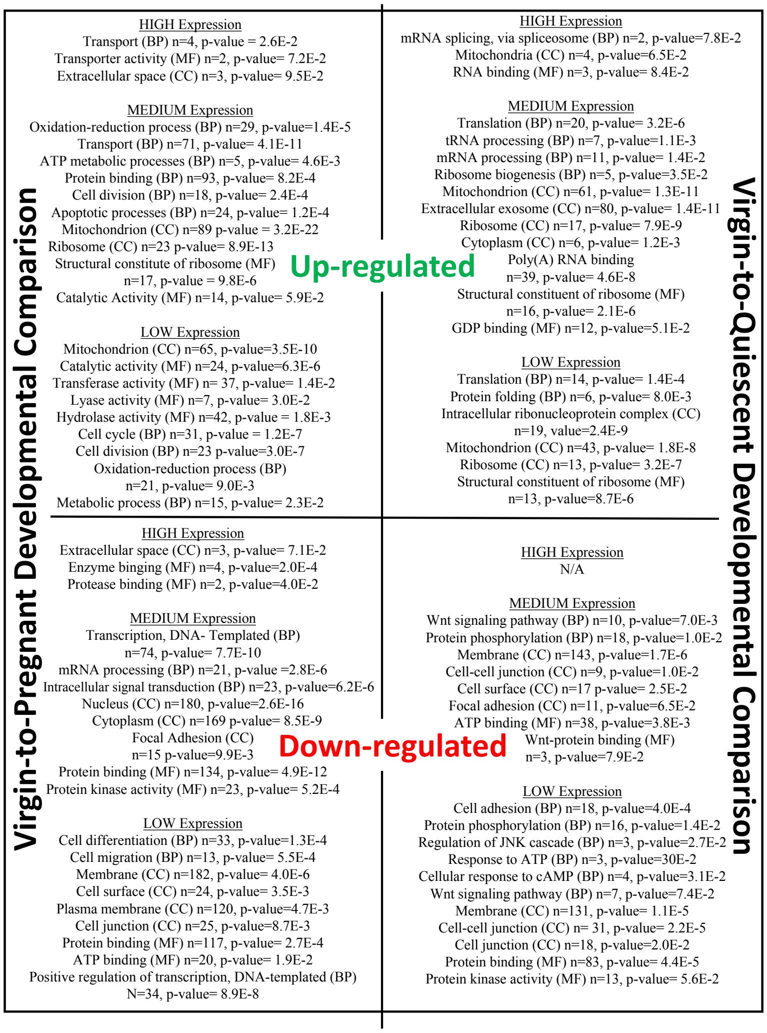
All up- and down-regulated genes differentially detected with high, medium, and low expression between the virgin and pregnant developmental comparison and all up- and down-regulated genes differentially detected with high, medium, and low expression between the virgin and post-involutional quiescent developmental comparison were individually uploaded into DAVID for GO term identification and pathway analysis. The number of high, medium, and low genes for each developmental comparison summed to more than the number of differentially detected genes considering a given gene may be expressed with a specific strength in one developmental state and a different strength in the other developmental state. GO term identification was utilized for a general description of the underlying themes influencing the MEC phenotype. Export from the DAVID database into KEGG was utilized for pathway analysis to provide a comprehensive illustration of the molecular mechanisms influencing the MEC phenotype.

3.3.2.1 Gene Ontology Analysis

For the virgin-to-pregnant developmental comparison, 756 differentially down-regulated and 715 differentially up-regulated genes were detected by the Green Line Analysis ($p<0.05$). Of the 8 down-regulated genes detected with high expression 3, 2, and 3 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 372 down-regulated genes detected with medium expression, 43, 65, and 170 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 694 down-regulated genes detected with low expression, 33, 49, and 121 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 8 up-regulated genes detected with high expression, 1, 1, and 2 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 369 up-regulated genes detected with medium expression, 741 32, and 70 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 442 up-regulated genes detected with low expression 20, 17, and 28 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively.

For the virgin-to-post-involutional quiescent developmental comparison, 725 differentially down-regulated genes and 667 up-regulated genes were detected by the CLC Genomics Workbench ($p<0.05$). Of the 13 down-regulated genes detected with high expression, no ontological records were identified for cellular component, molecular function, or biological process functional annotation terms. Of the 340 down-regulated genes detected with medium, expression, 37, 25, and 105 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 315 down-regulated genes detected with low expression, 32, 22, and 66 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 14 up-regulated genes detected with high expression, 2, 1, and 1 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 332 up-regulated genes detected with medium expression, 32, 26, and 38 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively. Of the 249 up-regulated genes detected with low expression, 33, 28, and 29 ontological records were identified for cellular component, molecular function, and biological process functional annotation terms, respectively.

For each comparison made, considering the biological relevance and associated p-value for that GO term as determined using Benjamini statistics, an exploratory approach was taken to extract biological meaning and consideration for mapping to pathway analysis. A snapshot of applicable and compelling GO terms within each annotation category have been depicted (Figure 8), along with the corresponding number of genes pertaining to that GO term and the calculated p-value.



Virgin-to-Pregnant Developmental Comparison

Up-regulated

HIGH Expression
Transport (BP) n=4, p-value = 2.6E-2
Transporter activity (MF) n=2, p-value= 7.2E-2
Extracellular space (CC) n=3, p-value= 9.5E-2

MEDIUM Expression
Oxidation-reduction process (BP) n=29, p-value=1.4E-5
Transport (BP) n=71, p-value= 4.1E-11
ATP metabolic processes (BP) n=5, p-value= 4.6E-3
Protein binding (BP) n=93, p-value= 8.2E-4
Cell division (BP) n=18, p-value= 2.4E-4
Apoptotic processes (BP) n=24, p-value= 1.2E-4
Mitochondrion (CC) n=89 p-value = 3.2E-22
Ribosome (CC) n=23 p-value= 8.9E-13
Structural constitute of ribosome (MF)
n=17, p-value = 9.8E-6
Catalytic Activity (MF) n=14, p-value= 5.9E-2

LOW Expression
Mitochondrion (CC) n=65, p-value=3.5E-10
Catalytic activity (MF) n=24, p-value=6.3E-6
Transferase activity (MF) n= 37, p-value= 1.4E-2
Lyase activity (MF) n=7, p-value= 3.0E-2
Hydrolase activity (MF) n=42, p-value = 1.8E-3
Cell cycle (BP) n=31, p-value = 1.2E-7
Cell division (BP) n=23 p-value=3.0E-7
Oxidation-reduction process (BP)
n=21, p-value= 9.0E-3
Metabolic process (BP) n=15, p-value= 2.3E-2

HIGH Expression
mRNA splicing, via spliceosome (BP) n=2, p-value=7.8E-2
Mitochondria (CC) n=4, p-value=6.5E-2
RNA binding (MF) n=3, p-value= 8.4E-2

MEDIUM Expression
Translation (BP) n=20, p-value= 3.2E-6
tRNA processing (BP) n=7, p-value=1.1E-3
mRNA processing (BP) n=11, p-value= 1.4E-2
Ribosome biogenesis (BP) n=5, p-value=3.5E-2
Mitochondrion (CC) n=61, p-value= 1.3E-11
Extracellular exosome (CC) n=80, p-value= 1.4E-11
Ribosome (CC) n=17, p-value= 7.9E-9
Cytoplasm (CC) n=6, p-value= 1.2E-3
Poly(A) RNA binding
n=39, p-value= 4.6E-8
Structural constituent of ribosome (MF)
n=16, p-value= 2.1E-6
GDP binding (MF) n=12, p-value=5.1E-2

LOW Expression
Translation (BP) n=14, p-value= 1.4E-4
Protein folding (BP) n=6, p-value= 8.0E-3
Intracellular ribonucleoprotein complex (CC)
n=19, value=2.4E-9
Mitochondrion (CC) n=43, p-value= 1.8E-8
Ribosome (CC) n=13, p-value= 3.2E-7
Structural constituent of ribosome (MF)
n=13, p-value=8.7E-6

Down-regulated

HIGH Expression
Extracellular space (CC) n=3, p-value= 7.1E-2
Enzyme binging (MF) n=4, p-value=2.0E-4
Protease binding (MF) n=2, p-value=4.0E-2

MEDIUM Expression
Transcription, DNA- Templated (BP)
n=74, p-value= 7.7E-10
mRNA processing (BP) n=21, p-value =2.8E-6
Intracellular signal transduction (BP) n=23, p-value=6.2E-6
Nucleus (CC) n=180, p-value=2.6E-16
Cytoplasm (CC) n=169 p-value= 8.5E-9
Focal Adhesion (CC)
n=15 p-value=9.9E-3
Protein binding (MF) n=134, p-value= 4.9E-12
Protein kinase activity (MF) n=23, p-value= 5.2E-4

LOW Expression
Cell differentiation (BP) n=33, p-value=1.3E-4
Cell migration (BP) n=13, p-value= 5.5E-4
Membrane (CC) n=182, p-value= 4.0E-6
Cell surface (CC) n=24, p-value= 3.5E-3
Plasma membrane (CC) n=120, p-value=4.7E-3
Cell junction (CC) n=25, p-value=8.7E-3
Protein binding (MF) n=117, p-value= 2.7E-4
ATP binding (MF) n=20, p-value= 1.9E-2
Positive regulation of transcription, DNA-templated (BP)
N=34, p-value= 8.9E-8

HIGH Expression
N/A

MEDIUM Expression
Wnt signaling pathway (BP) n=10, p-value=7.0E-3
Protein phosphorylation (BP) n=18, p-value=1.0E-2
Membrane (CC) n=143, p-value=1.7E-6
Cell-cell junction (CC) n=9, p-value=1.0E-2
Cell surface (CC) n=17 p-value= 2.5E-2
Focal adhesion (CC) n=11, p-value=6.5E-2
ATP binding (MF) n=38, p-value=3.8E-3
Wnt-protein binding (MF)
n=3, p-value=7.9E-2

LOW Expression
Cell adhesion (BP) n=18, p-value=4.0E-4
Protein phosphorylation (BP) n=16, p-value=1.4E-2
Regulation of JNK cascade (BP) n=3, p-value=2.7E-2
Response to ATP (BP) n=3, p-value=30E-2
Cellular response to cAMP (BP) n=4, p-value=3.1E-2
Wnt signaling pathway (BP) n=7, p-value=7.4E-2
Membrane (CC) n=131, p-value= 1.1E-5
Cell-cell junction (CC) n= 31, p-value= 2.2E-5
Cell junction (CC) n=18, p-value=2.0E-2
Protein binding (MF) n=83, p-value= 4.4E-5
Protein kinase activity (MF) n=13, p-value= 5.6E-2

Virgin-to-Quiescent Developmental Comparison

FIGURE 8: DAVID GO Terms Originating from Green Line Analysis Differential Analysis of Genes Detected with High, Medium, and Low Expression

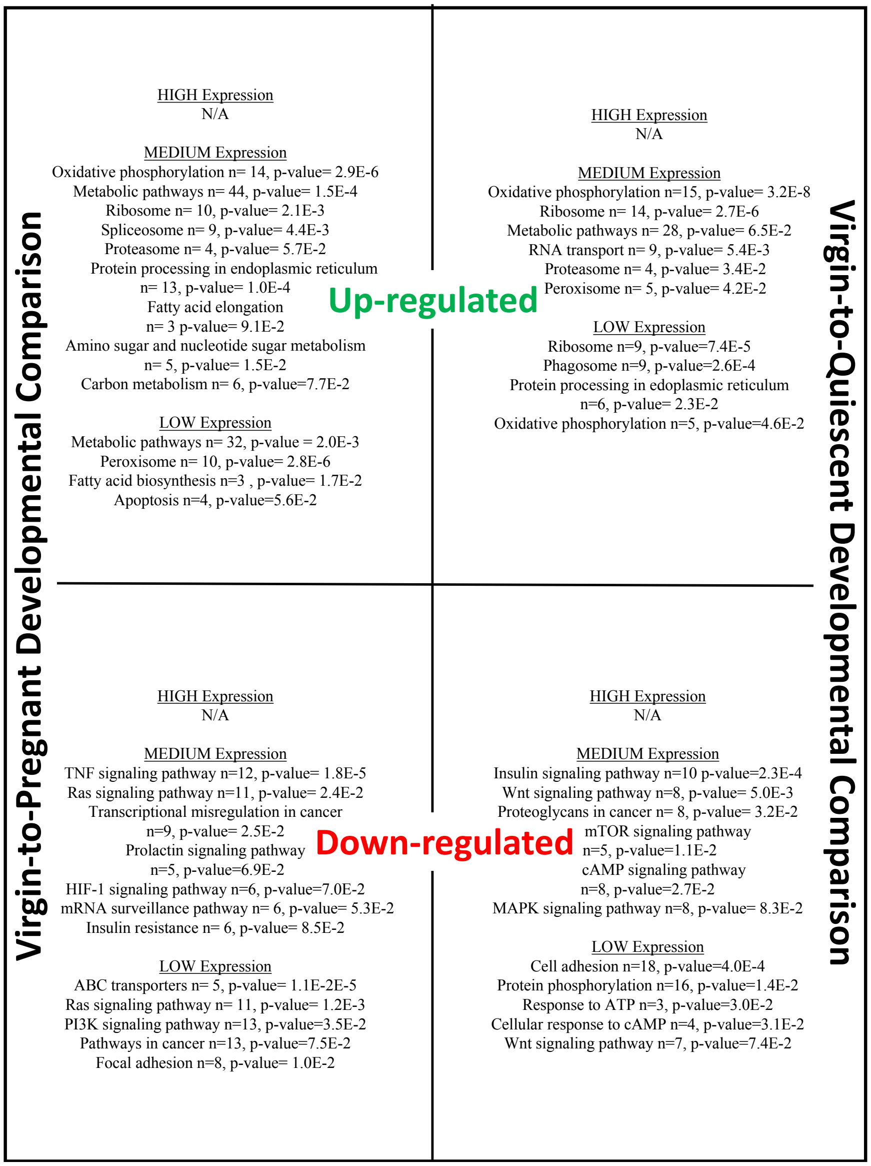
Detection of over-represented themes and classification into GO terms was performed using DAVID. The complete mouse transcriptome was used as background to calculate expected frequencies of over-represented themes as directed by the default parameters. GO terms were determined among genes detected with high, medium, and low differential expression for the virgin-to-pregnant developmental comparison and for the virgin-to-post-involutional quiescent developmental comparison. For each GO term depicted in the snapshot, the number of genes matching that term and the associated p-values are calculated to that expected by chance using Benjamini statistics. The corresponding GO domains representing the molecular biology concepts universal to all living systems are also indicated as either cellular component (CC), molecular function (MF), or biological process (BP).

3.3.2.2 Pathway Analysis

GO term identification was utilized for a general description of the underlying themes influencing the MEC phenotype. Export from the DAVID database into KEGG was utilized for pathway analysis to provide a comprehensive illustration of the molecular mechanisms significantly enriched and influencing the MEC phenotype. Following export into KEGG, for the virgin-to-pregnant developmental comparison, of the 8 down-regulated genes detected with high expression, 1 chart record was identified as compatible within a KEGG pathway. Of the 372 down-regulated genes detected with medium expression, 33 chart records were identified as compatible within a KEGG pathway. Of the 694 genes detected with low expression, 14 chart records were identified as compatible within a KEGG pathway. Of the 8 up-regulated genes detected with high expression, no chart records were identified as compatible within a KEGG pathway. Of the 369 up-regulated genes detected with medium expression, 26 chart records were identified as compatible within a KEGG pathway. Of the 422 genes detected with low expression, 9 chart records were identified as compatible within a KEGG pathway.

For the virgin-to-post-involutional quiescent developmental comparison, of the 13 down-regulated genes detected with high expression, no chart records were identified as compatible within a KEGG pathway. Of the 340 down-regulated genes detected with medium expression, 13 chart records were identified as compatible within a KEGG pathway. Of the 315 genes detected with low expression, 3 chart records were identified as compatible within a KEGG pathway. Of the 14 up-regulated genes detected with high expression, no chart records were identified as compatible within a KEGG pathway. Of the 332 up-regulated genes detected with medium expression, 13 chart records were identified as compatible within a KEGG pathway. Of the 249 genes detected with low expression, 12 chart records were identified as compatible within a KEGG pathway.

For each comparison made, considering the biological relevance and associated p-value identified for each pathway as determined using Benjamini statistics, an exploratory approach was taken to identify biological pathways significantly enriched within the data sets. A snapshot of applicable and compelling KEGG pathways have been depicted (Figure 9), along with the corresponding number of genes pertaining to that pathway and the calculated p-value.


Virgin-to-Pregnant Developmental Comparison
Virgin-to-Quiescent Developmental Comparison
Up-regulated
Down-regulated

HIGH Expression
N/A

MEDIUM Expression
Oxidative phosphorylation n= 14, p-value= 2.9E-6
Metabolic pathways n= 44, p-value= 1.5E-4
Ribosome n= 10, p-value= 2.1E-3
Spliceosome n= 9, p-value= 4.4E-3
Proteasome n= 4, p-value= 5.7E-2
Protein processing in endoplasmic reticulum
n= 13, p-value= 1.0E-4
Fatty acid elongation
n= 3 p-value= 9.1E-2
Amino sugar and nucleotide sugar metabolism
n= 5, p-value= 1.5E-2
Carbon metabolism n= 6, p-value=7.7E-2

LOW Expression
Metabolic pathways n= 32, p-value = 2.0E-3
Peroxisome n= 10, p-value= 2.8E-6
Fatty acid biosynthesis n=3 , p-value= 1.7E-2
Apoptosis n=4, p-value=5.6E-2

HIGH Expression
N/A

MEDIUM Expression
Oxidative phosphorylation n=15, p-value= 3.2E-8
Ribosome n= 14, p-value= 2.7E-6
Metabolic pathways n= 28, p-value= 6.5E-2
RNA transport n= 9, p-value= 5.4E-3
Proteasome n= 4, p-value= 3.4E-2
Peroxisome n= 5, p-value= 4.2E-2

LOW Expression
Ribosome n=9, p-value=7.4E-5
Phagosome n=9, p-value=2.6E-4
Protein processing in edoplasmic reticulum
n=6, p-value= 2.3E-2
Oxidative phosphorylation n=5, p-value=4.6E-2

HIGH Expression
N/A

MEDIUM Expression
TNF signaling pathway n=12, p-value= 1.8E-5
Ras signaling pathway n=11, p-value= 2.4E-2
Transcriptional misregulation in cancer
n=9, p-value= 2.5E-2
Prolactin signaling pathway
n=5, p-value=6.9E-2
HIF-1 signaling pathway n=6, p-value=7.0E-2
mRNA surveillance pathway n= 6, p-value= 5.3E-2
Insulin resistance n= 6, p-value= 8.5E-2

LOW Expression
ABC transporters n= 5, p-value= 1.1E-2E-5
Ras signaling pathway n= 11, p-value= 1.2E-3
PI3K signaling pathway n=13, p-value=3.5E-2
Pathways in cancer n=13, p-value=7.5E-2
Focal adhesion n=8, p-value= 1.0E-2

HIGH Expression
N/A

MEDIUM Expression
Insulin signaling pathway n=10 p-value=2.3E-4
Wnt signaling pathway n=8, p-value= 5.0E-3
Proteoglycans in cancer n= 8, p-value= 3.2E-2
mTOR signaling pathway
n=5, p-value=1.1E-2
cAMP signaling pathway
n=8, p-value=2.7E-2
MAPK signaling pathway n=8, p-value= 8.3E-2

LOW Expression
Cell adhesion n=18, p-value=4.0E-4
Protein phosphorylation n=16, p-value=1.4E-2
Response to ATP n=3, p-value=3.0E-2
Cellular response to cAMP n=4, p-value=3.1E-2
Wnt signaling pathway n=7, p-value=7.4E-2

FIGURE 9: KEGG Pathway Records Originating from Green Line Analysis Differential Analysis of Genes Detected with High, Medium, and Low Expression

To identify the biological pathways significantly enriched in the data sets of differentially down- and up-regulated genes between virgin and pregnant samples and between virgin and post-involutional quiescent samples as detected by the Green Line Analysis, genes were mapped to KEGG directly from DAVID. For each pathway identified, the number of genes matching that pathway and the percentage of those genes matching that pathway as determined from the number of differentially detected input genes have been listed. The associated p-values are calculated to that expected by chance using Benjamini statistics.

3.4 Discussion

In this experiment, which may be thought of as the subsequent analytical effort to the prior experiment, GO and pathway analysis were utilized to provide functional characterization and give biological meaning to the RNA-seq analysis of MEC isolated from virgin, pregnant, and post-lactational quiescent sibling mice. While DAVID interpreted the large differential data sets to provide relevant GO terms that identify the underlying themes of molecular biology, KEGG pathway analysis provided a graphical representation of the molecular systems that govern cellular processes and organism behavior (Huang et al., 2009a; Kanehisa et al., 2010).

Functional analysis for biological meaning was first performed for the commercially-available CLC Genomics Workbench and secondly performed for the publicly-available Green Line Analysis. Despite the differences in the differential gene lists generated by these two transciptomic platforms (discussed in the prior chapter), the overall depiction of the underlying themes and relevant pathways were surprisingly similar. Not only were numerous domains of gene ontology similarly represented within the differential gene lists generated by the CLC Genomics Workbench and the Green Line Analysis, but pathway analysis also produced similar representations of the molecular systems underlying the MEC phenotype. As it appears, in spite of the individual differences within the gene sets generated, the overall contribution and collaboration of those genes functioning together within each developmental stage analyzed are reflected and made apparent through systems biology. Therefore, the interpretations of these individual platform analyses may be considered simultaneously across the developmental comparison being made. The discussions that ensue are specific to those gleaned from the interpretations of the CLC Genomics Workbench to explain the possible events occurring at the molecular level in the isolated MEC.

Several considerations should be kept in mind when analyzing the global transcriptomic profiling of the isolated primary MEC. First, the focus of this experiment is to interpret those identified potential factors affecting mammary gland physiology and pathology by examining the transcriptomic global profiles of isolated MEC. Thus, the overall question being asked is *what changed*? To identify the underlying biological processes most pertinent to the biological phenotype being considered, the resulting analytical interpretations are based on all relevant differentially-detected genes instead of on a smaller set of restricted genes (Huang et al., 2009a). Second, the number of genes within a data set that are recognized as relating to a particular GO term or pathway do not directly affect the corresponding statistical significance of that term or pathway in describing the phenotype. Rather, the probability that an identified GO term meaningfully relates to the data set is calculated (Splendiani et al., 2012). This explains why some GO terms and pathways can have the same number of related genes yet differing calculated Benjamini statistics and p-values and vice versa. Third, it is assumed that all primary MEC analyzed were isolated luminal epithelial cells. Although differing buffer components, incubations, and washes were utilized in the MEC isolation protocol to limit the presence of fibroblasts, adipocytes, and erythrocytes, nevertheless there exists the potential for other cell type contamination, namely myoepithelial cells, endothelial cells, and leukocytes. Last, the estrous cycle was neither monitored nor considered in those mice from which the virgin and post-lactational quiescent mammary glands were harvested. As discussed in Chapter 1 of this thesis, while the mammary gland of the virgin and post-lactation female is relatively quiescent, minute morphological changes do take place in response to the cyclic endocrine regulation. This manifests primarily as a transitory appearance of alveolar buds that develop and regress in accordance to the four-to-five day murine estrous cycle. Thus, while there are undoubtedly specific effects the estrous cycle had on those MEC isolated from the virgin and post-lactational quiescent mice, it cannot be stated definitely what those effects are in relation to the subsequent RNA-seq and differential analyses performed. With these considerations in mind, a discussion of the down- and up-regulated profiling for both the virgin-to-pregnant and the virgin-to-quiescent comparisons and the possible events occurring at the molecular level in the isolated MEC will follow.

Virgin vs. Pregnant Comparison

Dramatic changes in cell composition and function occur in the mammary gland during pregnancy. Proper morphogenesis of a functional mammary gland is dependent upon the coordination of endocrine induction, signaling pathways, and molecular mediators to direct the extensive proliferation and then secretory differentiation of alveolar units capable of milk secretion. Of those genes detected as being differentially up-regulated from the virgin-to-pregnant developmental comparison, KEGG analysis identified two encompassing biological themes. First, there was enrichment for numerous metabolic pathways such as oxidative phosphorylation, citrate cycle, fatty acid metabolism, amino sugar and nucleotide sugar metabolism, and fatty acid biosynthesis. Second, pathways pertaining to the cell cycle and proliferation were likewise depicted as being enriched, as identified by those of the lysosome, ribosome, spliceosome, proteasome, protein processing, peroxisome, and cell cycle (Table 7).

Considering those pathways pertaining to cellular metabolism, GO analysis for cellular component strongly identified the mitochondria as the location where gene products were up-regulated in pregnant MEC relative to virgin MEC. GO analysis for molecular function and biological process likewise identified oxidation-reduction processes, metabolic processes, and translation as those biological objectives supporting cellular metabolism. Considering those pathways pertaining to cellular proliferation, GO analysis for biological process identified cell division and cell cycle as up-regulated in pregnant MEC relative to virgin MEC. GO analysis for molecular function identified both lyase and transferace activity as up-regulated in pregnant MEC relative to virgin MEC (Table 6).

These GO and KEGG results are in agreement with recent transcriptomic analyses on mammary gland development. Studies by Zhou and colleagues found that cells of the mammary gland from mice at day 12 of pregnancy were highly activated for pathways related to cell cycle control and proliferation (Zhou et al., 2014). From a metabolic standpoint, while glucose is required for the synthesis of lactose in the lactating mammary gland, cells of the pregnant mammary gland generally utilize glucose for the production of ATP through oxidative phosphorylation, which occurs in the mitochondrion. Accelerated metabolic processes ensure sufficient ATP and metabolic intermediaries that are essential for macromolecule biosynthesis compatible with the demands of cell growth and proliferation (Anderson et al., 2007). When exploring possible links between mitochondrial physiology and tumor cell maintenance, Fatin and colleagues have proposed lower oxidative phosphorylation processes are characteristic of carcinoma cells and are thought to result from dissemination of the mitochondrial proton gradient and resulting general inability of tumor cells to use mitochondria to meet their energetic needs (Fatin et al., 2006). It is therefore exciting to note that enriched mitochondrial processes not only make sense from a cellular proliferation standpoint, but also hold potential as a factor influencing the association of parity and protection against breast cancer. Although outside the scope of this global transcriptomic analysis, future efforts should be focused on the expression levels of individual mitochondrial gene products to further understand the cellular mechanisms occurring within MEC in preparation for milk synthesis and parity-induced protection.

Interestingly, the glutathione metabolic process was identified as an up-regulated biological objective in pregnant MEC relative to virgin MEC. Although glutathione is not coded for by a gene, it is an antioxidant associated with lactation as its deregulation has been documented in various pathologies (Zaragoza et al., 2015). Further investigation found that glutathione utilization is understood to play an indirect role in protein synthesis within MEC, and its decrease leads to apoptosis and involution of the mammary gland (Zaragoza et al., 2003). Glutathione is thought to be initially hydrolyzed extracellularly and further hydrolyzed intracellularly by nonspecific peptidases. The resulting constituents are then available for either synthesis of milk proteins or re-synthesis into intracellular glutathione (Baumrucker, 1985). Although glutathione is just one antioxidant influencing amino acid abundance, the availability of those amino acids within MEC is regarded as the limiting factor in the synthesis and secretion of milk proteins (Boisgard et al., 2001). Subsequent research regarding the specific mechanisms through which the glutathione metabolism influences the lactational capacity would therefore prove informative.

That pathways indicative of cellular proliferation were differentially up-regulated from the virgin-to-pregnant developmental comparison was expected, as the alveolar morphogenesis observed during pregnancy is orchestrated by progesterone (P) and prolactin (PRL) and marked by functional differentiation and proliferation. However, that Wnt4 and RankL signaling pathways specifically were not detected by KEGG in this developmental comparison was surprising, as they have been previously proposed to serve as the progesterone-induced mediators of alveolar differentiation (Oakes et al., 2006; Tanos et al., 2013). MEC were isolated from pregnant mice on day 10 of pregnancy, at which point pathways of the ribosome, splicosome, and proteasome were identified as those enriched in the virgin-to-pregnant developmental comparison. A possible explanation for this observation could be that the hormonal regulation of Wnt4 and RankL signaling pathways has already occurred by day 10 of pregnancy, resulting in those enriched pathways indicative of differentiation and proliferation.

The synthesis of proteins is important not only for proper cell function and survival but is also responsible for the generation of those secreted in the milk. Through the use of genetic information carried by RNA molecules, ribosomes execute the synthesis of proteins (Bionaz and Loor, 2011; Alberts et al., 2008). In contrast, the proteasome is a complex containing proteases that serve to degrade unneeded or damaged proteins tagged by ubiquitin (Alberts et al., 2008). Utilizing a microarray approach to study gene expression in murine mammary glands during day 12 of pregnancy, Rudolph and colleagues found that while milk protein gene expression increased throughout pregnancy, proteasomal gene expression correspondingly declined (Rudolph et al., 2003). Rudolph and colleagues proposed this decrease in proteasomal expression is a functional adaptation to conserve biosynthetic processes activated during lactation. Although the current RNA sequencing experiment found pathways of the proteasome up-regulated in MEC isolated from mice at day 10 of pregnancy compared to MEC isolated from virgin mice, perhaps this phenomenon is explained by the timing-specific transition into lactation initiated by P withdrawal in the presence of PRL and glucocorticoid (Anderson et al., 2007). Specific to murine mammary glands, PRL rapidly spikes as P decreases just prior to parturition (Neville et al., 2002). Thus, future efforts to better understand the molecular mechanisms of protein synthesis, degradation, and their influence within bovine MEC hold potential for capitalization on milk production in the dairy industry.

Still considering the virgin-to-pregnant developmental comparison, it is intriguing to note that cellular components such as the lysosome were not found to be similarly up-regulated within this developmental comparison, and yet pathways pertaining to the lysosme and ribosome were. While lysosomes are known to be involved in involution of the mammary gland, only recently have they been observed to be upregulated during lactation (Zhou et al., 2014). With the enriched pathways of the lysosome in mind, perhaps the mediators and inhibitors of the lysosome itself would be of clinical interest in future applications specific to the lactational capacity of the mammary gland. Furthermore, lysosomal proteins have been previously found to be activated during lactation and involution, suggesting that their function and dysfunction might also influence the prevalence of breast cancer (Boya, 2012).

Of those genes detected as being differentially down-regulated from the virgin-to-pregnant developmental comparison, KEGG analysis identified the biological theme of decreased <u>cell-cell communication and interaction</u>. Signaling and communicative pathways such as the TNF, Ras, MAPK, Rap1, Jak-STAT, cytokine-cytokine receptor interaction, focal adhesion, and ABC transporters were all depicted as being down-regulated (Table 7). Specific to these pathways, GO analysis for cellular component strongly identified the cell surface, cell membrane, and extracellular space as the locations where gene products were down-regulated in pregnant MEC relative to virgin MEC. GO analysis for molecular function and biological process likewise identified an observed decrease in immune response, cell adhesion, receptor binding, cell migration, and the transmembrane movement of substances (Table 6).

That a decrease in cellular communication was observed is surprising considering several prior *in vitro* experiments on the development and differentiation of the mammary gland. In their study of gap junctional communication in the CID-9 mammary cell line, El-Sabban and colleagues demonstrated the importance of cell-cell communication in mammary epithelial differentiation, where interactions with the ECM alone were unable to induce a differentiated phenotype (El-Sabban, 2003). A similar dependence on established cell-cell communication for optimal differentiation has also been described in the cell lines of epidermal cells (Alford and Rannels, 2001), lung alveolar cells (Alford and Rannels, 2001), and osteoblasts (Romanello et al., 2001). Studies within the MC3T3-E1 cell line likewise have shown that intercellular communication is integral to the development and differentiation of the mammary gland (McLachlan et al., 2007). Considering the current experiment utilized uncultured primary cells, perhaps the observed decrease in cellular communicative and interactive GO and KEGG terms highlights the idea that cellular communication may not be essential at all stages of MEC development. Further analysis on additional interactive and signaling mechanisms such as integrins and/or connexins is needed to determine if they compensate during this observed down-regulated phenomenon.

In contrast, the observed decrease in immune response is in agreement with analyses on both murine and bovine mammary gland development (Mowry et al., 2017; Mallard et al., 1998). While the increased metabolic demands within MEC is thought to partition energy resources that functionally enhance potential milk production, it does so by reducing those demands of immune responsiveness (Mowry et al., 2017). Pregnancy is typically characterized by an immune-tolerant microenvironment, and secretion of extracellular vesicles with immunosuppressant activities is increased in the pregnant state (Becker et al., 2016). Considering the immune response functions to provide a defense against invading pathogenic organisms, understanding the impact and influence of its down-regulation is therefore of economic relevance for future implications within the dairy industry. Mastitis often manifests during lactation and early involution, following the failure of macrophages to phagocytize the debris collected in the vasculature enveloping the alveolar secretory units. Specific to the bovine mammary gland, the prevalence of mastitis has numerous negative implications, including but not limited to reduced milk yield, increased antibiotic use, and premature culling (Pai and Horseman, 2011). In a microarray study by Clarkson and colleagues, gene expression profiling on murine mammary tissue found that the majority of those genes differentially detected following involution were also differentially detected during pregnancy (Clarkson et al., 2003). Thus, the immune response observed during pregnancy is influential upon the susceptibility to mastitis during lactation, and the mechanisms through which MEC govern this biological process holds potential for selection strategies that best support initial mammary gland development and subsequent milk production.

In summary, pregnancy is the developmental stage that enhances the lactational capacity. Development of the mammary gland during pregnancy occurs in two distinct phases—proliferation then differentiation. Taken together, these analyses identified significant differences in the functions performed by those genes with down- and up-regulated expression from the virgin-to-pregnant comparison of isolated MEC. Genes involved in cell-cell communication and interaction showed a decrease in expression in the pregnant state compared to that of the virgin. Yet for the same developmental comparison genes involved in metabolism and proliferation showed an increase in expression. This indicates that half-way through pregnancy, MEC are enhancing their mitochondrial functioning for energy production in preparation for milk synthesis and lactation, with differentiation and formation of the alveolus not yet occurring. The economic implications of mammary gland development and subsequent milk development would thus benefit from future studies restricted to those molecular systems that enhance mitochondrial processes, amino acid availability and utilization, ribosomal functioning, cellular communication, and the immune response within MEC.

Virgin vs. Post-Lactational Quiescent Comparison

Following the completion of lactation, involution is an essential process that returns the mammary gland to its pre-pregnant state. The morphology of the post-involutional mammary gland is similar to that of the virgin mammary gland, marked by a rudimentary ductal branching network (Pai and Horseman, 2011). Characteristic to involution, the cellular quiescence following lactation is a physiological state distinct from senescence, as the proliferative arrest is reversible in the former yet irreversible in the latter. The mammary gland remains dormant until the hormonal cues of pregnancy promote differentiation to reestablish the lactogenic alveoli in preparation for subsequent pregnancies (Harmes and DiRenzo, 2009). Of those genes detected as being differentially up-regulated from the virgin-to-post-involutional quiescent developmental comparison, KEGG analysis strongly identified pathways of the ribosome as being enriched, simply identified by those of the ribosome (Table 7).

Considering this specific ribosomal pathway, GO analysis for cellular component expectedly identified the ribosome as the location where gene products were up-regulated in quiescent MEC relative to virgin MEC. GO analysis for molecular function identified structural constituent of the ribosome as the activity of gene products up-regulated in quiescent MEC relative to virgin MEC. GO analysis for biological process likewise identified translation, rRNA processing, ribosomal large unit assembly, and ribosomal small unit assembly as biological entities up-regulated in post-involutional MEC relative to virgin MEC (Table 6).

These GO and KEGG results strongly indicate an extra-ribosomal function of the ribosome as a whole. Several hallmarks of cancer have been discussed in Chapter 1, including a self-sufficiency in growth signals, insensitivity to anti-proliferative signals, avoidance of apoptosis, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg, 2000). As an example, cancerous cells are able to avoid apoptosis, aiding their survival and proliferation. Tumorigenesis has been associated with alterations in the molecular mechanisms of the ribosome, where either the over- or under-expression of specific ribosomal proteins can impart ribosomal instability, cause nuclear stress, and ultimately have an effect on various oncogenes. For example, the ribosomal proteins RPL5, RPL23, and RPS7 have been shown to bind to MDM2, thereby blocking the degradation of p53 (Shenoy et al., 2012). As previously discussed in Chapter 1, p53 translocates from the cytoplasm into the nucleus following phosphorylation, stimulating the transcription of components for CDK inhibitors and halting the progression through the cell cycle. While cancerous cells are able to avoid apoptosis, increased expression of certain ribosomal proteins promote the extra-ribosomal functions of DNA repair and cellular homeostasis (Warner and McIntosh, 2009). Perhaps these post-involutional analyses exemplify the increased surveillance and monitoring capabilities of the primiparous cells of the mammary gland.

Of those genes detected as being differentially down-regulated from the virgin-to-post-involutional quiescent developmental comparison, KEGG analysis identified two encompassing biological themes. First, there was decreased protease activity, marked by the down-regulated pathways pertaining to focal adhesion. Second, various cancerous pathways were likewise depicted as being down-regulated, as identified by those of proteoglycans in cancer, pathways in cancer, Wnt signaling pathway, and mTOR signaling pathway (Table 7).

Considering the pathways pertaining to protease activity, GO analysis for cellular component identified the extracellular exosome and cell-cell adherens junction as the locations where gene products were down-regulated in quiescent MEC relative to virgin

MEC. GO analysis for molecular function identified protein binding and kinase activity as the activities of gene products down-regulated in quiescent MEC relative to virgin MEC. GO analysis for biological process likewise identified positive regulation of cell migration and cell adhesion as biological entities down-regulated in post-involutional MEC relative to virgin MEC. Considering those pathways pertaining to cancer, GO analysis for cellular component identified the cytoplasm as the location where gene products were down-regulated in quiescent MEC relative to virgin MEC. GO analysis for molecular function identified Wnt-protein binding as the activity of gene products down-regulated in quiescent MEC relative to virgin MEC. GO analysis for biological process likewise identified Wnt signaling as the biological objectives that are down-regulated in post-involutional MEC relative to virgin MEC (Table 6).

Through catabolic hydrolysis of peptide bonds, proteases enzymatically break down proteins and peptides. Normally cells are tethered to the ECM and to other cells by integrins and CAMs, respectively. However, when cancer cells alter these interactions, they often exhibit an increased transcription of extracellular protease genes concomitant with a decreased transcription of protease inhibitor genes (Alberts et al., 2008; Hanahan and Weinberg, 2000). Proteases degrade the surrounding matrix, thereby facilitating the invasion of cancerous cells into the stroma or blood supply. Invasive and metastatic capabilities result from the up-regulation of protease genes and down-regulation of protease inhibitors (Hanahan and Weinberg, 2000). Therefore, the decreased expression of protease genes observed in this global transcriptomic analysis holds potential as a factor influencing the association of parity and protection against breast cancer.

Proteases provide cancer a means through which they are able to break through the ECM and invade into other tissues. In *in* vitro studies, expression levels of protease-activated receptors (PAR) have previously been found to correlate with the degree of invasiveness in established cancer cell lines. In their study of PAR2, Morris and colleagues analyzed the signaling, migration, and invasion tendencies of these G-protein coupled receptors within MDA-MB-231 and BT549 human breast cancer cell lines. When cultured in NIH 3T3 fibroblast conditioned medium, depletion of PAR2 protein significantly reduced MDA-MB-231 and BT549 cell migration and invasion, suggesting PAR is a critical mediator of breast cancer tissue invasion and metastasis (Morris et at., 2006). However, the mechanism through which PAR2 promotes breast cancer cell migration and invasion is poorly understood, as whether PAR2 regulates effectors of malignant progression in cancer cells, such as Ras- and Rho-GTPases, has not been determined (Morris et at., 2006). Similarly, in their study on the activation of PAR1 in MCF7 and MDA-MB-231 human breast cancer cell lines, Kamath and colleagues determined that breast cancer cells express high levels of PAR1 (Kamath et al, 2001). *In vitro* studies by Yang et al. likewise found that ectopic expression of PAR1 induced an invasive phenotype representative of basal-like carcinoma that readily formed lesions in the lungs of mice (Yang et al., 2015). Interestingly, both findings suggest therapeutics targeted toward G_i/PI(3)kinase-dependent pathways expressed in breast cancers might prove beneficial in inhibiting the progression of tissue invasion and metastasis (Kamath et al., 2001; Yang et al, 2015). Simultaneously considering the decreased protease gene expression observed in this current study, these findings collaboratively highlight a need for additional *in vivo* experiments centered on the specific signaling pathways through which proteases influence the invasiveness of breast cancer.

The extracellular exosome was identified as a location where gene products were down-regulated in quiescent MEC relative to virgin MEC. Functioning through exocytosis, exosomes transfer various components such as bioactive molecules, proteins, and lipids, thereby mediating communication between cells (Becker et al., 2016; Simons and Raposo, 2009). While exosomes have been found to support the development and involution of the mammary gland, their biological activities have likewise been observed to contribute to patho-physiological processes, thereby mediating oncogenic signaling between cancer cells (Hendrix and Hume, 2011). Exosomes have consequently been proposed as viable biomarkers and therapeutic targets in both physiological and pathological processes (Simons and Raposos, 2009).

