Investigation of changes in histone modifications and factor recruitment to milk protein genes during mammary gland differentiation and identification of novel STAT5A-DNA binding sites in mammary gland tissue

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften (15)
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Stefanie Kliem
aus Bad Homburg vor der Höhe

Frankfurt am Main 2008
D30

vom Fachbereich Biowissenschaften der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. V. Müller
Gutachter:	Prof. Dr. A. Starzinski-Powitz
	PD Dr. E. Pfitzner

Datum der Disputation:

Eidesstattliche Erklärung

die selbstständige Anfertigung dieser Arbeit erkläre ich an Eides statt.

Frankfurt am Main, den 13 Juni 2008

(Stefanie Kliem)

Table of Contents

1. Introduction

1.1 Mouse mammary gland development

The development of the mammary gland is regulated by the endocrine system (Neville et al., 2002). Various steroid and peptide hormones coordinated the development of the mammary gland in mammals with respect to the reproductive status and the requirement for milk by the offspring (Hennighausen and Robinson, 2005; Neville et al., 2002). The major morphological developmental alterations of the mammary gland occur postnatal in the adult mammal during the distinct stages of puberty, pregnancy and lactation (reviewed in Hennighausen and Robinson, 2005, 2001, 1998).

During fetal development a mammary anlage is established that originates from the skin (Fig. 1.1). After birth a moderate ductal elongation and branching occurs from the nipples into the fat pad corresponding to the general growth rate of the mammal. With the onset of puberty and the secretion of gonadal hormones estrogen and progesterone the mammary ducts begin to elongate and branch out into the entire mammary gland fat pad. The mammary fat pad or stroma is a connective tissue that supports the epithelial tissue compartment. It harbors mainly adipocytes, fibroblasts and cells from the vasculature and haematopoietic system. During puberty bulb shaped structures at the ends of the ducts referred to as terminal end buds (TEB) are observed. These TEB structures contain highly proliferative precursor cells, which give rise to mammary epithelial and myoepithelial cells (Hennighausen and Robinson, 1998; Humphreys et al., 1996). The TEB structure disappears once the entire fat pad is filled with ducts and the virgin animal has matured with about 12 weeks of age. Additional ductal branching forms and disappears with each estrous cycle (Masso-Welch et al., 2000). Beginning with pregnancy an extensive ductal branching and lobulo-alveolar proliferation that is controlled mainly by prolactin, placental lactogens and progesterone occurs. This extensive proliferation of the ductal and alveolar cells continues throughout pregnancy and early lactation. The alveoli are ball like structures made up of luminal and basal epithelial cells that surround a central lumin. The luminal epithelial cells undergo functional differentiation during late pregnancy and lactation to produce and secrete the milk components into the alveoli lumen. The basal myoepithelial cells are contractile and thus able to transport the milk components into the alveoli lumen in direction of the nipples along the ducts during lactation. By parturition the mammary fat pad is completely filled with secretory lobulo-alveolar structures, that produce milk during lactation. Cytokines, like prolactin and possibly ERBB4 ligands confer the main signaling that controls the proliferation, differentiation and survival of the mammary

epithelium during pregnancy and lactation (Hennighausen and Robinson, 2005) (Fig. 1.1).

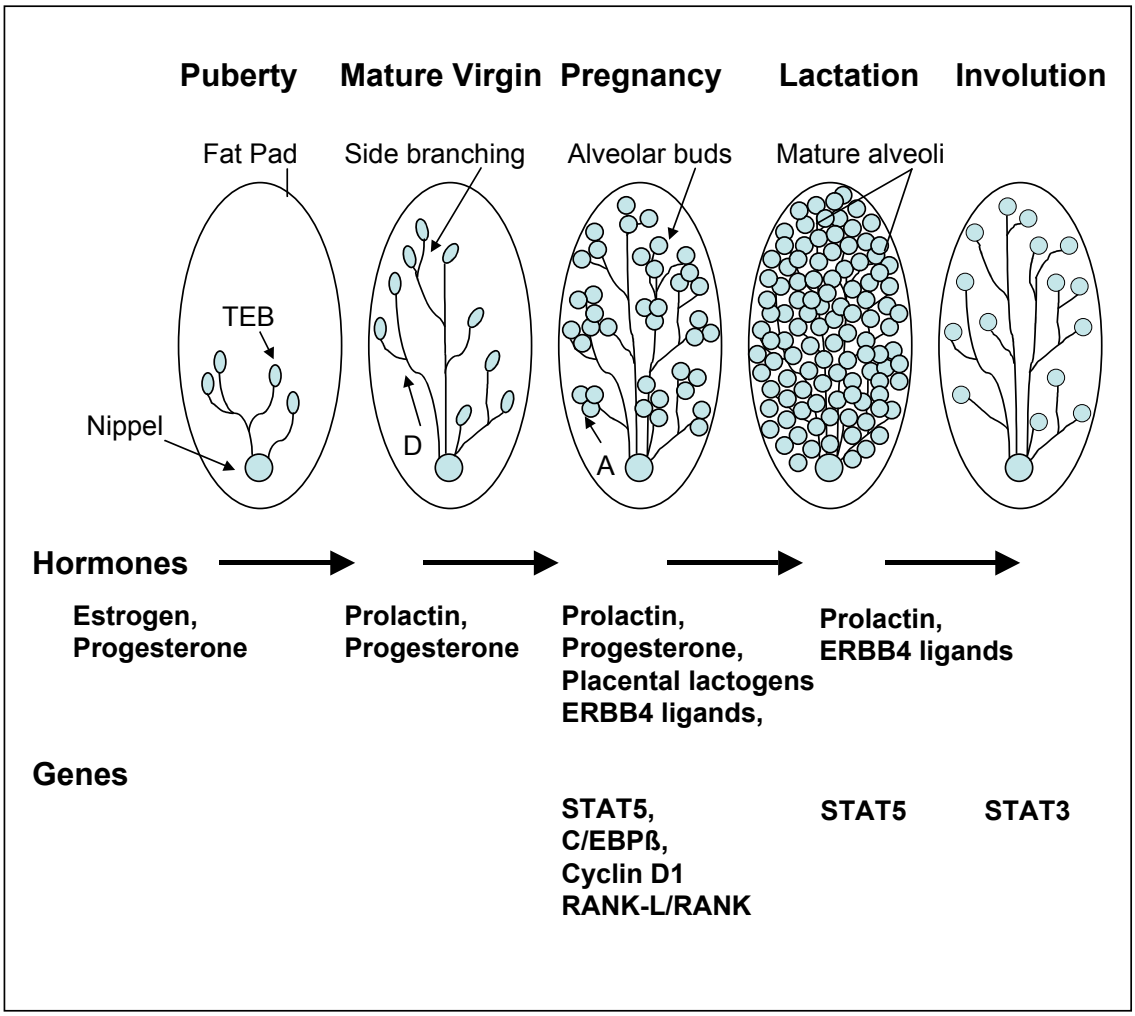


Figure 1.1 The different stages of mammary gland development are regulated by specific hormone and gene products.
Schematic picture depicting the postnatal developmental stages of the mammary gland. The oval shapes represent the mammary fat pad (stroma). During puberty the ovarian produced hormones estrogen and progesterone drive the ductal elongation and branching of the epithelial ducts from the nippel out into the fat pad. The terminal end buds (TEB) are depicted as bulbs at the tips of the growing and elongating epithelial ducts. In the mature virgin mouse the entire fat pad is filed out with extensively branched epithelial ducts (D).
With onset of pregnancy, the hormones prolactin, progesterone and placental lactogens as well as other signaling factors such as ERBB4 ligands and RANK-L control the proliferation and development of alveolar epithelial cell structures (A).
By lactation the entire mammary fat pad is filled with mature alveoli that contain differentiated luminal secretory cells, which produce and secrete the milk components into the lumen of the alveoli structures. The pituitary gland produced hormone prolactin is essential for maintaining lactation and the transcription factor STAT5 mediates the main prolactin and possibly ERBB4 ligand signaling. Apart from STAT5, RANK-L/RANK, Cyclin D1 and C/EBPß are gene products important in the proliferation and differentiation of mammary epithelial cells. During involution, tissue remodeling and cell death is controlled by another STAT member namely, STAT3. (Adapted and modified from Hennighausen and Robinson, 2005 and 2001).

After the young are weaned and suckling stops, there is a loss in the prolactin signal that is accompanied by extensive apoptosis and remodeling of the alveolar epithelium. This stage is called involution. The secretory alveolar epithelium of the mammary gland regresses and returns back to a near mature virgin state in appearance. With each following pregnancy then, the mammary gland undergoes repeated cycles of growth, differentiation and regression that is controlled mainly by steroid and peptide hormones in addition to other activated signaling factors (Liu et al., 1996).

With the use of gene targeting experiments it has become possible to identify the components in the signaling pathways and determine their contribution to the development of the mammary gland. A selection of those genes found to be important contributors to mammary gland development during pregnancy and lactation are shown (Fig. 1.1). Loss of these genes during mammary gland development results in the impairment of correct ductal branching or alveolar development (reviewed in Hennighausen and Robinson, 2001).