Recent studies have indicated a role for angiogenic signaling promoted by exosomes originating from hypoxic cancer cell. Sustained angiogenesis is an acquired capability that supports the abnormal metabolic needs of cancerous cells. Under hypoxic conditions, cancer cells secrete exosomes that modulate their local environment to facilitate tumor angiogenesis and maintain communication to metastatic sites (Hendrix and Hume, 2011; King et al, 2012). In their *in vitro* studies utilizing MCF7, SKBR3, and MDA-MB 231 breast cancer cell lines cultured under either moderate (1% O_2) or severe (0.1% O_2) hypoxia, King and colleagues isolated and quantitatively analyzed exosomes utilizing immunoblotting and qPCR techniques. All three cell lines showed a significant increase in the number of exosomes present in the hypoxic environment. Additionally, exosomes were observed to be released extracellularly when treated with dimethyloxalylglycine (DMOG), a HIF hydrolase inhibitor, however transfection of cells with HIF-1αsiRNA prior to hypoxic exposure prevented the hypoxic-induced exosome release (King et al., 2012). These findings provide evidence that hypoxia promotes the release of exosomes by breast cancer cells, and that this response may be mediated by HIF-1α. As cancer cells proliferate expression of HIF-1α is increased to promote the transcription of pro-angiogenic factors for the purpose of vascular development. That the extracellular exosome was found to be downregulated in the present global transcriptomic analysis holds potential as yet another factor influencing the association of parity and protection against breast cancer, however further investigations are needed to explore exosome biogenesis, release, and the mechanisms through these events are mediated (Becker et al., 2016; Hendrix and Hume, 2011).

The culmination of these ribosomal, proteasomal, and exosomal findings indicates an overall decrease in those factors influencing the development of cancer. Excitingly, this is supported by the observation that several pathways of cancer were likewise depicted as being down-regulated from the virgin-to-post-involutional quiescent developmental comparison.

Proteoglycans are glycosylated proteins that interact with growth factors, growth factor receptors, and cytokines in the ECM to influence the extracellular environment and govern cellular movement. Their effects have been observed in repair of the CNS, wound healing, and cell motility. As with most cellular factors previously discussed, this happens in both physiological and pathological states (Cattaruzza and Perris, 2005). For example, activated stromal and tumor cells secrete effectors that promote the reorganization of the ECM to facilitate tumor cell growth, migration, and invasion

(Theocharis et al., 2010). Specific proteoglycans such as decorin and syndecan-1 have previously been studied, and curiously influence different oncogenic environments. For example, syndecans are known to interact with integrins, promoting cell adhesion and migration. Syndican-1 expression by fibroblasts is thought to promote tumorigenesis by regulating tumor cell adhesion, proliferation, and angiogenesis (Theocharis et al., 2010). Conversely, low expression of decorin has prognostic significance in that it is associated with lower survival rates among female humans diagnosed with certain types of breast cancer (Troup et al., 2003) while administration of decorin can reduce breast cancer tumor growth and metabolism (Theocharis et al., 2010). Proteoglycans thus hold potential as mediators in pharmacological targets that modulate tumor progression.

Wnt signaling regulates numerous cellular processes such as cell fate and proliferation and is strongly established as an oncogenic factor in the murine mammary gland (Ayyanan et al., 2006; Howe and Brown, 2004; Klarmann et al., 2008). That pathways, biological processes, and molecular functioning of Wnt signaling were differentially down-regulated from the virgin-to-quiescent developmental comparison are not surprising from a biological standpoint and are in agreement with studies indicating either a genetically- or epigenetically-influenced disregulation of these controlling mechanisms contribute to breast cancer development (Karlmann et al., 2008). Wnt signaling is initiated by the interaction of a Wnt ligand and a Frizzled-related protein receptor that subsequently leads to the stabilization of β-catenin, permitting β-catenin to translocate into the nucleus and induce transcription. In the absence of Wnt signaling, the downstream effects of β-catenin are kept in check and targeted for degradation via phosphorylation (Ayyanan et al., 2006). As such, disregulation of the Wnt signaling pathway has drastic implications in cancer development (Ayyanan et al., 2006; Karlmann et al., 2008). A wide range of cancers displays mutations in β-catenin, rendering them resistant to phosphorylation, however the opposite is true in breast cancer. For example, although β-catenin has been found to be upregulated in over forty percent of human breast cancers, transgenic expression of stabilized β-catenin in murine mammary tissues results in tumor development. (Ayyanan et al., 2006). Furthermore, the loss of Wnt ligand antagonists lead to hyperactive Wnt signaling, thereby promoting tumorigenesis in human mammary tissues (Howe and Brown, 2004). Regardless of all the studies previously conducted concerning Wnt signaling and breast cancer development, much remains to be learned regarding the tumor microenvironment and how paracrine factors promote tumor propagation. Previous immunofluorescent studies concerning fibroblast-secreted exosomes suggest exosome-mediated Wnt signaling involvement in promoting breast cancer cell motility and metastasis (Luga et al., 2012).

Similar to the observed Wnt signaling pathway results, the mTOR signaling pathway was likewise differentially down-regulated from the virgin-to-quiescent developmental comparison. Utilizing immunoblotting techniques on MCF7 and T47D human breast carcinoma cell lines, Boulay and colleagues demonstrated cellular proliferation was dependent on mTOR signaling (Boulay et al., 2018). Additionally, phosphatase and tensin homolog (PTEN) has been proposed as a negative regulator of the PI3K/mTOR/STAT3 signaling pathway. When NOD/SCID mice were inoculated with MCF7 cells overexpressing PTEN, tumorigenicity was markedly decreased compared to control mice (Zhou et al., 2007). Curious to note considering the up-regulated pathways of the ribosome discussed above, when in the presence of mitogenic stimuli and sufficient nutritional requirements, mTOR relays a positive signal translational signal by activating the 40S ribosomal protein S6 kinase (Boulay et al., 2018). Perhaps the parity-induced increased stability of the ribosome contributes to extra-ribosomal functions such as mTOR regulation, ultimately promoting cellular homeostasis and the suppression of oncogenesis. The decrease in mTOR signaling observed in the current global transcriptonic analysis is thus in agreement with studies indicating its regulation is essential to tumor suppression.

Taken together, these analyses identified significant differences in the functions performed by those genes with down- and up-regulated expression from the virgin-to-pregnant comparison and from the virgin-to-post-lactational quiescent comparison of isolated MEC. Genes involved in several pathways influencing cell-cell communication and interaction showed a decrease in expression in the pregnant state compared to the virgin developmental state. Yet for the same developmental comparison genes involved in metabolism and proliferation showed an increase in expression, each with unique implications not only in milk production as it would relate to the dairy industry but also in future breast cancer studies. Genes involved in several pathways leading to cancer showed a decrease in expression in the quiescent state compared to that of the virgin developmental state. Yet for the same developmental comparison genes involved in the ribosome, its integrity, and its functioning showed an increase in expression. Recently, individual ribosomal proteins have been highlighted as having extra-ribosomal functions such as DNA repair, regulation of apoptosis, and autoregulation of ribosomal protein synthesis while disorders resulting from impaired ribosome biogenesis and function have been shown to be oncogenic and consequently detrimental to cellular homeostais (Shenoy et al., 2012; Warner and McIntosh, 2009). This indicates that perhaps the association of parity and protection against breast cancer are related to the ribosome and perhaps the association of risk for breast cancer is related to ribosomalpathies.

CHAPTER 4 – Comparison of Key Regulator Genes Affecting Developmental Stages in Mice to Factors Identified from the Parallel Proteomic Analysis

4.1 Introduction

The central dogma of biology states that DNA is encoded into mRNA for the production of proteins, the expression of which defines each cell (Pepke et al., 2009). RNA therefore influences the present and future activities of a cell and serves as the intermediary regulator between genotype and phenotype (Marguerat and Bahler, 2010; Mortazavi et al., 2008). Traditionally, mRNA concentrations have been used as proxies for the corresponding protein concentrations and activities (Gunawardana and Niranjan, 2013). However, this relationship in expression is not exact. The Pearson correlation coefficient from previous parallel transcriptomic and proteomic analyses range from 0.46 to 0.76, meaning approximately forty-six to seventy-six percent of the variation in protein abundance can be explained by knowing the mRNA abundance (Hack 2004; Vogel and Marcotte, 2012). Furthermore, while these correlation analyses have been studied in yeast and plant samples, the relationship has not been considered extensively in mammalian samples (Ghazalpour et al., 2011; Li et al., 2014).

Protein abundances are influenced primarily by four regulatory events: the rate at which genes are transcribed, the rate at which RNA is degraded, the rate at which proteins are translated, and the rate at which proteins are degraded. While the former two affect RNA abundance, the latter two affect the difference between RNA and protein abundance (Li et al., 2014). Synthesis of RNA itself is tightly controlled, yet through the actions of modifiers such as microRNA and binding proteins, transcript abundance is pliant and allows a cell to adapt rapidly to environmental or genetic changes (Vogel and

Marcotte, 2012; Marguerat and Bahler, 2010). The stability of proteins following translation depends on their biological role, where regulatory proteins that react to various stimuli are synthesized and degraded rapidly in contrast to structural proteins that are degraded less rapidly. However, considerable work remains to be accomplished for better understanding the molecular kinetics of transcription and translation (Vogel and Marcotte, 2012).

The regulation of gene expression is fundamental to the relationship between genotype and phenotype. Unfortunately, system based approaches have traditionally relied heavily on the interpretation of transcriptomic data for insight into cell physiology and pathology (Ghazalpour et al., 2011). The varying methods through which the transcriptome may be analyzed have been previously discussed. Various methods likewise exist through which the proteome may be analyzed, including isotope-coded affinity tag (ICAT), stable isotope labeling with amino acids in cell culture (SILAC), large-scale western blotting, multi-dimensional protein identification technology (MudPIT), and two-dimensional gel electrophoresis (2-DE) (Hack, 2004; Chandramouli and Qian, 2009). Although labor-intensive and mechanical in nature, 2-DE remains the primary method for separating proteins thanks in part to technical advances including the availability of pre-cast polyacrylamide gels and improvements in pH gradient strips (Hack, 2004). Specifically this method separates extracted proteins first by molecular charge and then by size. Following protein separation and gel analysis, mass spectrometry technologies such as mass-adsorption laser deionization time-of-flight (MALDI-TOF) provide a method through which peptide sequences can be detected for protein identification (Chandramouli and Qian, 2009). Interpretation of the spectra results thus generated is likewise dependent upon bioinformatics for the integration of experimental data with software programs and databases to allow for the illustration of the underlying molecular dynamics (Kitano, 2002).

In this experiment, results from the transcriptomic and proteomic global profiling of isolated MEC at key developmental stages are explored. Significant to the "-omics" experimental design was the use of the same sample source across all developmental comparisons. Thus, for every isolated MEC sample being considered, RNA and protein were extracted in parallel, with the subsequent differential analyses highlighting the respective transcriptomic and proteomic molecular mechanisms influencing mammary gland physiology and pathology. While this approach not only allowed for the comparison of two differing means of analyzing the molecular phenotype, it is also a novel joint approach unique to mammary gland development that has not yet been previously reported.

4.2 Methods

4.2.1 Proteomic Analysis of Isolated Primary Virgin, Pregnant, and Post-Involutional Quiescent Mammary Epithelial Cells

All procedures were performed as specified in Conly 2014 (see Appendix C).

4.2.2 Comparison of Proteomic Results to Identified Key Regulator Genes Affecting Developmental Stages in Mice

For every differentially detected protein identified between the down- and up-regulated virgin and pregnant protein sets and between the down- and up-regulated virgin and post-involutional quiescent protein sets by Conly 2014, the corresponding transcript was assessed for detection and fold change as identified by the CLC Genomics Workbench.

4.3 Results

Of the 31 protein spots differentially detected between the virgin and pregnant samples, 28 were down-regulated while the remaining 3 were up-regulated ($p < 0.02$). Of those proteins, 29 were detected as being expressed in the transcriptomic analysis (93.5% similarity), but only 6 of those 29 were differentially expressed (19.6% similarity detected proteins, 19.4% total proteins). Furthermore, differences existed in the direction of fold change. For these protein-gene pairs that were dually differentially detected, 1 agreed in the direction of change (16.7% agreement), with the other 5 showing a fold change in the opposite direction (83.3% disagreement). Of the 36 protein spots differentially detected between the virgin and post-involutional quiescent samples, 31 were down-regulated while the remaining 5 were up-regulated ($p < 0.02$). Of those proteins, 34 were detected as being expressed in the transcriptomic analysis (94.4% similarity), but only 7 of those 34 were differentially expressed (20.6% similarity detected proteins, 19.4% total proteins). Again, differences existed in the direction of fold change. For these protein-gene pairs that were dually differentially detected, 1 agreed in the direction of change (14.3% similarity), with the other 6 showing a fold change in the opposite direction (85.7% disagreement). The magnitude of change detected by both the proteomic and transcriptomic analyses have been listed (Tables 7 and 8).

Table 7: Fold Change Expression Values of Transcripts Corresponding to Differentially Detected Proteins in the Pregnant State Compared to the Virgin State Originating from Proteomic Analysis of Isolated MEC

Protein and RNA were extracted in parallel from all isolated murine MEC. For every previously identified differentially detected down- and up-regulated protein and that protein's detected fold change, the magnitude of fold change was identified for every corresponding gene according to the CLC Genomics Workbench. Detections of down-regulation from the virgin to the pregnant state are depicted in red, detections of up-regulation from the virgin to the pregnant state are depicted in green. If the detection in the transcriptomic analysis was not significant, those numbers were not color-coded, although the direction is indicated by the presence or absence of a negative (-) symbol.

Gene Symbol	Protein Name	Ensembl Gene ID	Proteomic Fold Change	CLC Genomics Workbench (p-value)
Acadl	Long-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	ENSMUSG00000026003	**1.65**	**1.57** (0.033)
Acads	Short-chain specific acyl-CoA dehydrogenase	ENSMUSG00000029545	**-0.75**	1.77 (0.112)
Acat1	Acetyl-CoA transferase, mitochondrial precursor	ENSMUSG00000032047	**-0.71**	1.68 (0.110)
Aco2	Aconitate hydratase, mitochondrial precursor	ENSMUSG00000022477	**-0.66**	1.06 (0.453)
Aco2	Aconitate hydratase, mitochondrial precursor	ENSMUSG00000022477	**-0.56**	1.06 (0.456)
Afp	Alpha-fetoprotein, partial	ENSMUSG00000054932	**-0.58**	-1.04 (0.775)
Alb	Serum albumin	ENSMUSG00000029368	**-0.53**	N/A
Alb	Serum albumin	ENSMUSG00000029368	**-0.52**	N/A
Aldoa	Fructose-bisphophate aldolase A isoform precursor	ENSMUSG00000030695	**-0.68**	1.09 (0.240)
Cat	Catalase	ENSMUSG00000027187	**-0.47**	1.20 (0.457)
Etfa	Electron transfer flavoprotein subunit alpha, mitochondrial	ENSMUSG00000032314	**-0.46**	**1.38** (0.047)
Gapdh	Glyceraldehydes-3-phosphate dehydrogenase	ENSMUSG00000057666	**-0.64**	-1.24 (0.339)
Hspa8	Heat shock protein 70 cognate	ENSMUSG00000015656	**3.74**	-1.23 (0.244)
Hspa1a	Heat shock protein 1A	ENSMUSG00000091971	**-0.40**	-1.34 (0.440)

Ighvdj	Immunoglobulin heavy chain variable region	ENSMUSG00000096767	**-0.73**	2.76 (0.288)
Krt1	Keratin, type II cytoskeletal 1	ENSMUSG00000046834	**-0.35**	-5.64 (0.0051)
Krt16	Keratin, type I cytoskeletal 16	ENSMUSG00000053797	**-0.59**	1.43 (0.281)
Krt19	Keratin, type I cytoskeletal 19	ENSMUSG00000020911	**1.39**	1.41 (0.245)
Lasp1	LIM and SH3 domain protein 1	ENSMUSG00000038366	**-0.63**	1.33 (0.327)
Lmna	Prelamin-A/C isoform A precursor	ENSMUSG00000028063	**-0.49**	-1.06 (0.193)
Lmnb1	Lamin-B1	ENSMUSG00000024590	**-0.69**	1.24 (0.067)
Mdh2	Malate dehydrogenase	ENSMUSG00000019179	**-0.55**	**1.21** (0.036)
Pdlim1	PDZ and LIM domain protein 1	ENSMUSG00000055044	**-0.56**	-1.16 (0.472)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.53**	-8.76 (0.103)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.50**	-8.76 (0.103)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.41**	-8.76 (0.103)
Tf	Serotransferrin precursor	ENSMUSG00000032554	**-0.46**	1.16 (0.646)
Tkt	Transketolase	ENSMUSG00000021957	**-0.79**	**0.63** (0.004)
Trap1	Heat shock protein 75 kDa, mitochondrial	ENSMUSG00000005981	**-0.56**	1.24 (0.104)
Tufm	Elongation factor Tu, mitochondrial isoform 2	ENSMUSG00000073838	**-0.72**	**0.16** (0.039)
Vdac2	Voltage-dependent anion channel 2	ENSMUSG00000021771	**-0.64**	**1.48** (0.006)

Table 8: Fold Change Expression Values of Transcripts Corresponding to Differentially Detected Proteins in the Post-Involutional Quiescent State Compared to the Virgin State Originating from Proteomic Analysis of Isolated MEC

Protein and RNA were extracted in parallel from all isolated murine MEC. For every previously identified differentially detected down- and up-regulated protein and that protein's detected fold change, the magnitude of fold change was identified for every corresponding gene according to the CLC Genomics Workbench. Detections of down-regulation from the virgin to the pregnant state are depicted in red, detections of up-regulation from the virgin to the pregnant state are depicted in green. If the detection in the transcriptomic analysis was not significant, those numbers were not color-coded, although the direction is indicated by the presence or absence of a negative (-) symbol.

Gene Symbol	Protein Name	Ensembl Gene ID	Proteomic Fold Change	CLC Genomics Workbench (p-value)
Acads	Short-chain specific acyl-CoA dehydrogenase	ENSMUSG00000029545	**-0.63**	N/A
Aco2	Aconitate hydratase, mitochondrial precursor	ENSMUSG00000022477	**-0.61**	1.06 (0.456)
Actb	Actin, cytoplasmic 1	ENSMUSG00000029580	**2.34**	1.19 (0.415)
Actb	Actin, cytoplasmic 1	ENSMUSG00000029580	**1.47**	1.19 (0.415)
Afp	Alpha-fetoprotein, partial	ENSMUSG00000054932	**-0.60**	-1.04 (0.775)
Alb	Serum albumin	ENSMUSG00000029368	**-0.60**	N/A
Aldh2	Aldehyde dehydrogenase, mitochondrial precursor	ENSMUSG00000029455	**-0.48**	-1.19 (0.116)
Atp5b	ATP synthase subunit beta, mitochondrial precursor	ENSMUSG00000025393	**-0.55**	1.07 (0.291)
CDC42	Cell division control protein homolog 42	ENSMUSG00000006699	**-0.36**	1.04 (0.212)
DLD	Dihydrolipoamide dehydrogenase precursor	ENSMUSG00000020664	**-0.46**	1.18 (0.119)
Eef1g	Elongation factor 1-gamma	ENSMUSG00000071644	**-0.42**	**1.18** (0.016)
Etfa	Electron transfer flavoprotein subunit alpha, mitochondrial	ENSMUSG00000032314	**1.12**	**1.38** (0.047)
Fh	Fumarate hydratase, mitochondrial precursor	ENSMUSG00000026526	**-0.81**	1.05 (0.516)
Hnrnph1	Heterogeneous nuclear ribonuclearprotein H	ENSMUSG00000007850	**-0.57**	1.00 (0.899)
Hnrnpa3	Heterogeneous nuclear ribonuclearprotein A3	ENSMUSG00000059005	**-0.55**	-1.09 (0.276)

Hsp90ab1	Heat shock protein 84	ENSMUSG00000023944	**-0.48**	-1.01 (0.774)
Hsp90b1	Endoplasmin	ENSMUSG00000020048	**-0.42**	**1.12** (0.015)
Hspd1	60 kDa heat shock protein, mitochondrial	ENSMUSG00000025980	**-0.60**	**1.33** (0.039)
Hspa8	Heat shock protein 70 cognate	ENSMUSG00000015656	**3.93**	-1.23 (0.244)
Ighvdj	Immunoglobulin heavy chain variable region	ENSMUSG00000096767	**-0.60**	2.76 (0.288)
Khsrp	Far upstream element-binding protein 2	ENSMUSG00000007670	**-0.65**	-1.01 (0.815)
Krt1	Keratin, type II cytoskeletal 1	ENSMUSG00000046834	**-0.59**	-5.64 (0.051)
Krt16	Keratin, type I cytoskeletal 16	ENSMUSG00000053797	**-0.51**	1.43 (0.281)
Krt19	Keratin, type I cytoskeletal 19	ENSMUSG00000020911	**1.87**	1.41 (0.245)
Lmna	Prelamin-A/C isoform A precursor	ENSMUSG00000028063	**-0.43**	-1.06 (0.193)
Pdlim1	PDZ and LIM domain protein 1	ENSMUSG00000055044	**-0.55**	-1.16 (0.472)
Sdha	Succinate dehydrogenase flavoprotein subunit	ENSMUSG00000021577	**-0.34**	1.21 (0.257)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.72**	-8.76 (0.103)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.40**	-8.76 (0.103)
SERPINA1	Alpha-1-antiproteinase precursor	ENSMUSG00000066366	**-0.33**	-8.76 (0.103)
Tf	Serotransferrin precursor	ENSMUSG00000032554	**-0.50**	1.16 (0.646)
Tf	Serotransferrin precursor	ENSMUSG00000032554	**-0.47**	1.16 (0.646)
Tf	Serotransferrin precursor	ENSMUSG00000032554	**-0.44**	1.16 (0.646)
Tkt	Transketolase	ENSMUSG00000021957	**-0.66**	**1.64** (0.001)
Tufm	Elongation factor Tu, mitochondrial isoform 2	ENSMUSG00000073838	**-0.45**	**1.19** (0.007)
Vdac2	Voltage-dependent anion channel 2	ENSMUSG00000021771	**-0.68**	**1.48** (0.006)

4.4 Discussion

System-level research and advances in “-omic” technologies have enabled the analyses of thousands of biomolecules simultaneously, providing a unique approach to better understanding the biology of the organism of interest (Klopfleisch and Gruber, 2012). Such global approaches are aimed at comprehensively illustrating the complex molecular mechanisms underlying cell physiology and pathology. Traditionally, transcriptomic results and analyses have been used as proxies for the corresponding interpretation into protein concentrations and activities (Gunawardana and Niranjan, 2013). However, not only is this relationship in expression inexact, but it also has been limited primarily to yeast and plant samples (Ghazalpour et al., 2011; Li et al., 2014).

The Peterson Lab's comparative analyses of developmental stages in isolated MEC have generated a relatively large transcriptomic data set and a relatively small proteomic data set. Both data sets identify the names of either genes or proteins that are differentially expressed between key developmental stages, the magnitude in fold change of that differential expression, and the corresponding statistical significance. Unique to this experiment, the analyses of the transcriptome and proteome were performed in parallel, using the same sample source. This novel joint approach has allowed for an exploration in the relationship of the governing transcriptomic and proteomic mechanisms within isolated murine MEC, where the corresponding gene expression profile was assessed for each previously identified differentially expressed protein.

As previously discussed, the transcriptomic analysis of mammary gland development identified specific molecular mechanisms regulating cell metabolism, communication, pathways of cancer, and ribosomal function. Although similar inferences were made from the proteomic analysis between the nulliparous and primiparous states, a greater abundance of proteins was detected in the virgin MEC compared to both other developmental stages investigated. Specifically, the identification of those proteins expressed differentially suggests a greater level of molecular activity in MEC isolated from the virgin mammary gland (Conly, 2014). While the transciptomic and proteomic analyses were conducted in parallel from the same sample course, it is interesting that distinct abundances of proteins yet somewhat similar abundances of genes were detected as being differentially up- and down-regulated across the developmental stages being compared.

There are several factors and variables to consider when comparing transcriptomic and proteomic data. First, although a small proportion of proteins were unable to be compared to their corresponding transcripts, this does not mean the transcripts were absent within the isolated MEC. Rather, the differential analysis performed by the software applications was unable to detect a difference in expression between the two developmental states being considered. Second, although transcriptomic technologies produce a greater amount high-throughput data compared to the limitations in depth and coverage of proteomic technologies, each provides a necessary and unique perspective for analyzing the molecular phenotype (Nagalakshmi, et al., 2010; Hegde et al., 2003). Specific to the 2-DE methods utilized for the proteomic analysis, while the data sets thus generated are not complete lists of all differential protein products influencing MEC physiology and pathology, they nevertheless highlight those events occurring and provide a framework for further exploration and validation (Conly, 2014). Unfortunately, although 2-DE is extensively used for qualitative proteomic experiments, the analysis of hydrophobic proteins remains a challenge unique to this approach due to their poor solubility in aqueous buffers (Chandramouli and Qian, 2009). Third, biological events and practical aspects may be accounted for in any observed differences in detected abundances. For example, the diverse chemical nature of proteins compared to RNA is important to note, where global analyses are complicated by the various ways the twenty different amino acids versus only the four nucleotide bases may be combined (Hegde et al., 2003). Likewise, post-transcriptional and post-translational modifications such as alternative splicing, polyadenylation, RNA degredation, allosteric protein interactions, phosphorylation, glycosylation, and proteolysis affect transcript and protein abundance, stability, and turnover (Vogel and Marcotte, 2012; Marguerat and Bahler, 2010). Fourth, many transcripts could have encoded for either relatively large, small, or highly insoluble proteins that are difficult to detect and analyze through proteomic technologies. For example, the transcriptomic analysis differentially detected an up-regulation in transcripts corresponding to the whey acidic protein (*Wap)* gene in the virgin-to-pregnant comparison (fold change = 129.14, p-value = 0.02); an up-regulation in transcripts corresponding to the immunoglobulin kappa chain variable 8-30 (*Igkv8-30*) gene in the virgin-to-post-involutional quiescent comparison (fold change = 96.34, p-value = 0.01); and a down-regulation in transcripts corresponding to the fibrillin (*Fnb2*) gene in the virgin-to-post-involutional quiescent comparison (fold change = -20.24, p-value = 0.04); however none of these proteins were detected in the corresponding proteomic analyses. That the proteins for which the former two genes encode are relatively small at 14.423 kDa and 14.529 kDa, respectively, and that the protein for which the latter gene encodes is relatively large at 313.818 kDa might explain the lack of detection within the 2-DE analysis. Additionally, the RNA-seq technology utilized sequenced single-ended reads as opposed sequencing reads in pairs. While paired-end RNA-seq can help detect alignment errors and improve sequencing sensitivity and specificity, such an experimental approach is necessary only when isoform annotation and exploration of the genetic architecture are the primary goals (Li and Homer, 2010). As these were not the objectives of the previous comparative transcriptomic analyses discussed, the approach did not consider isoform specific expression. Here, the inability to measure isoform expression may be impacting the correlation between abundance results for certain peptides that represent specific isoforms, however this cannot be definitively stated (Ghazalpour et al., 2011). With these considerations in mind, the observed lack of correlation between protein and transcript abundance in relation to those of isolated virgin MEC can be explained by the technical differences in the methods utilized and the various biological events and practical aspects that influence RNA stability and the potential for protein degradation.

Of the 6 protein-gene pairs that were dually differentially detected in the virgin-to-pregnant developmental comparison, insights into the biological mechanisms of MEC can be found by considering the functions of those identified proteins. For example, transketolase (TKT) is a pentose-phosphate enzyme whose overexpression leads to increased production of glyceraldehyde-3-phosphate for augmented fermentation of glucose to lactate. In the histological analysis of breast cancer samples for transkelotase-like-1 (TKTL1), a mutated transkelotase enzyme, Foldi and colleagues demonstrated TKTL1 is overexpressed in tumor cells yet treatment with specific transkelotase inhibitors led to a reduction in tumor cell proliferation, indicating this enzyme holds potential as a targeted biomarker for tumor growth maintenance (Foldi et al., 2006).

Another metabolic protein-gene pair identified in the present study is malate dehydrogenase (MDH2), an enzyme that catalyzes the oxidation of malate to oxaloacetate for the generation of NADPH. Both TKT and MDH2 were found to be down-regulated in the proteomic analysis of MEC in the pregnant state yet up-regulated in that of the transcriptomic analysis. In a unique analysis of MALDI mass spectrometry of proteins excised from gel spots of liver and mammary samples collected from lactating Friesian cows, Rawson and colleagues found an overall greater abundance of both TKT and MDH2 in the liver tissue compared to the mammary tissue. Findings from this proteomic analysis supported the hypothesis that gluconeogenesis and β-oxidative pathways should predominate in the liver during lactation while fat synthesis should predominate in the mammary gland (Rawson et al., 2012). Although this was not a developmental comparison on isolated MEC, and although drastic differences do exist in the regulatory mechanisms between bovine and murine mammary glands, the proteomic findings by Rawson and colleagues do emphasize the consideration of metabolic outputs of hepatic and mammary tissues.

Voltage-dependent anion channel 2 (VDAC2) was likewise another protein-gene pair identified in the present study. With the exception of a few membrane-permeable lipophilic compounds, hydrophilic metabolites and respiratory substrates such as ATP, ADP, and inorganic phosphate that enter and exit the mitochondria must pass through the outer mitochondrial membrane through a VDAC. In addition to ATP generation during oxidative phosphorylation, VDAC2 is significant in enhancing glycolysis for the synthesis of lipids, proteins, and nucleotides (Maldonado et al., 2013). Recently, VDAC2 has also been shown to have inhibitory effects upon the Bak-mediated apoptotic response as Bak is typically inactive when bound to VDAC2 and localized in the outer mitochondrial membrane. When no longer sequestered to VDAC2, Bak is able to carry out its pro-apoptotic functions, suggesting that in addition to its metabolic functions

VDAC2 also plays a role in the regulation of controlled cell death and may serve as a target for drug discovery (Chandra et al., 2005).

Still pertaining to metabolism, the electron transfer flavorprotein subunit alpha (ETFA), together with the beta subunit, is localized within the mitochondrial matrix and serves as an obligatory electron acceptor during fatty acid β-oxidation. This protein is thought to be dependent upon GH signaling, as prior studies in GHR knockout mice ($GHR^{-}/^{-}$) showed a notable reduction in ETFA content, granted these studies were focused on the proteomic activity influencing murine lung development (Beyea et al., 2006).

Long-chain specific acyl-CoA dehydrogenase (ACADL) was the only protein-gene dually differentially detected in the virgin-to-pregnant developmental comparison that agreed in the direction of fold change. Long-chain acyl-CoA esters not only serve as important intermediates in lipid biosynthesis and fatty acid degradation but are also known to regulate metabolism and gene expression (Faergeman and Knudsen, 1997). Accordingly, the ACADL enzyme plays a pivotal role in lipogenesis within the mammary gland. Through mitochondrial fatty acid β-oxidation and the degradation of fatty acids of different chain lengths, each cycle of β-oxidation by ACADL generates a two-carbon-chain shortened acyl-CoA and acetyl-CoA (Hunt and Alexson, 2002). These newly formed fatty acid chains are esterified to a glycerol-3-phosphate backbone by the actions of glycerol-3-phosphate acyl transferase and diacylglycerol acyltransferase enzymes located on the endoplasmic reticulum, thereby completing the synthesis of triacylglycerol (TAG) (Bernard et al., 2008). During lactation, individual TAG molecules combine and incorporate themselves into cytoplasmic lipid droplets that are ultimately secreted in a membrane-enveloped lipid particle known as the milk fat globule (MFG) (Neville and Picciano, 1997). Interestingly, studies on ACADL knockout mice (ACADL^{-}/$^{-}$) have shown that decreased mitochondrial fatty acid oxidation results in an increased content of intracellular diacylglycerol, the activation of protein kinase C (PKC), and consequently decreased insulin signaling and action within hepatic and skeletal muscle tissues (Zhang et al., 2007).

Taken together, the differences in the transcriptomic and proteomic expression found in the pregnant state relative to the virgin state highlight a hormonal phenomenon influencing the metabolic regulation of MEC. Analyses on murine mammary tissue utilizing electron microscopy have not only shown that during pregnancy there is a notable increase in the number of mitochondria per secretory cell but also an increase in the activities of numerous mitochondrial enzymes, suggesting that the mitochondrial activity within MEC is correlated to milk production (Hadsell et al., 2010). These findings have similarly been noted in humans and dairy cows (Laubenthal et al., 2016). The increased mitochondrial processes observed in the transcriptomic analyses of isolated MEC in the pregnant state relative to the virgin were therefore to be expected and suggest that half-way through pregnancy, MEC are enhancing their mitochondrial functioning for energy production in preparation for milk synthesis and lactation. Yet curiously, the overall trend noted from the proteomic analyses of MEC isolated in parallel was a decrease in metabolic activity in the pregnant state of the cell. All proteins involved in metabolic processes were downregulated in the pregnant state compared with the virgin state, suggesting less energy generation is occurring in the pregnant state than the virgin state and perhaps that the differentiated state of the cell is more energy efficient (Conly, 2014).

Similarly, of the 7 protein-gene pairs that were dually differentially detected in the virgin-to-post-involutional quiescent developmental comparison, biological significance within MEC can be found by considering those functions of the identified proteins. VDAC2, TKT, ETFA, and TUFM, which were all dully differentially detected in the prior developmental comparison, were again dully detected in the primiparous developmental comparison. Again, only one protein-gene pair dually differentially detected in agreed in the direction of fold change, this time ETFA. Unique to the post-involutional developmental comparison are 60 kDa heat shock protein (HSPD1), elongation factor 1-gamma (EEF1G), and enoplasmin (HSP90B1).

HSPD1 is a specific heat shock protein weighing 60 kDa. Heat shock proteins are known to be involved in protein synthesis and folding through their contributions to protein synthesis, secretion, trafficking, degradation, and regulation of transcription factors. By preventing the formation of nonspecific protein aggregates, they maintain proteostasis and have come to be known as "protein chaperones" that enhance protein stability. Conversely, heat shock proteins are characteristically over-expressed in cancer. By preventing the translocation of Bax to induce apoptosis, HSPD1 also has the potential to promote cell survival can detrimentally contribute to tumor cell proliferation, invasion, differentiation, and metastasis (Lianos et al., 2015; Swindell et al., 2009). Specifically, by recognizing various exposed hydrophobic amino acid side-chains HSPD1 assists in the transport and folding of mitochondrial proteins through ATP-regulated cycles of binding and hydrolysis (Hartl and Hayer-Hartyl, 2009). HSPD1 expression has been found to be elevated in breast cancer tissues, and thus holds potential as a molecular marker of cancer and in drug targeting (Lianos et al., 2015).

Another heat shock protein, SHP90B1, functions in a similar manner as HSPD1, except it is located in the endoplasmic reticulum (ER). Under normal circumstances this endoplasmic chaperone not only assists in protein folding but also targets misfolded proteins for ER-associated degradation. In cancerous cells, HSP90B1 expression is upregulated in an effort to combat the accumulation of misfolded and damaged proteins that accumulate in the lumen of the ER (Kumar et al., 2018). HSP90B1 has recently been found to be a chaperone for the group of pathogenic receptors known as Toll-like receptors (TLR's), implying HSP90B1 plays a critical role in the immune response against infection (Liu et al., 2010).

The three steps of protein translation—initiation, elongation, and termination—are mediated by several factors. The cycles of elongation repeat a number of times that corresponds to the number of amino acids comprising the protein of interested (Kavaliauskas et al., 2012). Amino acids destined for protein synthesis are coupled to their conjugate tRNA and selected according to the correct base pairing match between the codon exposed in the A site on the small ribosomal subunit and the anticodon of the incoming tRNA. During elongation, the amino-bound tRNA is delivered to the A site of the ribosome, GTP hydrolysis is activated, and a peptide bond is formed (Alberts et al., 2008). Specific to this process is elongation factor 1γ (EEF1γ), which functions in the guanine nucleotide exchange following the delivery of the aminoacyl-tRNA (Al-Maghrebi et al., 2005). Prior *in vitro* studies have reported altered and upregulated expression of EEF1γ in T47D (Al-Maghrebi et al., 2005), MDA-MA-231 (Al-Maghrebi et al., 2005), and MCF-7 cancer cell lines (Joseph et al., 2004). Distinct from EEF1γ is TUFM, the elongation factor that functions in the selection of the correct amino acid to be incorporated into the growing peptide chain (Kavaliauskas et al., 2012). TUFM is additionally understood to inhibit serine proteases, presumably allowing for increased protein degradation and decreased protein production (Conly, 2014).

Taken together, the differences in the transcriptomic and proteomic expression found in the post-involutional quiescent state relative to the virgin state highlight a hormonal phenomenon influencing the translational regulation of MEC. In the global transcriptomic analysis, it was surprising to observe increased expression in those genes pertaining to the ribosome, its integrity, and its functioning. These findings suggest that perhaps the increase in ribosomal integrity may be associated with the parity-induced protection against breast cancer. Yet curiously, the overall trend noted from the proteomic analyses of MEC isolated in parallel was a decrease in RNA processing in the post-lactational quiescent state of the cell. All proteins involved in transcriptional regulatory processes were downregulated in the quiescent state compared with the virgin state, suggesting a decrease in production of transcripts and proteins in the primiparous MEC relative to virgin MEC (Conly, 2014)

A comparative analysis performed in parallel of the transcriptome and proteome has allowed for an exploration in the relationship of the governing mechanisms within isolated murine MEC of developmental stages. Although a Pearson correlation coefficient could not be calculated from the dual analysis, confidence does exists in describing those results obtained from these two differing means of analyzing the molecular phenotype. Results generated from this analysis included a relatively large transcriptomic data set and a relatively small proteomic data set, yet upregulation in mRNA did not necessarily reflect the expression pattern of the corresponding protein. This lack of correlation between protein and transcript abundance in relation to those of isolated virgin MEC was surprising. Technical differences in the methods utilized and the various biological events influencing RNA stability and the potential for protein degradation were previously listed as possible considerations to explain this phenomenon. However, the data further indicate a unique mathematical phenomenon occurring within MEC.

Of the 31 protein spots differentially detected between the virgin and pregnant samples, 6 were detected as being differentially expressed in the transcriptomic analysis (19.4% total proteins). Of the 36 protein spots differentially detected between the virgin and post-involutional quiescent samples, 7 were detected as being differentially expressed in the transcriptomic analysis (19.4% total proteins). That the same percentage of dually detected protein-gene pairs was identified for both developmental comparisons was fascinating and warranted further investigation.

There are numerous regulatory events occurring after mRNA translation that influence protein abundance. mRNA is less stable than protein, and accordingly suggest a marked decrease in protein concentration could be explained by preparations for cellular division (Vogel and Marcotte, 2012). This would support the observed differences in the virgin-to-pregnant developmental comparison, where MEC are supposedly devoted to growth and proliferation, but not the virgin-to-post-translational developmental comparison. A study examining the comparisons between these stages of development in both mammary and hepatic samples would be useful in categorizing these observed trends of transcript and protein expression profiles in MEC. The kinetics of transcription and translation also deserve consideration, as mRNAs are produced at a much slower rate than proteins, approximately two copies of mRNA per hour versus multiple corresponding proteins per hour, respectively, in mammalian cells (Vogel and Marcotte, 2012). This might support the observed differences in both developmental comparisons, where hormonal influences have encouraged transcriptional factors to promote the generation of specific transcripts, yet the translation of which is strictly monitored. Perhaps the 19.4% protein expression observed in this study implies even with sufficient transcript abundance approximately only twenty percent of genes are actively expressed at a given time. Indeed, it has been previously documented that variations in mRNA and protein abundances are often uncorrelated and a specifically protein expression is thought to be buffered with respect to the variation introduced transcriomically (Battle et al, 2015).

In conclusion, from the transcriptomic and proteomic profiles differentially detected the comparative analyses of developmental stages in isolated MEC have identified several molecular mechanisms influencing murine mammary gland physiology and pathology. Yet by no means do these investigations provide the complete story to the molecular happenings that can be described. They do, however, uniquely contribute to the global understanding of the biological systems influencing mammary gland development. Through this novel join approach, the identification of transcriptomic and proteomic effectors in cell metabolism, communication, pathways in cancer, and ribosomal function may guide further analyses related to enhanced lactational capacity and breast cancer development.

LIST OF REFERENCES

Akers, R.M., McFadden, T.B., Purup, S., Vestergaard, M., Sejrsen, K., and Capuco, A.V. 2000. Local IGF-I axis in peripubertal ruminant mammary development. Mammary Gland Biology and Neoplasia 5(1): 43-51.

Akers, R.M. 2006. Major advances associated with hormone and growth factor regulation of mammary growth and lactation in dairy cow. Dairy Science 89(4): 1222-1234.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. 2008. In *Molecular Biology of the Cell*, Fifth Edition. Garland Science, Taylor & Francis Group, LLC. New York, NY.

Alford, A.I. and Rannels, D.E. 2001. Extracellular matrix fibronectin alters connexin43 expression by alveolar epithelial cells. American J Physiology Lung Cellular and Molecular Physiology 280: 680-688.

Anders, S. and Huber, W. 2010. Differential expression analysis for sequence data count. Genome Biology 11: 106-117.

Anderson, S.M., Rudolph, M.C., McManaman, J.L., and Neville, M.C. 2007. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Research 9: 204 DOI: 10.1186/bcr1653.

Ansorge, W.J. 2009. Next-generation DNA sequencing techniques. New Biotechnology 25(4): 195-203.

Auer, P.L. and Doerge, R.W. 2010. Statistical design and analysis of RNA sequencing data. Genetics 185: 405-416.

Ayyakannu, A., Civenni, G., Ciarloni, L., Morel, C., Mueller, N., Lefort, K., Mandinova, A., Raffoul, W., Fiche, M., Dotto, G. P., and Brisken, C. 2006. Increased Wnt signaling triggers oncogenic conversion of human breast epithelial cells by a Notch-dependent mechanism. PNAS 103(10): 3799-3804.

Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., and Bissell, M.J. 1989. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development 105(2): 223-235.

Battle, A., Kahn, A., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. 2015. Impact of regulatory variation from RNA to protein. Science 347(6222): 664-667).

Bauman, D.E. 1999. Bovine somatotropin and lactation: from basic science to commercial application. Domestic Animal Endocrinology 17(2-3): 101-116.

Baumrucker, C.R. 1985. Nutrient uptake across the mammary gland. J Dairy Science 68: 2436-2451.

Becker, A., Thakur, B.K., Weiss, J.M., Kim, H.S., Peinado, H., and Lyden, D. 2016. Etracellular vesicles in cancer: cell-to-cell mediators of metastasis. Cancer Cell 30: 836-848.

Bernard, L., Leroux, C., and Chillard, Y. 2008. Expression and nutritional regulation of lipogenic genes in the ruminant lactating mammary gland. Advances in Experimental Medicine and Biology 606: 67-108.

Beyea, J.A., Sawicki, G., Olson, D.M., List, E., Kopchick, J.J., and Harvey, S. 2006. Growth hormone (GH) receptor knockout mice reveal actions of GH in lung development. Proteomics 6: 341-348.

Bionaz, M. and Loor, J.J. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinformatics and Biology Insights 5: 83-98.

Boisgard, R., Chanat, E., Lavialle, F., Pauloin, A., and Ollivier-Bousquet, M. 2001. Roads taken by milk proteins in mammary epithelial cells. Livestock Production Science 70: 49-61.

Bose, I. and Ghosh, B. 2007. The p53-MDM@ network: from oscillations to apoptosis. Bioscience 32(5): 991-997.

Boulay, A., Rudloff, J., Ye, J., Zumstein-Mecker, S., O'Reilly, T., Evans, D.B., Chen, S., and Lane, H.A. 2018. Dual inhibition of mTOR and estrogen receptor signaling *in vitro* induces cell death in models of breast cancer. Clinical Cancer Research 11(14): 5319-5328.

Boya, P. 2012. Lysosomal function and dysfunction: mechanism and disease. Antioxidants and Redox Signaling 17(5): 766-774.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liecht, A., Aximu-Petri, A., Kircher, M., Albert, F.W., Zeller, U., Khaitovich, P., Grutzner, F., Bergmann, S., Nielsen, R., Paabo, S., and Kaessmann, H. 2011. The evolution of gene expression levels in mammalian organs. Nature 478: 343-348.

Brisken, C., Park, S., Vass, T., Lydon, J.P., O'Malley, B.W., and Weinberg, R.A. 1998. A paracrine role for the epithelial progesterone receptor in mammary gland development. Proceedings of the National Academy of Science 95: 5076-5081.

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S.K., McMahon, J.A., McMahon, A.P., and Weinberg, R.A. 2000. Essential function of *Wnt-4* in mammary gland development downstream of progesterone signaling. Genes and Development 14: 650-654.

Brisken, C. and Rajaram, R.D. 2006. Alveolar and lactogenic differentiation. Mammary Gland Biology and Neoplasia 11(3-4): 239-248.

Britt, K., Ashworth, A., and Smalley, M. 2007. Pregnancy and the risk of breast cancer. Endocrine-Related Cancer 14: 907-933.

Capuco, A.V. and Akers, R.M. 1999. Mammary involution in dairy animals. Mammary Gland Biology and Neoplasia 4(2): 137-144.

Cattaruzza, S. and Perris, R. 2005. Proteoglycan control of cell movement during wound healing and cancer spreading. Matrix Biology 24: 400-417.

Chandra, D., Choy, G., Daniel, P.T., and Tang, D.G. 2005. Bax-dependent regulation of Bak by voltage-dependent anion channel 2. Biological Chemistry 280(19): 19051-19061.

Chandramouli, K. and Qian, P. 2009. Proteomics: challenges, techniques, and possibilities to overcome biological sample complexity. Human Genomics and Proteomics 1: DOI:10.4061/2009/239204.

Chua, A.C.L., Hodson, L.J., Moldenhauer, L.M., Robertson, S.A., and Ingman, W.V. 2010. Dual roles for macrophages in ovarian cycle-associated development and remodeling of the mammary gland epithelium. Development 137: 4229-4238.

Clarkson, R.W.E., Wayland, M.T., Lee, J., Freeman, T., and Watson, C.J. 2003. Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Research 6(2): 92-109.

Colletta, R.D., Christensen, K., Reichenberger, K.J., Lamb, J., Micomonaco, D., Huang, L., Wolf, D.M., Muller-Tidow, D., Golub, T.R., Kawakami, K., and Ford, H.L. 2004. The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. Proceedings of the National Academy of Science 101: 6478-6483.

Conly, A.K. 2014. Proteomic Analysis of Mammary Epithelial Cell Development. Unpublished thesis.

Conneely, O.M., Mulac-Jericevic, B., and Lydon, J.P. 2003. Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptors. Steroids 68(10-13): 771-778.

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. 2010. Uncovering the complexity of transcriptomes with RNA-seq. Biomedicine and Biotechnology 2010: 853916 DOI: 10.1155/2010/853916.

Couse, J.F. and Korach, K.S. 1999. Estrogen receptor null mice: what have we learned and where will they lead us? Endocrine Reviews 20: 358-417.

Cunha, G.R., Young, P., Christov, K., Guzman, R., Nandi, S., Talamantes, F., and Thordarson, G. 1995. Mammary phenotypic expression induced in epidermal cells by mammary mesenchyme. Acta Anatomica 152: 195-204.

Cunha, G.R., Young, P., Hom, Y.K., Cooke, P.S., Taylor, J.A., and Lubahn, D.B. 1997. Elucidation of a role for stromal steroid hormone receptors in mammary gland growth and development using tissue recombinations. Mammary Gland Biology and Neoplasia 2: 393-402.

Datta, S.R., Dudek, H., Tao, X., Masters, S., Fu, H., Gotog, Y., and Greenberg, M.E. 1997. Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. Cell 91: 231-241.

Eccles et al., 2013. Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Research 15(5): R92. DOI: 10.1186/bcr3493.

El-Sabban, M., Sfeir, A.J., Daher, M.H., Kalaany, N.Y., Bassam, R.A., and Talhouk, R.S. 2003. ECM-induced gap junctional communication enhances mammary epithelial cell differentiation. J Cell Sciences 116(17): 3531-3541.

Faergeman, N.J. and Knudsen, J. 1997. Role of long-chain fatty acyl-CoA esters in the regulation of metabolism and cell signaling. Biochemistry 323: 1-12

Fata, J.E., Werb, Z., and Bissell, M.J. 2004. Regulation of mammary gland branching morphogenesis by extracellular matrix and its remodeling enzymes. Breast Cancer Research 6: 1-11.

Fatin, V.R., St-Pierre, J., and Leder, P. 2006. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. Cancer Cell 9: 425-434.

Faupel-Badger, J.M., Arcaro, K.F., Balkam, J.J., Eliassen, H., Hassiotou, F., Lebrilla, C.B., Michels, K.B., Palmer, J.R., Schedin, P., Stuebe, A.M., Watson, C.J., and Sherman, M.E. 2012. Postpartum remodeling, lactation, and breast cancer risk: summary of a national cancer institute-sponsored workshop. National Cancer Institute 105(3): 166-74.

Fernandez-Valdivia, R. and Lydon, J.P. 2012. From the ranks of mammary progesterone mediators, RANKL takes the spotlight. Molecular and Cellular Endocrinology 357(1-2): 91-100.

Foldi, M., Stickeler, E., Bau, L., Kretz, O., Watermann, D., Gitsch, G., Kayser, G., Hausen., A., and Coy, J. 2006. Transketolase protein TKTL1 overexpression: a potential biomarker and therapeutic target in breast cancer. Oncology Reports 17: 841-845.

Fonesca, N.A., Rung, J., Brazma, A., and Marioni, J.C. 2012. Tools for mapping high-throughput sequencing data. Bioinformatics 28: 3169-3177.

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. Nature Methods 8(6): 469-478.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P., Brewer, H., Weitz, K., Camp II, D.G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirschgessner, T., Smith, D.J., Smith, R.D., and Lusis, A.J. 2011. Comparative analysis of proteome and transcriptome variation in mouse. PLos Genetics 7(6): e1001393.

Giacinti, C. and Giordano, A. 2006. Rb and cell cycle progression. Nature 25: 5220-5227.

Giancotti, F.G. and Ruoslahti, E. 1999. Integrin signaling. Science 285: 1028-1032.

Gors, S., Kucia, M., Langhammer, M., Junghans, P., and Metges, C.C. 2009. Technical note: milk composition in mice—methodological aspects and effects of mouse strain and lactation day. Dairy Science 92(2): 632-637.

Gouon-Evans, V., Rothenberg, M.E., and Pollard, J.W. 2000. Postnatal mammary gland development requires macrophages and eosinophils. Development 127: 2269-2282.

Grafen, A. and Hails, R. 2002. In *Moderns Statistics for the Life Sciences*, First Edition. Oxfird University Press Inc., New York, NY.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Varen, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., and Dudoit, S. 2011. The developmental transcriptome of *Drosphilia melanogaster*. Nature 471: 473-479.

Gunawardana, Y. and Niranjan, M. 2013. Bridging the gap between transcriptome and proteome measurements identifies post-translationally regulated genes. Bioinformatics 29(32): 3060-3066.

Hack, C.J. 2004. Integrated tanscriptome and proteome data: the challenges ahead. Breifings in Functional Genomics and Proteomics 3(3): 212-219.

Hadsell, D.L., Olea, W., Wei, J., Fiorotto, M.L., Matsunami, R.K., Engler, D.A., and Collier, R.J. 2010. Developmental regulation of mitochondrial biogenesis and function in the mouse mammary gland during a prolonged lactation cycle. Physiological Genomics 43: 271-285.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., Kraus, W.L. 2011. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell 145: 622-634.

Hanahan, D. and Weinberg, R.A. 2000. The hallmarks of cancer. Cell 100: 57-70.

Harmes, D.C. and DiRenzo, J. 2009. Cellular quiescence in mammary stem cells and breast tumor stem cells. J Mammary Gland Biology and Neoplasia 14(1): 19-27.

Harris, M.A., et al. 2004. The Gene Ontology (GO) database and informatics resource. Nucelic Acids Research 32: 258-261.

Harris, J., Stanford, P.M., Sutherland, K., Oakes, S.R., Naylor, M.J., Robertson, F.G., Blazek, K.D., Kazlauskas, M., Hilton, H.N., Wittlin, S., Alexander, W.S., Lindeman, G.J., Visvader, J.E., and Ormandy, C.J. 2006. Socs2 and Elf5 mediate prolactin-induced mammary gland development. Molecular Endocrinology 20(5): 1177-1187.

Hart, S.N., Therneau, T.M., Zhang, Y., Poland, G.A., and Kocher, J.P. 2013. Calculating sample size estimates for RNA sequencing data. J Computational Biology 20(12): 970-978.

Hartl, F.U and Hayer-Hartl, M. 2009. Converging concepts of protein folding *in vitro* and *in vivo*. Nature Structural and Molecular Biology 16(6): 574-581.

Hegde, P.S., White, I.R., and Debouck, C. 2003. Interplay of transcriptomics and proteomics. Current Opinion in Biotechnology 14: 647-651.

Hendrix, A. and Hume, A.N. 2011. Exosome signaling in mammary gland development and cancer. Developmental Biology 55: 879-887.

Hennighausen, L. and Robinson G.W. 1998. Think globally, act locally: the making of a mouse mammary gland. Genes and Development 12: 449-455.

Hens, J.R. and Wysolmerski, J.J. 2005. Molecular mechanisms involved in the formation of the embryonic mammary gland. Breast Cancer Research 7(5): 220-224.

Herman, A., Bignon, C., Daniel, N., Grosclaude, J., Gertler, A., and Djiane, J. 2000. Functional heterodimerization of prolactin and growth hormone receptors by ovine placental lactogen. Biological Chemistry 275(9): 6295-6301.

Heuberger, B., Fitzka, I., Wasner, G., and Kratochwil, K. 1982. Induction of androgen receptor formation by epithelium-mesenchyme interaction in embryonic mouse mammary gland. Proceedings of the National Academy of Science 79: 2957-2961.

Hill, D.P., Blake, J.A., Richardson, J.E., and Ringwald, M. 2002. Extension and integration of the Gene Ontology (GO): combining GO vocabularies with external vocabularies. Genome Research 12: 1982-1991.

Hinck, L. and Silberstein, G.B. 2005. The mammary end bud as a motile organ. Breast Cancer Research 7: 245-251.

Horseman, N.D. 1999. Prolactin and mammary gland development. Mammary Gland Biology and Neoplasia 4(1): 79-88.

Hovey, R.C., McFadden, T.B., and Akers, R.M. 1999. Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. Mammary Gland Biology and Neoplasia 14(1): 53-68.

Hovey, R.C., Trott, J.F., and Vonderhaar, B.K. 2002. Establishing a framework for the functional mammary gland: from endocrinology to morphology. Mammary Gland Biology and Neoplasia 7(1): 17-38.

Howe, L.R. and Brown, A.M.C. 2004. Wnt signaling and breast cancer. Cancer Biology and Threapy 3(1): 36-41.

Howlin, J., McBryan, J., and Martin, F. 2006. Pubertal mammary gland development: insights from mouse models. Mammary Gland Biology and Neoplasia 11: 283-297.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research 37(1): 1-13 DOI:10.1093/nar/gkn923.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4(1): 44-57.

Hunt, M.C. and Alexson, S.E.H. 2002. The role of acyl-CoA thioesterases play in mediating intracellular lipid metabolism. Progress in Lipid Research 41: 99-130.

Imagawa, W., Pedchenko, V.K., Helber, J., and Zhang, H. 2002. Hormone/growth factor interactions mediating epithelial/stromal communication in mammary gland development and carcinogenesis. Steroid Biochemistry and Molecular Biology 80: 213-230.

Ioannidis, J.P.A., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Cilhane, A.C., Falchi, M., Furlanello, C., Game., L., Jurman, G., Mangion, J., Mehta, T., Nitzberg, M., Page, G.P., Petretto, E., and Van Noort, V. 2009. Repeatability of published microarray gene expression analyses. Nature Genetics 41(2): 149-155.

Jiline, M., Matwin, S., and Turcotte, M. 2011. Annotation concept synthesis and enrichment analysis: a logic-based approach to the interpretation of high-throughput experiments. Bioinformatics 27(17): 2391-2398.

Joseph, P., O'Kernick, C.M., Othumpangat, S., Lei, Y., Yuan, B.Z., abd Ong, T. 2004. Expression profile of eukaryotic translation factors in human cancer tissues and cell lines. Molecular Carcinogenesis 40:171-179.

Kamath, L., Meydani, A., Foss, F., and Kuliopulos, A. 2001. Signaling from protease-activated receptor-1 inhibits migration and invasion of breast cancer cells. Cancer Research 61(15): 5933-5940.

Kanehisa, M. and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28(1): 27-30.

Kanehisa, M., Goto, S., Murumichi, M., Tanabe, M., and Hirakawa, M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research 38(1): 355-360.

Klarmann, G.J., Decker, A., and Farrar, W.L. 2008. Epigenetic gene silencing in the Wnt pathway in breast cancer. Epigenetics 3(2): 59-63.

Kavaliauskas, D., Nissen, P., and Knudsen, C.R. 2012. The busiest of all ribosomal assistants: elongation factor Tu. Biochemistry 51: 2642-2651.

Kelly, P.A., Bachelot, A., Kedzia, C., Hennighausen, L., Ormandy, C.J., Kopchick, J.J., and Binart, N. 2002. The role of prolactin and growth hormone in mammary gland development. Molecular and Cellular Endocrinology 197: 127-131.

King, H.W., Michael, M.Z., and Gleadle, J.M. 2012. Hypoxic enhancement of exosome release by breast cancer cells. BMC Cancer 12: 421-431.

Kitano, H. 2002. Computational systems biology. Nature 420: 208-210.

Klopfleisch, R. and Gruber, A.D. 2012. Transcriptome and proteome research in veterinary science: what is possible and what questions can be asked? Scientific World Journal 2012: DOI: 10.1100/2012/254962.

Kratochwil, K. 1977. Development and loss of androgen responsiveness in the embryonic rudiment of the mouse mammary gland. Developmental Biology 61: 358-365.

Klinowska, T.C.M., Soriano, J.V., Edwards, G.M., Oliver, J.M., Valentijn, A.J., Montesano, R., and Streuli, C.H. 1999. Laminin and β1 integrins are crucial for normal mammary gland development in the mouse. Developmental Biology 215: 13-32.

Kumar, B.V.S., Bhardwaj, R., Mahajan, K., Kashyap, N., Kumar, A., and Verma, R. 2018. The overexpression of Hsp90B1 is associated with tumorigenesis of canine mammary glands. Mol Cell Biochem 440: 23-31.

Lamote, I., Meyer, E., Massart-Leen, A.M., and Burvenich, C. 2004. Sex steroids and growth factors in the regulation of mammary gland proliferation, differentiation, and involution. Steroids 69(3): 145-159.

Lascelles, A.K. and Lee, C.S. 1978. Involution of the mammary gland. In Larson, B.L and Smith, V.R. (eds.) *Lactation: A Comprehensive Treatise, Volume 4*. New York, NY: Academic Press.

Laubenthal, L., Hoelker, M., Danicke, S., Gerlach, K., Sudekum, K.H., Sauerwein, H., and Haussler, S. 2016. Mitochondrial DNA copy number and biogenesis in different tissues of early- and late-lactating dairy cows. Dairy Science 99(2): 1571-1583.

Levine, J.F. and Stockdale, F.E. 1985. Cell-cell interactions promote mammary epithelial cell differentiation. Cell Biology. 100: 1415-1422.

Li., H. and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5): 589-595.

Li, H. and Homer, N. 2010. A survey of sequence alignment algorithms for next-generation sequencing. Briefings in Bioinformatics 11(5): 473-483.

Li, M., Hu, J., Heermeier, K., Hennighausen, L., and Furth, P.A. 1996. Apoptosis and remodeling of mammary gland tissue during involution proceeds through p53-independent pathways. Cell Growth 7: 13-20.

Li., C., Su., P., and Shyr, Y. 2013. Sample size calculation based on exact test for assessing differential expression analysis in RNA-seq data. BioMed Central

Bioinformatics 14: 357-363.

Li, J.J., Bickel., P.J., and Biggin, M.D. 2014. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. PeerJ 2(207): DOI: 10.7717/peerj.270.

Lianos, G.D., Alexiou, G.A., Mangano, Alb., Mangano, Ale., Rausei, S., Boni, L., Gianlorenzo, D., and Roukos, D.H. 2015. The role of heat shock proteins in cancer. Cancer Letters 360: 114-118.

Liao, D. and Johnson, R.S. 2007. Hypoxia: a key regulator of angiogenesis in cancer. Cancer Metastisis Reviews 26: 281-290.

Liu, X., Robinson, G.W., Wagner, K., Garrett, L., Wynshaw-Boris, A., and Hennighausen, L. 1996. STAT5A is mandatory for adult mammary gland development and lactogenesis. Genes and Development 11: 179-186.

Liu, B.L., Qui, Z., Staron, M., Hong, F., Li, Y., Wu, S., Li, Y., Hao, B., Bona R., Han, D., and Li, Z. 2010. Folding of Toll-like receptors by the HSP90 paralogue gp96 requires a substrate-specific cochaparone. Nature Communications 79: DOI: 10.1038/ncomms1070.

Liu, Y., Zhou, J., and White, K.P. 2014. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30(3): 301-304.

Luga, V., Zhang, L., Viloria-Petit, A.M., Ogunjimi, A.A., Inanlou, M.R., Chiu, E., Buchanan, M., Hosein, A.N., Basik, M., and Wrana, J.L. 2012. Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. Cell 151: 1542-1556.

Maas, J.A., France, J., and McBride, B.W. 1997. Model of milk protein synthesis. A mechanistic model of milk protein synthesis in the lactating bovine mammary gland. Theoretical Biology 187: 363-378.

Maldonado, E.N., Shaldon, K.L., DeHart, D.N., Patnaik, J., Manevich, Y., Townsend, D.M., Bezrukov, S.M., Rostovtseva, T.K., and Lemasters, J.L. 2013. Voltage-dependent anion channels modulate mitochondrial metabolism in cancer cells. Biological Chemistry 288(17): 11920-11929.

Mallard, B.A., Dekkers, J.C., Ireland, M.J., Leslie, K.E., Sharif, S., Vankampen, C.L., Wagter, L., and Wilkie, B.N. 1998. Alteration in immune responsiveness during the peripartum period and its ramifications on dairy cow and calf health. J Dairy Science 81: 585-595.

Marguerat, S. and Bahler, J. 2010. RNA-seq: from technology to biology. Cellular and Molecular Life Sciences 67: 569-579.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9: 387-402.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M.S., and Gilad, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research 18: 1509-1517.

McLachlan, E., Shao, Q., and Laird, D.W. 2007. Connexins and Gap Junctions in Mammary Gland Development and Breast Cancer Progression. J. Membrane Biol 218: 107-121.

McManaman, J.L. and Neville, M.C. 2003. Mammary physiology and milk secretion. Advanced Drug Delivery Reviews 55: 629-641.

Mele, V., Muraro, M.G., Calabrese, D., Pfaff, D., Amatruda, N., Amicarella, F., Kvinlaug, B., Bocelli-Tyndall, C., Martin, I., Resink, T.J., Heberer, M., Oertli, D., Terracciano, L., Spagnoli, G.C., and Iezzi G. 2013. Mesenchymal stromal cells induce epithelial-to-mesenchymal transition in human corectal cancer cells through the expression of surface bound TGF-β. International Journal of Cancer 134(7): DOI: 10.1002/ijc.28598.

Morris, D.R., Ding, Y., Ricks, T.K., Gullapalli, A., Wolfe, B.L., and Trejo, J. 2006. Protease-activated receptor-2 is essential for Factor VIIa and Xa-induced signaling, migration, and invasion of breast cancer cells. Cancer Research 66(1): 307-314.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nature Methods 5(7): 621-628.

Mowry, A.V., Donoveil, Z.S., Kavazis, A.N., Hood, W.R. 2017. Mitochondrial function and bioenergetics trade-offs during lactation in the house mouse (*Mus musculus*). Ecology and Evolution 7: 2994-3005.

Mulac-Jericevic, B., Mullinax, R.A., DeMayo, F.J., Lydon, J.P., and Conneely, O.M. 2000. Subgroup of reproductive functions of progesterone mediated by progesterone receptor-B isoform. Science 289: 1751-1754.

Nagalakshmi, U., Waern, K., and Snyder, M. 2010. RNA-seq: a method for comprehensive transcriptome analysis. Current Protocols in Molecular Biology Chapter 4 (Unit 4.11): 1-13.

Nakayama, K.I. and Nakayama, K. 2006. Ubiquitin ligases: cell-cycle control and cancer. Nature Reviews Cancer 6: 369-381.

Neville, M.C., McFadden, T.B., and Forsyth, I. 2002. Hormonal regulation of mammary differentiation and milk secretion. Mammary Gland Biology and Neoplasia 7(1): 49-66.

Neville, M.C. and Picciano, M.F. 1997. Regulation of milk lipid secretion and composition. Annual Review of Nutrition. 17: 159-184.

Oakes, S.R., Hilton, H.N., and Ormandy, C.J. 2006. The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Research 8(207): DOI: 10.1186/bcr1411.

Ozsolak, F. and Milos, P.M. 2011. RNA sequencing: advances, challenges, and opportunities. Nature Reviews Genetics 12(2): 87-98.

Pai, V.P. and Horseman, N.D. 2011. Mammary gland involution: events, regulation, and influences on breast disease. In Carrasco, J. and Mota, M. (eds.) *Endothelium and Epithelium*. Nova Science Publishers, Inc. Hauppauge, NY.

Parmar, H. and Cunha, G.R. 2004. Epithelial-stromal interactions in the mouse and human mammary gland *in vivo*. Endocrine-Related Cancer 11: 437-458.

Pareek, C.S., Smoczynski, R., and Tretyn, A. 2011. Sequencing technologies and genome sequencing. Applied Genetics 52: 413-435.

Pepke, S., Wold, B., Mortazavi, A. 2009. Computation for ChIP-seq and RNA-seq studies. Nature 6(11): 522-532.

Pop, M. and Salzberg, S.L. 2007. Bioinformatics challenges of new sequencing technology. Trends in Genetic 24(3): 142-149.

Porterfield, S.P. and White, B.A. 2007. In *Endocrine Physiology,* Third Edition. Mosby, Inc. Philadelphia, PA.

Propper, A.Y. and Gomot, L. 1973. Control of chick epidermis differentiation by rabbit mammary mesenchyme. Experientia 29: 1543-544.

Radisky, D.C. and Hartmann, L.C. 2009. Mammary involution and breast cancer risk: transgenic models and clinical studies. Mammary Gland Biology and Neoplasia 14: 181-191.

Rawson, P., Stockum, C., Peng, L., Manivannan, B., Lehnert, K., Ward, H.E., Berry, S.D., Davis, S.R., Snall, R.G., McLauchlan, D., and Jordan, T.W. 2012. Metabolic proteomics of the liver and mammary gland during lactation. Proteomics 75: 4429-4435.

Raynaud, A. 1950. Experimental research on the development of the reproductive and functioning of the endocrine glands of fetal mouse and field mouse. Archives of Microscopic Anatomy and Experimental Morphology 39: 518-576.

Richert, M.A., Schwertfeger, K.L., Ryder, J.W., and Anderson, S.M. 2000. An atlas of mouse mammary gland development. Mammary Gland Biology and Neoplasia 5(2): 227-241.

Robinson, G.W., Karpf, A.B.C., and Kratochwil, K. 1999. Regulation of mammary gland development by tissue interaction. Mammary Gland Biology and Neoplasia 4(1): 9-19.

Romanello, M., Moro, L., Pirulli, D., Crovella, S., and D'Andrea, P. 2001. Effects of cAMP on intracellular coupling and osteoblast differentiation. J Biochemical and Biophysical Research Communications 282: 1138-1144.

Ronnov-Jessen, L., Petersen, O.W., and Bissell, M.A. 1996. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiological Reviews 76(1): 69-125.

Rudolph, M.C., McManaman, J.L., Hunter, L., Phang, T., and Neville, M.C. 2003. Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. J. Mammary Gland Biology and Neoplasia 8(2): 287-307.

Russo, J. and Russo, I.H. 1987. Development of the human mammary gland. In Neville, M.C. and Daniel, C.W. (eds.) *The Mammary Gland: Development, Regulation and Function*. Plenum, New York.

Russo, J., Hu, Y.F., Silva, I.D.C.G., and Russo, I.H. 2001. Cancer risk related to mammary gland structure and development. Miscroscopy Research and Tequnique 52: 204-223.

Sakakura, T., Nishizuka, Y., and Dawe, C.J. 1976. Mesenchyme-dependent morphogenesis and epithelium-specific ctyodifferentiation in mouse mammary gland. Science 194: 1439-1441.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. 2006. Geration of a functional mammary gland from a single stem cell. Nature 439(5): 84-88.

Shamay, A., Shapiro, F., Leitner, G., and Silanikove, N. 2003. Infusion of casein hydrolyzates into the mammary gland disrupt tight junction intergrity and induce involution in cows. Dairy Science 86(4): 1250-1258.

Shennan, D.B. and Peaker, M. 2000. Transport of milk constituents by the mammary gland. Physiological Reviews 80(3): 925-951.

Shenoy, N., Kessel, R., Bhagat, T.D., Bhattacharyya, S., Yu, Y., McMahon, C., and Verma, A. 2012. Alterations in the ribosomal machinery in cancer and hematologic disorders. Hematology and Oncology 5(1): 32-40.

Silberstein, G.B. 2001. Role of the stroma in mammary development. Breast Cancer Research 3: 218-224.

Simons, M. and Raposos, G. 2009. Exosomes- vesicular carriers for intracellular communication. Current Opinion in Cell Biology 21: 575-581.

Soneson, C. and Delorenzi, M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. BioMed Central Bioinformatics 14: 91-108.

Splendiani, A., Donato, M., and Draghici, S. 2014. In *Ontologies for Bioinformatics*. Springer Hanbook of Bio/Neuroinformatic. Springer. Berlin, Heidelberg.

Storey, J.D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100(16): 9440-9445.

Strange, R., Li, F., Saurer, S., Burkhardt, A., and Friis, R.R. 1992. Apoptotic cell death and tissue remodeling during mouse mammary gland involution. Development 115: 49-58.

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.G., Lander, E.S., and Mesirov, J.P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wise expression profiles. Proceedings of the National Academy of Sciences 120(43): 15545-15550.

Svennersten-Sjaunja, K. and Olsson, K. 2005. Endocrinology of milk production. Domestic Animal Endocrinology 29: 241-258.

Swindell, W.R., Masternak, M.M., Kopchick, J.J., Conover, C.A., Bartke, A., and Miller, R.A. 2009. Endocrine regulation of heat shocl protein mRNA levels in long-lived dwarf mice. Mechanisms of Aging and Development 130: 393-400.

Tanos, T., Sflomos, G., Echeverria, P.C., Ayyanan, A., Guiterrez, M., Delaloye, J., Raffoul, W., Fiche, M., Dougall, W., Schneider, P., Yalcin-Ozuysal, O., and Brisken, C. 2013. Progesterone/RankL is a major regulatory axis in the human breast. Science Translational Medicine 5(182): DOI: 10.1126/scitranslmed.3005654.

Theocharis, A.D., Skandalis, S.S., Tzanakakis, G.N., and Karamanos, N.K. 2010. Proteoglycans in health and disease: novel roles for proteoglycans in malignancy and their pharmacological targeting. FEBS Journal 277: 3904-3923.

Thompson, A., Brennan, K., Cox, A., Gee, J., Harcourt, D., Harris, A., Harvie, M., Holen, I., Howell, A., Nicholson, R., Steel, M., Streuli, C. 2008. Evaluation of the current knowledge limitations in breast cancer research: a gap analysis. Breast Cancer Research 10: R26 DOI: 10.1186/bcr1983.

Trapnell, C., Hendrickson, D.G., Sauvagea, M., Goff, L., Rinn, J.L., and Patcher, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology 31(1): 46-54.

Trapnell, C., Roberts, A. Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel., H., Slazberg, S.L., Rinn, J.L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7(3): 562-578.

Troup, S., Njue, C., Kliewer, E.V., Parisien, M., Roskelley, C., Chakravarti, S., Roughley, P.J., Murchey, L.C., Watson, P.H. 2003. Reduced expression of the small leucine-rich proteoglycans, lumican, and decorin is associated with poor outcome in node-negative invasive breast cancer. Clinical Cancer Research 9: 207-214.

Tucker, H.A. 1981. Physiological control of mammary growth, lactogenesis, and lactation. Dairy Science 64: 1403-1421.

Turkington, R.W. and Hill, R.L. 1969. Lactose synthetase: progesterone inhibition of the induction of α-lactalbumin. Science 163(3874): 1458-1460.

Uvnas-Moberg, K. and Eriksson, M. 1996. Breastfeeding: physiological, endocrine and behavioral adaptions caused by oxytocin and local neurogenic activity in the nipple and mammary gland. Acta Paediatrica 85: 525-530.

Vogel, C. and Marcotte, E.M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews Genetics 13(4): 227-232.

Wadden, P.D., Ruan, W., Feldman, M., and Kleinberg, D.L. 1998. Evidence that the mammary fat pad mediates the action of growth hormone in mammary gland development. Endocrinology 139: 659-662.

Wall, E.H., Crawford, H.M., Ellis, S.E., Dahl, G.E., and McFadden, T.B. 2006. Mammary response to exogenous prolactin or frequent milking during early lactation in dairy cows. Dairy Science 89(12): 4640-4648.

Warner, J.R. and McIntosh, K.B. 2009. How common are extraribosomal functions of ribosomal proteins? Molecular Cell 34: 3-11.

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

Wickramasinghe, S., Rincon, G., Islas-Trejo, A. and Medrano, J.F. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BioMed Central Genomics 13(1): 45-58.

Wysolmerski, J.J., Philbrick, W.M., Dunbar, M.E., Lanske, B., Kronenberg, H., and Broadus, A.E. 1998. Rescue of the parathyroid hormone-related protein knockout mouse demonstrates that parathyroid hormone-related protein is essential for mammary gland development. Development 125:1285-1294.

Yang, E., Cisowski, J., Nguyen, N., O'Callaghan, K., Xu, J., Agarwal., A., Kuliopulos, A., and Covic, L. 2015. Dysregulated protease activated receptor 1 (PAR1) promotes metastatic phenotype in breast cancer through HMGA2. Oncogene 35: 1529-1540.

Zaragoza, R., Garcia, C., Rus, A.D., Pallardo, F.V., Barber, T., Torres, L., Miralles, V.J., and Vina, J.R. 2003. Inhibition of liver trans-sulphuration pathway by proparglycine mimics gene expression changes found in the mammary gland of weaned lactating rats: role of glutathione. J Biochem 373: 825-834.

Zaragoza, R., Garcia-Trevijano, E.R., Lluch, A., Ribas, G., and Vina, J.R. 2015. Involvement of different networks in mammary gland involution after the pregnancy/lactation cycle: implications in breast cancer. Intl Union Biochemistry Molecular Biol 37(4): 227-238.

Zhang, D., Liu, Z., Choi, C.S., Tian, L., Kibbey, R., Dong, J., Cline, G.W., Wood, P.A., and Shulman, G.I. 2007. Mitochondrial dysfunction due to long-chain Acyl-CoA dehydrogenase deficiency causes hepatic steatosis and hepatic insulin resistance. PNAS 104(43): 17075-17080.

Zhou, J., Chehab, R., Tkalcevic, J., Naylor, M.J., Harris, J., Wilson, T.J., Tsao, S., Zavarsek, S., Xu, D., Lapinskas, E.J., Visvader, J., Lindeman, G.J., Thomas, R., Ormandy, C.J., Hertzog, P.J., Kola, I., and Pritchard, M.A. 2005. Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. European Molecular Biology Organization Journal 24: 635-644.

Zhou, J., Wulfkuhle, J., Zhang, H., Gu, P., Yang, Y., Deng, J., Margolick, J.B., Liotta, L.A., Petricoin III, E., and Zhang, Y. 2007. Activation of the

PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance. PNAS 104(41): 16158-16163.

Zhou, Y.Y., Gone, W., Xiao, J.F., Wu, J.Y., Pan, L.L., Li, X.N., Wang, X.M., Wang, W.W., Hu, S.N., and Yu, J. 2014. Transcriptomic analysis reveals key regulators of mammogenesis and the pregnancy-lactation cycle. Life Sciences 57(3): 340-355.

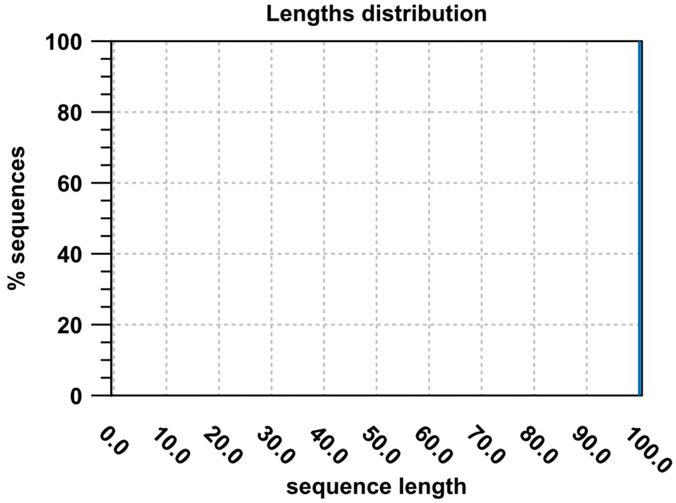
Zinser, G., Packman, K., and Welsh, J. 2002. Vitamin D_3 receptor ablation alters mammary gland morphogenesis. Development 129: 3067-3076.