The development of the mammary gland is characterized by the coordinated interaction between complex signaling networks of different hormones and signaling molecules leading to a functional differentiated mammary gland that secretes milk at the end of pregnancy. The STAT transcription factors, mostly STAT5 and STAT3, are essential signaling molecules.

1.2 STAT signal pathway activation and structure

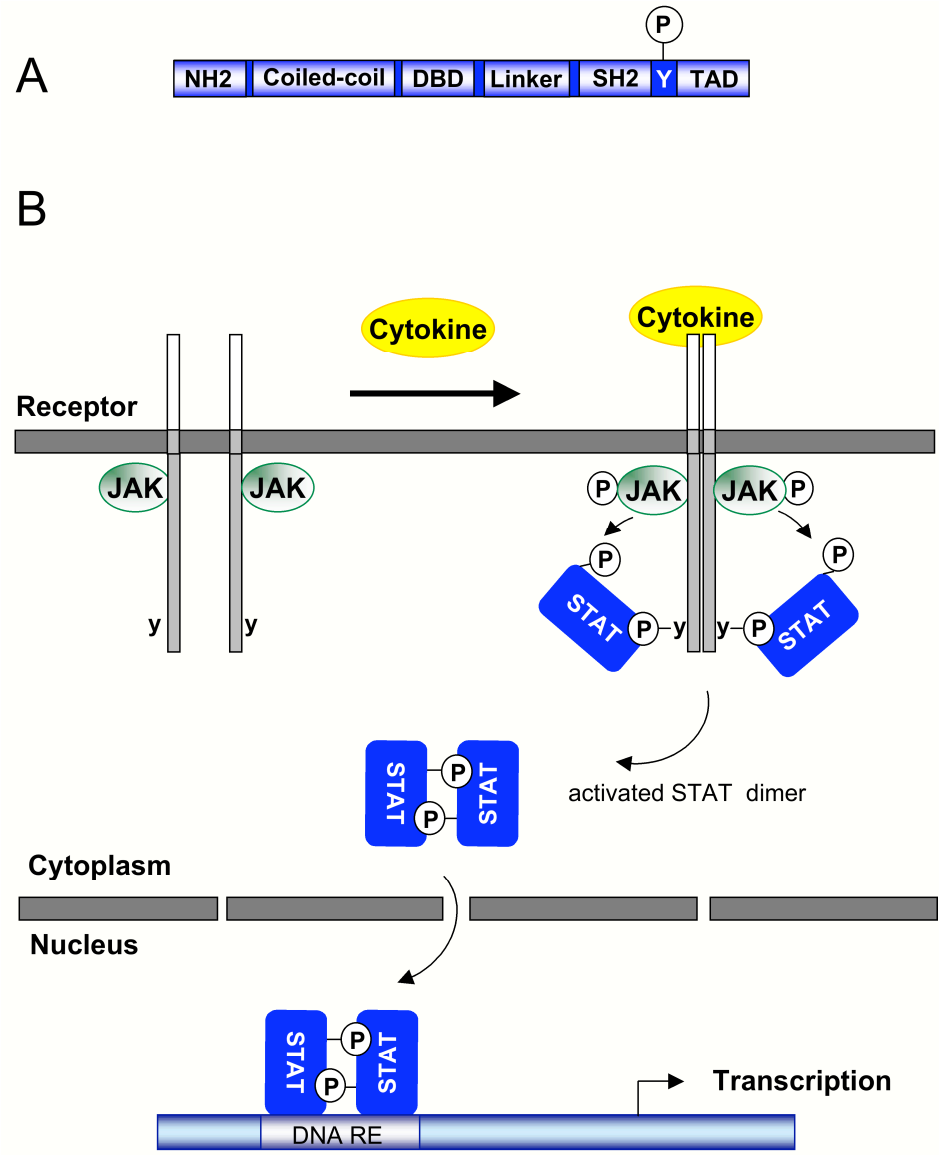
There are seven mammalian Signal Transducers and Activators of Transcription (STAT) family members (STAT 1, 2, 3, 4, 5A, 5B and 6) each encoded by different genes, which share similarities in structure and function.

The STATs are transcription factors that are activated by a wide variety of cytokines, like interferons and interleukins as well as by hormones and growth factors (reviewed in Schindler et al., 2007; Kisseleva et al., 2002; Levy and Darnell, 2002; Bromberg, 2001). There are several modes by which STAT proteins can be activated. Activation of STATs can be mediated through cytokine receptors, receptor tyrosine kinases (growth factors), G-protein-coupled receptors (peptides/chemokines) and non-receptor tyrosine kinases (oncoproteins, v-src) (Mertens and Darnell, 2007; Lim and Cao, 2006).

A, well known mode of STAT activation is the JAK-STAT signaling pathway that mediates cytokine signaling. This pathway is involved in the regulation of various cellular processes including development, proliferation, differentiation and apoptosis. The JAK-STAT signaling pathway transmits information from extracellular signals directly to the gene promoters in the nucleus without the need for second messengers.

Cytokine ligands binding to the cell surface receptor result in the dimerization or oligomerization of the receptor and initiate activation of Janus kinases (JAKs) that are associated with the intracellular domain of the receptor (Fig. 1.2 B).

Figure 1.2 The structure of STAT and the JAK/STAT signaling pathway.



(A) The schematic figure depicts the structural and functional domains of STAT.
The major functional domains are displayed: amino-terminal domain (NH2), Coiled- coil domain, the DNA-binding domain (DBD), the linker domain, the SH2 domain (src homology domain), the conserved tyrosine (Y) residue, that is phosphorylated during activation and the carboxy-terminal transactivation domain (TAD).
(B) Schematic diagram of the cytokine induced JAK/STAT signaling pathway.
Binding of a cytokine to its cognate receptor allows dimerization of the cytokine receptor and transphosphorylation and activation of the receptor associated Janus kinases (JAKs). The activated JAKs phosphorylate specific tyrosine residues on the cytoplasmic tails of the cytokine receptor which serve as docking sites for the STAT proteins via their SH2 domain. The recruited STAT proteins are then tyrosine phosphorylated by the JAKs. This induces the release and dimerization of the activated (tyrosine phosphorylated) STAT protein followed by translocation into the nucleus. In the nucleus the activated STATs bind to specific DNA recognition elements (DNA RE) in the promoters of target genes and regulate transcription.

The activated JAKs trans-phosphorylate specific tyrosine residues on the intracellular domains of the receptor. These phospho-tyrosine residues on the receptor then can serve as docking sites for the latent cytoplasmic STATs. The STAT proteins are recruited to the receptor via their SH2 domain and the JAKs phosphorylate the STATs on the single specific tyrosine residue found in the STATs C-terminal tyrosine activation domain. Homo and heterodimers of STATs are then formed by reciprocal binding of the phospho-tyrosine of one STAT monomer with the SH2 domain of the other partner STAT. The activated and dimerized STATs translocate to the nucleus and bind to specific DNA-binding elements and regulate the transcription of responsive target genes.

All seven mammalian STAT family members have in common six conserved structural domains (reviewed in Schindler et al., 2007; Kisseleva et al., 2002; Levy and Darnell, 2002; Bromberg, 2001) namely, the amino-terminal (NH_2), coiled-coil, DNA-binding (DBD), linker, Src homology domain 2 (SH2) and transcriptional activation domain (TAD) (Fig. 1.2A). The amino-terminal (NH2) domain is important for the tetramer formation between two STAT dimmers at genomic regions with two adjacent DNA-binding elements and for the translocation into and out of the nucleus. The adjacent coiled-coil domain functions as a protein-protein interaction domain. The coiled-coil domain is followed by the DNA-binding domain (DBD), which is essential for binding to specific DNA-binding element in promoter or enhancer regions of target genes. The adjacent linker domain is important for maintaining the correct conformation between the DBD and the adjacent SH2 domain. The SH2 domain is a phospho-tyrosine binding domain, which is important for receptor recruitment and dimerization. The SH2 domain allows for the binding of phospho-tyrosine residues on the activated cytokine receptors and the phospho-tyrosine residue on other STAT molecules to form STAT dimers. Following the SH2 domain is a tyrosine residue that requires phosphorylation for activation of STAT proteins. The carboxyl-terminal (C) residues are the least conserved among STAT family members and make up the transcriptional activation domain (TAD), which is important for transcriptional activation and interaction with diverse transcriptional regulators.

1.3 STAT5

Initially, STAT5 was identified as a mammary gland factor (MGF) for which the tyrosine-phosphorylation dependent activation is induced by prolactin (PRL). STAT5 is important in the terminal differentiation of mammary gland epithelial and expression of milk proteins such as ß-casein and WAP during pregnancy and lactation (Liu et al.,

1995; Wakao et al., 1992). Inactivation of the genes that belong to the PRL signaling pathway in the mouse genome, revealed that the PRL receptor and STAT5 are both absolutely essential components for mammary gland development and function (Liu et al., 1997; Ormandy et al., 1997). Meanwhile, it has become clear that the STAT5 mediated signaling is not only relevant in mammary gland development, but also in haematopoisis and immune regulation (Levey and Darnell 2002).