APPENDIX A – CLC Genomics Workbench Sequencing QC Reports

JFM49A, Sequencing Index 2, Virgin Mouse 1

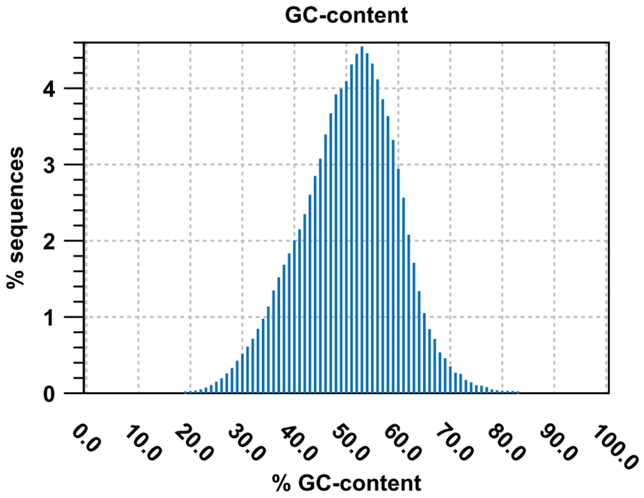
Sequencing QC Report
Based upon: 31,241,022 sequences in 9 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 17:47:47 PST 2013
Software: CLC Genomics Workbench 6.5

2.1 Lengths distribution



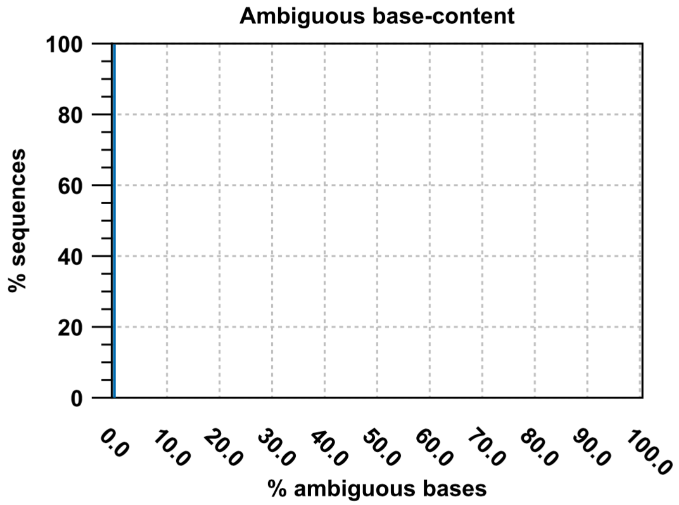
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



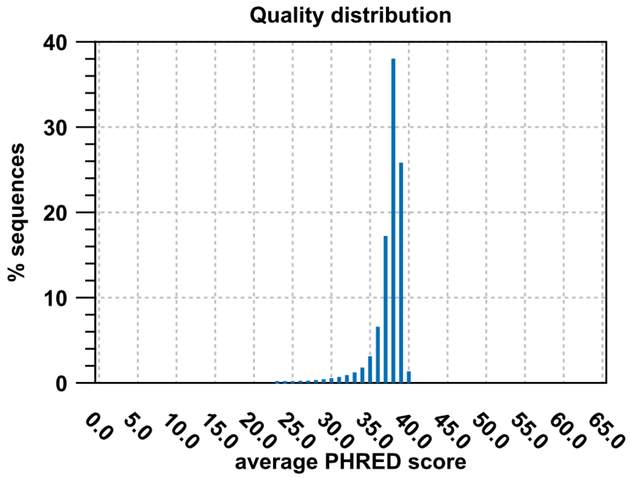
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



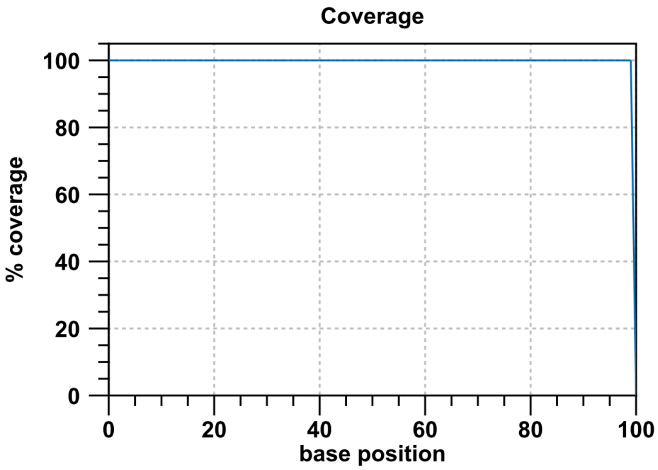
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences

2.4 Quality distribution



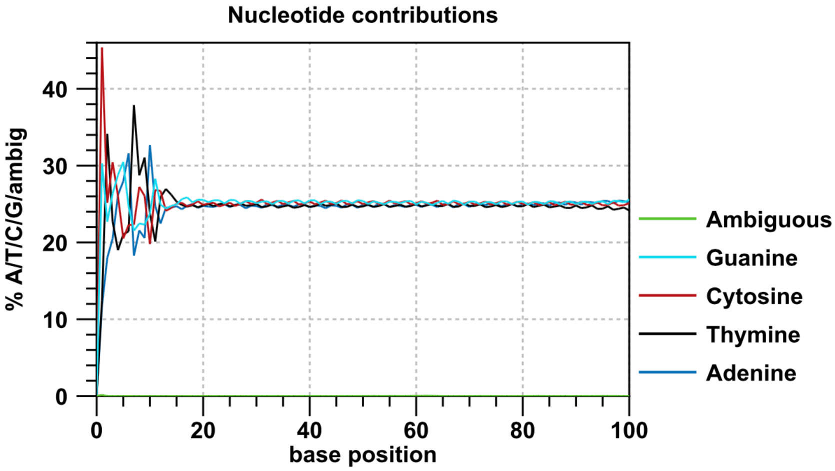
Distribution of average sequence qualitie scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences

3.1 Coverage



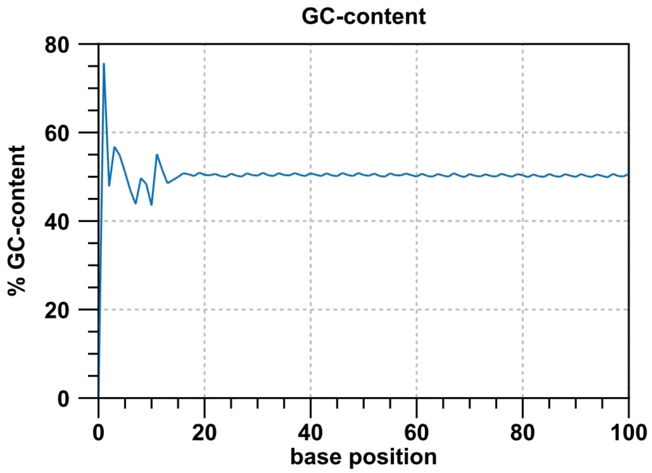
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



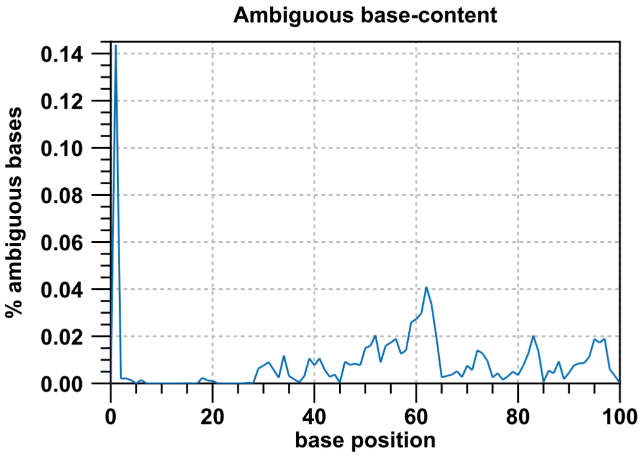
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



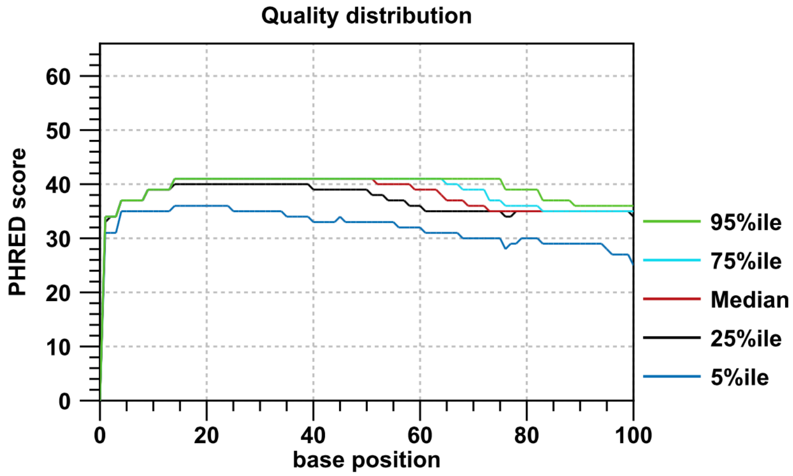
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



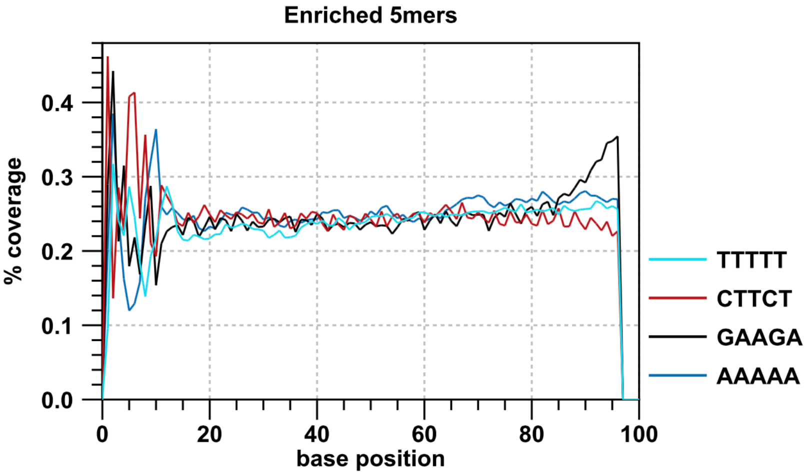
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at that position

3.5 Quality distribution



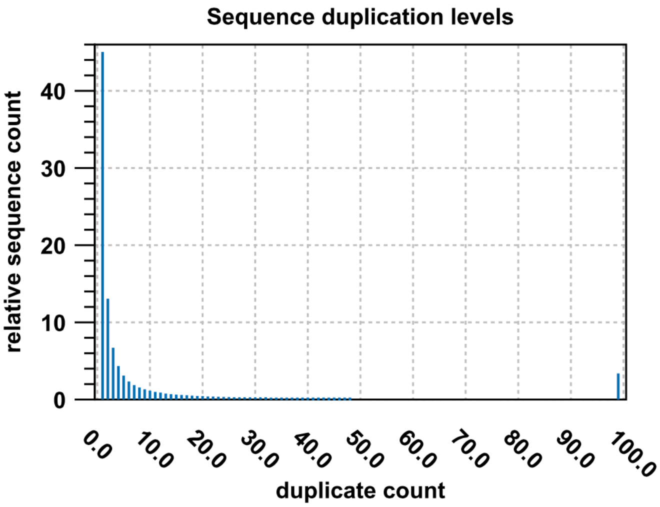
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

4.1 Enriched 5mers



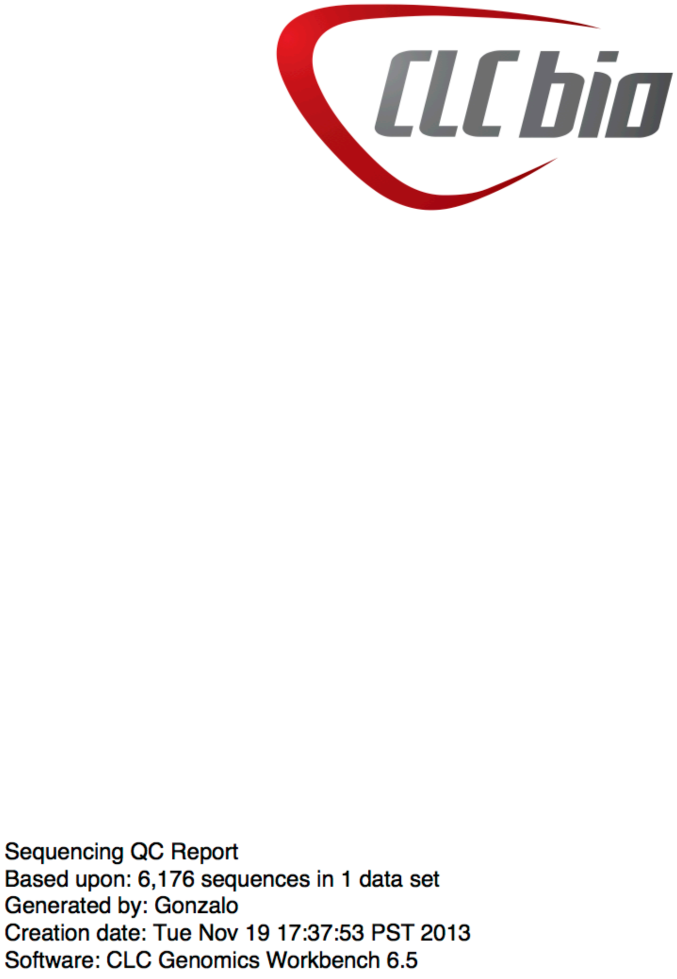
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the observed and expected 5mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5mer. (5mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion

4.2 Sequence duplication levels



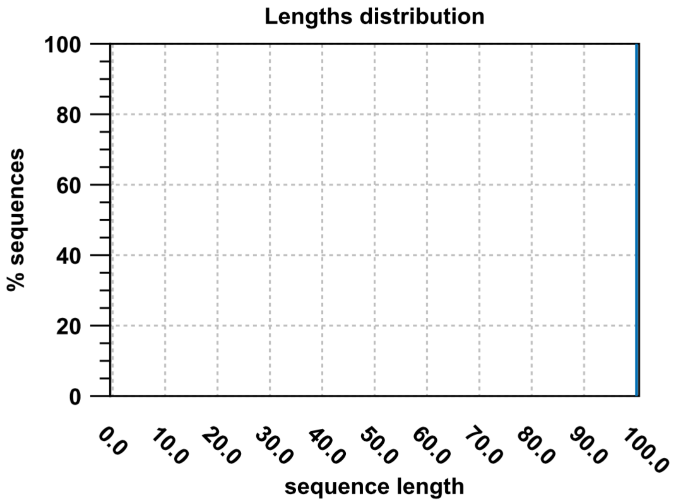
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences

JFM49B, Sequencing Index 4, Virgin Mouse 2



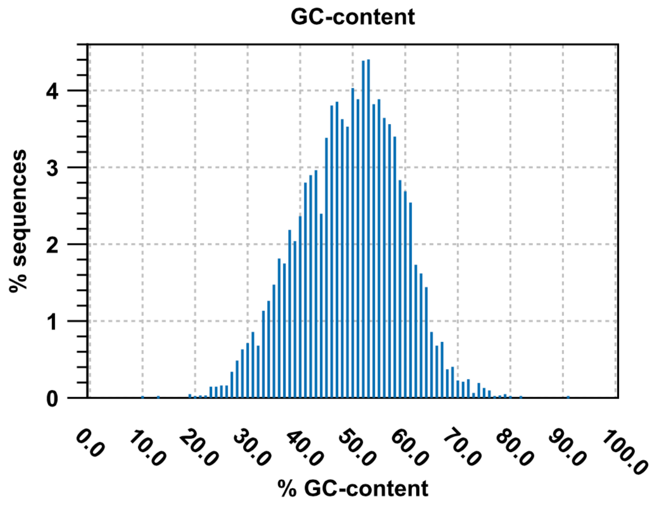
Sequencing QC Report
Based upon: 6,176 sequences in 1 data set
Generated by: Gonzalo
Creation date: Tue Nov 19 17:37:53 PST 2013
Software: CLC Genomics Workbench 6.5

2.1 Lengths distribution



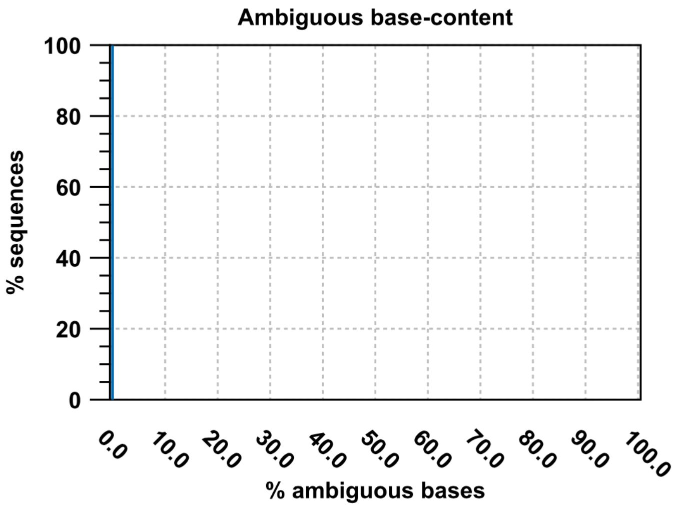
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



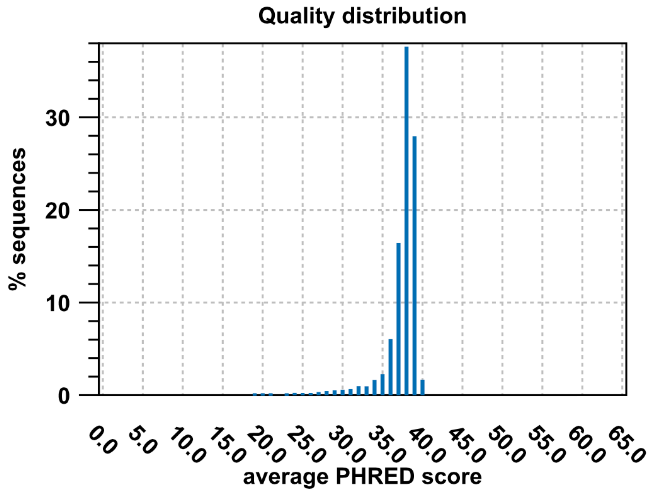
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



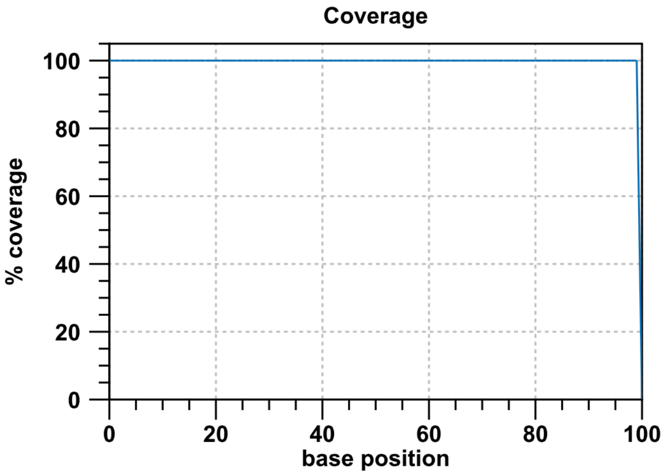
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



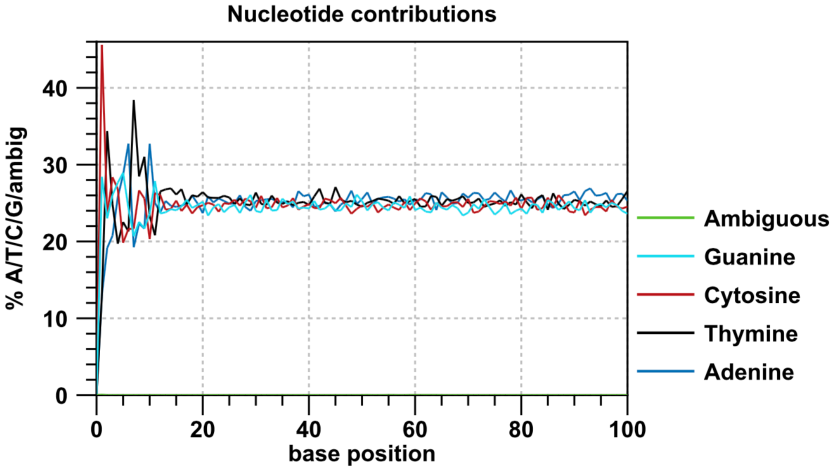
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



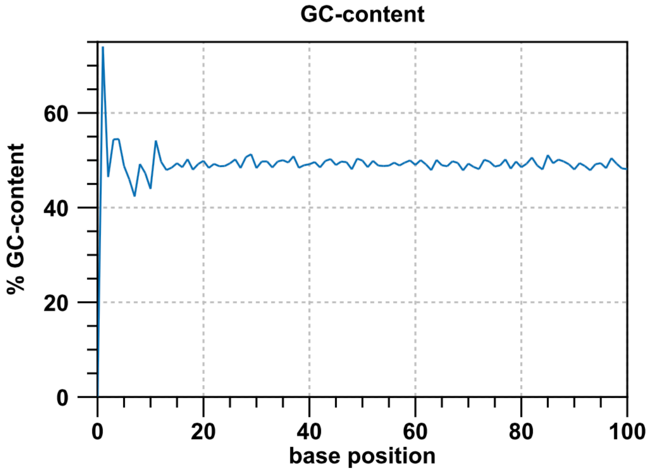
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



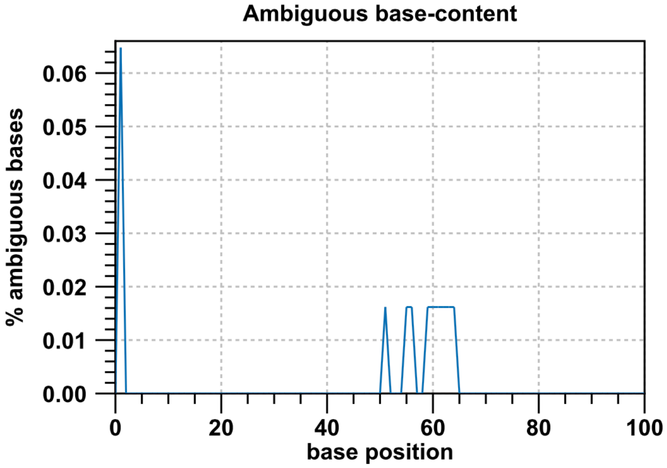
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



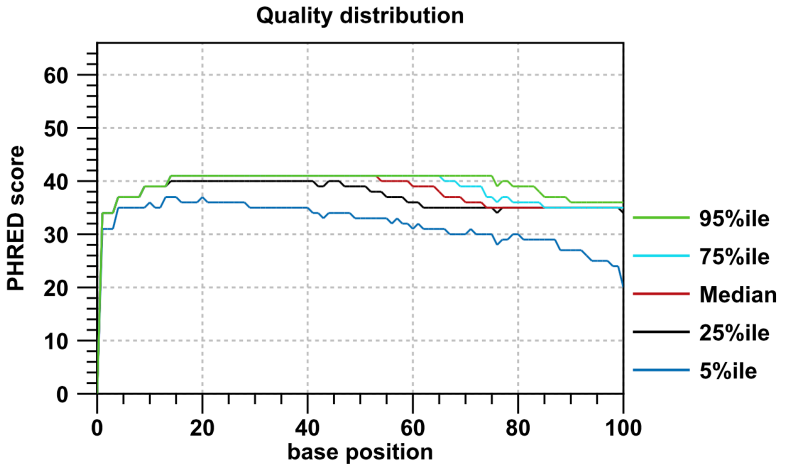
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



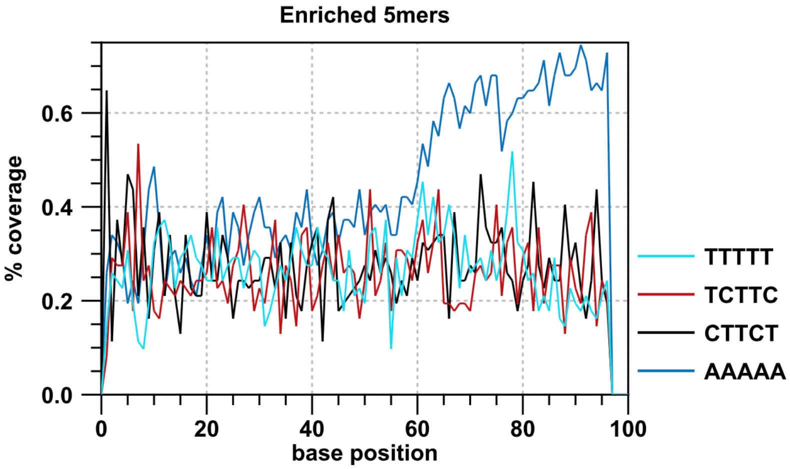
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



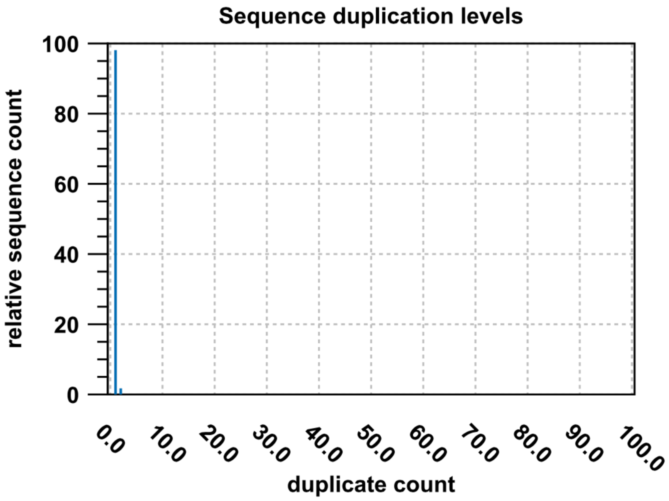
```
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position
```

4.1 Enriched 5mers



```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

4.2 Sequence duplication levels

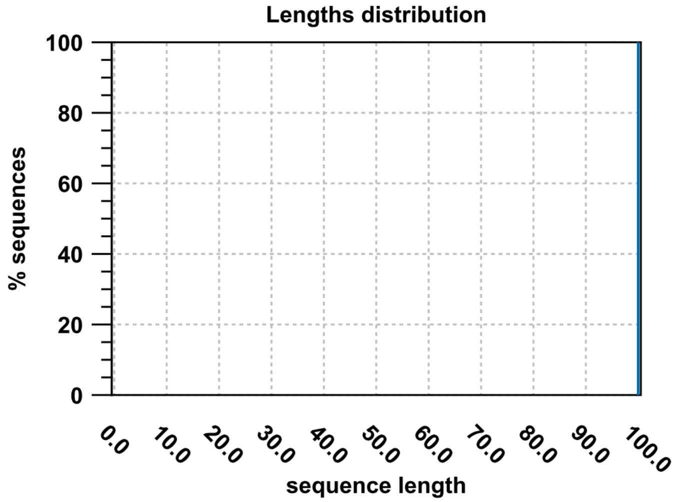


```
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has
been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences
```

JFM49C, Sequencing Index 5, Virgin Mouse

Sequencing QC Report
Based upon: 16,864,550 sequences in 5 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 17:47:29 PST 2013
Software: CLC Genomics Workbench 6.5

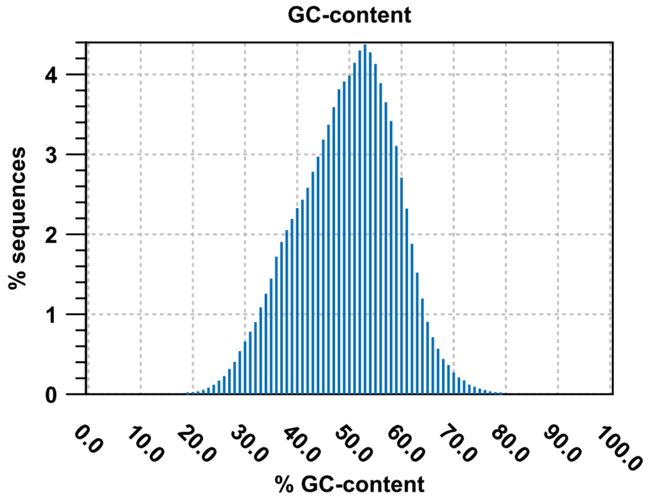
2.1 Lengths distribution



Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs

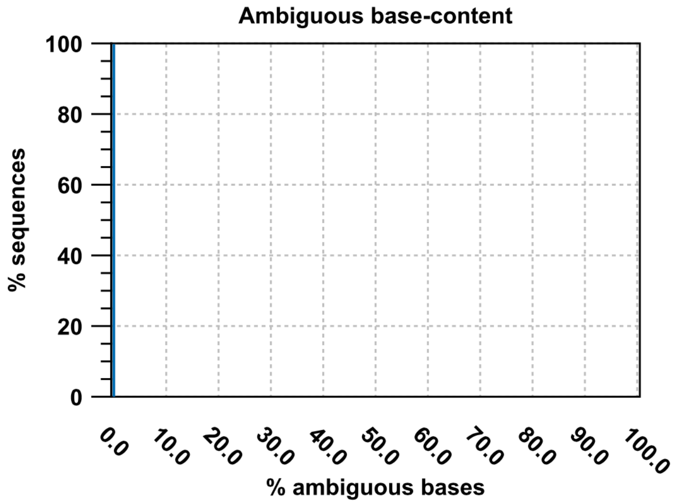
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



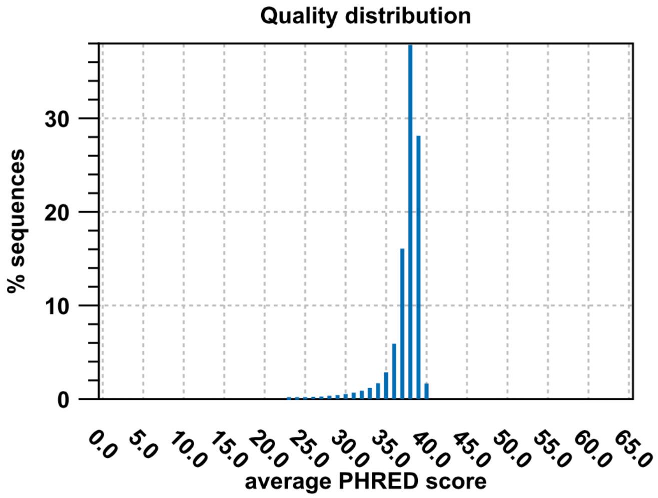
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



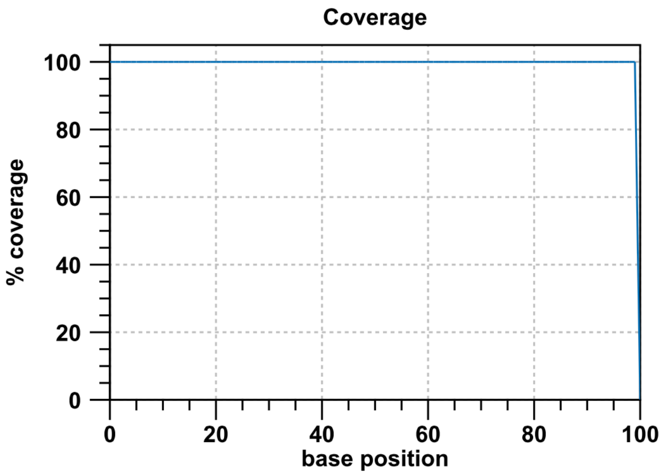
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences

2.4 Quality distribution



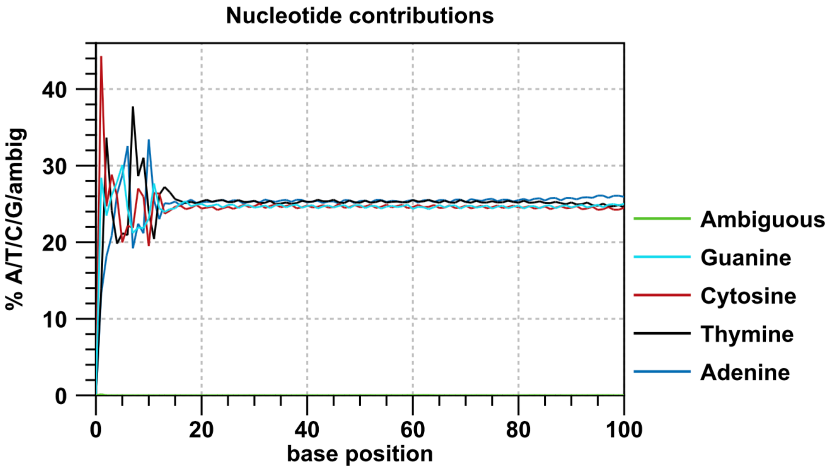
Distribution of average sequence qualitie scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences

3.1 Coverage



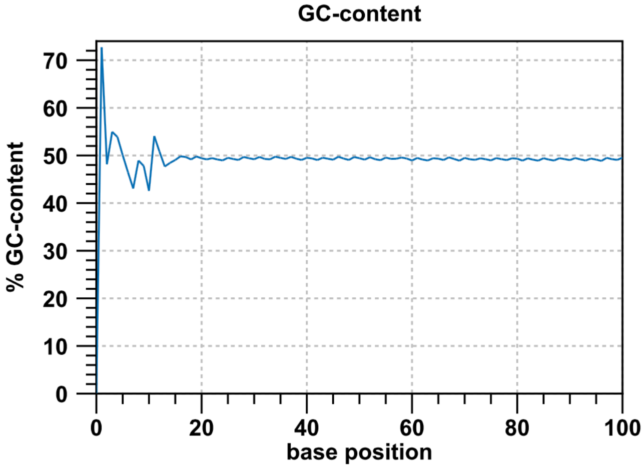
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



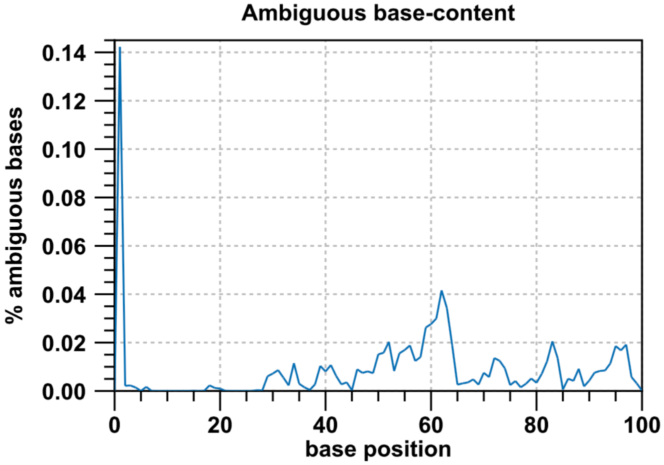
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



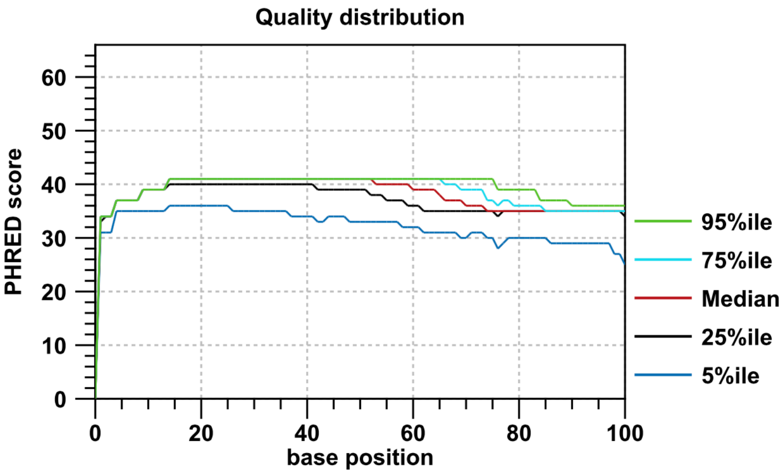
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



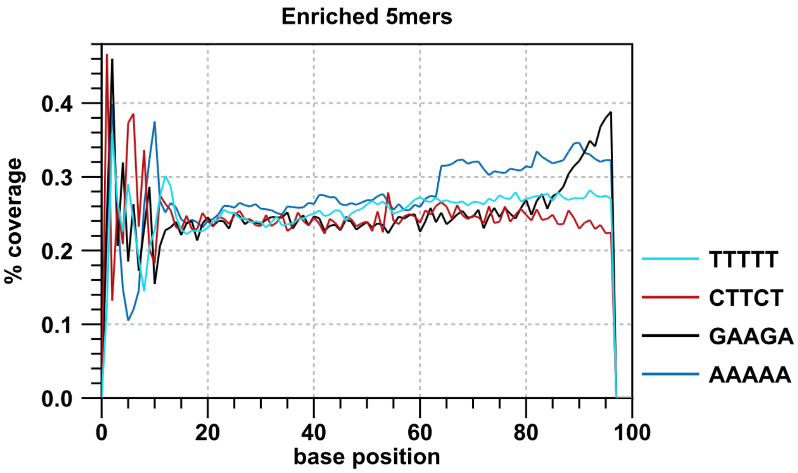
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



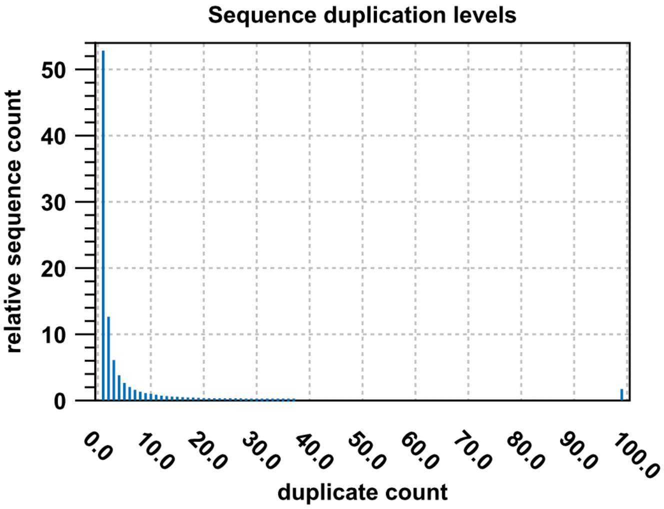
```
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position
```

4.1 Enriched 5mers



```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

4.2 Sequence duplication levels

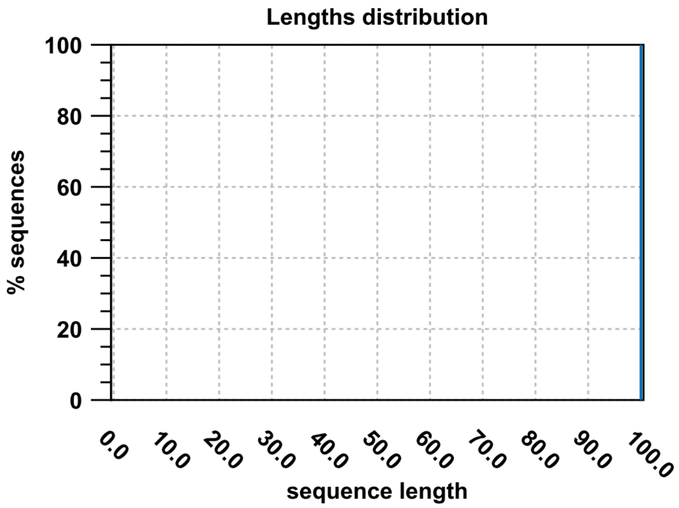


```
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has
been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences
```

JFM49D, Sequencing Index 6, Pregnant Mouse 1

Sequencing QC Report
Based upon: 16,666,247 sequences in 5 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 17:45:40 PST 2013
Software: CLC Genomics Workbench 6.5

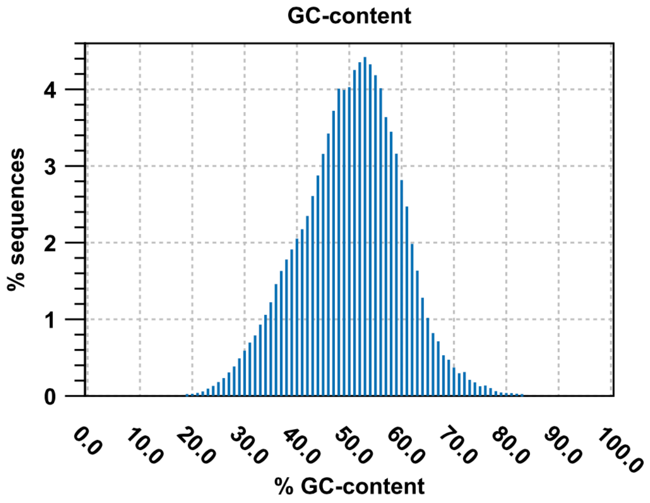
2.1 Lengths distribution



Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs

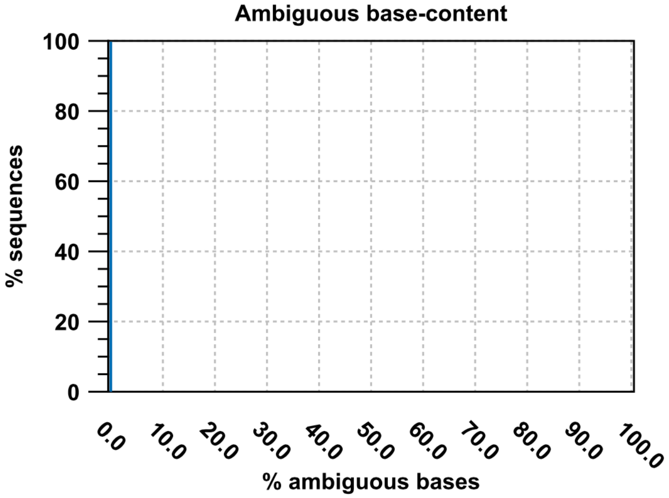
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



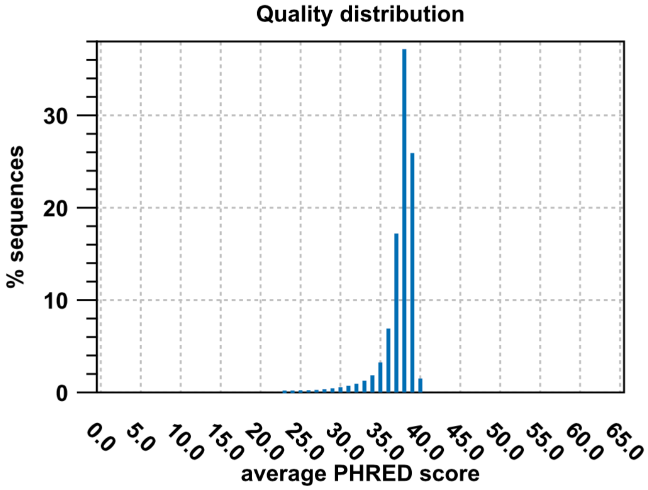
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



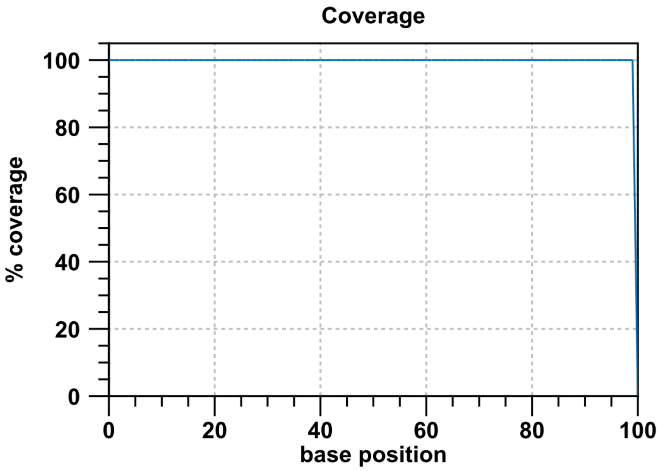
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



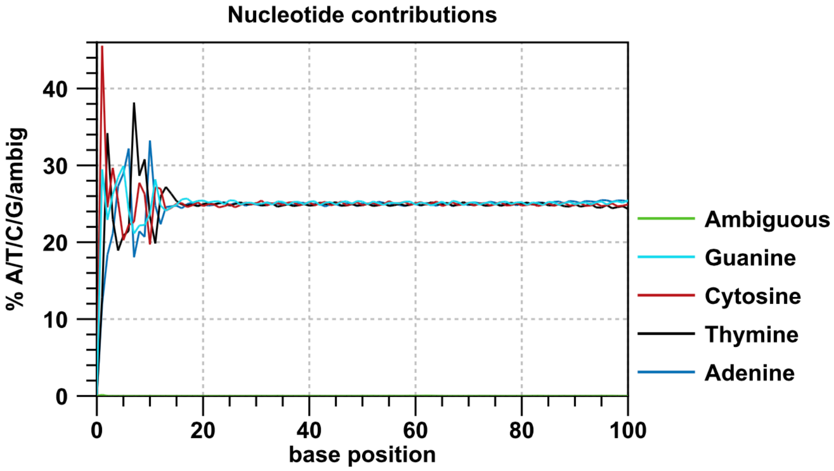
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



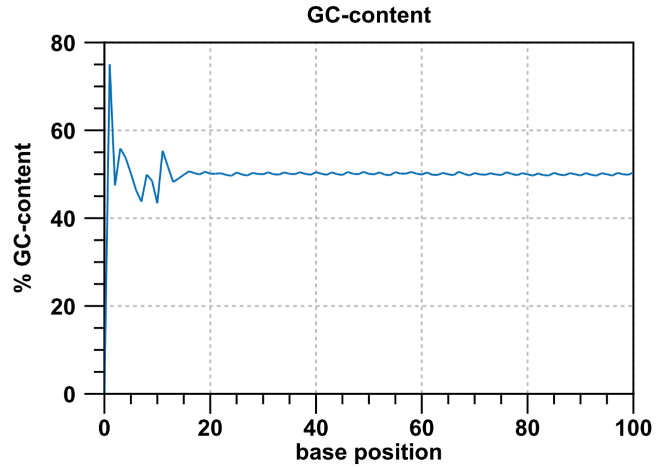
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