Two independent genes arranged in tandem, that are localized to chromosome 11 in the mouse genome, code for the highly homologous proteins STAT5A and STAT5B. The murine STAT5A and STAT5B proteins, are 96% identical at an amino acid level and differ mainly at the C-terminus (Mui et al., 1995). The high sequence homolgy between the two *Stat5* genes indicate that they arose from a recent tandem gene duplication event (Copeland et al., 1995).

STAT5A and STAT5B can be activated by a wide variety of cytokines, hormones and growth factors such as interleukins (IL-2, IL-7, IL-9, IL-15, IL-3, IL-5), GH (Growth Hormone), PRL (Prolactin), EGF (Epidermal Growth Factor), EPO (Erythropoietin) and GM-CSF (Granulocyte/Macrophage Colony Stimulating Factor) (Bromberg, 2001). Both are expressed in various tissues and cell lines at comparatively equal levels (Mui et al., 1995), but subtle differences in tissue distribution of STAT5A and STAT5B mRNAs imply that they have distinct functions (Liu et al., 1995). STAT5A is more abundant in the mammary gland, while in the liver STAT5B is more prevalent (Brisken and Rajaram, 2006).

1.3.1 *In vivo* function of STAT5A and STAT5B determined by gene targeting deletion in mice

Although, STAT5A and STAT5B proteins are structurally very similar and display functional redundancy they also have functional differences *in vivo* as demonstrated by the phenotypes observed in various STAT5-deficient mice.

In vivo knockout experiments have shown that STAT5A is needed for mammopoiesis and lactogenesis. STAT5A knockout mice display a defect in the development in the mammary gland and are not able to secrete sufficient milk (Liu et al., 1997). In female STAT5A deficient mice the mammary gland development was altered, since there was insufficient milk to sustain the young even though low levels of milk proteins were detected. Following pregnancy the alveolar development in the mammary gland is strongly reduced in comparison to wild-type animals. Milk protein expression such as alpha-lactalbumin and WAP (whey acidic protein) expression was barely detectable.

On the other hand, STAT5B deficient mice display a different phenotype, which appears most prominent in the male mouse. There is a loss of sexual dimorphism in body growth rates of male mice that is attributed to a failure in the growth hormone response and changed expression of genes in the liver (Teglund et al., 1998; Udy et al., 1997). In both male and female mice, there were effects observed on fat deposition and delays in hair growth (Udy et al., 1997). This indicates that depending on the tissue and cell type specific context, the functional STAT5A was not able to compensate for the loss of *stat5b* gene. This suggests that STAT5A and STAT5B also have distinct functions *in vivo*. In general, STAT5A is involved mainly in prolactin mediated signaling, while STAT5B is more important for mediating growth hormone signaling. Mice deficient in both STAT5A and STAT5B display defects in various organs (Teglund et al., 1998). The STAT5 deficient female mice are infertile and therefore mammary gland development could not be investigated. The infertility was not observed in the STAT5A or STAT5B deficient mice indicating functional redundancy in the STAT5A and STAT5B proteins. In addition these mice were much smaller in body size and weight than their wild-type littermates (Teglund et al., 1998). Also reduced numbers in bone marrow progenitors hint at defects in hematopoiesis within these STAT5 deficient mice (Teglund et al., 1998).

The original STAT5 deficient mice where created by using a gene targeting strategy to delete the first protein-coding exon (Teglund et al., 1998). These original STAT5 knockout mice are believed to still encode a N-terminal truncated, but possibly still functional STAT5 protein at extremely low levels (Hoelbl et al., 2006; Yao et al., 2006). Therefore more recent studies have used *stat5* knockout mice where the entire *stat5a* and *stat5b* locus was deleted (Cui et al., 2004). These mice display a more severe phenotype. *Stat5*-null mice, that have the entire *stat5* locus deleted in their germ line, died mostly perinatal due to acute anemia and other still to be defined physiological defects (Cui et al., 2004). Studies with the *stat5*-null mice have further confirmed the importance of STAT5 in erythropoiesis and lymphopoiesis (Yao et al., 2006) and shown that STAT5 is required for normal lymphoid development and differentiation. Conditional mouse mutant models enable the study of later stages in development and allow for stage and cell specific gene deletion. During pregnancy conditional gene inactivation show that STAT5 is required for mammary gland epithelial cell proliferation, differentiation and survival at this stage of mammary gland development (Cui et al., 2004).

1.3.2 Regulation and function of STAT5 in the mammary gland

In the mammary gland both the prolactin receptor and ERBB4 receptor, a receptor tyrosine kinase, can activate STAT5 by phosphorylation of specific tyrosine residues to convey the proliferative and differentiation signals to the mammary epithelial cells (Hennighausen and Robinson, 2005) (Fig. 1.3).

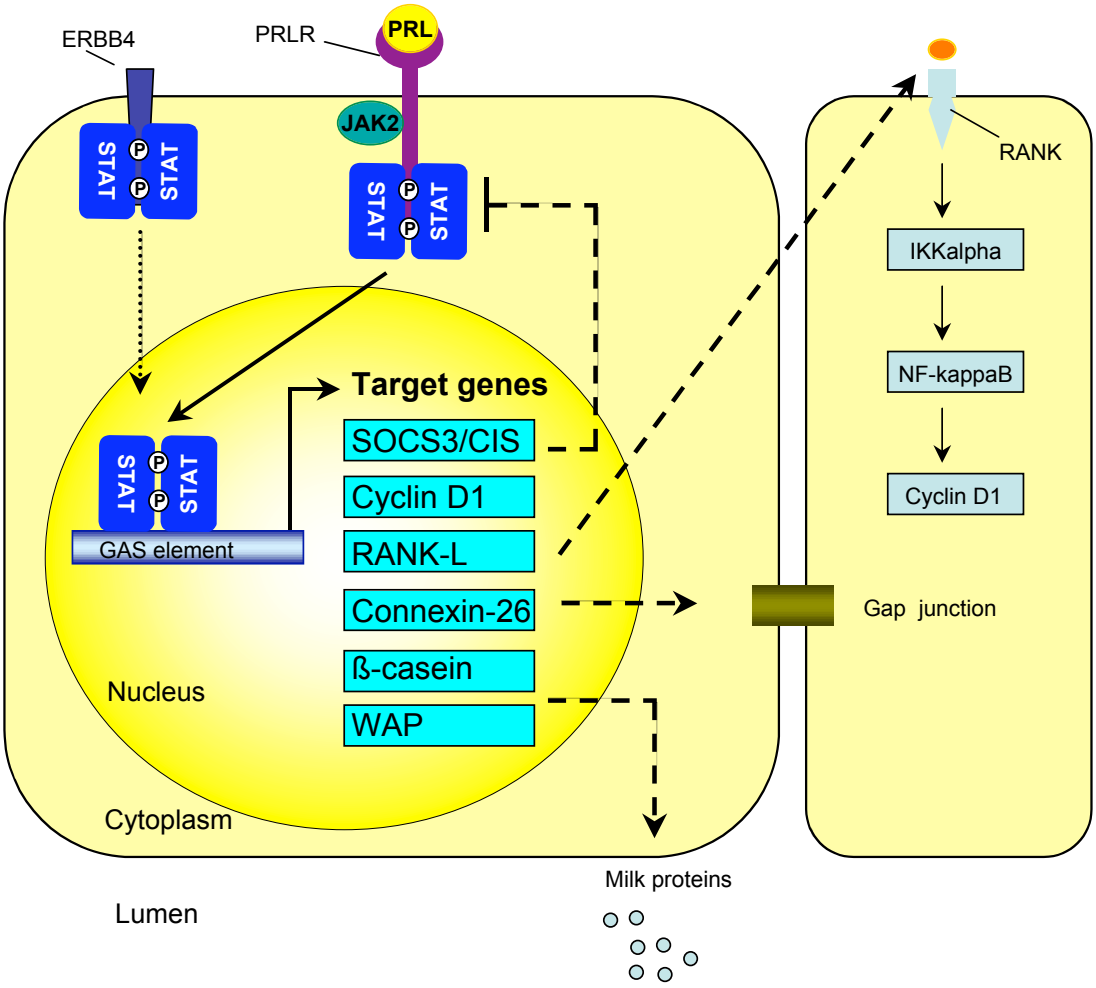


Figure 1.3 Bona fide STAT5 target genes that are induced by STAT5 signaling in mammary epithelial cells.

Binding of prolactin during pregnancy and lactation to the prolactin receptor activates the receptor-associated Janus Kinases 2 (JAK2) and result in the phosphorylation and activation of STAT5. Signaling through the ERBB4 is also able to phosphorylate and activate STAT5 and is thus thought to influence the functional differentiation of mammary epithelial cells. Activated STAT5 dimers bind the GAS motif in the DNA of the target genes and regulate transcription. Direct target genes that contain the GAS binding motif are shown. SOCS3 and CIS are both negative regulators of cytokine signaling pathway. When the RANKL growth factor ligand binds its receptor it activates the NF-KappaB signaling pathway, which in turn influences the regulation of Cyclin D1, that controls cell proliferation. Connexin 26 is a structural protein that functions as a component of the gap junction. ß-casein and WAP are the classic milk proteins that reflect the terminal differentiation state of secretory epithelial cells during pregnancy and lactation (Adapted and modified from Hennighausen and Robinson, 2008 and 2005).

The ERBB4 receptor is required for the phosphorylation of STAT5 during late pregnancy (Long et al., 2003) and in addition also induces phosphorylation of specific serine residues of STAT5A (Clark et al., 2005; Long et al., 2003). The activation of STAT5 by the two different signaling pathways can potentially lead to activation of distinct downstream targets (Brisken and Rajaram 2006).

PRL activated STAT5 dimerizes and translocates into the nucleus where it activates target genes like WAP and ß-casein milk protein genes that have GAS (gamma-interferon activated sequence) DNA-binding elements in the promoters (Watson and Burdon, 1996). The typical GAS DNA recognition element found in the promoters of these STAT target genes is characterized by a TTC(N3)GAA sequence. Although both STAT5A and STAT5B can bind the same DNA-binding sites differences can be attributed to subtle differences in DNA-binding affinities and the difference found in the (TAD) transcriptional activation domain. Differences in expression levels of the two genes as well as different isomers in different cells and tissues can contribute to the functional differences in gene regulation by the two STAT5 proteins observed (Grimley et al., 1999).

Although STAT5 is expressed in many tissues and is activated by a variety of cytokines the number of direct STAT5 target genes in the mammary gland during pregnancy and lactation is only beginning to be explored. DNA microarray analysis and differential expression profiling have been used to identify genes expressed in mammary gland tissue during pregnancy and lactation. Using expression profiles from transplanted wild-type and prolactin receptor null mammary epithelial cells, genes were identified that are dependent on the PRL signaling pathway during pregnancy (Harris et al., 2006). This list includes the milk proteins caseins alpha, beta and kappa as well as the extracellular proteinase inhibitor (Expi) and the ets family transcription factor E74-like factor 5 (Elf5). Another study by Clarkson et al. (2006) also utilized microarray analysis to identify STAT5-regulated genes in a mammary epithelial cell line (KIM-2) that was induced to differentiate using a conditional activated dimer of STAT5A. Here, about 25 principle transcripts were induced, such as the milk proteins ß-casein and alpha-casein, WAP, and SOCS2. The expression of these genes here is definitely regulated by STAT5, but these genes do not necessarily need to be direct target genes bound by STAT5, but may be regulated indirectly.

The majority of real STAT5 target genes identified so far in mammary gland tissue are few and variable in their function. Some bona fide target genes of STAT5 are ß-casein, Whey acidic acid (WAP), SOCS3, CIS, RANK-L (receptor activator of nuclear factor NF-*k*B)-ligand), and Cyclin D1 (Fig. 1.3). The milk protein genes ß-casein and WAP are probably the best known STAT5 target genes in the mammary gland. The expression of these genes marks the terminal differentiation stage of a mature secretory mammary

epithelial cell. Both ß-casein and WAP contain GAS motifs in their promoters through which STAT5 mediates their transcriptional activation in the mammary gland epithelium during pregnancy and lactation (Li and Rosen, 1994; Pittius et al., 1988). There is a severe reduction in the expression of both milk proteins when STAT5A is not present and when both STAT5A and STAT5B are deleted, expression of both milk protein ceases (Miyoshi et al., 2001; Liu et al., 1997). Receptor activator of NF-kappaB ligand (RANKL) is required for the alveolar development during pregnancy (Fata et al., 2000) and is a direct STAT5 target gene, since RANKL has a gamma-interferon activating sequence (GAS element) in its promoter that confers PRL responsiveness (Srivastava et al., 2003). Binding of the growth factor RANKL to its receptor activates the NF-kappaB signaling pathway. A downstream target gene of this signaling pathway is the cell cycle regulator Cyclin D1, which is also another bona fide STAT5 target gene. Cyclin D1 contains one STAT5 responsive GAS motif in its promoter and is important for the normal proliferation and development of mammary alveolar cells (Brockman et al., 2002). Apart from the Cyclin D1, which is involved in regulating cell proliferation, there are also some negative regulators of the JAK/STAT signaling pathway that are STAT5 target genes. The members of the suppressors of cytokine signaling (SOCS) are induced by STATs and regulate cytokine signaling in a negative feedback loop (Kubo et al., 2003). There are eight members of the SOCS family (SOCS1-7) and the cytokine-inducible Src-homology-2 (SH2) protein (CIS) and some SOCS members bind to the phosphotyrosines on the cytokine receptors via their SH2 domain and thus directly compete with the STAT-binding and activation. SOCS proteins can also target their cytokine receptor/JAK substrate for degradation by the ubiquitin proteosome pathway via their SOCS box and thus inhibit signaling (Zhang et al., 1999).

Both the CIS promoter (Matsumoto et al., 1997) and SOCS3 promoter (Auernhammer et al., 1999) contain the classic GAS STAT binding motifs. In mammary epithelial cells both CIS and SOCS3 are inducible by PRL and expression of either inhibits STAT5 tyrosine phosphorylation in mammary epithelial cells (Tonko-Geymayer et al., 2002). An additional gene that has been implicated to be important in mammary gland development is the gap junction component Connexin 26. Connexin 26 contains a conserved STAT5 consensus site in its promoter (Kiang et al., 1997) and is essential for alveolar development in early pregnancy as shown by use of conditional connexin 26 null mice (Bry et al., 2004).

Still much needs to be learned about the program of altered gene expression that controls the developmental events during mammary gland differentiation. Since STAT5 plays a major role in mediating the developmental signals of the hormone prolactin during pregnancy and lactation it is crucial to identify the *in vivo* bona fide genomic targets of STAT5

1.4 Regulation of transcription

1.4.1 Basal transcription

Initiation of transcription is a main step in controlling the expression of protein-coding genes. The initiation of transcription requires the formation of the pre-initiation complex (PIC) on the promoter of the gene. The PIC includes RNA polymerase II and the general transcription factors (GTFs) TFIIA, TFIID, TFIIB, TFIIF, TFIIE and TFIIH. The current model, based on *in vitro* studies and structural analysis, suggest the following stepwise mechanism (reviewed in Hahn, 2004; Woychik and Hampsey, 2002). In the initial steps there is the binding of the TATA-box binding protein (TBP), a subunit of TFIID, to a TA-rich region (TATA-box) in the promoter region of the gene. TFIIA is involved in the stabilization of TBP and TFIID binding to DNA and is involved in recruitment of TFIIB. TFIIB stabilizes the binding of the TFIID complex to the promoter and together with TFIIF recruits the unphosphorylated RNA polymerase II to the promoter. After the recruitment of TFIIE and TFIIH, the ATP dependent TFIIH helicase activity initiates initial unwinding of DNA near the transcription start site. Initiation of transcription begins with the phosphorylation of the C-terminal domain (CTD) of RNA polymerase II, and allows for the release of the polymerase from the PIC and elongation to begin.

The CTD is found in the C-terminus of the largest subunit of RNA polymerase II and contains a highly conserved repeat sequence (YSPTSPS) that varies in the number of repeat occurrences, between species (Dahmus, 1995). The CTD of the elongating RNA polymerase II is highly phosphorylated and therefore can serve as a mark for RNA polymerase II that is transcribing actively (Dahmus, 1995). RNA polymerase II beginning initial transcription is predominantly phosphorylated on the serine in the fifth position of the repeat sequence while during later elongation the serine in the second position of the repeat sequence is mainly phosphorylated (Komarnitsky et al., 2000). The CTD serves as a scaffold for binding of factors that are involved in the regulation of transcription. The binding of these factors is influenced by the CTD phosphrylation (Hirose and Manley, 2000).

Even though the general transcription factors are sufficient for the initiation of the basal transcription on a naked DNA template *in vitro*, additional factors are required for specific gene transcription in native chromatin context. In eukaryotes the regulation of gene expression needs exact spatial and temporal coordination of a large number of general and specific transcription factors at *cis*–regulatory elements such as promoters and enhancers (reviewed in West and Fraser, 2005; Smale and Kadonaga, 2003;

Orphanides and Reinberg, 2002). The *cis*–regulatory elements serve as center stage for integrating all the regulatory inputs from various signal transduction pathways to regulate the transcription of the target gene. Transcription factors bind sequence specific promoter or enhancer DNA elements and recruit additional regulatory factors. The recruited factors can be chromatin remodeling enzymes and histone modifying enzymes that alter the chromatin structure to make it permissive for the binding of the basal transcription machinery and subsequent transcription. In which order and timing the recruitment of factors and the restructuring of the chromatin occurs, varies depending on the promoter, stimuli, cell type and developmental context (Cosma, 2002).

1.4.2 Chromatin structure

On a basic level, the transcription of genes is dependent on the DNA accessibility. The basic repeating unit of chromatin is the nucleosome (Kornberg, 1974). The nucleosome consists of 147 base pairs of DNA wrapped around a set of eight core histone proteins, which is composed of two copies each of the four histones H2A, H2B, H3 and H4. In the native chromatin context the chromatin structure hampers formation of the PIC and the binding of transcription factors.