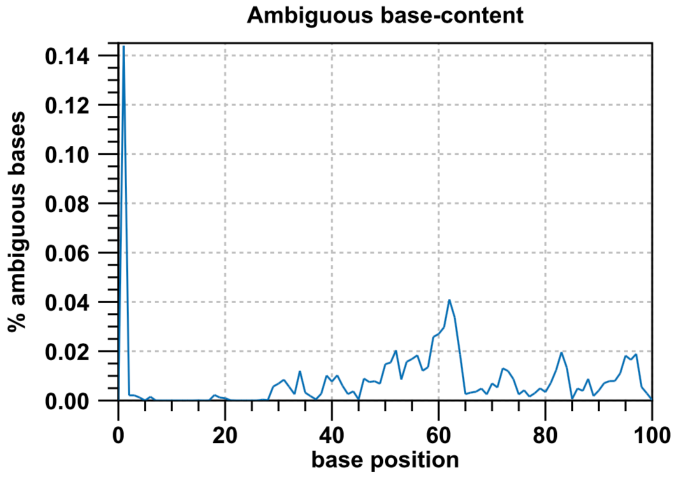
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



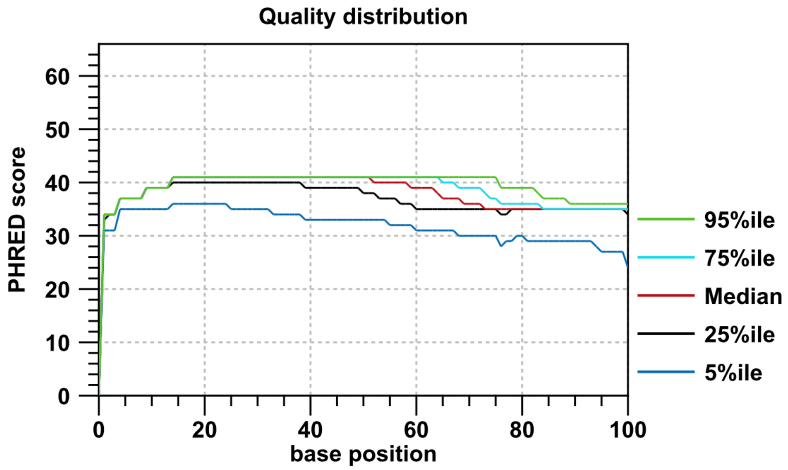
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



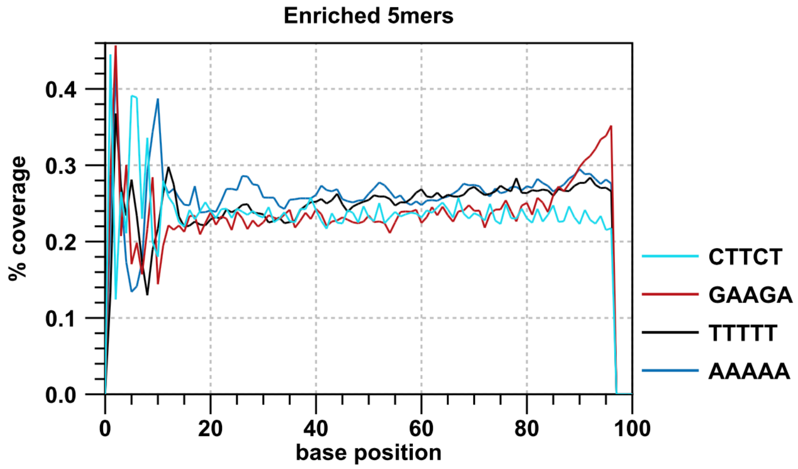
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at that position

3.5 Quality distribution



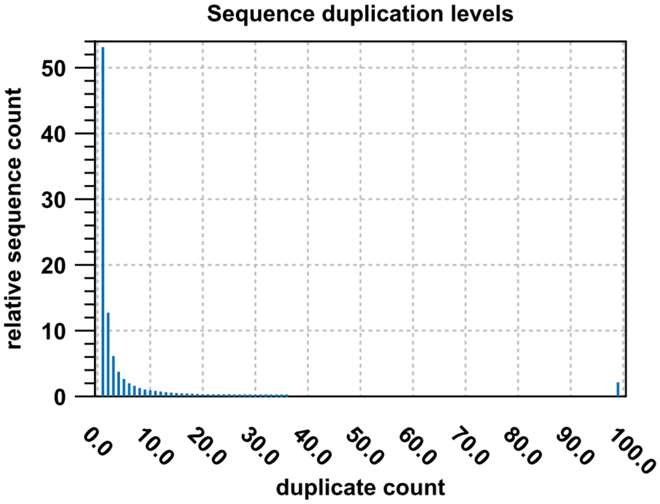
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

4.1 Enriched 5mers



The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the observed and expected 5mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5mer. (5mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion

4.2 Sequence duplication levels

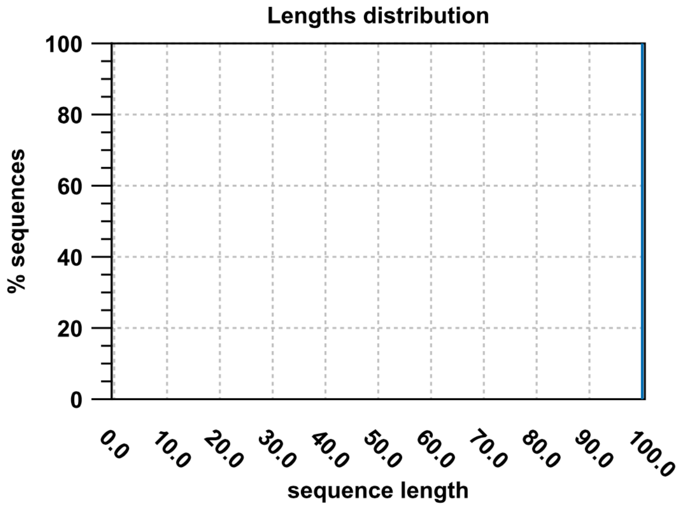


```
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has
been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences
```

JFM49E, Sequencing Index 7, Pregnant Mouse 2

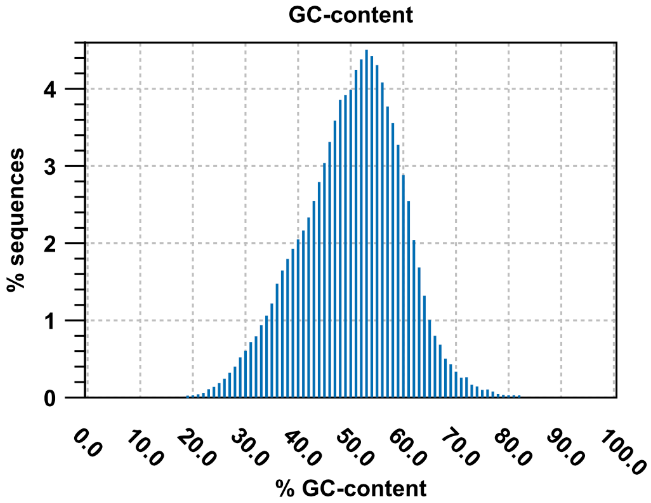
Sequencing QC Report
Based upon: 11,861,826 sequences in 4 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 18:01:20 PST 2013
Software: CLC Genomics Workbench 6.5

2.1 Lengths distribution



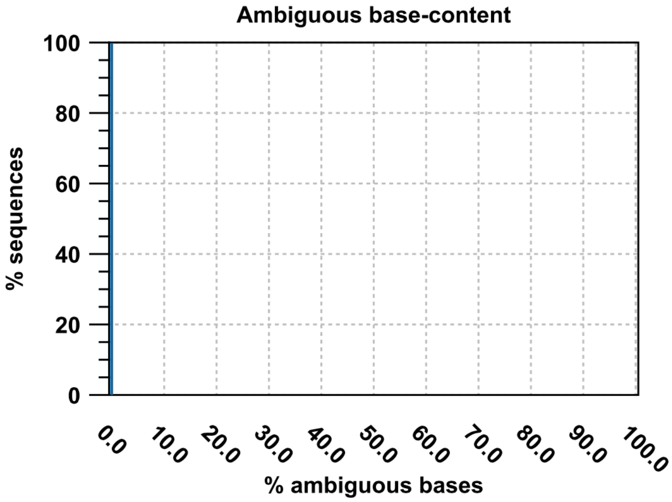
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



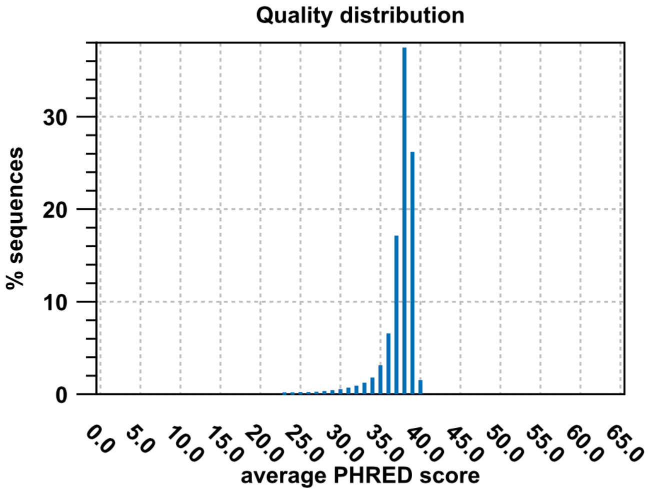
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



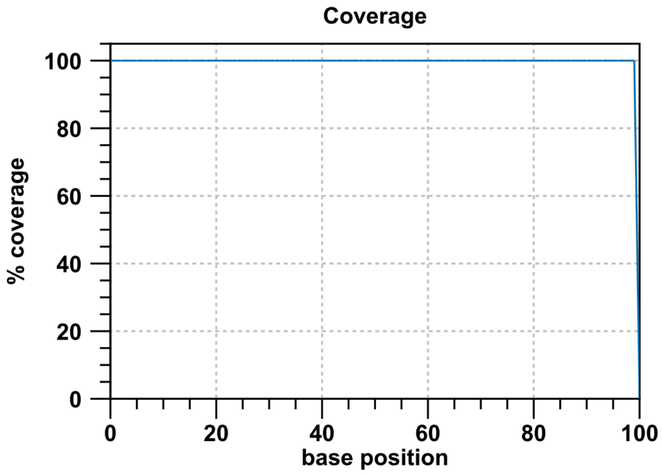
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



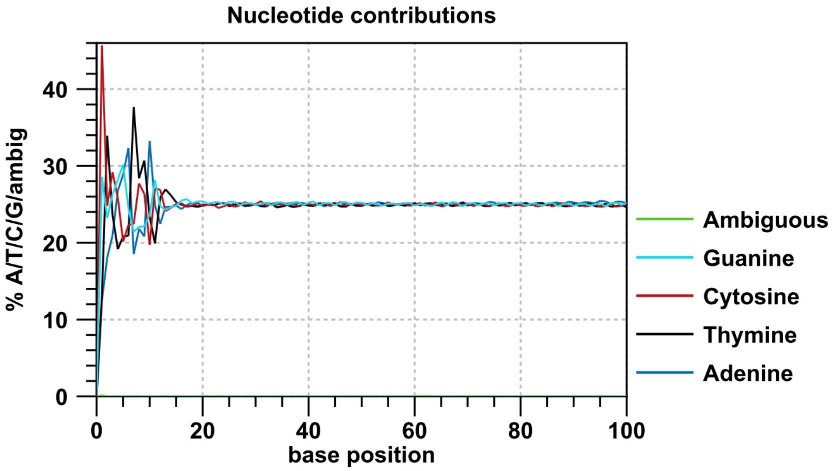
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



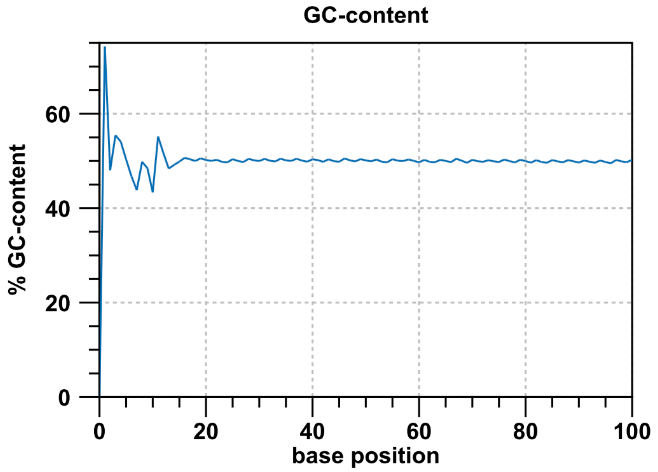
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



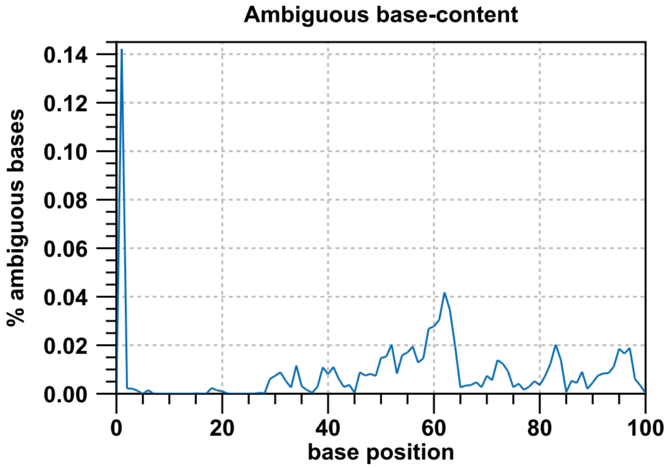
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



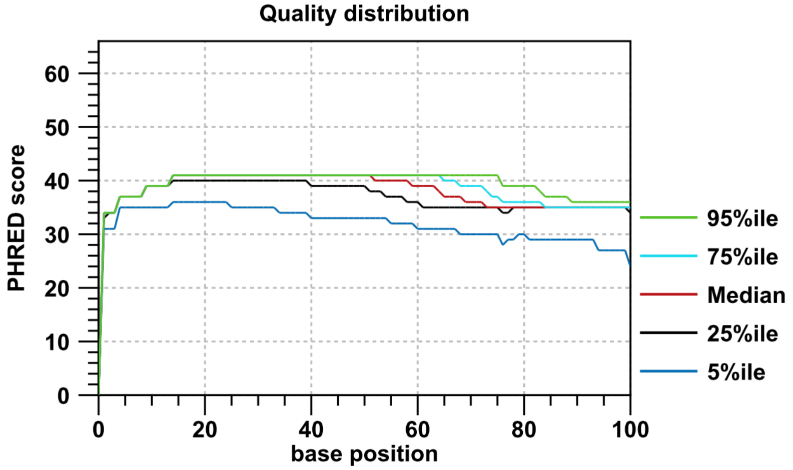
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



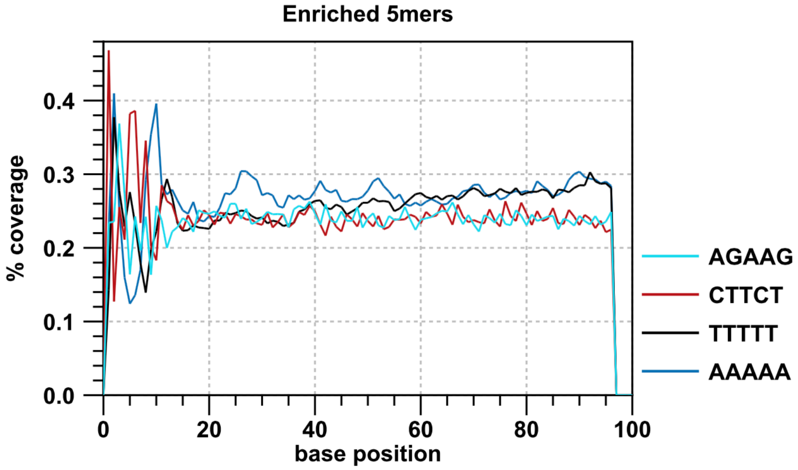
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



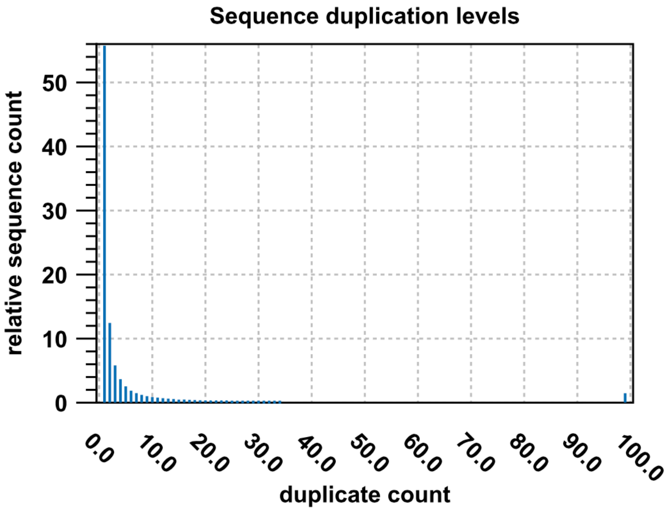
```
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position
```

4.1 Enriched 5mers



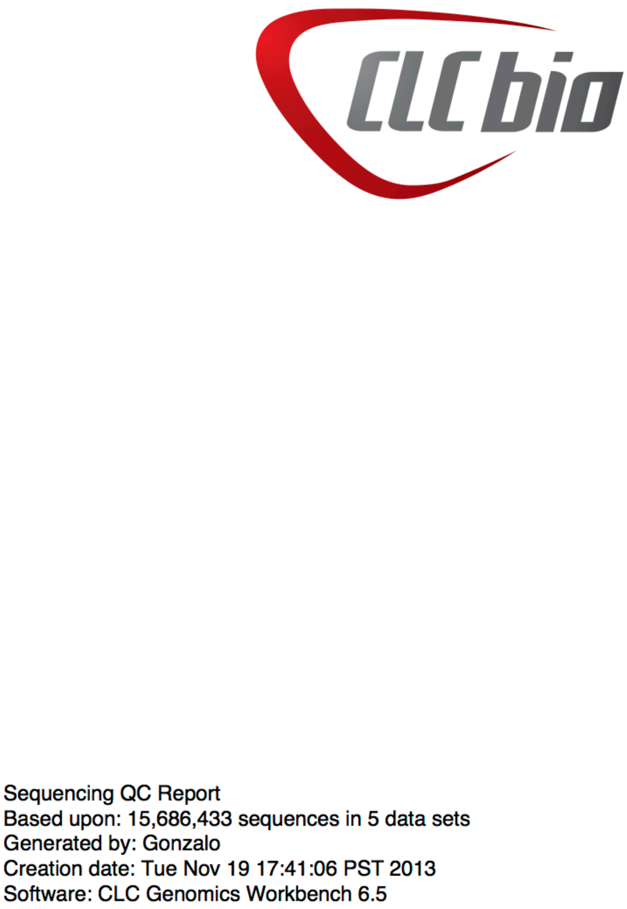
```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

4.2 Sequence duplication levels



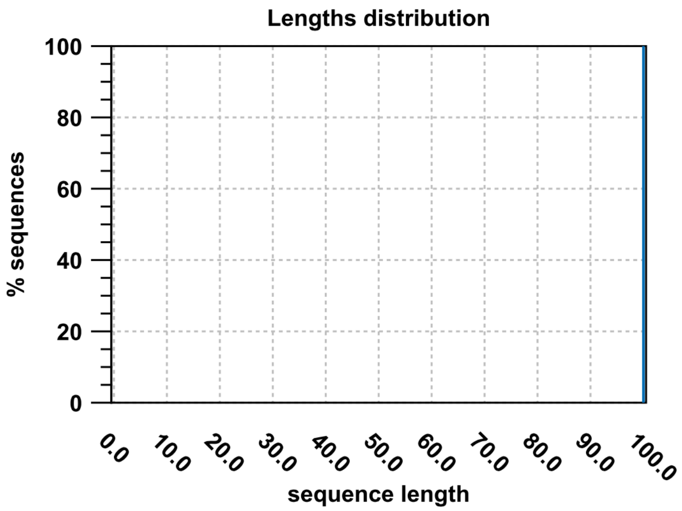
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences

JFM49F, Sequencing Index 12, Pregnant Mouse 3



Sequencing QC Report
Based upon: 15,686,433 sequences in 5 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 17:41:06 PST 2013
Software: CLC Genomics Workbench 6.5

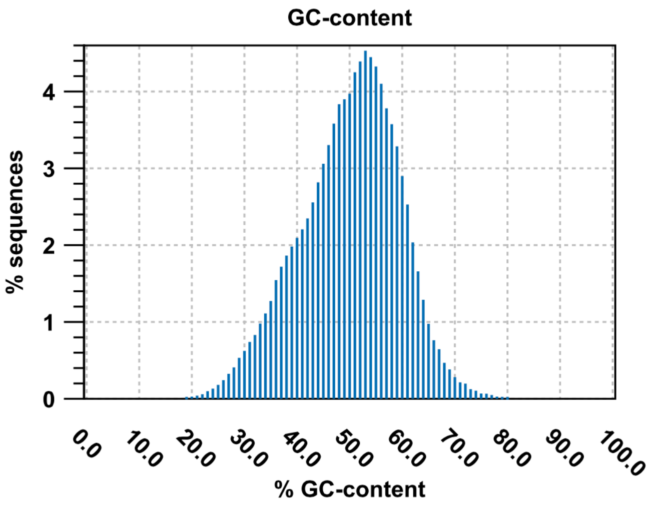
2.1 Lengths distribution



Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs

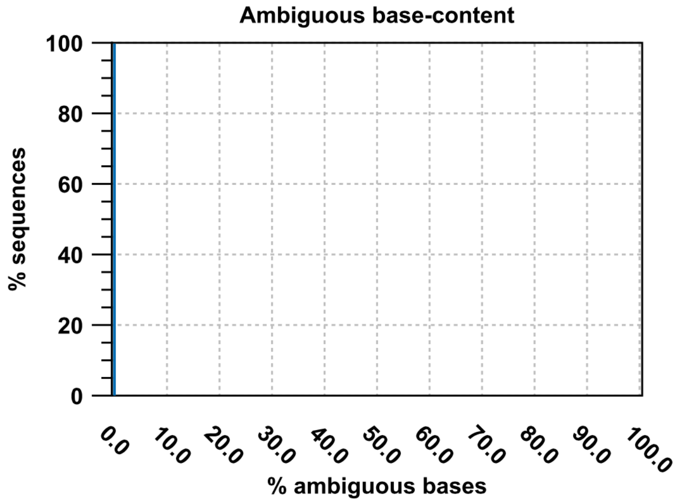
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



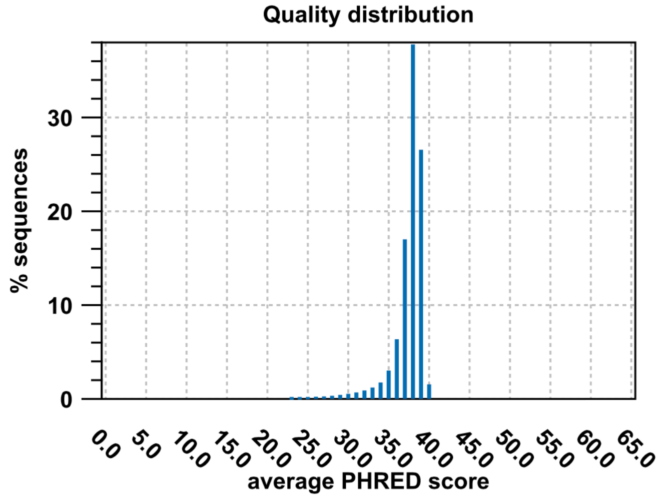
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



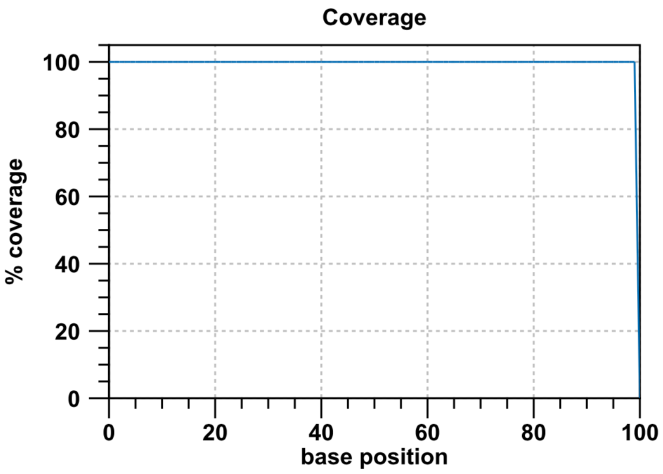
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



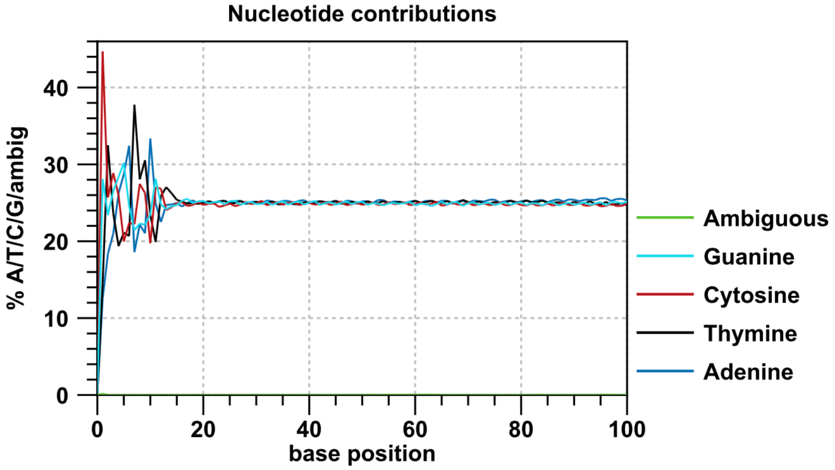
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



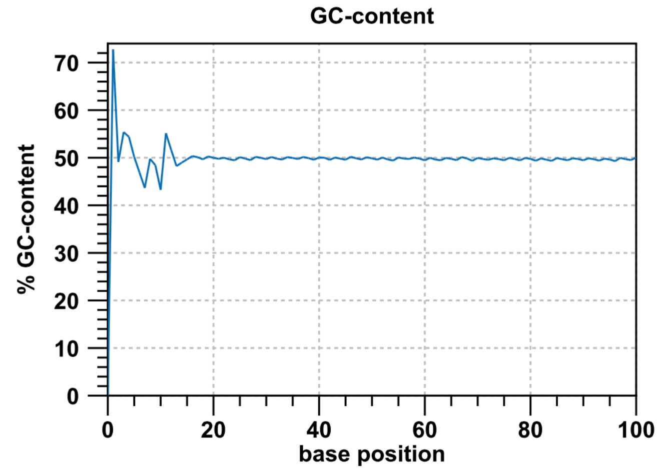
```
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or
SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences
```

3.2 Nucleotide contributions



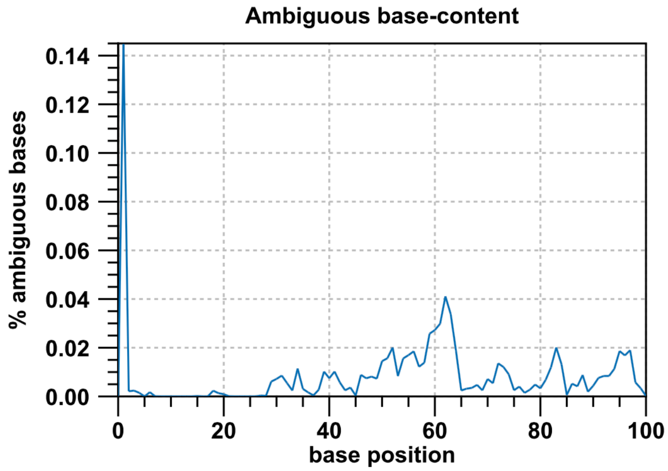
```
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that
position
```

3.3 GC-content



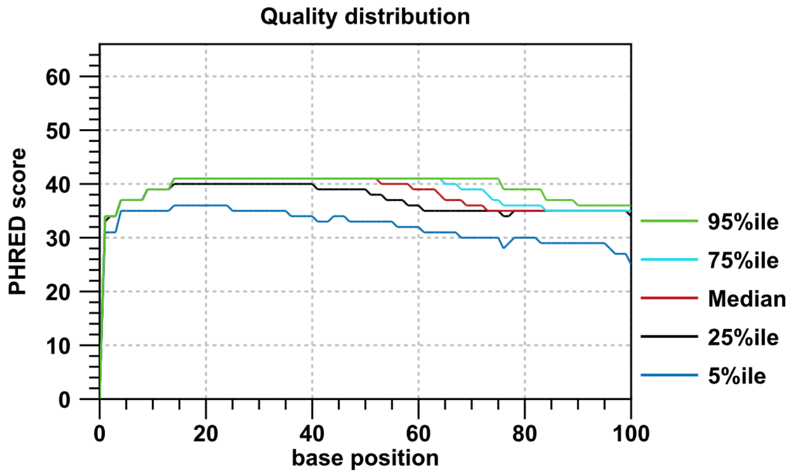
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



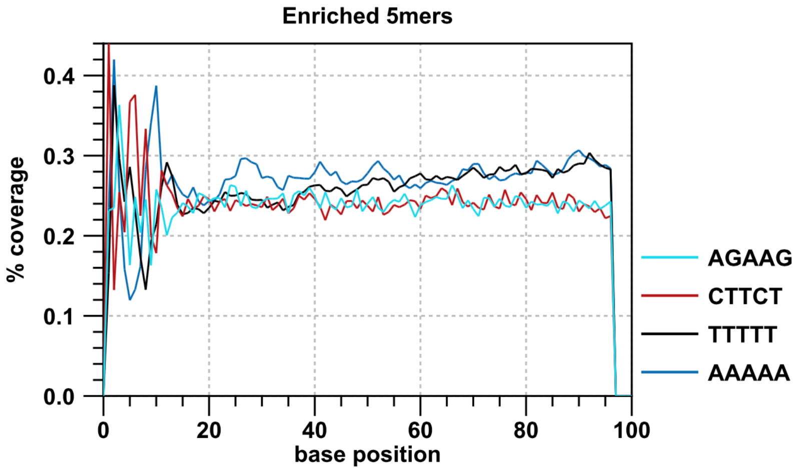
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



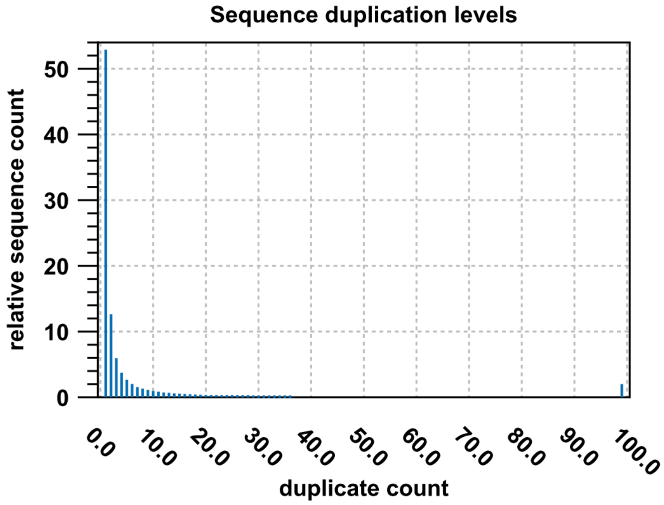
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

4.1 Enriched 5mers



The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the observed and expected 5mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5mer. (5mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion

4.2 Sequence duplication levels

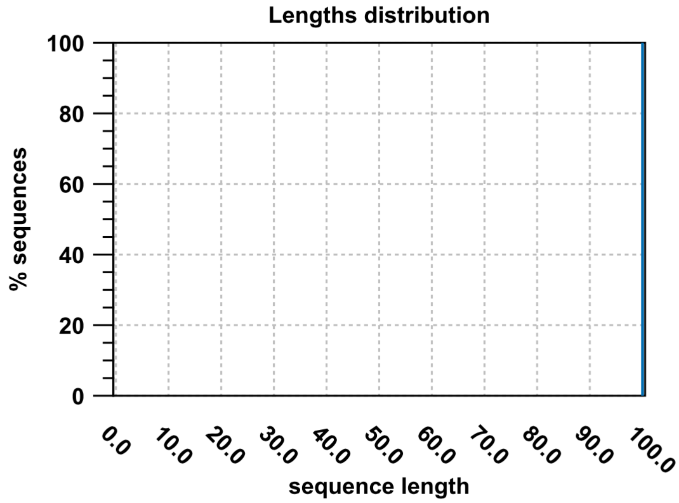


Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences

JFM49G, Sequencing Index 13, Quiescent Mouse 1

Sequencing QC Report
Based upon: 18,222,830 sequences in 5 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 18:05:02 PST 2013
Software: CLC Genomics Workbench 6.5

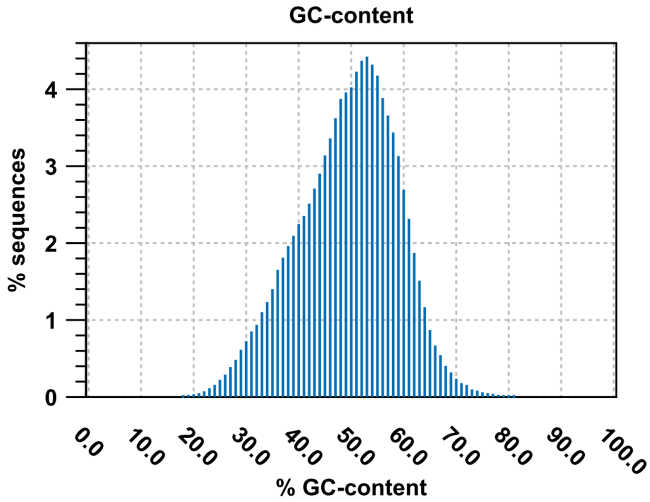
2.1 Lengths distribution



```
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single
peak.
x: sequence length in base-pairs
```

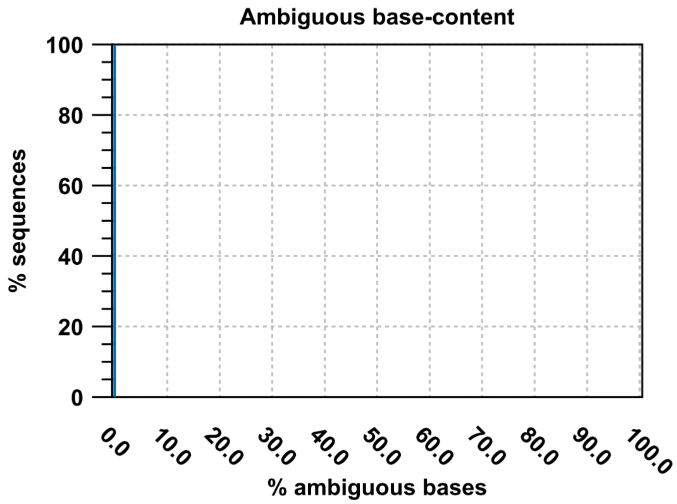
```
y: number of sequences featuring a particular length normalized to the total number of seq uences
```

2.2 GC-content



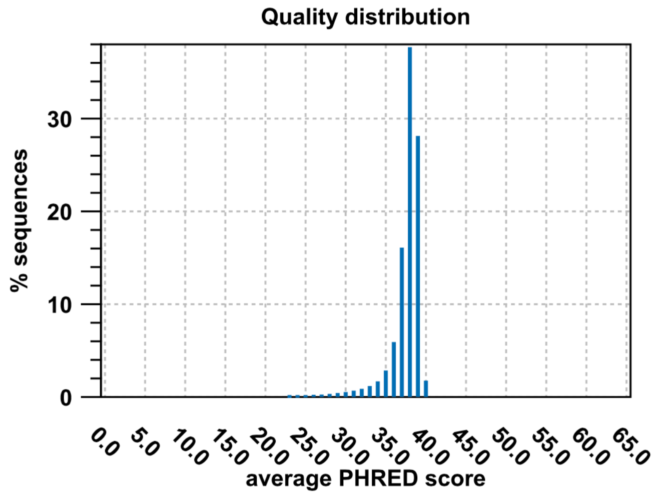
```
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all
bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number  of sequences
```

2.3 Ambiguous base-content



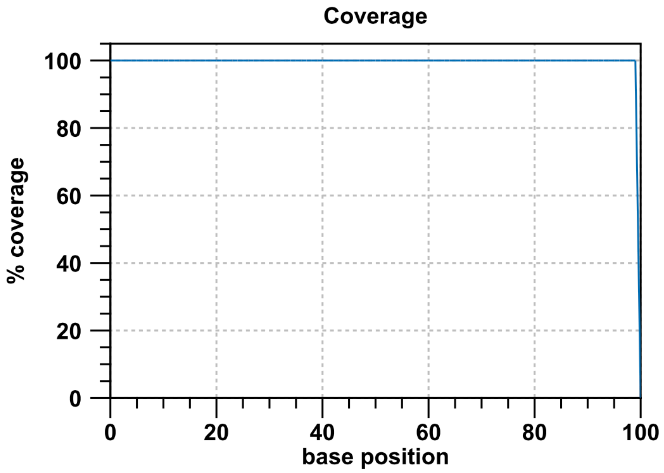
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



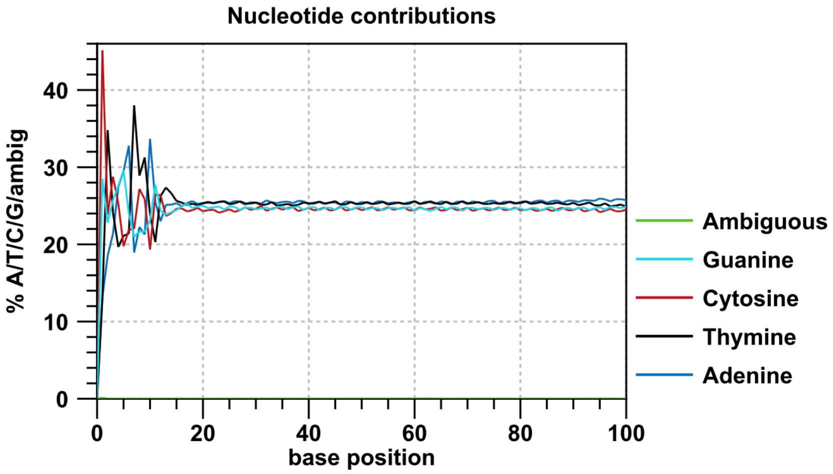
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



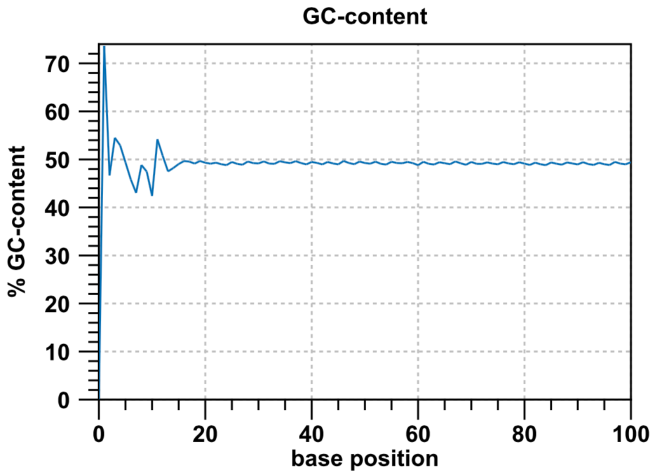
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



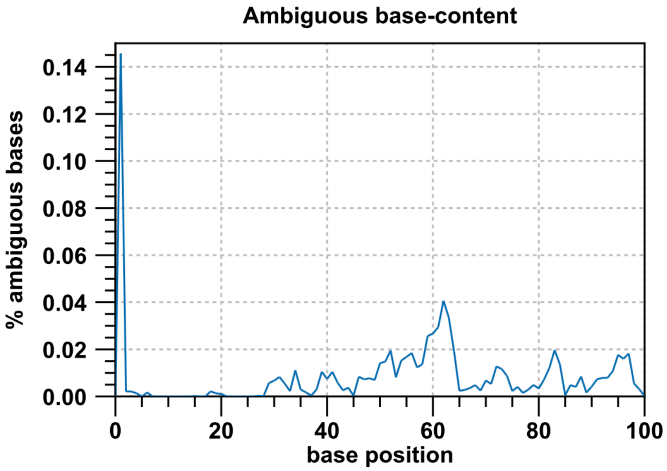
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



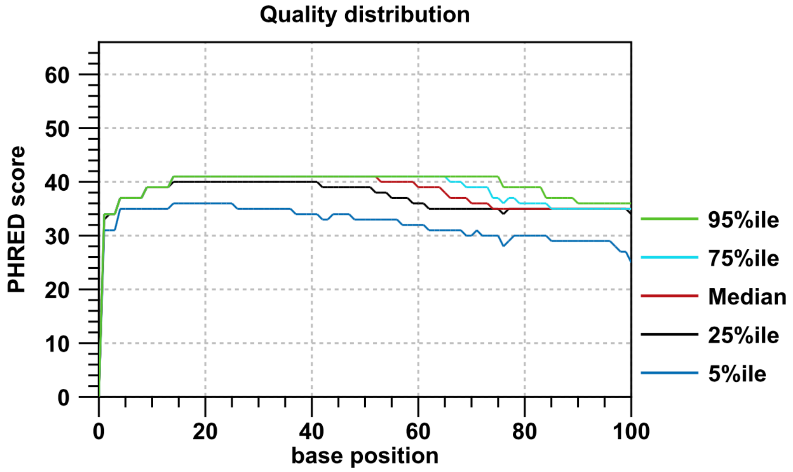
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



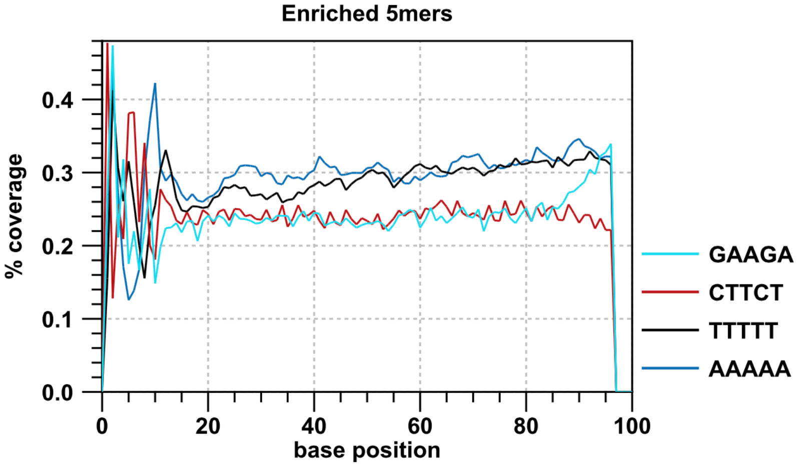
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



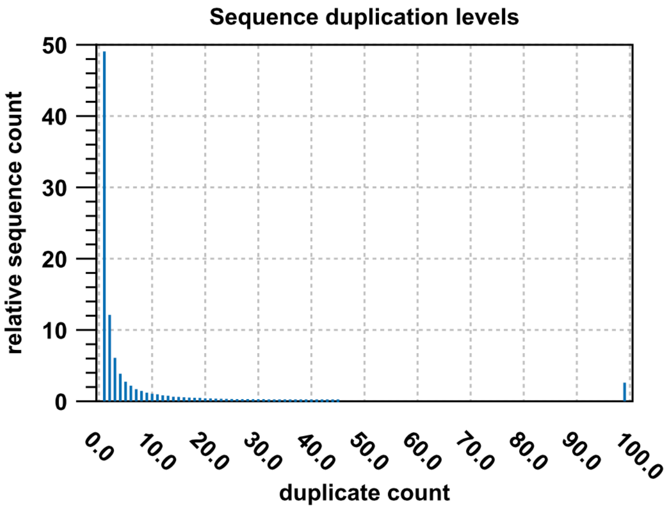
```
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position
```

4.1 Enriched 5mers



```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

4.2 Sequence duplication levels

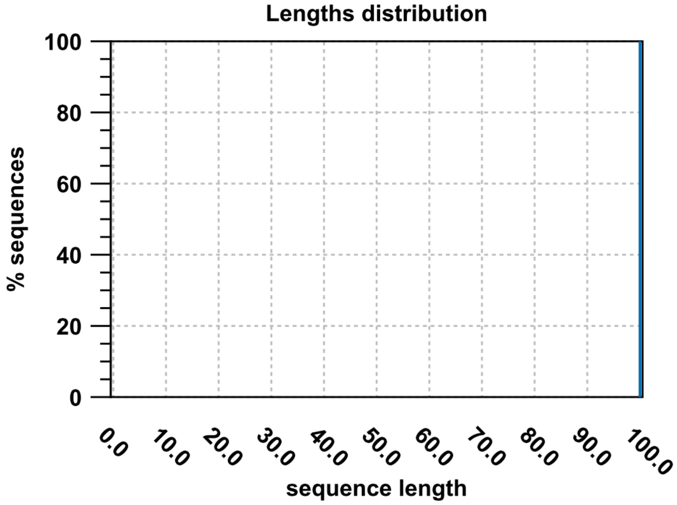


```
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has
been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences
```

JFM49H, Sequencing Index 14, Quiescent Mouse 2

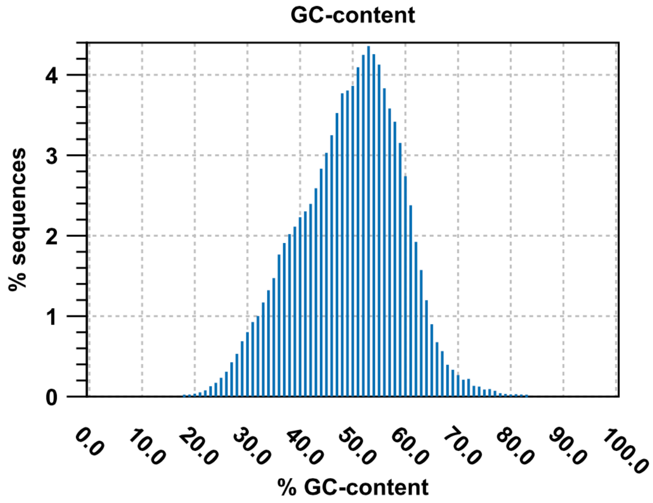
Sequencing QC Report
Based upon: 20,998,909 sequences in 6 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 18:05:09 PST 2013
Software: CLC Genomics Workbench 6.5

2.1 Lengths distribution



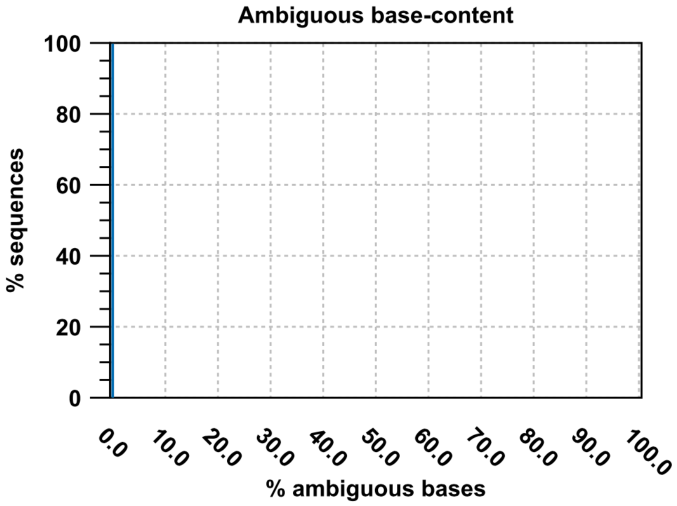
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



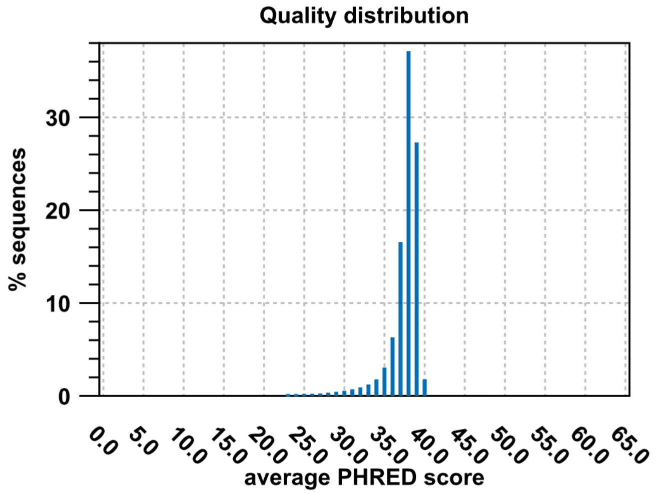
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



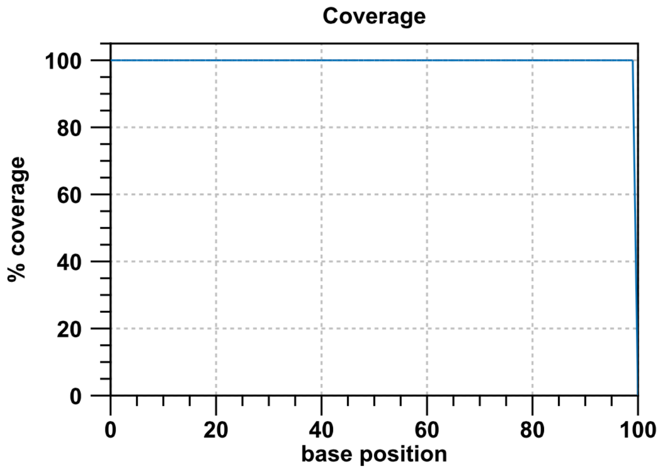
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



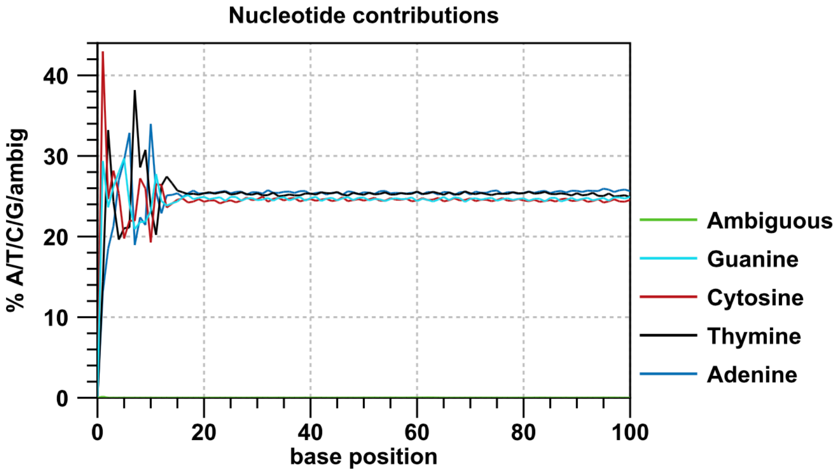
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



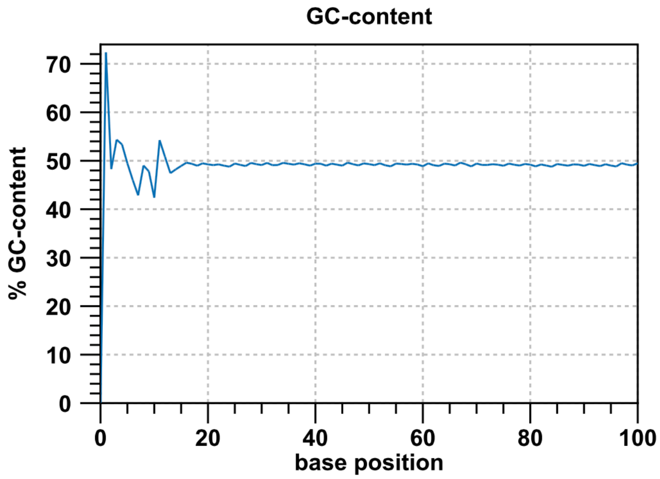
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



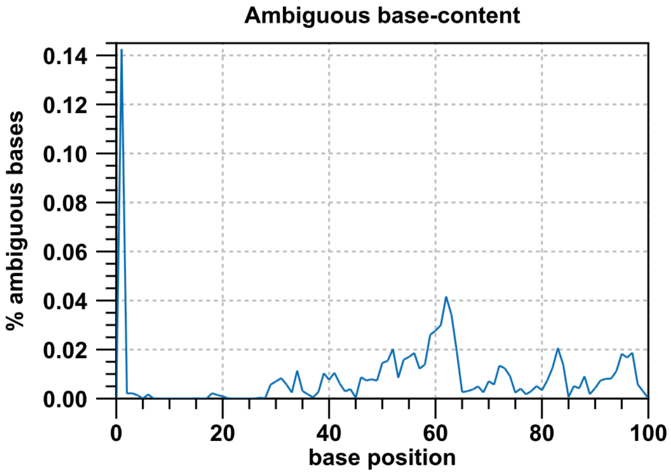
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



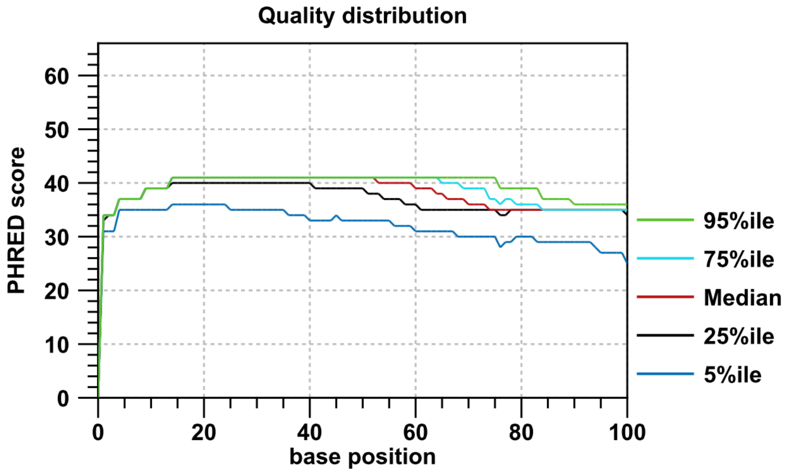
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



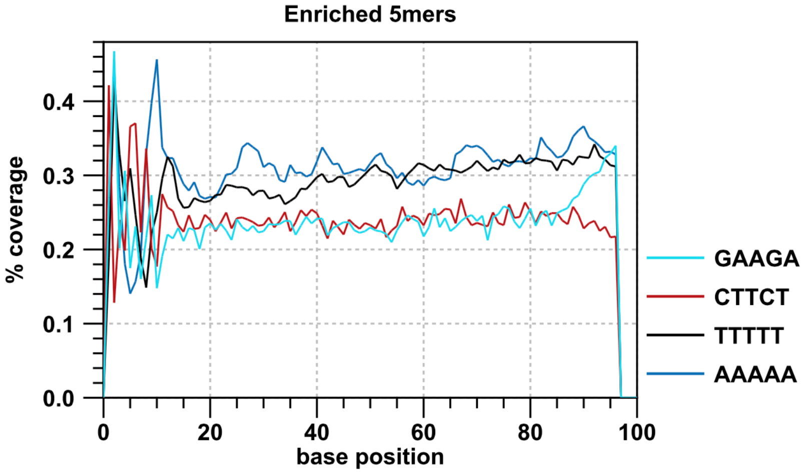
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at that position

3.5 Quality distribution



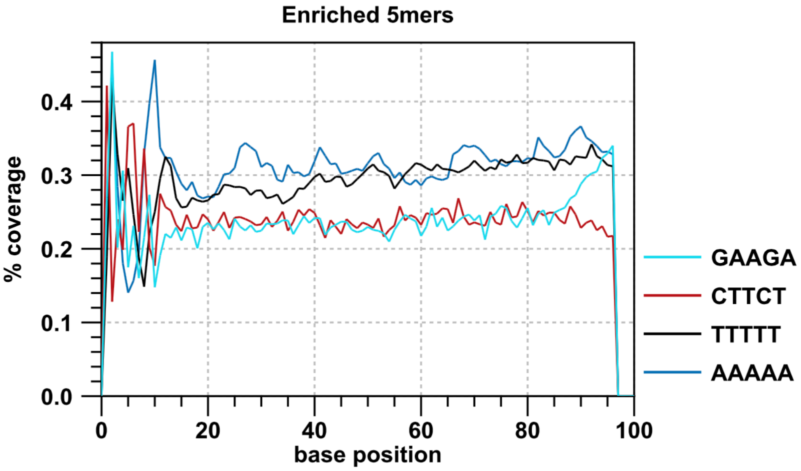
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

4.1 Enriched 5mers



The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the observed and expected 5mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5mer. (5mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion

4.1 Enriched 5mers

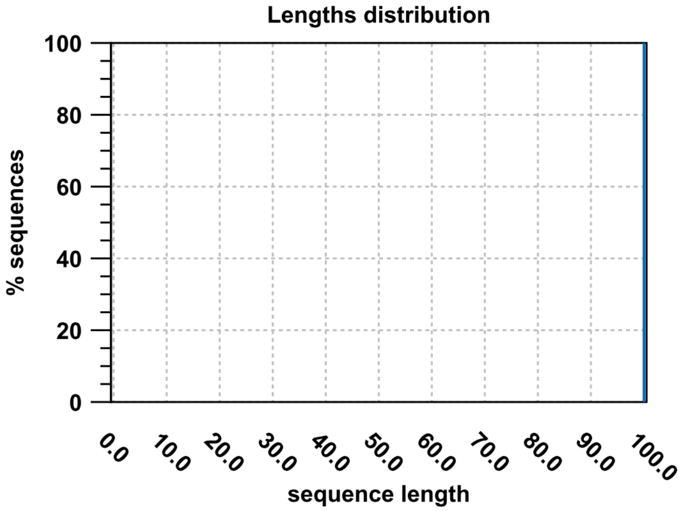


```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

JFM49I, Sequencing Index 15, Quiescent Mouse 3

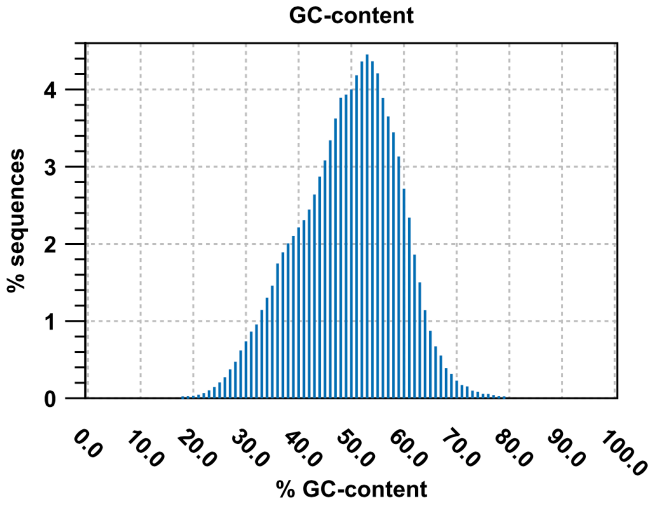
Sequencing QC Report
Based upon: 20,109,186 sequences in 6 data sets
Generated by: Gonzalo
Creation date: Tue Nov 19 18:06:03 PST 2013
Software: CLC Genomics Workbench 6.5

2.1 Lengths distribution



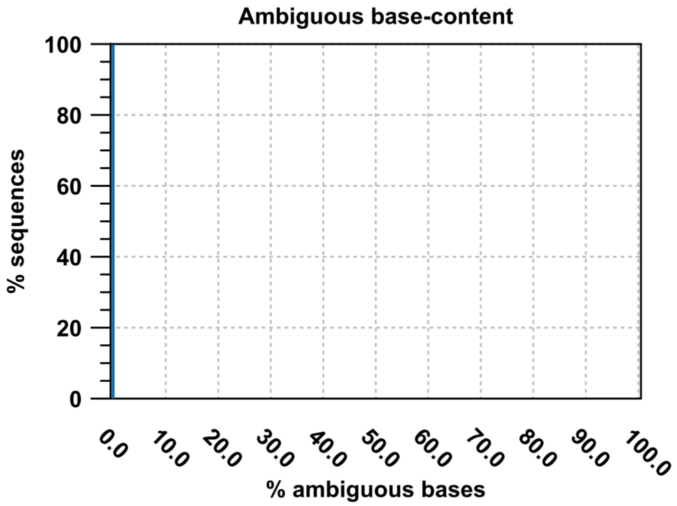
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



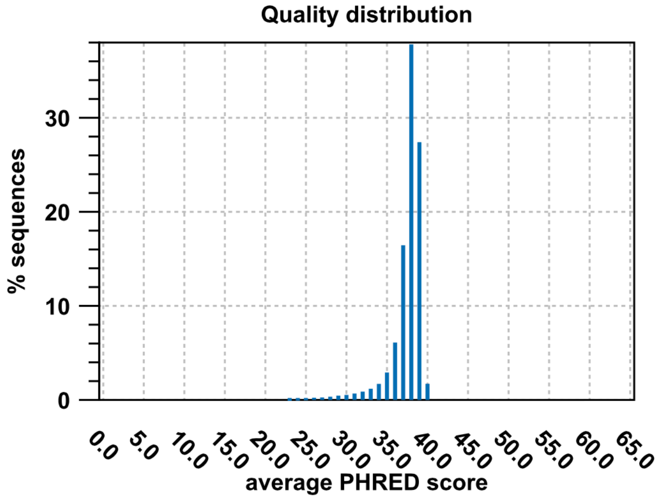
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



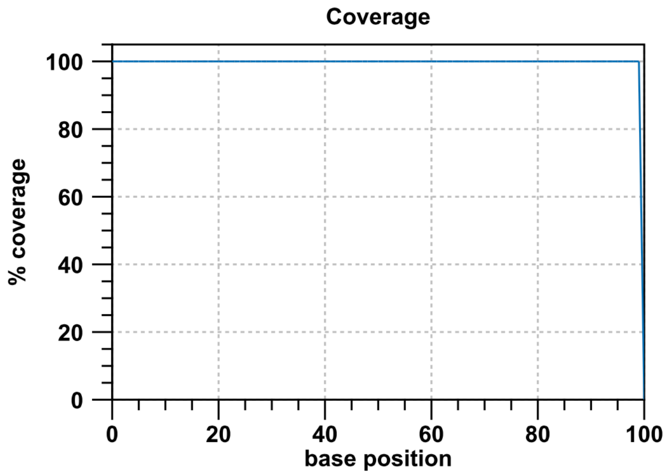
```
Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared
to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences
```

2.4 Quality distribution



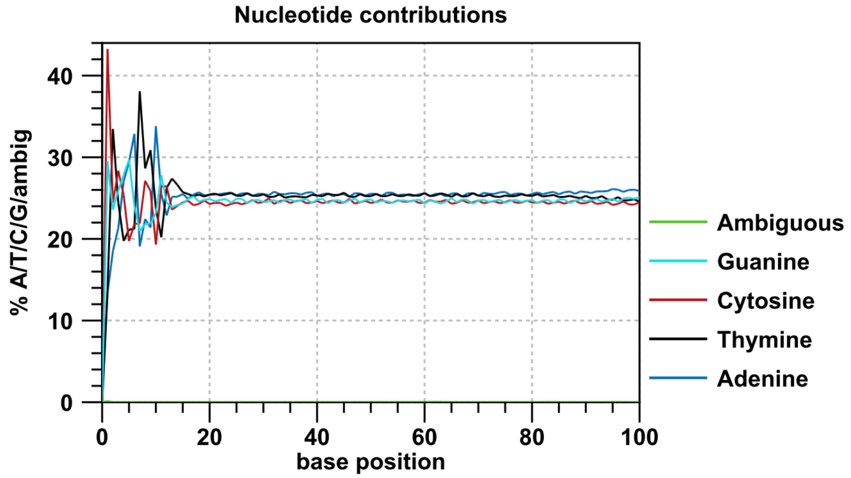
```
Distribution of average sequence qualitie scores. The quality of a sequence is calculated  as the arithmetic mean
of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences
```

3.1 Coverage



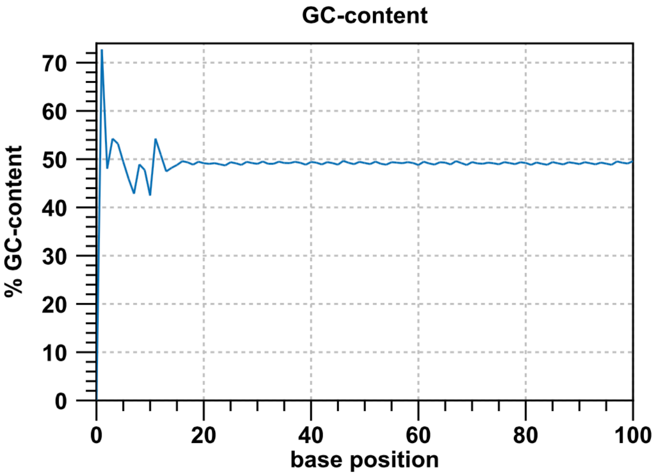
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



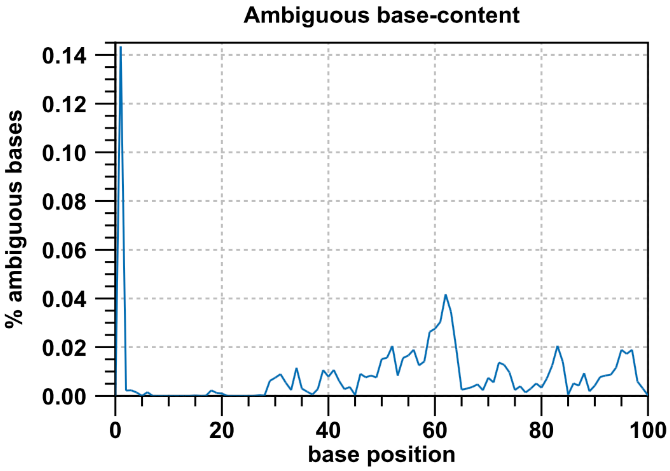
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



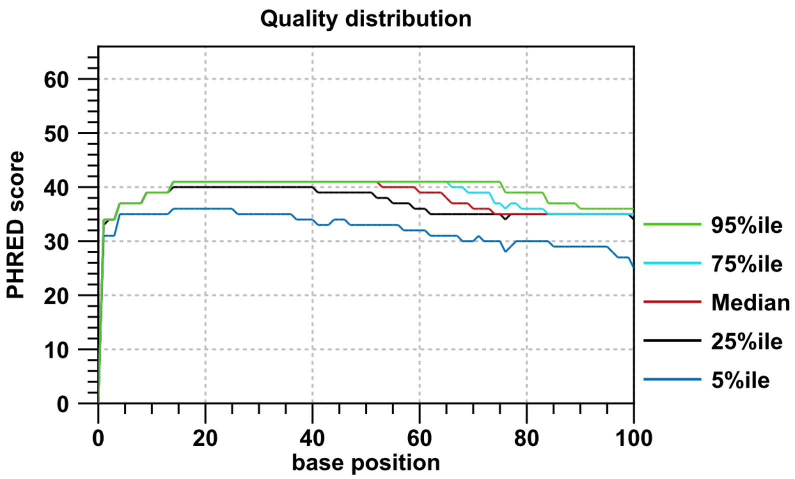
```
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of  bases observed at that
position
```

3.4 Ambiguous base-content



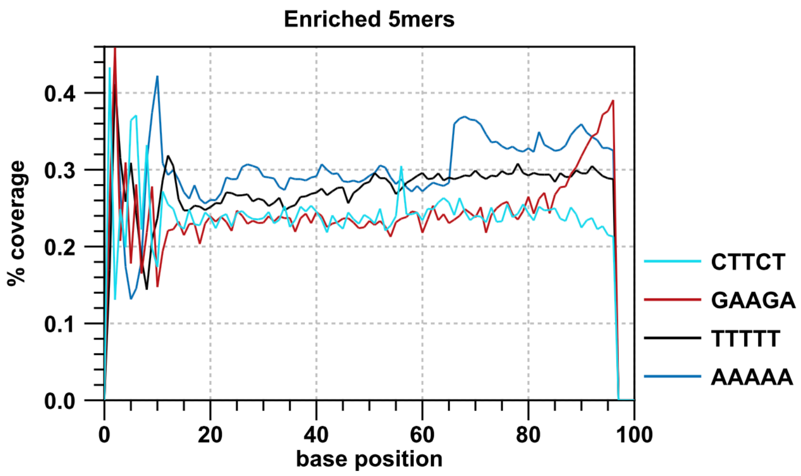
```
Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at
that position
```

3.5 Quality distribution



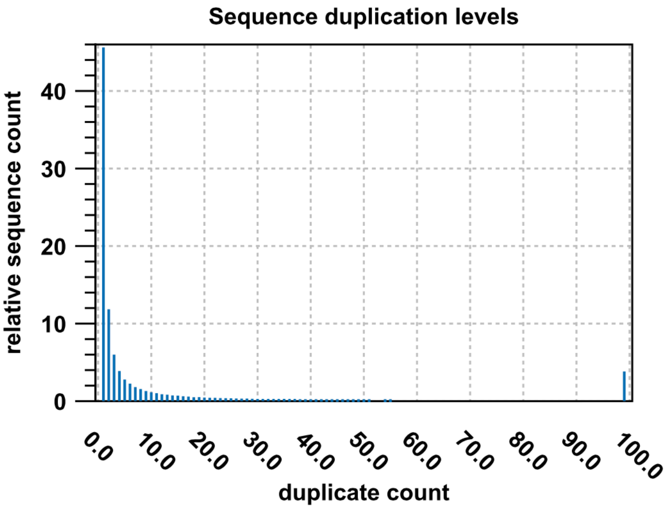
```
Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position
```

4.1 Enriched 5mers



```
The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the
observed and expected 5mer frequency. The expected frequency is calculated  as product of the empirical nucleotide
probabilities that make up the 5mer. (5mers that  contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion
```

4.2 Sequence duplication levels

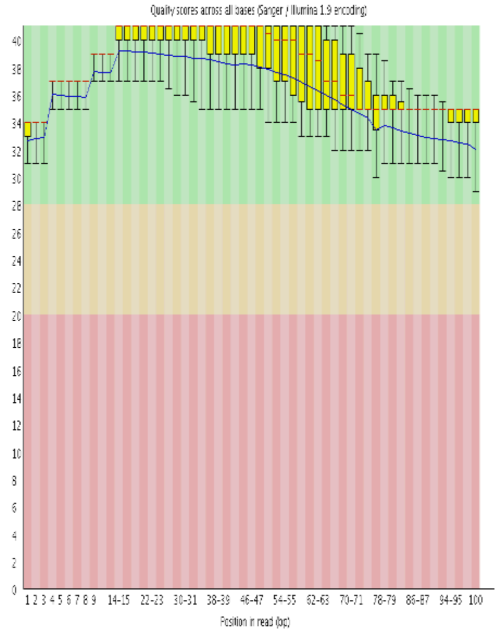