A change in chromatin structure can be achieved by the two main mechanisms of chromatin remodeling as well as through specific histone modifications. First chromatin remodeling is achieved by enzymes, that use ATP hydrolysis to alter nucleosome structure and positioning on the DNA, thus making previously shielded DNA accessible to regulatory factors. And second, histone modifications are generated by enzymes that can covalently modify the histones at specific residues mainly on the amino-terminal ends (tails) of the histones (Kouzarides, 2002; Strahl and Allis, 2000). These histone modifications themselves can influence the stability of the histone-DNA and inter-nucleosomal interaction and/or serve as binding sites for the recruitment of additional regulatory factors.

1.4.3 Function of histone modifications in transcription

The histones are basic proteins with amino-terminal tails that are highly conserved and extend from the nucleosome. These histone tails are subject to various post-translational modifications, such as acetylation methylation, phosphorylation, ubiquitination and ADP-ribosylation (Spencer and Davie, 1999; Van Holde, 1988). The

most well studied histone modifications are lysine acetylation and methylation (Fig. 1.4).

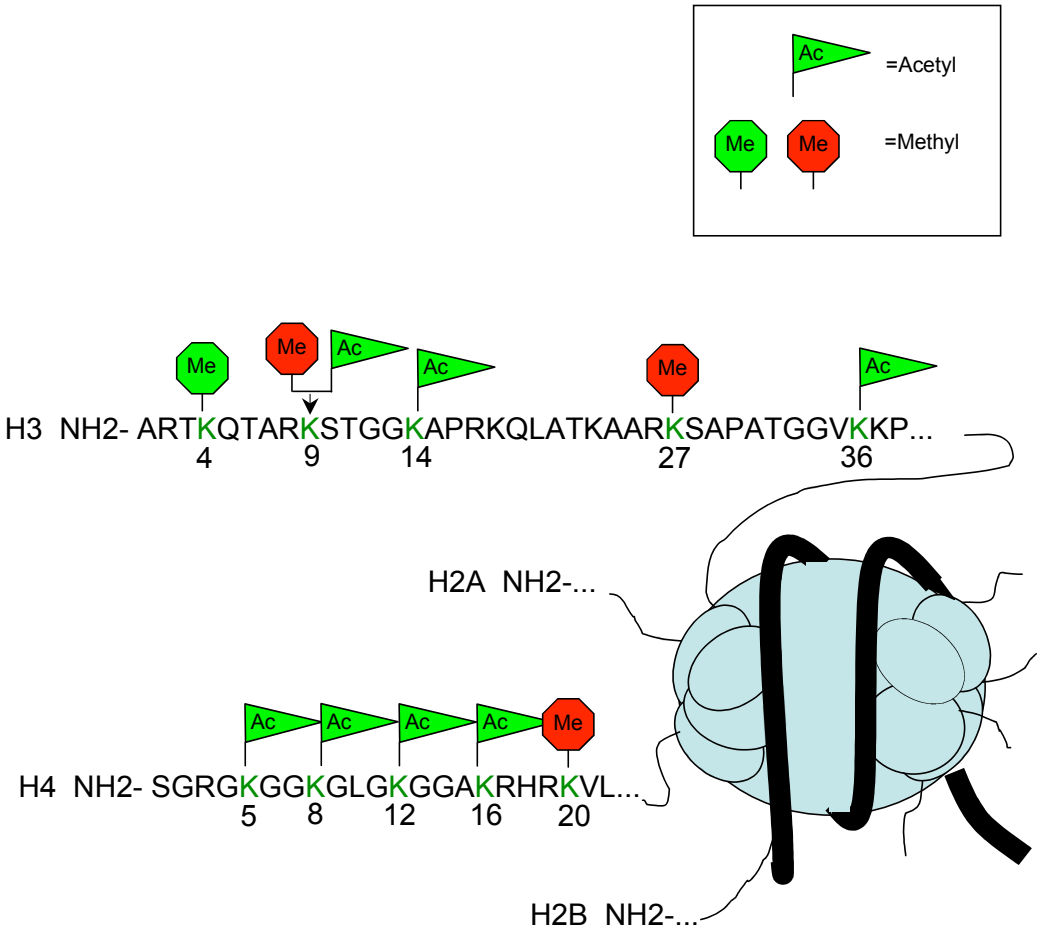


Figure 1.4 Acetylation and methylation on specific lysine residues of Histone H3 and H4. The nucleosome with DNA (Black ribbon) wrapped around it and the N-terminal tails extending from the individual core histones are show on the right. Depicted are the N-terminal tail amino acid sequences of human histone H3 and H4. Lysine acetylation and methylation on the N-terminal tails of histone H3 and H4 are marked by flags. Lysines (K) are in green and their amino-acid positions are shown below the respective modified lysine residue. Green Flags indicate modification associated with transcriptional active chromatin and red with repressed chromatin (Adapted and modified from Khorasanizadeh, 2004; Rice and Allis, 2001).

Distinct enzymes are responsible for the generation of various histone modifications (reviewed in Kouzarides, 2007). In addition, there are enzymes that can remove or modify the histone modifications, making the modification of histone tails a dynamic regulated process. An example for a very dynamic histone modification is the acetylation of lysine residues on the N-terminal tails of Histone H3 and H4, which is mediated by enzymes containing histone acetyltransferases (HATs) activity (Brown et al., 2000). Histone H3 and H4 can be acetylated on various lysine residues and this histone modification is found to correlate with chromatin regions that are transcriptional

active (Turner, 1993). In contrast to the HATs, the histone deacetylases (HDACs) are responsible for the removal of the acetylation.

In the past a charge neutralization model and the histone code hypothesis have been proposed to explain the function of these histone modifications in gene regulation (Schones and Zhao, 2008). Acetylation and phosphorylation are examples of two post-translational histone modifications that can change the overall charge distribution on the mostly positively charged N-terminal histone tails. Acetylation of the lysine residues in the histone tails neutralizes the positive charge. Phosphorylation of the conserved serine (10) in the N-terminal tail of histone H3 also results in the addition of a negative charge, thus decreasing the overall positive charge (Workman and Kingston, 1998). *In vitro* experiments indicate that histone acetylation leads to a relaxation and opening of chromatin structure (Wolffe and Hayes, 1999). These observations support a charge-neutralization model, in which these histone modifications can lead to an overall relaxation of the chromatin fiber (Turner, 2000).

With the Identification of enzymes, that modify the N-terminal histone tails and proteins that can recognize and bind specific histone modifications, the hypothesis of a 'histone code' evolved. The histone code proposes that multiple histone modifications work together sequentially or in combination to control downstream biological functions like gene regulation. Although this is found in some cases, deciphering a simple histone code, which allows the precise prediction of the functional consequence turns out to be much more complex and dynamic than initially thought (Kouzarides, 2007; Liu et al., 2005).

Thus histone modifications have two main functional consequences in that they can cause not only a direct structural change in the chromatin structure, but also indirectly act as binding surfaces for the recruitment of effector proteins (reviewed in Kouzarides 2007; Berger, 2007) that affect various downstream cellular processes, such as transcription, DNA repair and replication (Strahl and Allis, 2000; Turner, 2000). The proteins recruited bind to the histone modifications via specific domains. Modifications like acetylation are recognized by proteins containing bromodomains and methylation is recognized via proteins with chromodomains. The binding of these proteins to the modified histone tails is a mechanism by which proteins with enzymatic activities can be recruited into close proximity of the chromatin.

Gene expression in general is regulated by stimulus activated transcription factors that bind regulatory regions of genes and recruit other chromatin-modifying enzymes leading to the transcriptional activation or silencing of the target gene (Kouzarides, 2007). For the biological consequence of transcription, the post-translational histone modifications have initially been divided into two groups correlating with either

transcriptional active or repressed chromatin (Kouzarides, 2002; Rice and Allis, 2001; Zhang and Rheinberg, 2001) (Table 1).

Table 1 Histone modifications correlate with transcription.
Overview of selected modifications identified on histones. The histone modification type and the residue position on the histones are listed. The enzymes implicated in the particular histone modification are shown. Their function in transcription associated with each modification is listed (Adapted and modified from Li et al., 2007; Kouzarides, 2007).

Modification		Residues modified	Histone-modifying enzymes	Recognition modules	Functions in transcription
Methylation	H3	K4	MLL; SET1	Chromodomain	Activation
		K27	E(Z)	Ezh2	Repression
		K20	SET8	Tudor	Silencing
		K9	Suv39h	Chromodomain	Repression, activation
Acetylation	H3	K9	PCAF/GCN5	Bromodomain	Activation
	H4	K5, K8, K12	TIP60	Bromodomain	Activation
		K16	hMOF/TIP60	Bromodomain	Activation

Analysis using the chromatin immunoprecipition (ChIP) methodology combined with high resolution DNA-microarrays and sequencing techniques has provided major genome-wide information on the location and distribution of histone modifications. Initial genome-wide studies in yeast indicated that certain histone modifications like acetylation and methylation correlate with distinct genomic regions and specific transcriptional states (reviewed in Schones and Zhao, 2008).