```
Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has
been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences
```

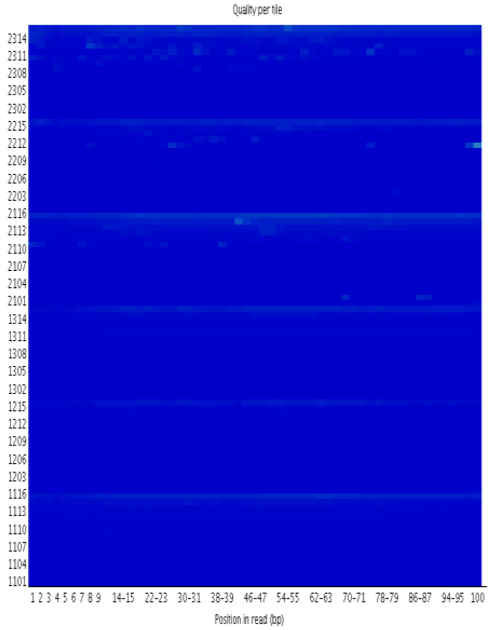
APPENDIX B – Green Line Analysis Sequencing QC Reports

Virgin Mouse 1

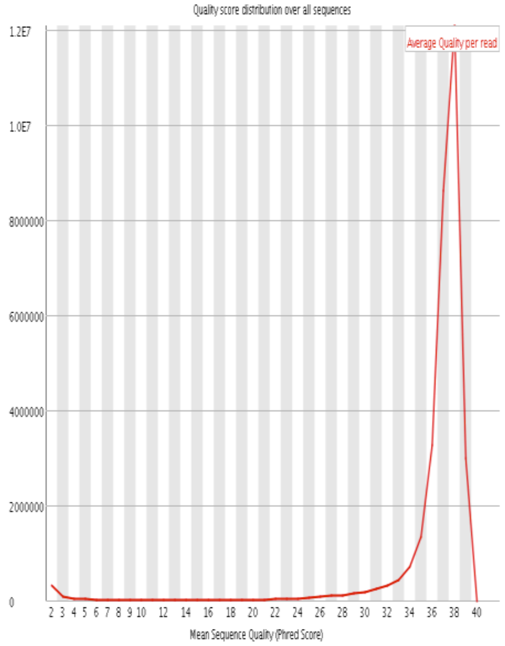
✓ Per base sequence quality



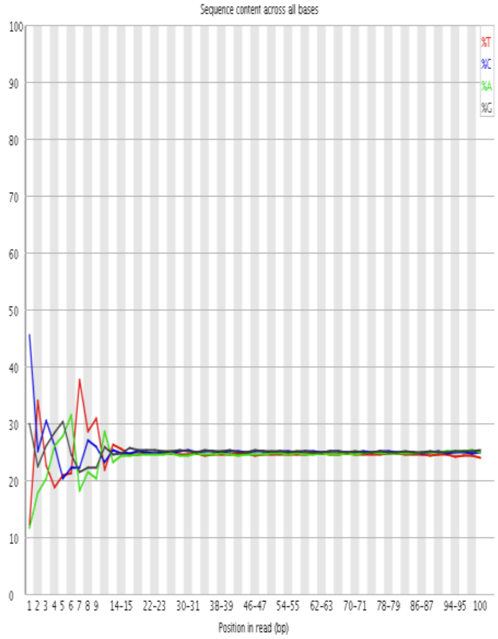
✓ Per tile sequence quality



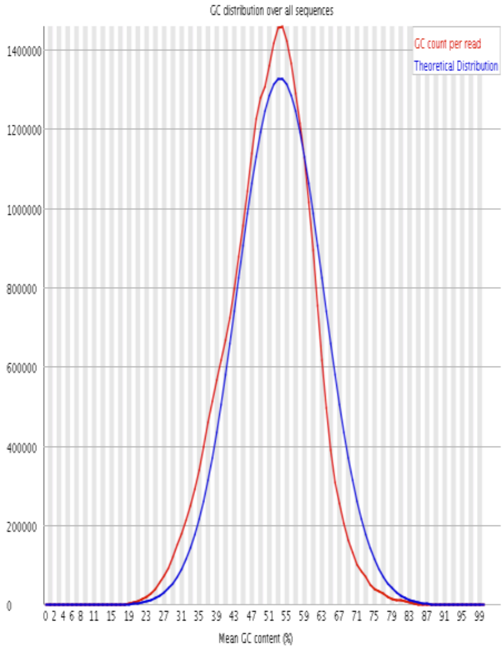
✓ Per sequence quality scores



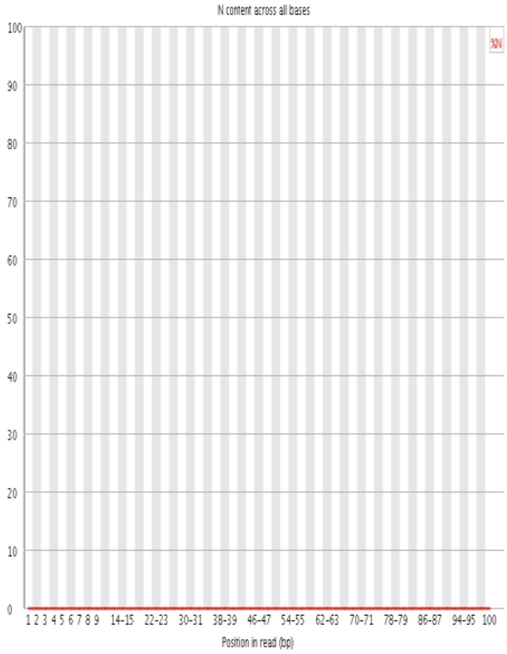
✗ Per base sequence content



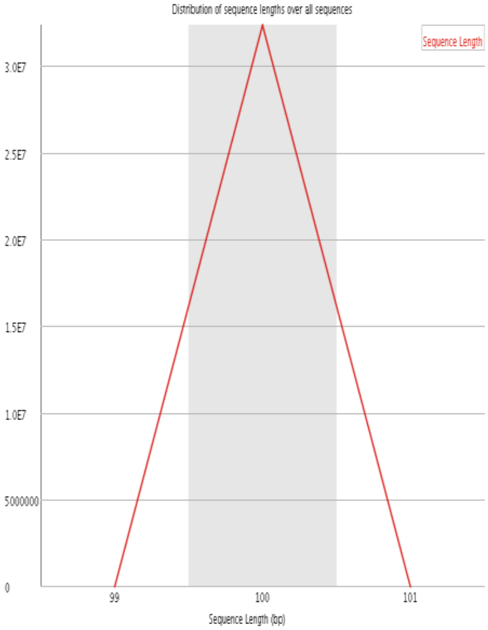
Per sequence GC content



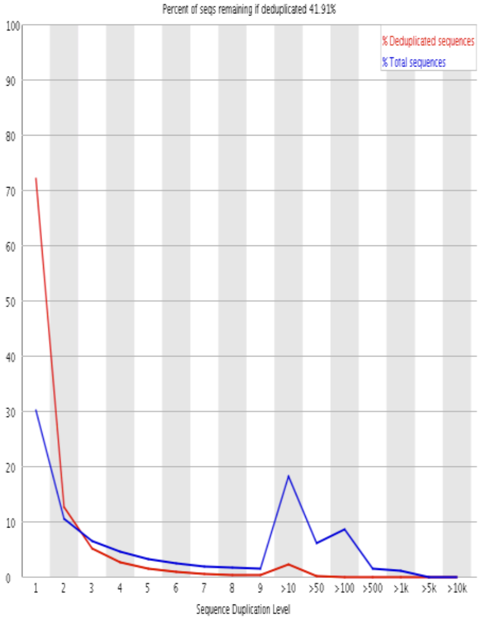
Per base N content



Sequence Length Distribution



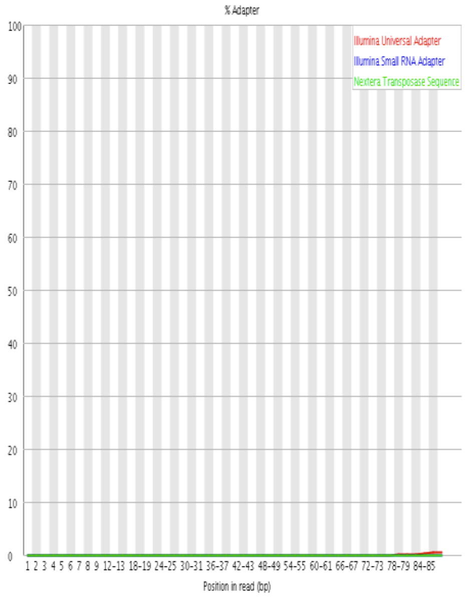
Sequence Duplication Levels



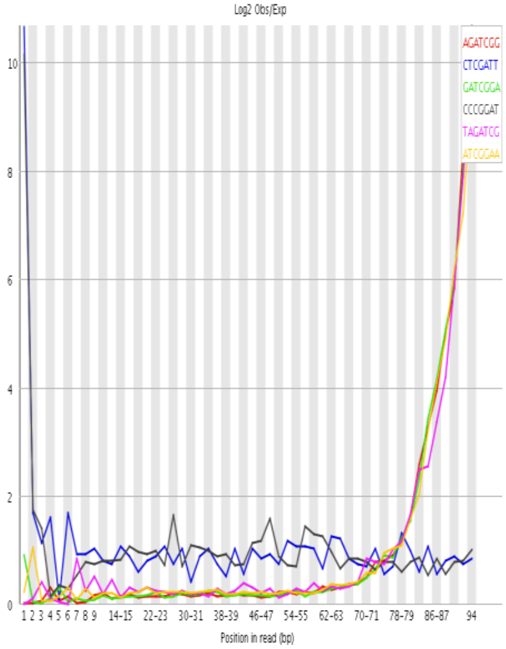
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

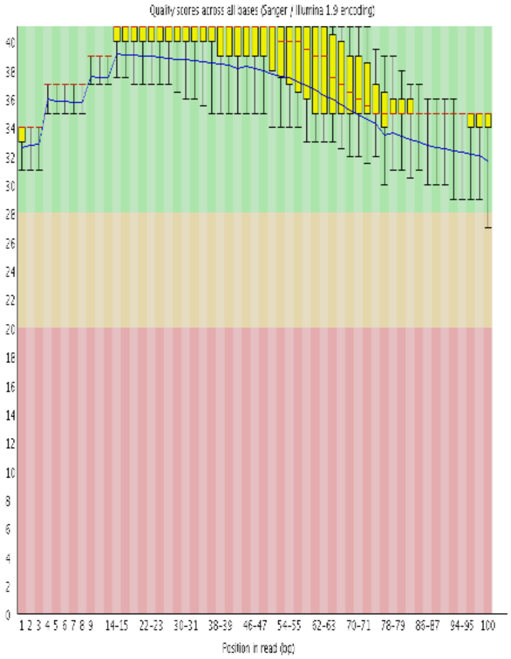


✗ Kmer Content

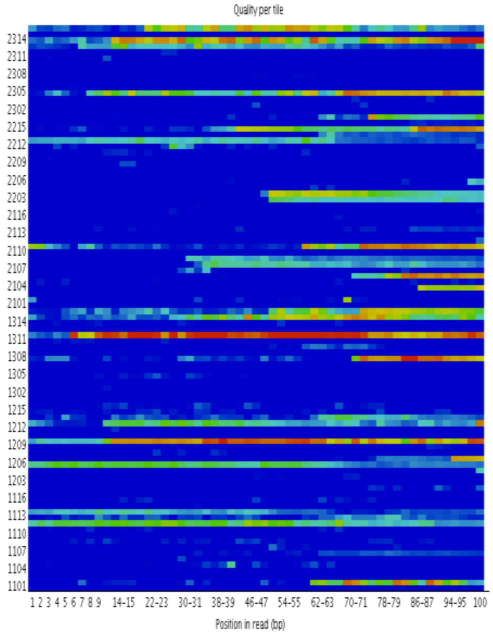


Virgin Mouse 2

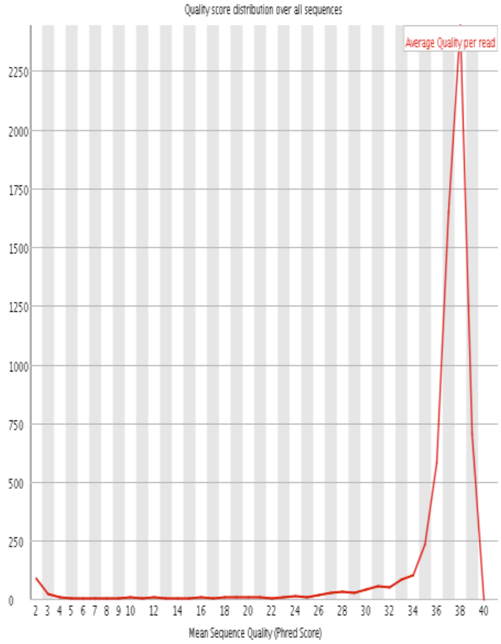
✓ Per base sequence quality



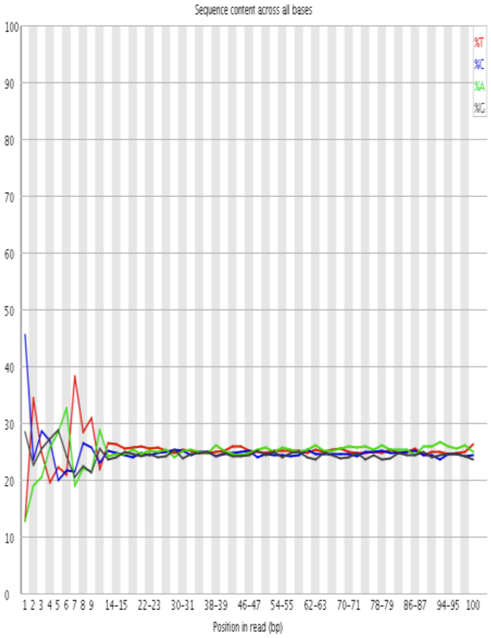
✗ Per tile sequence quality



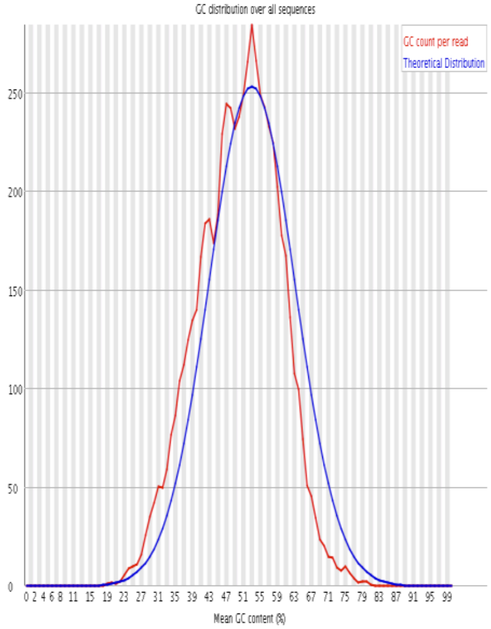
✓ Per sequence quality scores



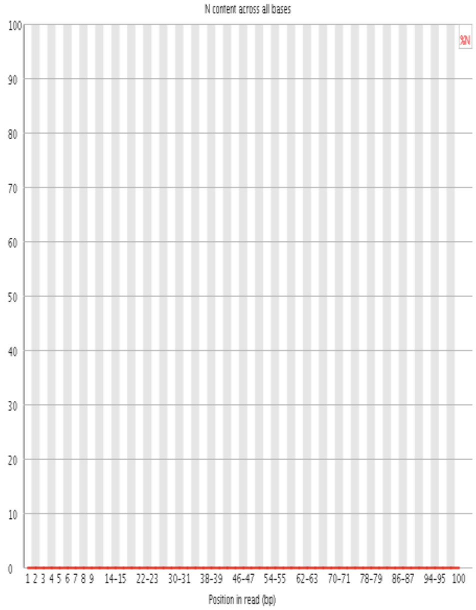
✗ Per base sequence content



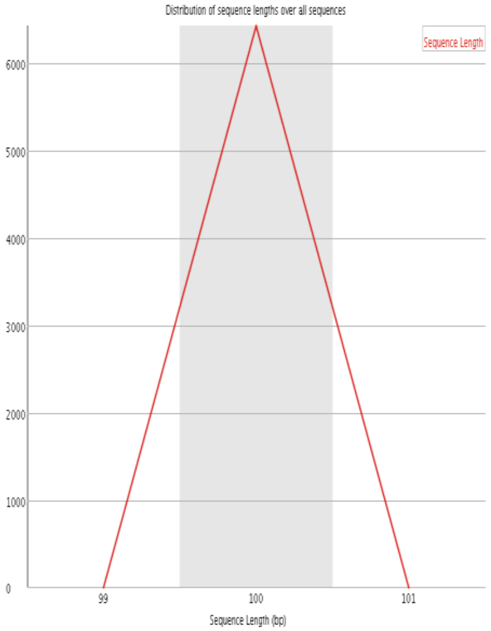
! Per sequence GC content



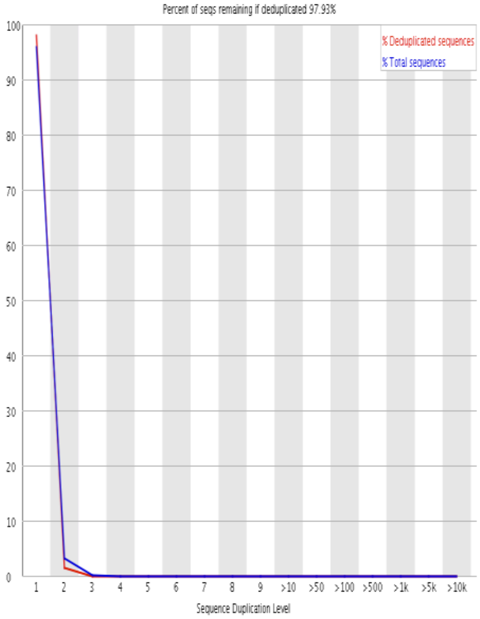
✓ Per base N content



Sequence Length Distribution



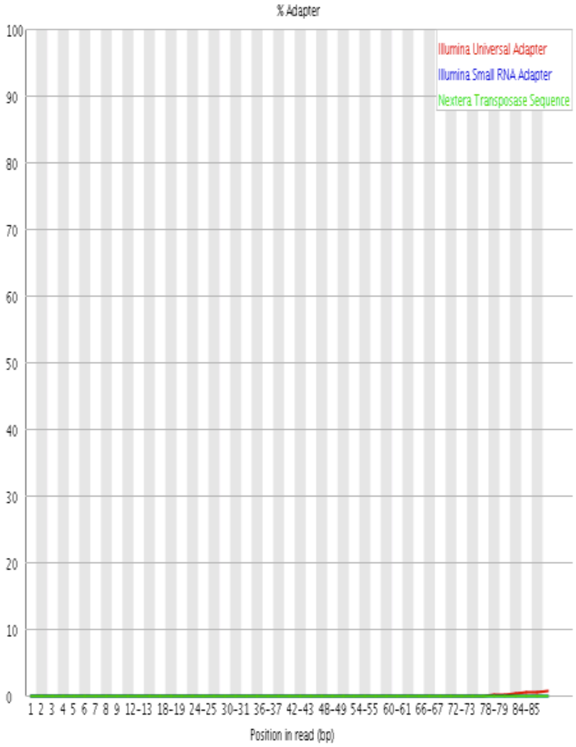
Sequence Duplication Levels



Overrepresented sequences

No overrepresented sequences

Adapter Content

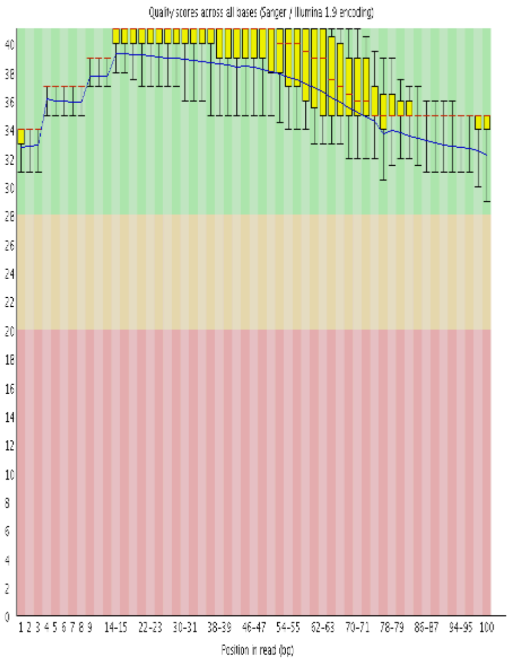


Kmer Content

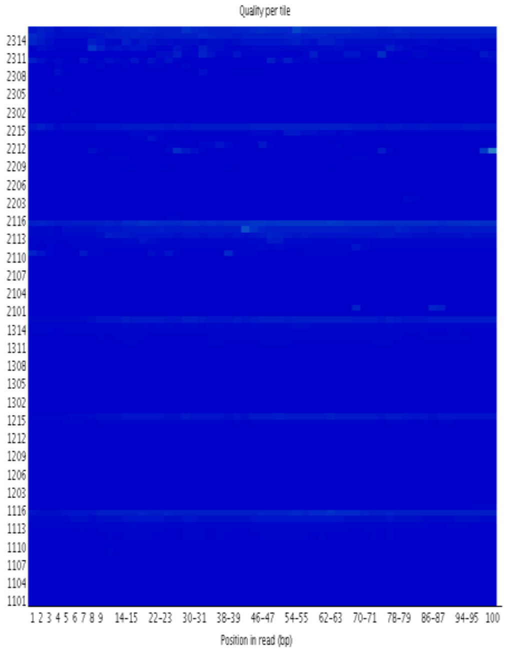
No overrepresented Kmers

Virgin Mouse 3

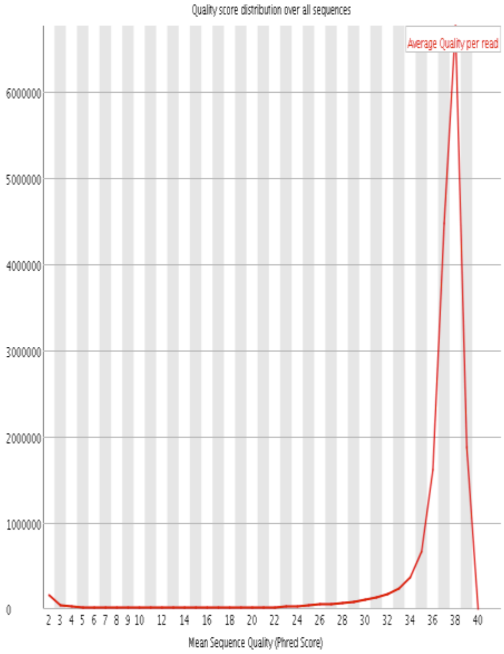
✓ Per base sequence quality



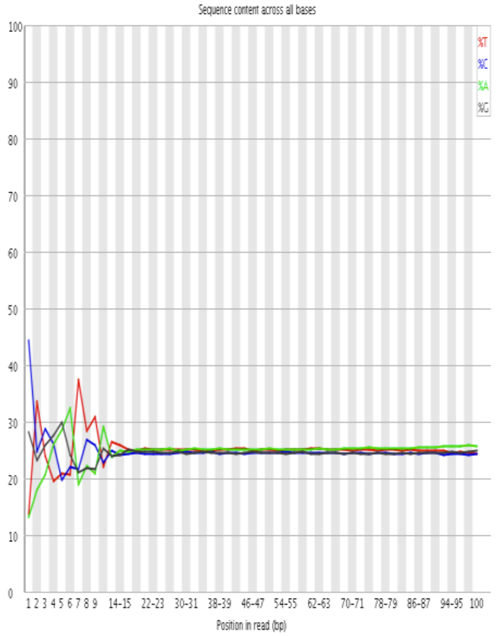
✓ Per tile sequence quality



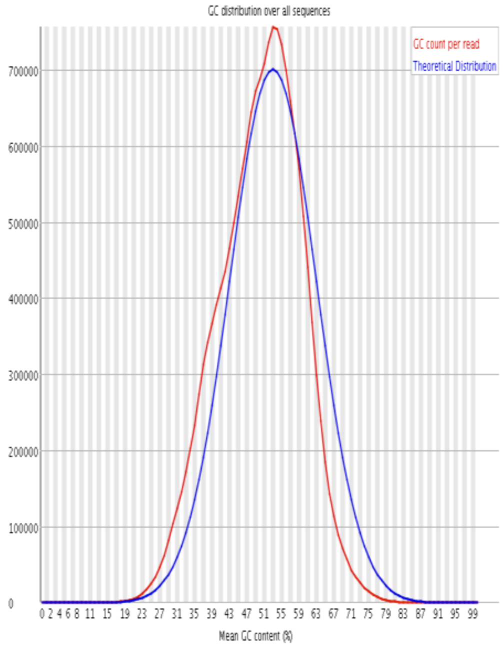
✓ Per sequence quality scores



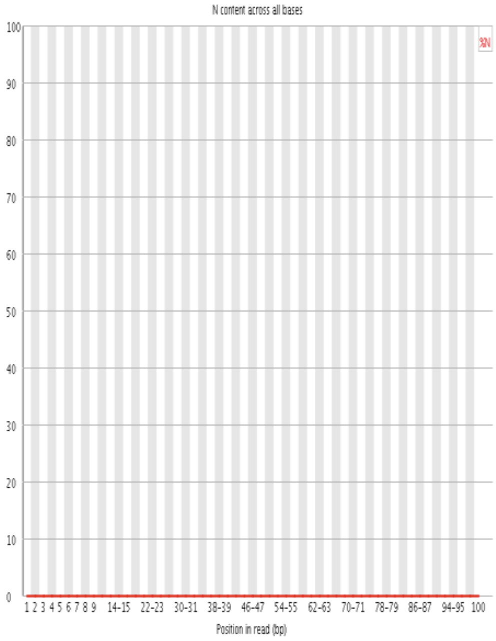
✗ Per base sequence content



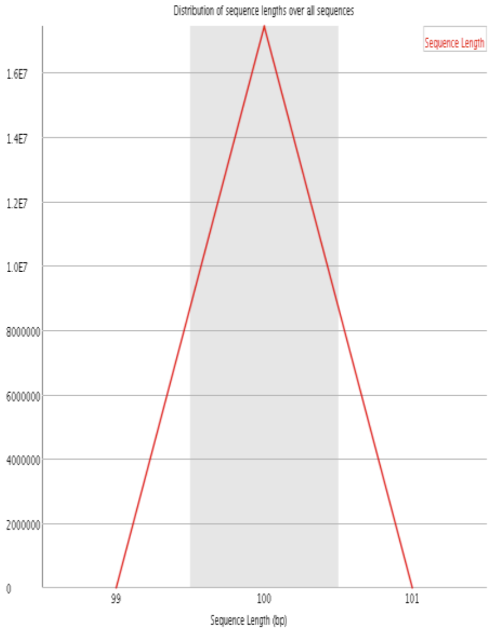
⚠ Per sequence GC content



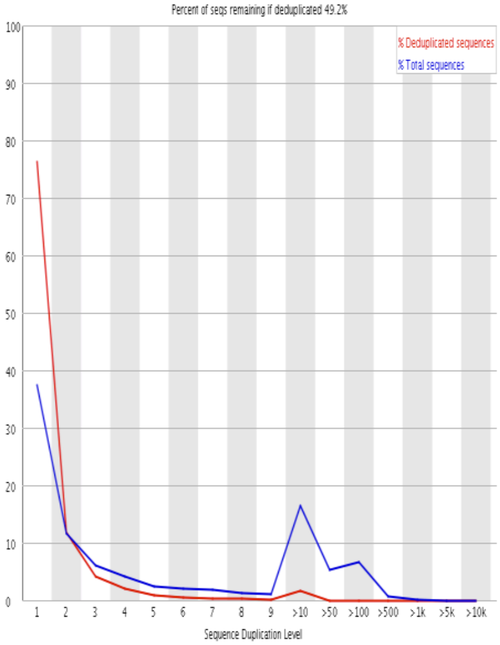
✓ Per base N content



✓ Sequence Length Distribution



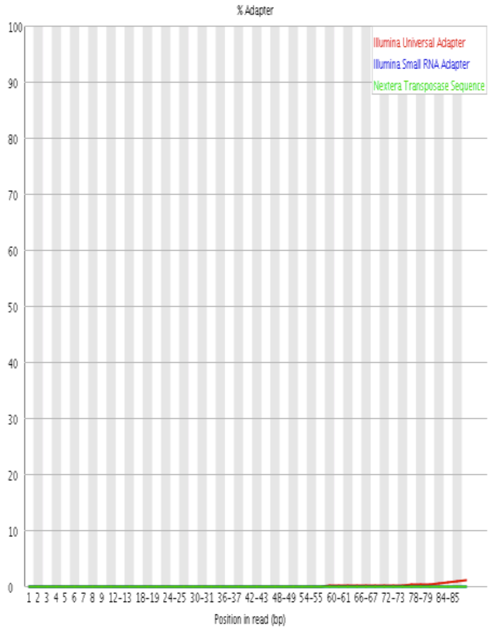
✗ Sequence Duplication Levels



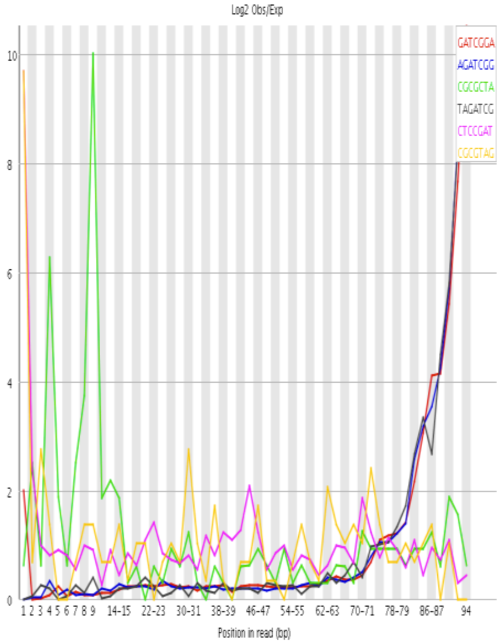
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

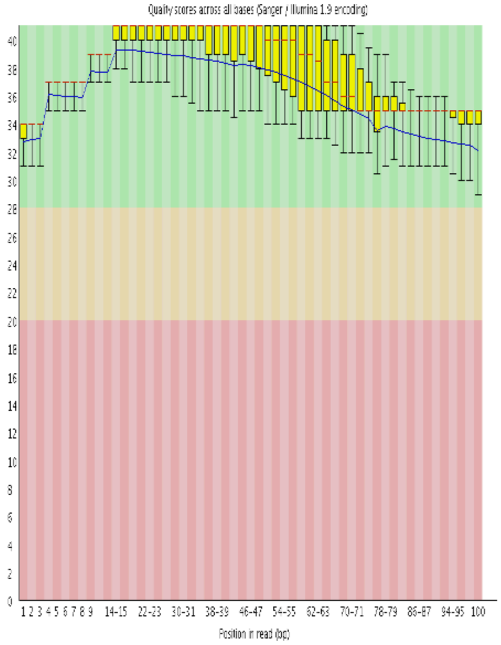


☒ Kmer Content

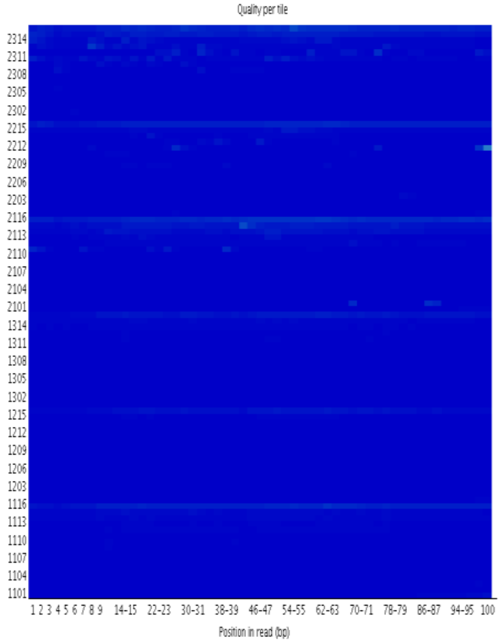


Pregnant Mouse 1

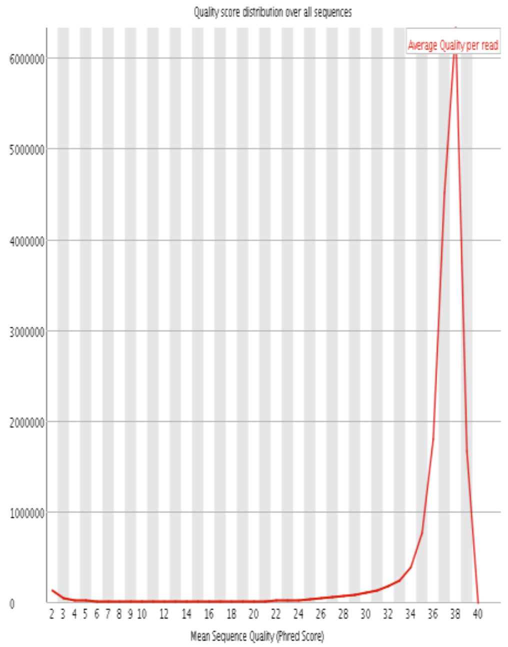
✓ Per base sequence quality



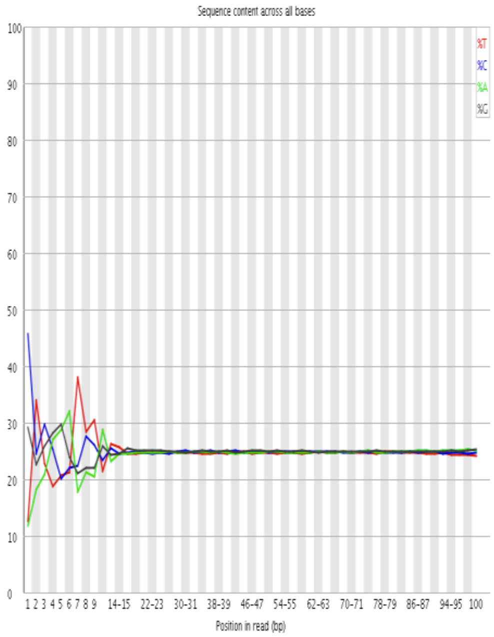
✓ Per tile sequence quality



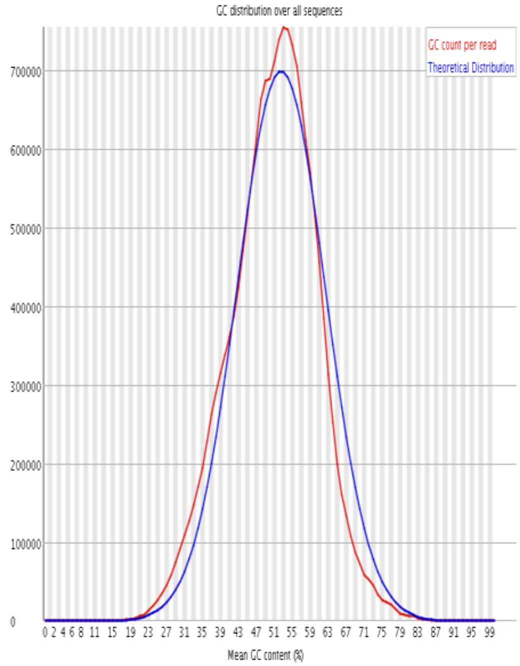
✓ **Per sequence quality scores**



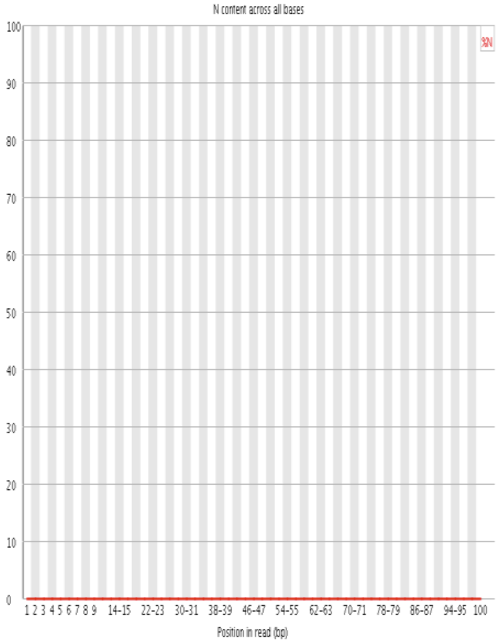
✗ **Per base sequence content**



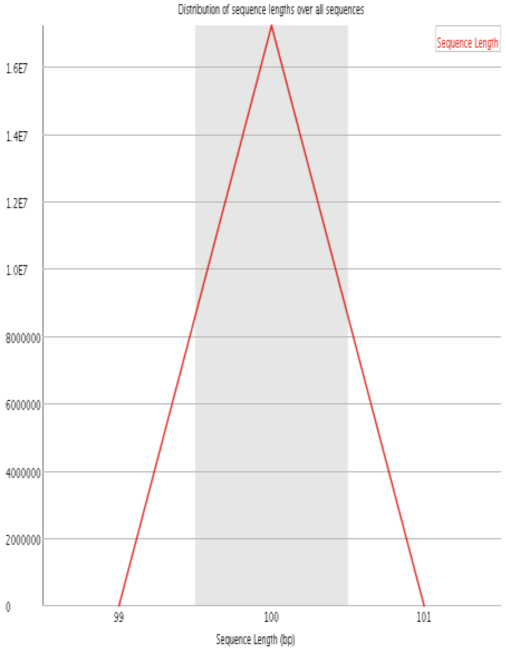
✓ **Per sequence GC content**



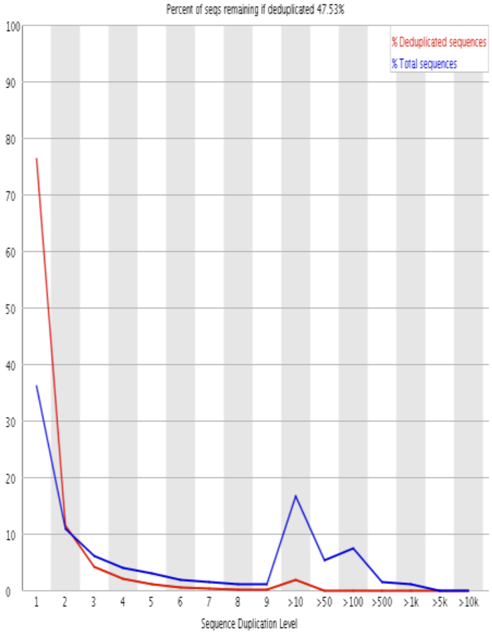
✓ **Per base N content**



✓ Sequence Length Distribution



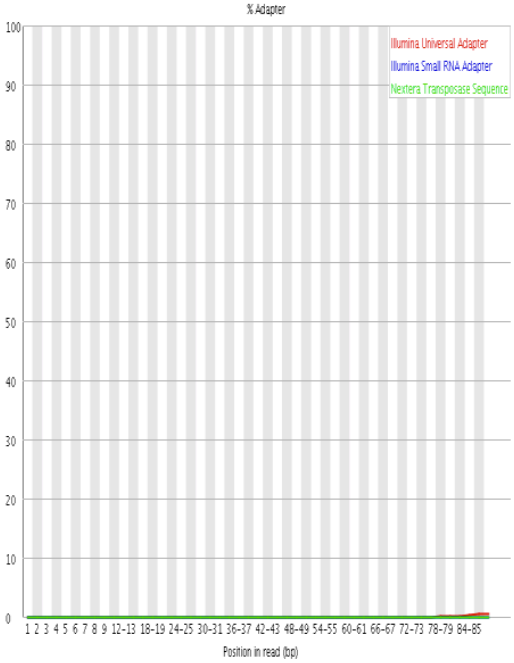
✗ Sequence Duplication Levels



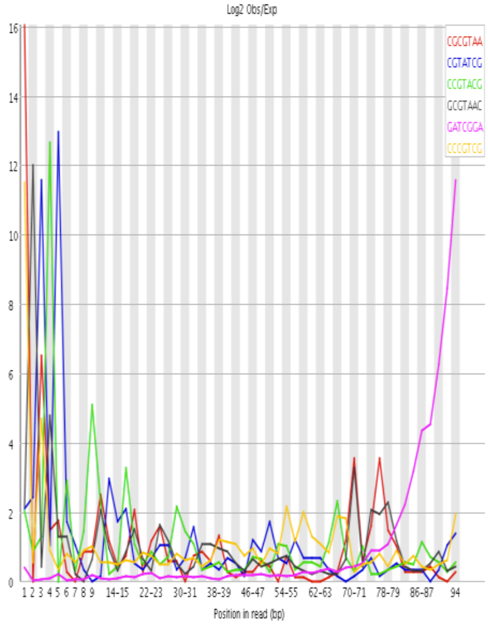
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

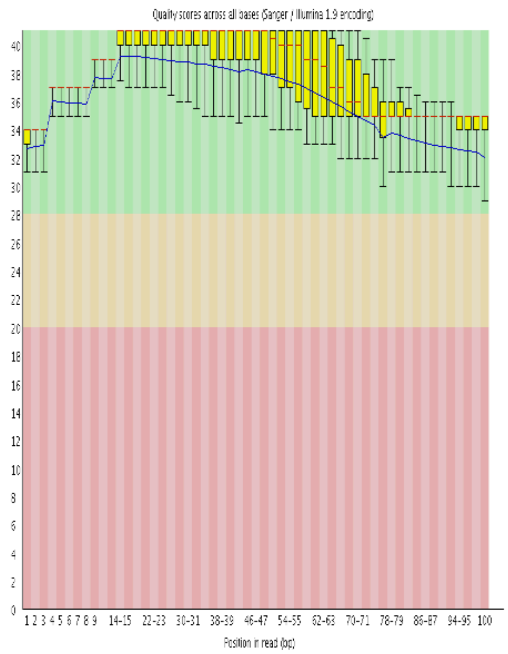


✗ Kmer Content

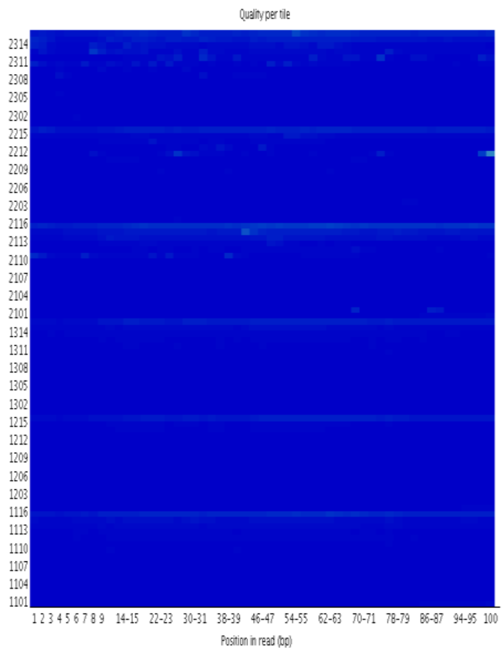


Pregnant Mouse 2

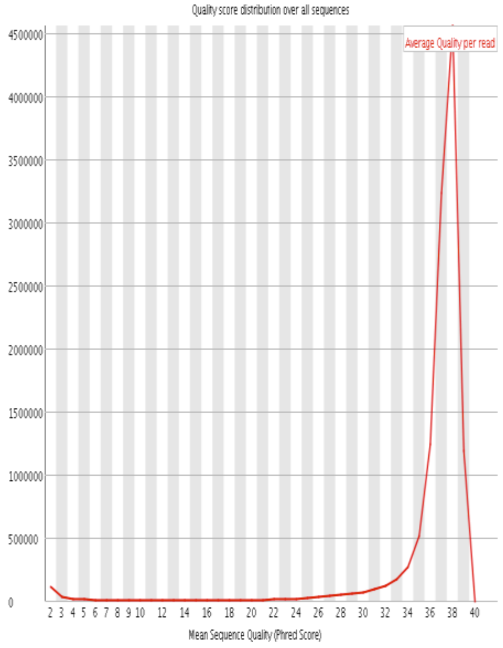
✓ Per base sequence quality



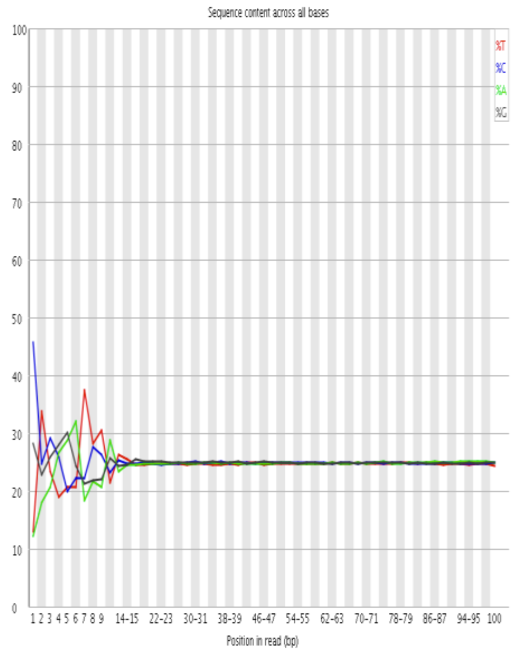
✓ Per tile sequence quality



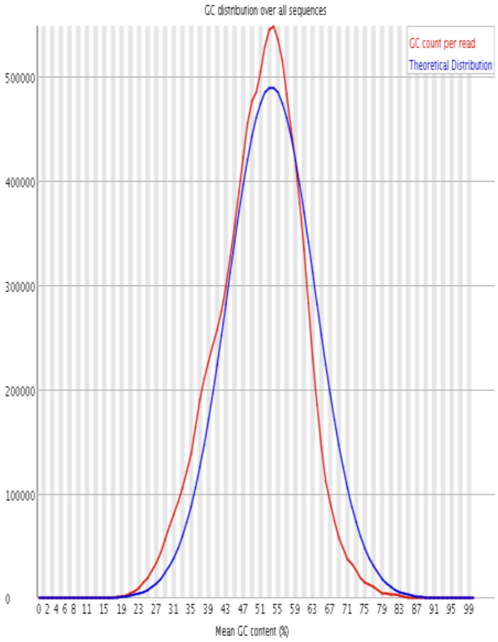
✓ Per sequence quality scores



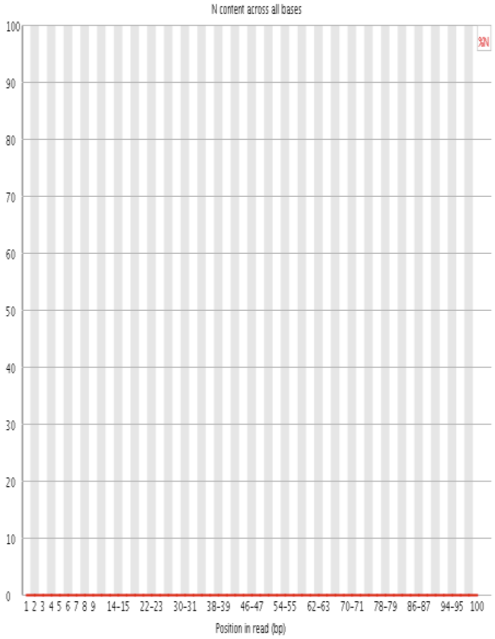
✗ Per base sequence content



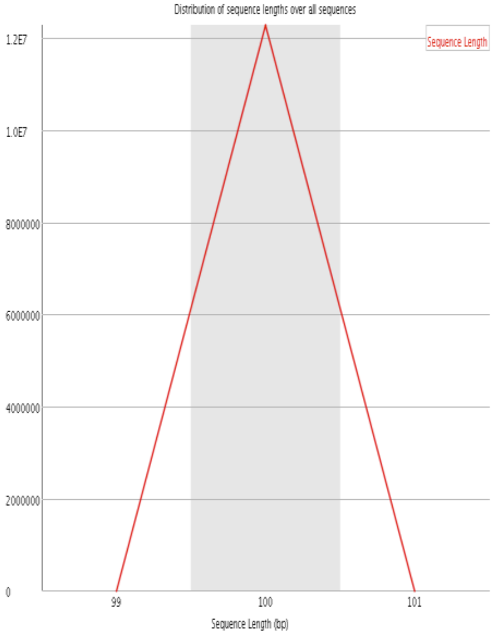
! Per sequence GC content



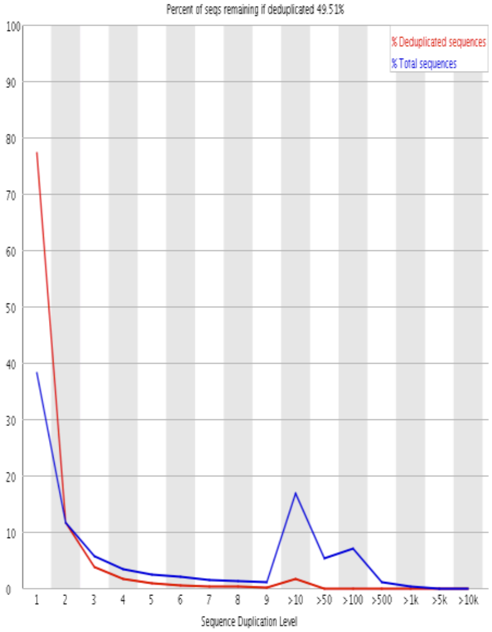
✓ Per base N content



✓ Sequence Length Distribution



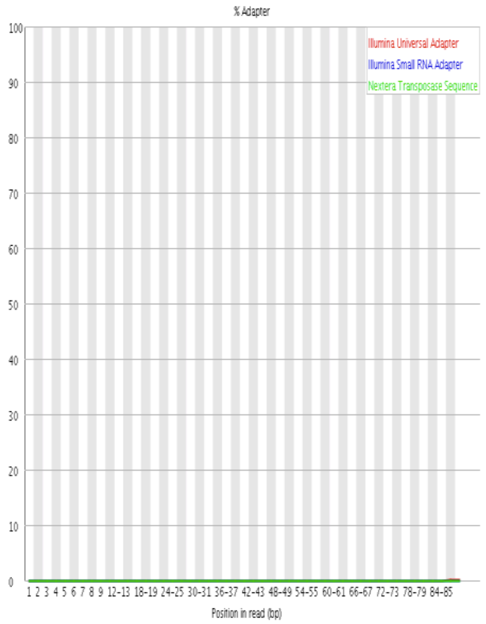
✗ Sequence Duplication Levels



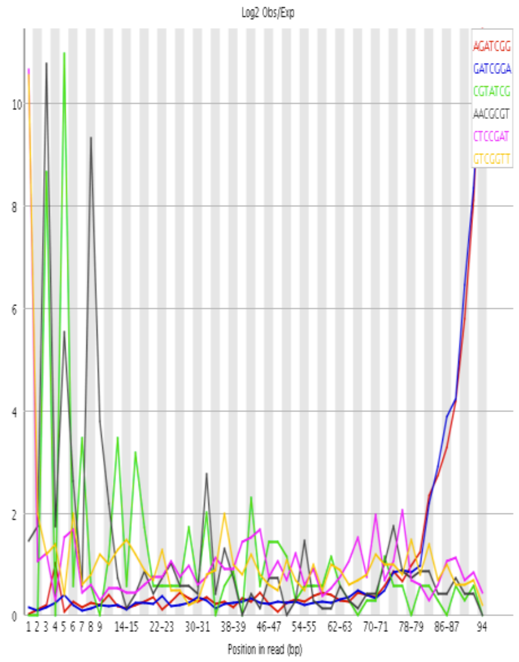
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

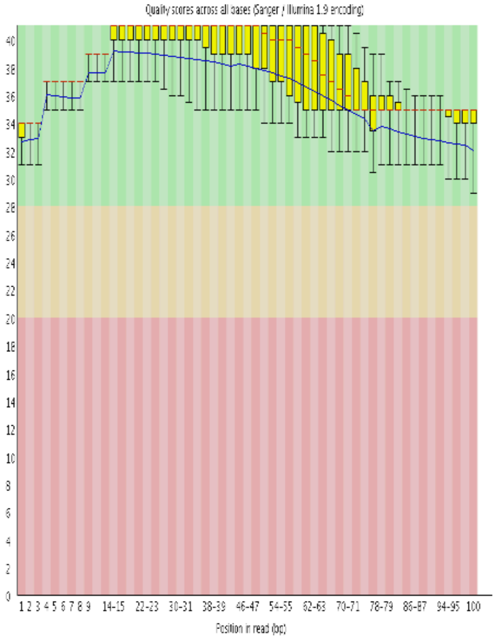


✗ Kmer Content

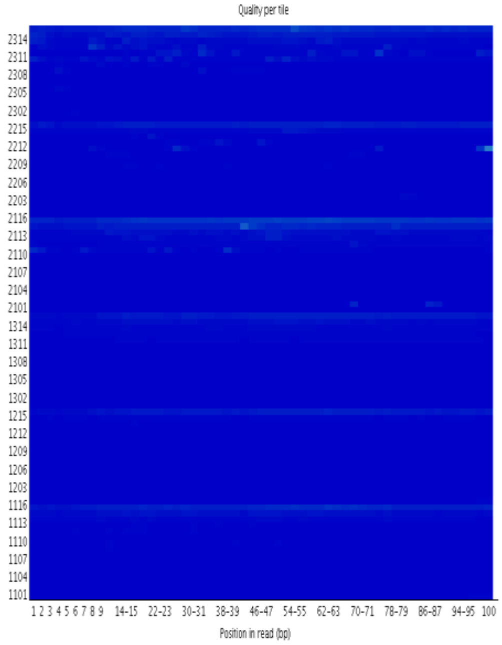


Pregnant Mouse 3

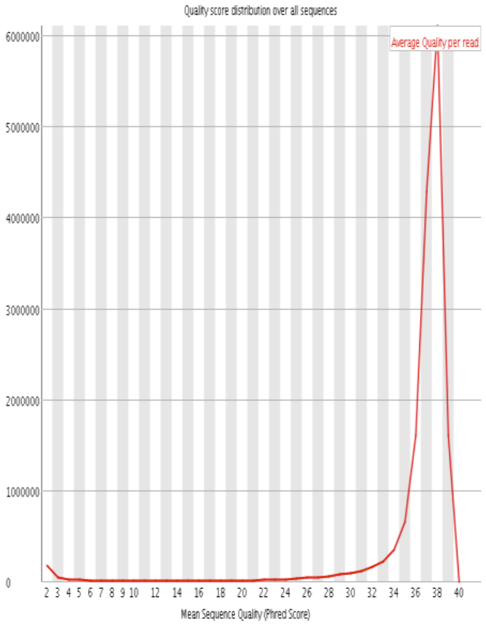
✓ Per base sequence quality



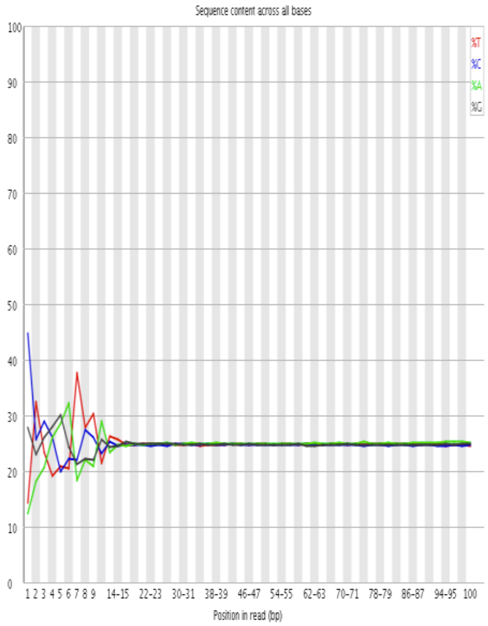
✓ Per tile sequence quality



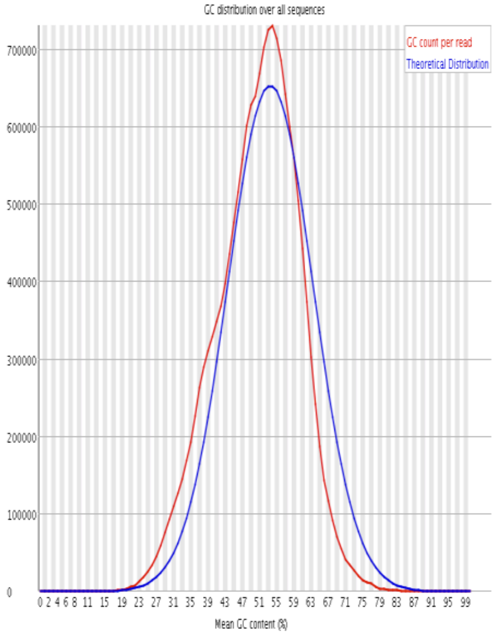
✓ Per sequence quality scores



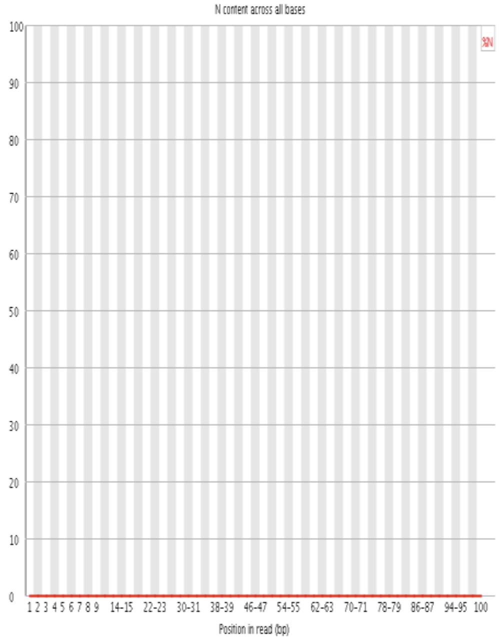
✗ Per base sequence content



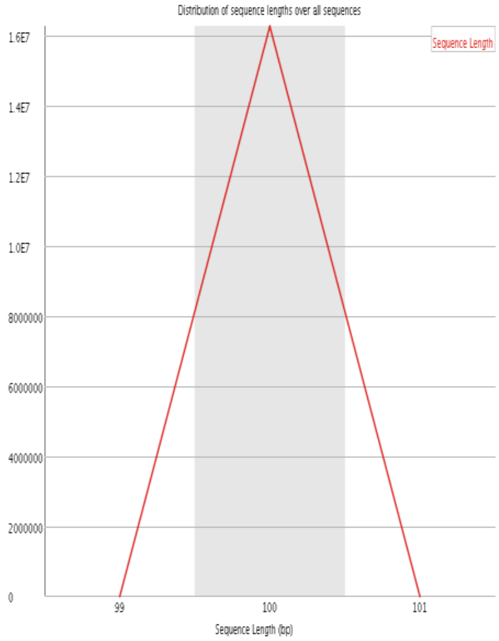
! Per sequence GC content



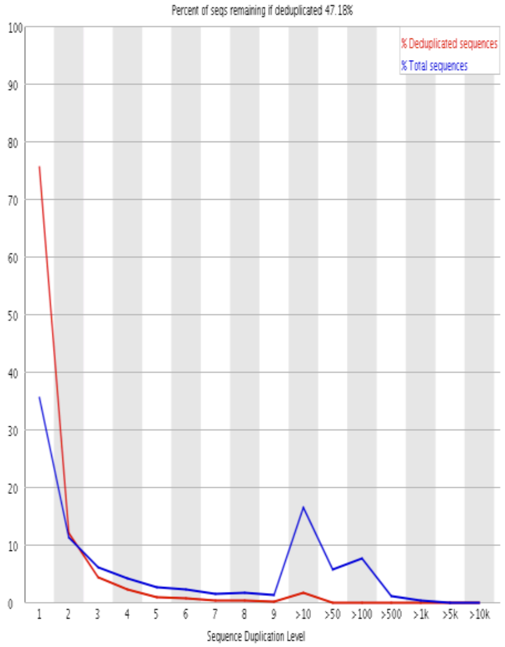
✓ Per base N content



✓ Sequence Length Distribution



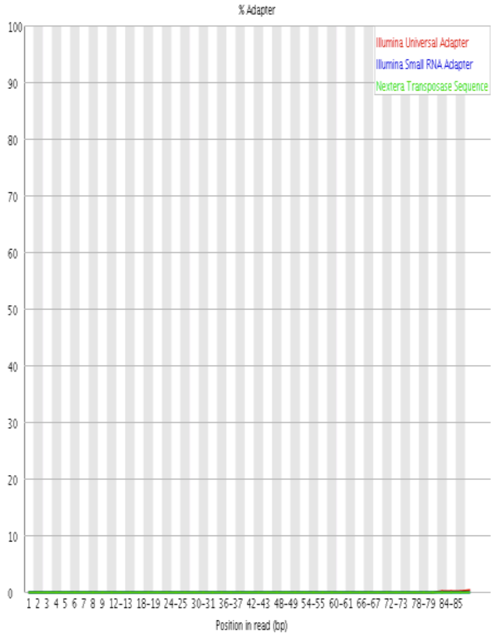
✗ Sequence Duplication Levels



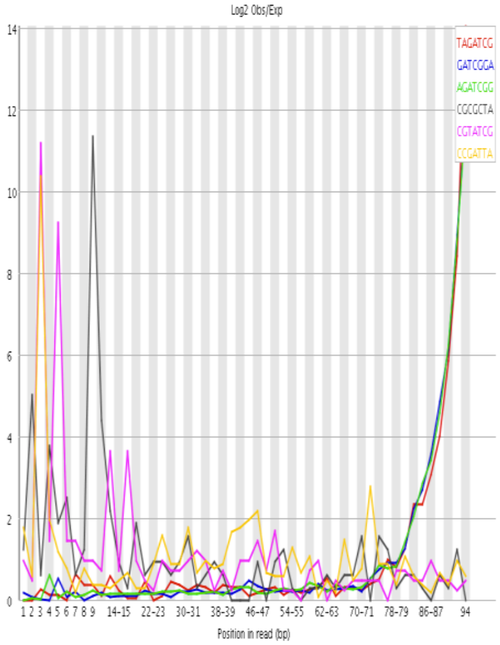
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

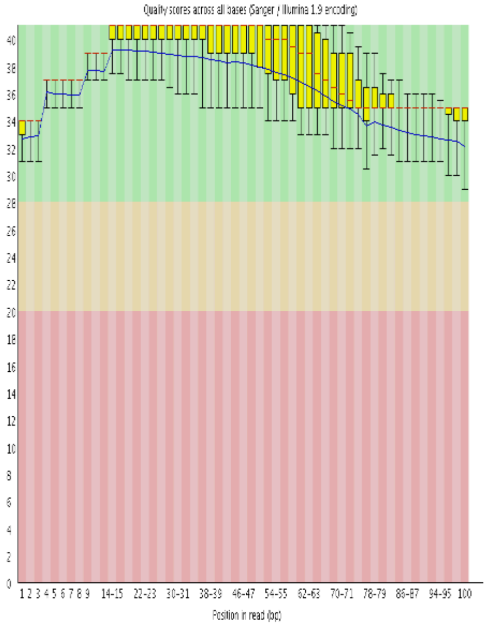


✗ Kmer Content

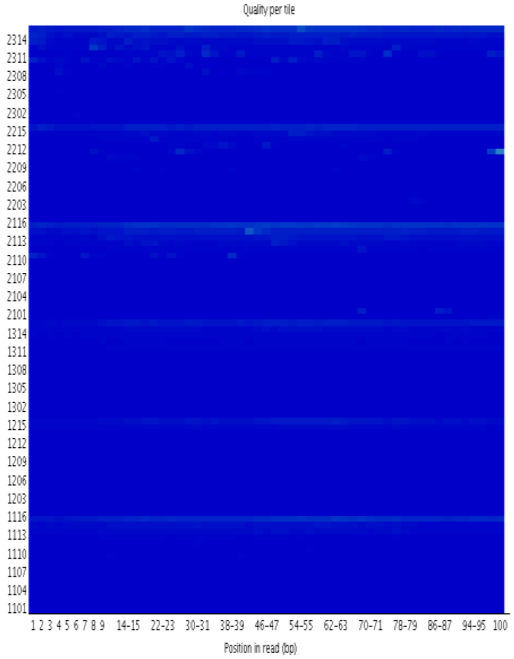


Quiescent Mouse 1

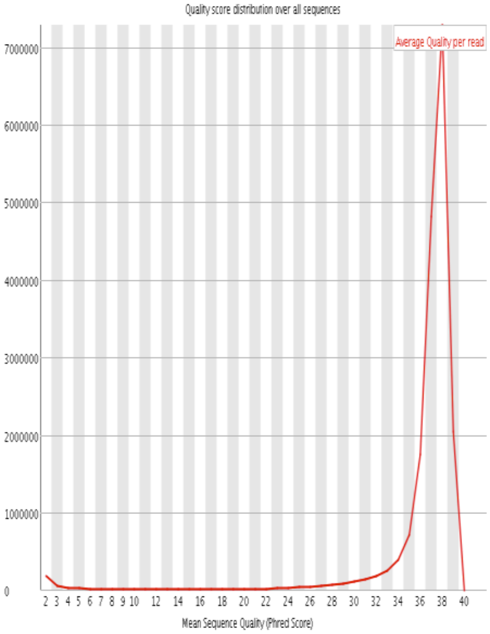
✓ Per base sequence quality



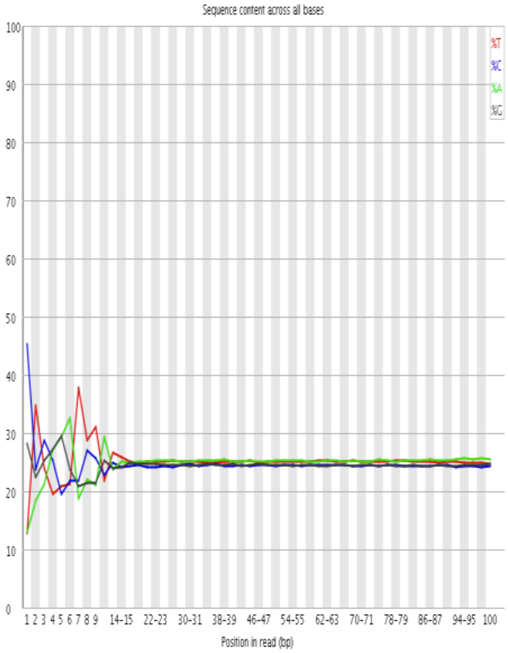
✓ Per tile sequence quality



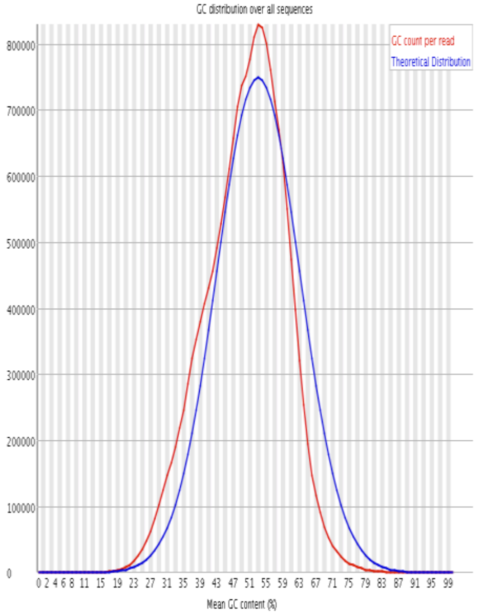
✓ Per sequence quality scores



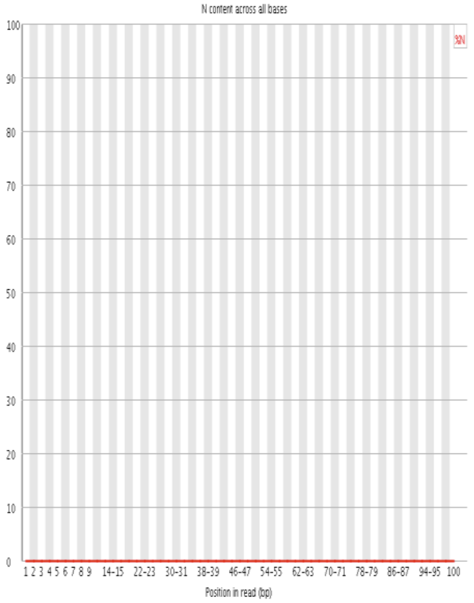
☒ Per base sequence content



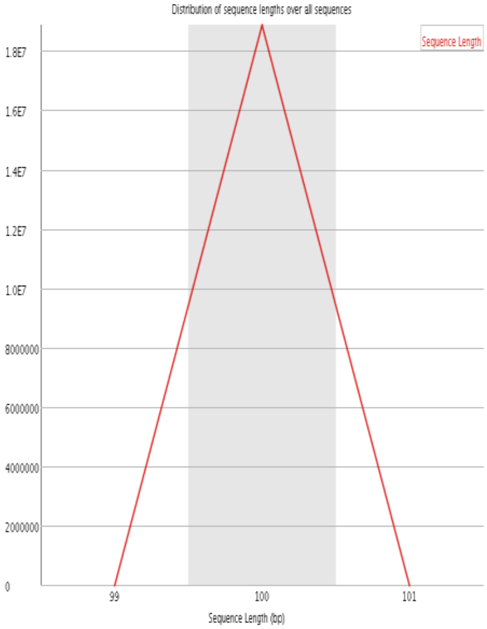
Per sequence GC content



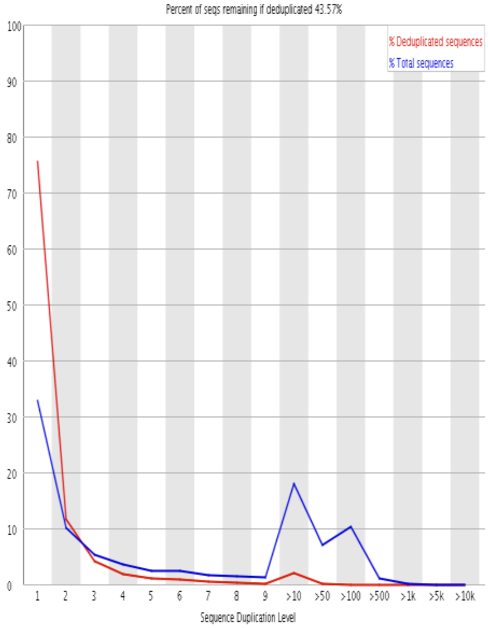
Per base N content



Sequence Length Distribution



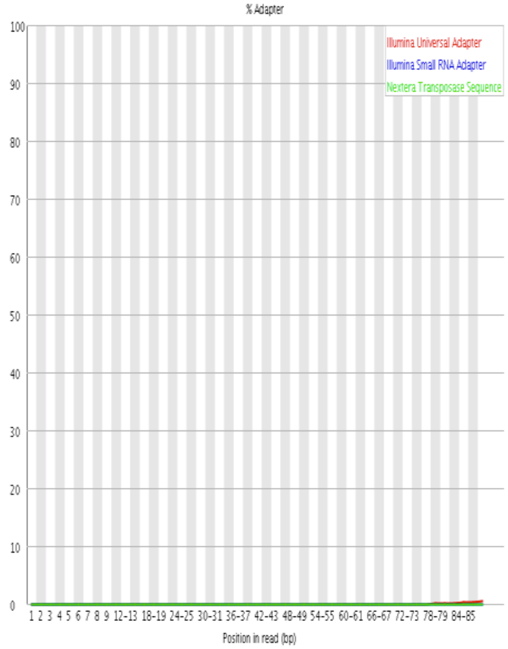
Sequence Duplication Levels



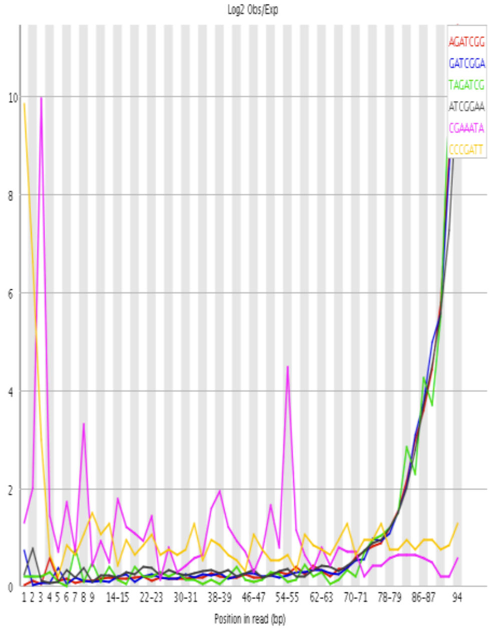
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content

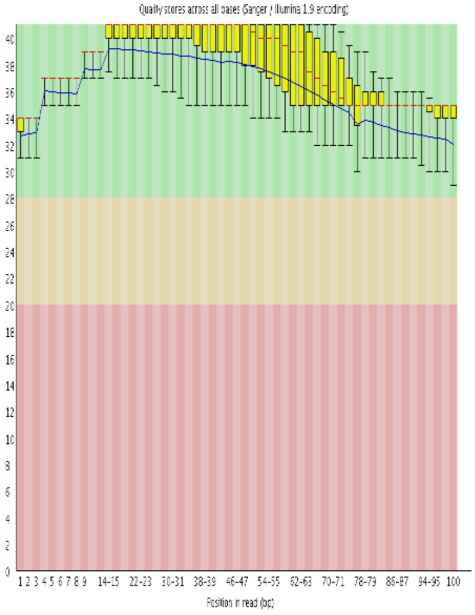


✗ Kmer Content

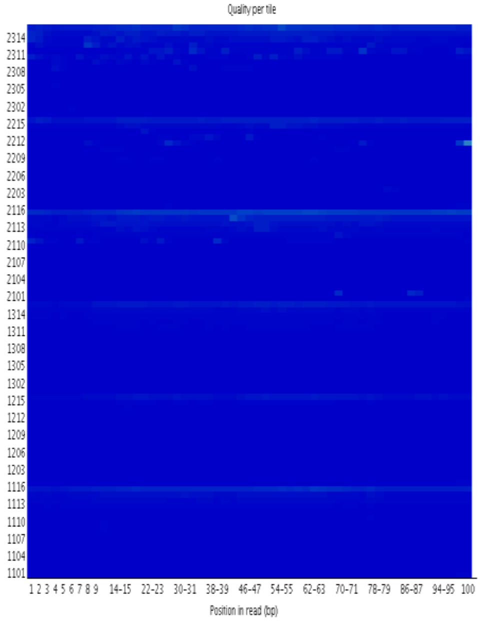


Quiescent Mouse 2

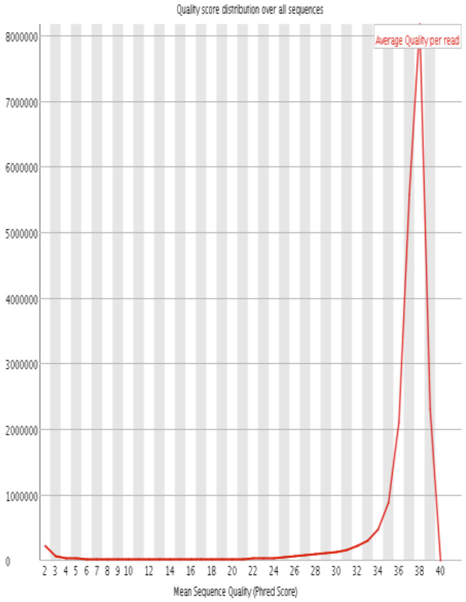
✓ Per base sequence quality



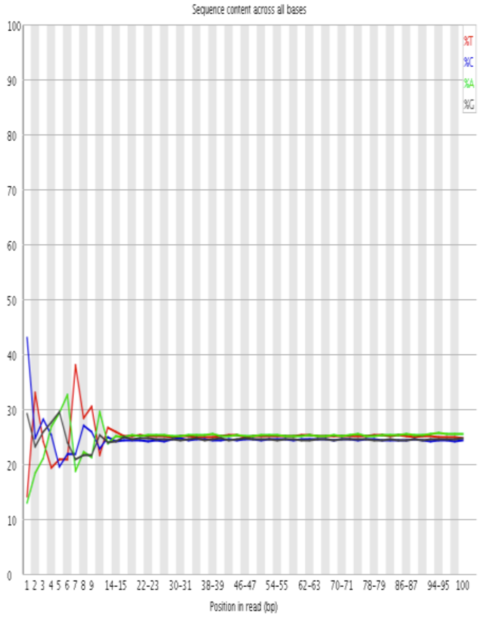
✓ Per tile sequence quality



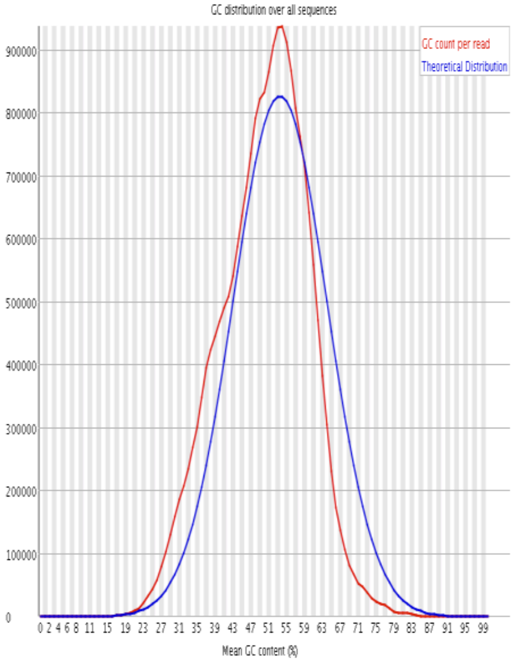
✓ Per sequence quality scores



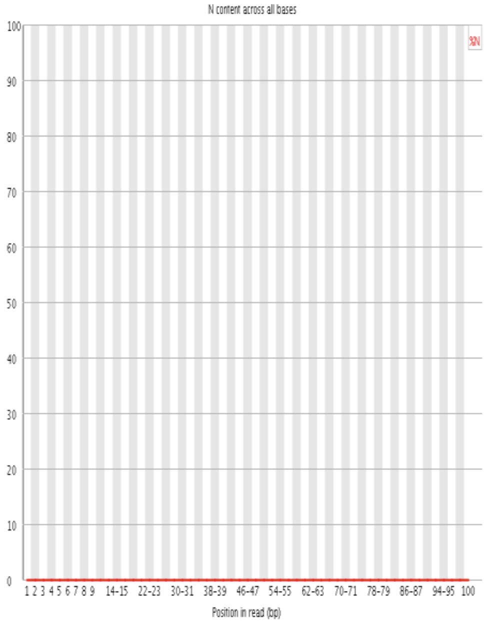
✗ Per base sequence content



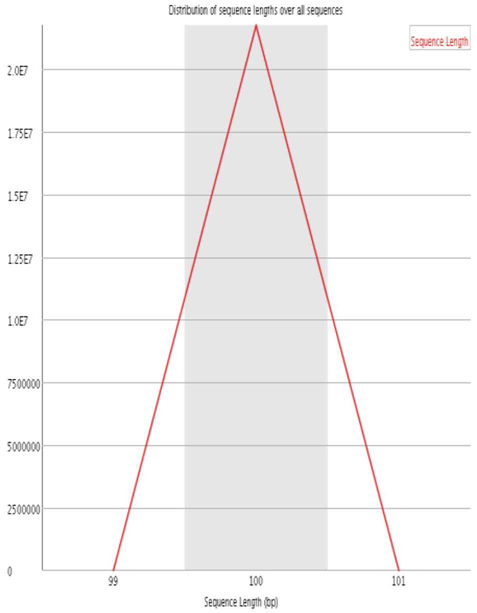
! Per sequence GC content



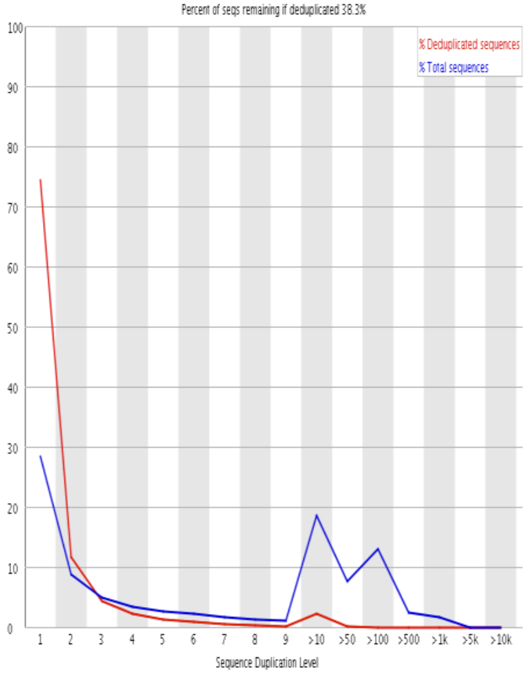
✓ Per base N content



Sequence Length Distribution



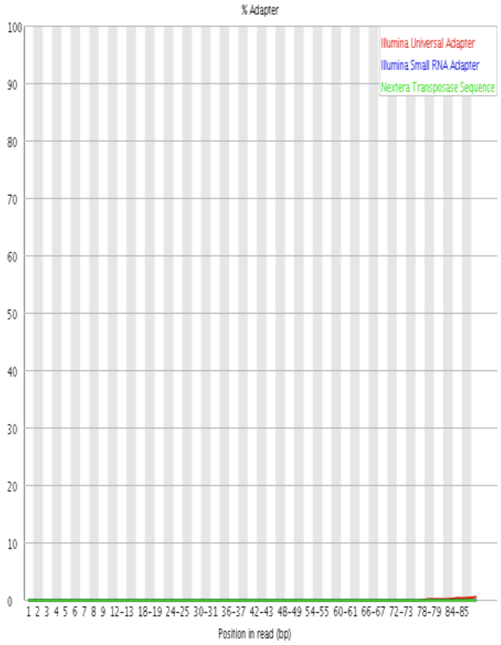
Sequence Duplication Levels



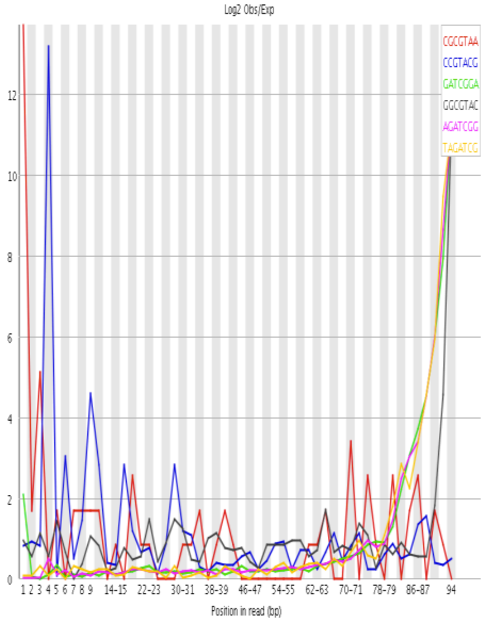
Overrepresented sequences

No overrepresented sequences

Adapter Content

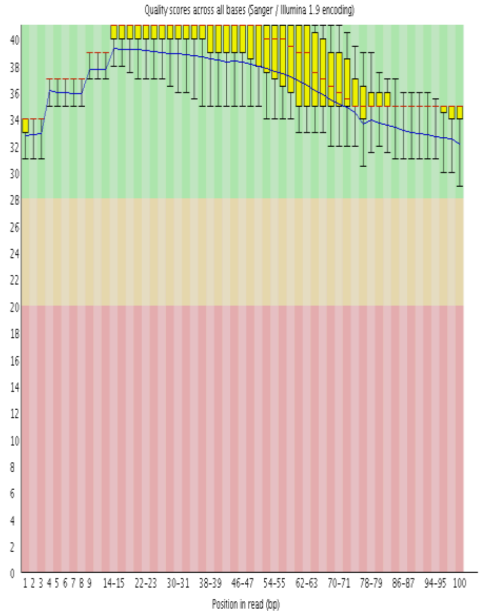


Kmer Content

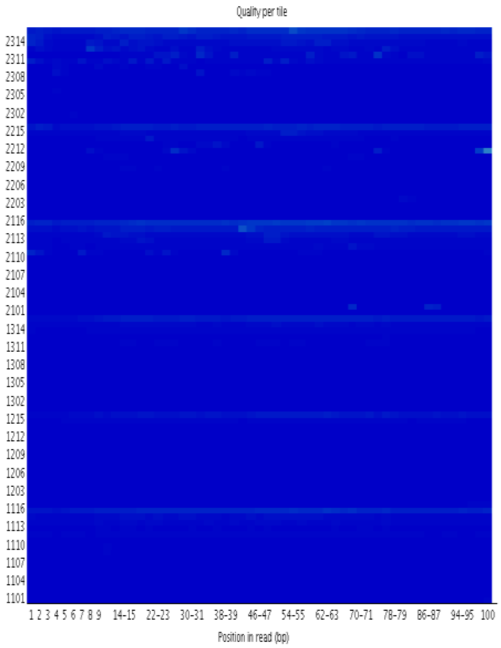


Quiescent Mouse 3

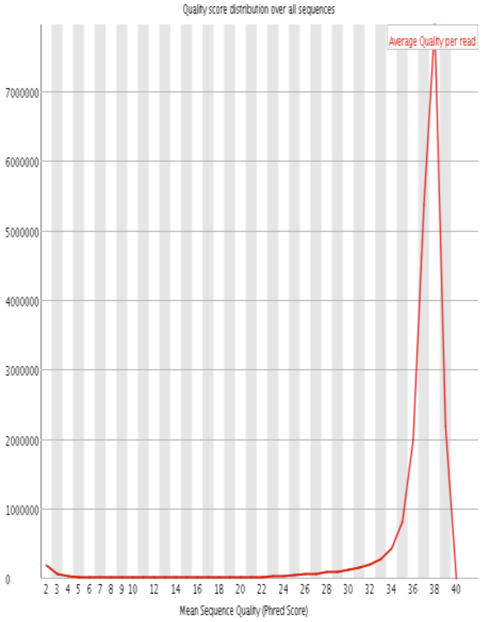
✓ Per base sequence quality



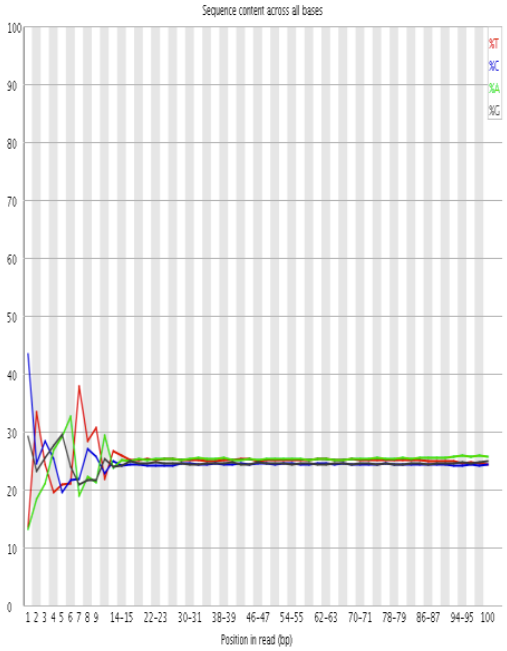
✓ Per tile sequence quality



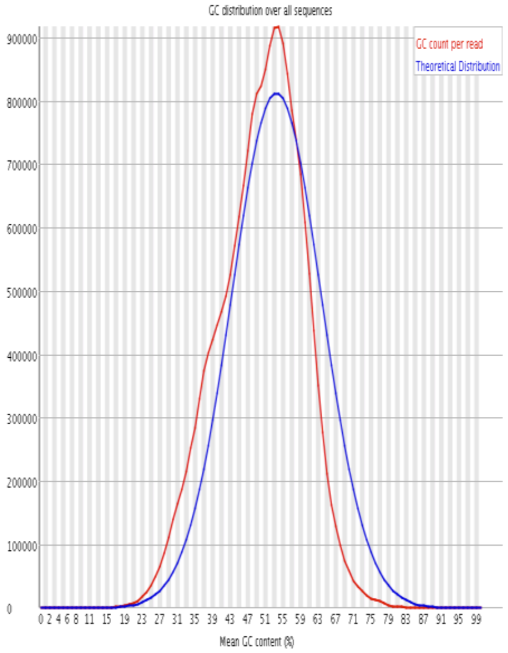
✓ Per sequence quality scores



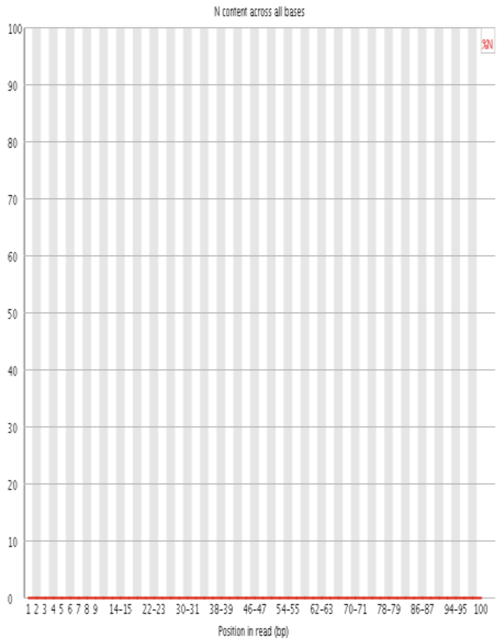
✗ Per base sequence content



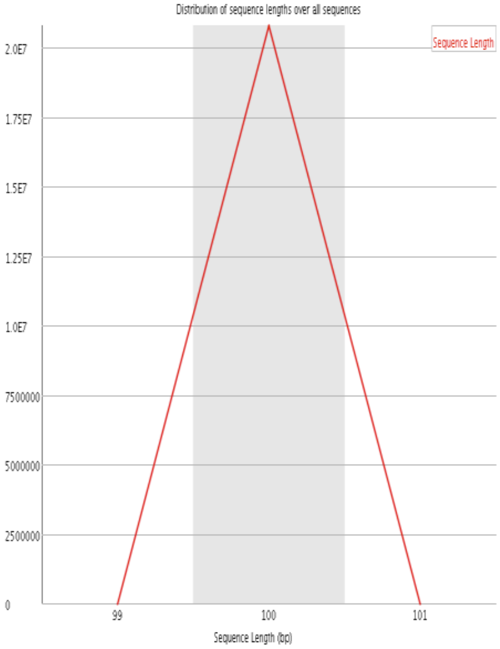
⚠ Per sequence GC content



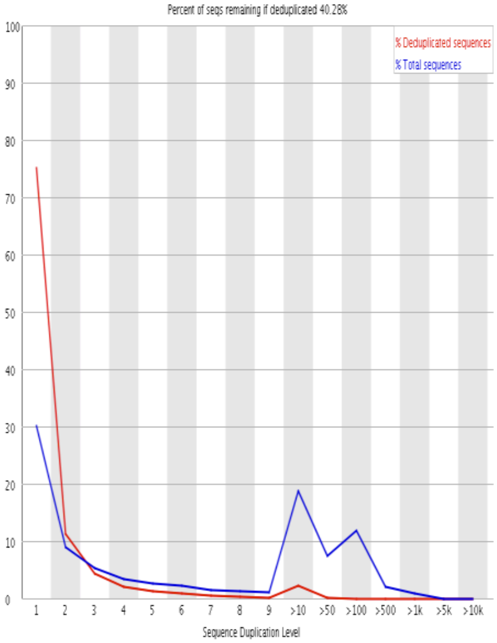
✓ Per base N content



✓ Sequence Length Distribution



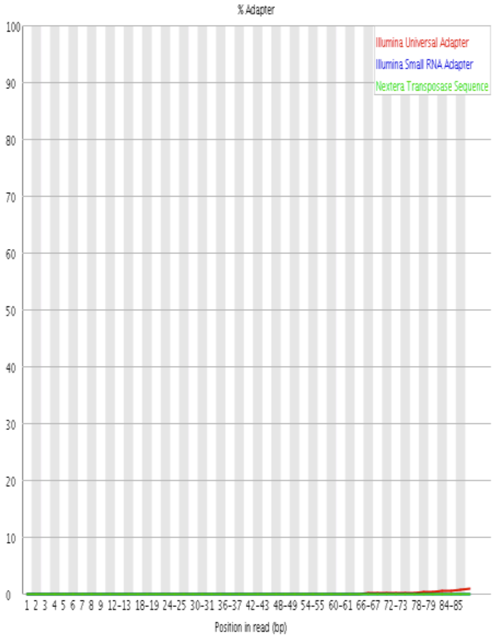
✗ Sequence Duplication Levels



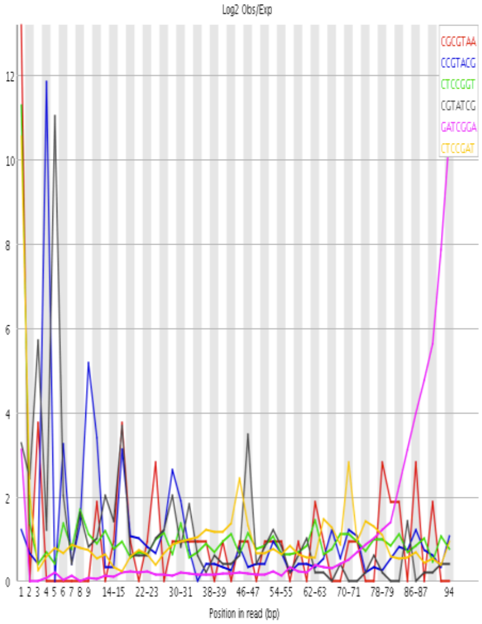
Overrepresented sequences

No overrepresented sequences

Adapter Content



Kmer Content



APPENDIX C-- Methods for the Proteomic Analysis of Isolated Virgin, Pregnant, and Primiparous Quiescent Mammary Epithelial Cells

Protein Extraction

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

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

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

Two-dimensional gel electrophoresis (2DGE)

All equipment and materials used for 2DGE were purchased from Bio-Rad

(Hercules, CA) unless otherwise stated. All buffer reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated.

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

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

Gel analysis, spot picking and trypsin digestion

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

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

Mass spectrometry and protein identification

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