In yeast, genome-wide determination of methylation patterns showed that methylation of histone H3 lysine residues at positions K4, K36 and K79 associate with transcriptional activity, whereas methylation of K9, K27 as well as histone H4 K20 correlate with transcriptional silent chromatin (Fischle et al., 2003; Lachner et al., 2003). In addition, the lysine residues on N-terminal tails of histones can be either mono-, di-, or tri-methylated which adds an additional layer of complexity. An example is the di- and tri-methylation of histone H3 that is associated with transcriptional permissive and transcriptional active euchromatin, respectively (Santos-Rosa et al., 2002). In addition, posttranslational modifications of histone tails are found in distinct localized patterns within the 5' up-stream region from the transcription start site, the core promoter region, the transcribed region and beyond the 3' downstream end of the transcript (reviewed in Li et al., 2007; Barrera and Ren, 2006). Genome-wide studies in yeast as well as mammals such as mouse and human, have shown that histone

acetylation and histone H3 lysine 4 methylation are associated with transcriptional active loci and are specifically found enhanced 5' of the transcription start site and extend well into the transcribed region of the gene (Roh et al., 2006; Liu et al., 2005; Pokholok et al., 2005; Bernstein et al., 2005; Kim et al., 2005).

Because it is anticipated that histone modifications mark regulatory regions such as enhancers and promoters and transcribed regions, histone marks have been utilized to predict regulatory regions (Heintzman et al., 2007; Barski et al., 2007; Birney et al., 2007). In addition, binding of members belonging to the transcriptional apparatus, such as RNA polymerase II, correlate with specific histone modifications and the transcriptional activity of the genes (reviewed in Barrera and Ren, 2006) (Fig. 1.5). It will be necessary to map specific histone modifications, which generally associate with transcriptional active or inactive chromatin across the different regions of a gene, to learn more about the mechanisms involved in regulating the transcription of genes. This is specifically important for genes regulated by complex signaling pathways upon differentiation.

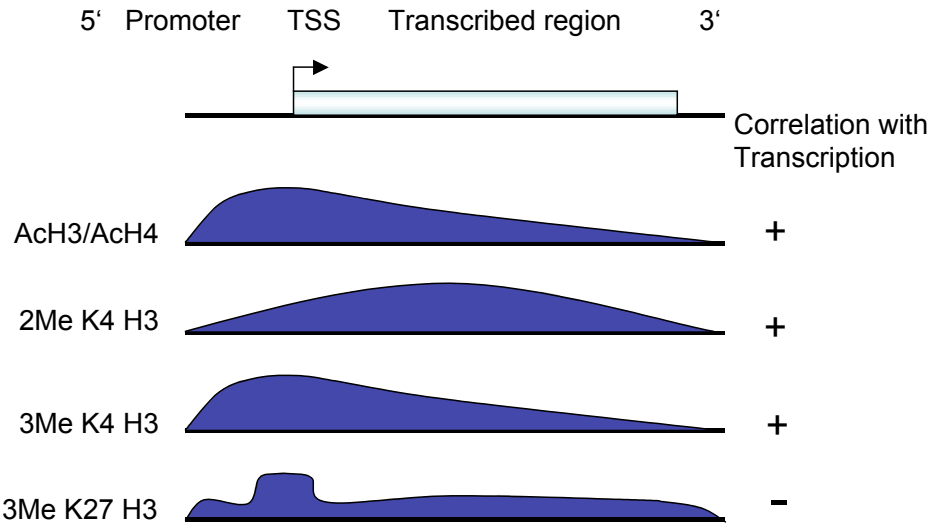


Figure 1.5 Distribution patterns of certain histone modifications.

The distribution patterns across a hypothetical gene that correlate with active or inactive transcription are shown. Distribution of histone modifications are marked in relation to an theoretical body of a gene, were the arrow symbolizes the transcription start site (TSS), the light colored rectangle the transcribed region across the body of the gene. Upstream of the TSS, the promoter and far 5' region are located. The 3' region is defined downstream of the end of the transcribed region. The curves below the gene correspond to the distributions of histone modifications (Ac= acetylation. Me= methylation, K= lysine, H3= histone H3 and H4= histone H4) that were collected mainly from genome-wide studies in yeast (Adapted and modified from Li et al., 2007).

1.5 Aim of project

The initiation of gene transcription requires precise interaction between the sequence-specific transcription factor and their target DNA regulatory sequence. Transcription correlates with dynamic changes in the chromatin structure and associated histone modifications, whereby specific post-translational histone modifications mark the regulatory and transcribed genomic regions.

Increased knowledge about the integration of signaling pathways at the level of binding patterns of sequence-specific transcription factors, components of the general transcription machinery (such as RNA polymerase II) and dynamic histone modification patterns will lead to a better understanding in transcription regulation.

The STAT5 transcription factor is required for the proliferation, differentiation and survival of the differentiated mammary epithelium during pregnancy and lactation. It has been proposed that this action of STAT5 in maintaining differentiation of mammary epithelial cells is achieved in part by activation of target gene transcription. Two classic known STAT5 target genes are the milk proteins ß-casein and WAP, which are differentiation specifically expressed by epithelial cells in mammary gland tissue during the developmental stages of pregnancy and lactation. In this thesis the *in vivo* recruitment of transcription factors STAT5 and the RNA polymerase II, as well as changes in histone modifications on different regions of the milk protein genes ß-casein and WAP are to be characterized with help of the ChIP method during different stages of mammary gland development. Discovery of novel DNA regulatory elements bound directly by STAT5A is essential to identifying additional STAT5 target genes involved in the maintenance and survival of differentiated mammary epithelial cells in mammary gland tissue during lactation. The second aim of this work is to identify novel genomic binding sites of the transcription factor STAT5A in mouse mammary gland tissue from lactating mice. The ChIP cloning method will be utilized to create a library containing the chromatin fragments enriched with an antibody specific against STAT5A. The cloned DNA fragments will then be sequenced to allow for the identification of STAT5-binding motifs and determination of their location in the genome using bioinformatics tools. Select novel DNA STAT5A-binding regions will be verified for STAT5 binding in a differentiation specific manner and are then subjected to subsequent experimental and computational analysis to further characterize the novel STAT5-binding motifs. Reporter gene assays and annotated transcript RNA levels are to be analyzed for determining the functional contribution of the novel STAT5-binding elements in the regulation of putative target genes.

2. Materials and Methods

2.1 Materials

2.1.1 General chemicals and Materials

General chemicals were obtained in molecular grade quality, if not noted otherwise through the companies Applichem (Darmstadt), Amersham Biosciences (Freiburg), Fluka (Buchs, Schweiz), Peqlab (Erlangen), Merck (Darmstadt), New England Biolabs (Frankfurt am Main), Roche (Mannheim), Roth (Karlsruhe) and Sigma (Taufkirchen). For inhibition of phosphatases and proteases, the Phosphatase Inhibitor Cocktail I and II from Sigma (Taufkirchen) and the Protease Inhibitor Mix (Complete, EDTA free) from Roche (Mannheim) were used.
Consumables and plastic ware used for laboratory and tissue culture were obtained from one of the following companies Biozyme (Oldendorf), BD Bioscience (Heidelberg), Costar (Bodenheim), Greiner Bio-one (Frickenhausen) and Millipore (Eschborn).

2.1.2 Solutions and Buffers

All solutions were made up in deionized water. The solutions were either autoclaved (20 minutes (min) at 121°C and 2 bar) or filter sterilized. All other solutions and buffers used are mentioned in the corresponding method sections.

Solution	**Components**
10x TBE (electrophoresis buffer)	890 mM Tris base, 890 mM boric acid, 20 mM EDTA, pH 8.0
TE, pH 8.0	10 mM Tris•HCL pH 8.0, 1 mM EDTA
DEPC-water	0.2 % (v/v) DEPC

2.1.3 Enzymes

All restriction endonucleases and their corresponding buffers were obtained from New England Biolabs (NEB) (Frankfurt am Main). Other DNA modifying enzymes used are listed below.

Enzyme	Company
T4 DNA Ligase	NEB, Frankfurt am Main
Taq DNA polymerase	Invitrogen, Karlsruhe
Alkaline Phosphatase, Calf Intestinal (CIP)	NEB, Frankfurt am Main

2.1.4 Kits and molecular weight markers

	Company
1kb plus DNA ladder	Invitrogen, Eggenstein
QIAGEN® Plasmid Maxi Kit	Qiagen, Hilden
QIAquick® PCR Purification Kit	Qiagen, Hilden
QIAquick® Gel Extraction Kit	Qiagen, Hilden
RNeasy® Mini Kit	Qiagen, Hilden
Omniscript™ Reverse Transcriptase	Qiagen, Hilden
TA Cloning® Kit	Invitrogen, Eggenstein
TripleMaster® PCR mix	Eppendorf AG, Hamburg

2.1.5 Antibodies

The listed antibodies were used in Chromatin Immunoprecipitation (ChIP).
Depending on the Antibody two different TSE 2 wash Buffers were used (see ChIP protocol). Immunoprecipitations washed with TSE 2 wash Buffer (TSE 2-II, 250 mM NaCl) that contain no SDS detergent and a lower NaCl concentration are marked with a star symbol. Immunoprecipitations were performed with protein A agarose (sc-2001) when using rabbit IgG antibodies and protein G (sc-2002) or A/G PLUS agarose (sc-2003) were used for capturing mouse IgG antibodies.

Name/ specificity	Host species and clonal type	µg or µl/ChIP	Ref. Number	Company
IgG (preimmune serum)	normal rabbit IgG	4 µg	sc-2027	Santa Cruz, Heidelberg
STAT5A (a.a. 775-794)	rabbit antiserum	2-4 µg	STAT5A	R & D Systems, Wiesbaden
STAT5B (a.a. 769-782)	rabbit antiserum	2-4 µg	STAT5B	R & D Systems, Wiesbaden
RNA POL II P-CTD (phospho-CTD of	mouse monoclonal IgG1	0.5-2 µg	Upstate 05-623	BIOMOL, Hamburg

RNA polymerase II)				
2M(K4)H3 (histone H3 dimethylated at lysine 4)	rabbit antiserum	5 µl	Upstate 07-030	BIOMOL, Hamburg
3M(K4)H3 (histone H3 trimethylated at lysine 4)	rabbit polyclonal IgG	2.5 µg	ab8580	Abcam, Cambridge (UK)
3M(K27)H3 (histone H3 trimethylated at lysine 27)	mouse ascites	5 µl	Upstate 05-851	BIOMOL, Hamburg
Ac-H3 (acetylated histone H3)	rabbit polyclonal IgG	2-4 µg	Upstate 06-599	BIOMOL, Hamburg
Ac-H4* (acetylated histone H4)	rabbit polyclonal IgG	5-10 µl	Upstate 06-866	BIOMOL, Hamburg
Ac(K8)H4* (histone H4 acetylated at lysine 8)	rabbit polyclonal IgG	2 µg	ab1760	Abcam, Cambridge (UK)
Ac(K12)H4* (histone H4 acetylated at lysine 12)	rabbit polyclonal antiserum	10 µl	ab1761	Abcam, Cambridge (UK)
Ac(K16)H4* (histone H4 acetylated at lysine 16)	rabbit antiserum	10 µl	Upstate 07-329	BIOMOL, Hamburg
H3 (N-terminal H3)	rabbit polyclonal IgG	5 µg	Upstate 06-755	BIOMOL, Hamburg

2.1.6 Oligonucleotides

Oligonucleotides were ordered from MWG Biotech (München) or Invitrogen (Karlsruhe). All PCR Primer stocks were resuspended at 100 pmol/µl and used at a working dilution of 10 pmol/µl. Unless stated otherwise, all primers were designed based on mouse *Mus musculus* sequences

All Primers were run in quantitative (q) PCR with an annealing temperature of 58°C, unless stated otherwise.

2.1.6.1 Oligonucleotides for reverse transcript quantitative RT-PCR

Primer Name	RNA	Sequence (5'→3')	Reference /GenBank Acc.
18S For	18S RNA	CGGCTACCACATCCAAGGA	(Baus and Pfitzner, 2006)
18S Rev		CCAATTACAGGGCCTCGAAA	
YWHAZ_F	YWHAZ	GAAAAGTTCTTGATCCCCAATGC	PrimerBank ID:6756041a1/ NM_011740
YWHAZ_R		TGTGACTGGTCCACAATTCCTT	(Wang and Seed, 2003)
CyclophilinBF	cyclophilin B	TGCAGGCAAAGACACCAATG	PrimerBank ID:6755142a3/ NM_011149
CyclophilinBR		GTGCTCTCCACCTTCCGTA	(Wang and Seed, 2003)

ß-caseinF	ß-casein	TCACTCCAGCATCCAGTCACA	Feraci E., PhD thesis
ß-caseinR		GGCCCAAGAGATGGCACCA	
WAP_zhu_F	WAP	GGAGCATTCTATCTTCATTGG	(Zhu et al., 2004)
WAP_zhu_R		TCAGTTCAGTCCATGTTCCA	
SOCS3F_Doppler	SOCS3	GCTCCAAAAGCGAGTACCAGC	(Tonko-Geymayer et al., 2002)
SOCS3R_Doppler		AGTAGAATCCGCTCTCCTGCAG	
NFIB_F	NFI/B	AGCACCATCATCCCGGAATAC	PrimerBank ID: 31982161a1/ NM_008687
NFIB_R		GTACCAGGACTGGCTCGTTTG	(Wang and Seed, 2003)
AW106080est_140	AW106080/ EST Exon1/2	TGGATGAATTCCACCCATTT	This thesis / AW106080
AW106080est_290		TCGTCTTTGACTGCCCTTTC	
Bu610001_15	BU610001/ EST Exon1/2	GTCTTCTGTGGAACCCAAGC	This thesis / BU610001
Bu610001_211		GAGTTCAGTGTATCCGAGTACTGCT	
Zswim6_F	Zswim6/ AK030163	AGACACCAACAACCACTACACC	PrimerBank ID: 26326145a1
Zswim6_R	Exon 14 and 3'	GGTTCTGGCACACTGTATGTTC	(Wang and Seed, 2003)
AK122528_1525	Zswim6/ Exon 6/7	TCGGACAGTGTTCACGAGAG	This thesis / XM_358311
AK122528_1699		GGCTGTAGGGACGTGTTCAT	
XM358311_632	Zswim6/ Exon1/2	CTGGAGAGTGGCTGCGTAGA	This thesis / XM_358311
XM358311_821		CCACAACATGGGCACAATAA	

2.1.6.2 Oligonucleotides used for ChIP qRT-PCR

A) Primers used in ChIP

The transcribed gene region refers to the region within the genomic sequence of the gene that is transcribed (mRNA). The ß-casein enhancer primers require an annealing temperature of 55°C.

Name	Gene region	Sequence (5'→3')	GenBank Acc./ GI Nr.	5` Primer Nucleotide Position	Source
ß-casein RosenF	Enhancer	agtctcaaggaaatactggatctattg			(Kabotyanski and Rosen 2006)
ß-casein RosenR	Enhancer	gagtttgtgaaccatcttttactaacc			
ß-casein 1312F	5'	tccttccatgggtgttttgt	NT_039308/ 51711197	11667321	This thesis
ß-casein 1561R	5'	gcaaggtaatgggaacagga		11667072	
ß-casein 199F	Promoter	gcttctgaattgctgccttg		11663963	
ß-casein 308R	Promoter	caaaaagtcccttcaattcca		11663854	
ß-casein 6601F	Transcribed	aggtgagcaggtgacgtctt		11662038	
ß-casein 7200R	Transcribed	tgtttgatgagtggcaaaagtt		11661882	

ß-casein 510F	3'	tgctgcctaatgagaacgtg	NT_039308/ 51711197	11656781	This thesis
ß-casein 728R	3'	accaagatttccaggagcaa		11656563	
ß-casein 1403F	3'	tcacccctttcttcgtcatc		11655888	
ß-casein 1606R	3'	catgtgccctcctcttctgt		11655685	
SOCS3 2778F	Promoter	cacagccttcagtgcagag	AF117732.1/ 4336990	2778-2797	
SOCS3 2946R	Promoter	agagacagcggtcgtaggag		2946-2927	
WAP 1059F	very 5'	caagcctcacttgtcttactttga	NT_039515/ 63615460 or 51766363	3537382	
WAP 1247R	very 5'	gagtacgtgtgtttgtgtgtgtgt		3537194	
WAP 1991F	5'	tagagggatgagtttccacgat		3536450	
WAP 2139R	5'	tcctcgaggtttttgtttttgt		3536302	
WAP 1772F	Promoter	agtgccagcaggacatctct		3534063	
WAP 1891R	Promoter	tgtcgggcatacattgaaaa		3533944	
WAP 1270F	Transcribed	tctgagaagccttagaagtgaggt		3532178	
WAP 1478R	Transcribed	gggacttactgttgacaggagttt		3531970	
WAP 210F	3'	agctcctgaaggaacgacaa		3530241	
WAP 419R	3'	cagacgatgcccatttcttt		3530032	
WAP 1236F	3'	acaagcagcataacgcacag		3529215	
WAP 1401R	3'	aagggaggggtcatgagtct		3529049	

B) Primers used in ChIP for STAT5A library clones

NAME	Sequence (5'→3')	GenBank Accesion/ Build	Primer Nucleotide Position	STAT5 site Position	Source
c153_526	CTTGTGGCATGTTATGATTGCT	NT_039260.6/ Mm4_39300_36	25716165-25716186	25716096-25716104	This thesis
c153_700	CAAATCCAGATTGCCCTTTTAC		25716340-25716319		
C194_339	TTGCTCAGTTCTTGGAAACAGA	NT_162143.2/ Mm3_159849_36	19409623-19409644	19409851-19409860	
C194_533	GAGAAGACCCATGCTTGGTAAC		19409817-19409795		
c236_236F	ACTCAATGGGTTTTTGTTTTGTTT	NT_165760.1/ Mm5_163274_36	7795792-7795815	7795735-7795727	
c236_425R	AGAAAAATTGCCATCTATTTGAGC		7795981-7795958	7795792-7795815	
c831swi281	GGTCTTGGCTTCTGCAGTTAGT	NT_039590.6/ Mm13_39630_36	15821474-15821453	15821352-15821344	
c831swi503	CGAAATGTTGCAATGAATCTGT		15821252-15821274		
c20_1527	ACTGTGTGTGGCAGCCCTCT	NT_165773.2/ Mm11_163287_37	9467848-9467867	9467876-9467980	Xin Fu,
c20_1646	GAACTAAGGGCTGGTTTCCAAAA		9467967-9467945		Pfitzner
c34_2108	CCGGCAGGCCTCCTATCTAT	NT_039606.7/ Mm14_39646_37	599689-599708	599435-599866	group
c34_2233	ATCTGCCTGCCCAACTTCAA		599814-599795		

c73_707	GCAGGAGGCTAATCCAGCAA	NT_078575.6/ Mm8_78640_37	9471062-9471081	9470473-9471984	Xin Fu,
c73_824	CCTGGGAGCTTGGCTTTTCT		9471179-9471160		Pfitzner
c398_3_604	TGAACGCCGATCATCAAAAT	NT_039674.7/ Mm18_39714_37	39434516-39434535	39434662-39434846	group
c398_3_744	CTCCGATGTCCCGAAATGAC		39434656-39434637		
c882_2028	GTCAGAGCCTTCCACGCTGT	NT_039240.7/ Mm3_39280_37	50594949-50594968	50594847-50595127	
c882_2162	CACACTCCTCTCTGCCAGAAAAA		50595083-50595061		
c721_155	GAAGCAGACTTAACTTGAAAAACAAGG	NT_039433.7/ Mm7_39473_37	15228875-15228901	15228869- 15229238	
c721_295	TGACTTTGGCAAGCAAGCAA		15229015-15228996		
c878_1220	AATCAAACTTGGGGCATCAG	NT_165773.1/ Mm11_163287_36	27179216-27179234	27179208-27179216	This thesis
c878_1389	ACACACACGCCTCCTAATCC		27179389-27179369	27179347-27179356	

2.1.6.3 Oligonucleotides for sequencing and chromatin amplification

Primer Name	Description of use	Sequence (5'→3')
RV3_pGL3	primer for sequencing of inserts	CTAGCAAAATAGGCTGTCCC
GL2_pGL3	in MCS of pGL3 basic vector (Promega)	CTTTATGTTTTTGGCGTCTTCCA
JW102	unidirectional linker oligo's used to amplify ChIP chromatin	GCGGTGACCCGGGAGATCTGAATTC
JW103	(Oberley and Farnham Lab Protocol)	GAATTCAGATC

2.1.6.4 Oligonucleotides for cloning

A) Primers to amplify the fragment regions containing the STAT motifs corresponding to STAT5A library clone out of mouse genomic DNA.

Primer Name	Description of use	Sequence (5'→3')	GenBank Acc. / Primer nt position
c153gen_1284	5'-primer to clone c153 (750 bp)	CACACACAATCAGACTTTCATCC	NT_039260.6 / 25716487-25716465
c153gen_2033	3'-primer to clone c153	TAAAGAAAAGAATGGGCTTGCT	25715738-25715759
c194 680F_gDNA	5'-primer to clone c194 (800 bp)	AGGTCTGAAAACTGTTGGTTGAG	NT_162143.2/ 19409504-19409526
c194 1479R_gDNA	3'-primer to clone c194	AAGCTCTCCCAATAAACGTGTG	19410303-19410282
c831 694F_gDNA	5'-primer to clone c831 (748 bp)	GCTTTGAAGACACTGCTTTGAAT	NT_039590.6/ 15821681-15821659
c831 1441R_gDNA	3'-primer to clone c831	AAGGTAAGGGCTACACCCTATCA	15820934-15820956

B) Nested primers with restriction enzyme overhangs to sub-clone the regions amplified from genomic DNA into pGL3 basic vector multiple cloning sites (MCS).

Primer Name	Description of use	Sequence (5'→3')	GenBank Acc. / Primer nt position
c153_1284F_gBamH1	5'-primer with BamHI (cgGGATCC) to subclone c153	CGGGATCCCACAATCAGACTTTCATCC	NT_039260.6/ 25716483 -25716465
c153_2033R_gBamH1	3'-primer with BamHI (cgGGATCC) to subclone c153	CGGGATCCAGAAAAGAATGGGCTTGCT	25715741-25715759
c194_680F_gBglII	5'-primer with Bgl II (gaAGATCT) to subclone c194	GAAGATCTCTGAAAACTGTTGGTTGAG	NT_162143.2/ 19409508-19409526
c194_1479R_gBglII	3'-primer with Bgl II (gaAGATCT) to subclone c194	GAAGATCTCTCTCCCAATAAACGTGTG	19410300-19410282
c831_694F_gBamH1	5'-primer with BamHI (cgGGATCC) to subclone c831	CGGGATCCTGAAGACACTGCTTTGAAT	NT_039590.6/ 15821677 -15821659
c831_1441R_gBamH1	3'-primer with BamHI (cgGGATCC) to subclone c831	CGGGATCCTAAGGGCTACACCCTATCA	15820938-15820956
c236_kpn1Fg	5' primer with KpnI (cggGGTACC) to amplify c236 in pCR2.1	CGGGGTACCGGCTTTCAGAGTTGC	NT_165760.1/ 7795845 -7795824
c236_kpn1Rg	3' primer with KpnI (cggGGTACC) to amplify c236 in pCR2.1	CGGGGTACCAACAGTAACAAGGGC	7795553 -7795572
c153_622R_kpn1	5' primer with Kpn1 (cggGGTACC) to subclone 339 bp of c153	CGGGGTACCGGTCAGGTTATTCACA	NT_039260.6/ 25716261 -25716246
c153_283F_kpn1	3' primer with Kpn1 (cggGGTACC) to subclone 339 bp c153	CGGGGTACCTGACCAAACAATGAG	25715922 -25715936

2.1.6.5 Oligonucleotides for site-directed mutagenesis

Oligonucleotides for site-directed mutagenesis of STAT5 binding sites from clones located in pGL3 luciferase reporter vectors. The manufactures protocol for primer design was followed as outlined in the Quikchange Site-Directed Mutagenesis kit (Stratagene). The STAT motifs are underlined and the bases that were changed from wild type (wt) are highlighted.

Primer Name	Description of use	Sequence (5'→3')
c236_M1TT_F	sense primer for mutagenesis of wt STAT5 binding site I from AA to TT in mouse clone c236	ACTCTGCACGCC**TT**GGGGCTTTGTT
c236_M1TT_R	anti-sense primer for mutagenesis of wt STAT5 binding site I in mouse clone c236	AACAAAGCCCC**AA**GGCGTGCAGAGT
c236_M2TT_F	sense primer for mutagenesis of wt STAT5 binding site II from AA to TT in mouse clone c236	GGAGCTCTCTTTTCCAAG**TT**TAGACTACAC
c236_M2TT_R	anti-sense primer for mutagenesis of wt STAT5 binding site II in mouse clone c236	GTGTAGTCTA**AA**CTTGGAAAAGAGAGCTCC
c153_S5_mut1F	sense primer for mutagenesis of wt STAT5 binding site I from TT to AA in mouse clone c153	GTTATTCTACAAAACGGCTT**AA**CATAGAAACCTAATTC
c153_S5_mut1R	anti-sense primer for mutagenesis of wt STAT5 binding site I in mouse clone c153	GAATTAGGTTTCTATG**TT**AAGCCGTTTTGTAGAATAAC
c153_S6_mut1F	sense primer for mutagenesis of wt STAT binding site II from TT to AA in mouse clone c153	CTTCCTTTATCC**AA**CCCTAGAAGGCTTCAGGAGTTTCCCGG
c153_S6_mut1R	anti-sense primer for mutagenesis of wt STAT binding site II	CCGGGAAACTCCTGAAGCCTTCTAGGG**TT**GGATAAAGGAAG

	in mouse clone c153	
c153_M1gc_M3gcF	sense primer for mutagenesis of M1 STAT5 binding site I from AA to GC and wt STAT binding site III from AA to GC in mouse clone c153	GTTATTCTACA**GC**ACGGCTT**GC**CATAGAAACCTAATTC
c153_M1gc_M3gcR	anti-sense primer for mutagenesis of M1 STAT5 binding site I and wt STAT binding site III in mouse clone c153	GAATTAGGTTTCTATG**GC**AAGCCGT**GC**TGTAGAATAAC
c153_M1gc_M4cgF	sense primer for mutagenesis of M1 STAT5 binding site I from AA to GC and wt STAT binding site IV from TT to CG in mouse clone c153	CGGCTT**GC**CATAGAAACCTAATTCAAT**CG**TAAGAAAATG
c153_M1gc_M4cgR	anti-sense primer for mutagenesis of M1 STAT5 binding site I and wt STAT binding site IV in mouse clone c153	CATTTTCTTA**CG**ATTGAATTAGGTTTCTATG**GC**AAGCCG
c831_S5_mut1F	sense primer for mutagenesis of wt STAT5 binding site from TT to AA in mouse clone c831	GGTTTACTCGGGAGG**AA**CATGGAACTGGCGTGGC
c831_S5_mut1R	anti-sense primer for mutagenesis of wt STAT5 binding site in mouse clone c831	GCCACGCCAGTTCCATG**TT**CCTCCCGAGTAAACC

2.1.7 Plasmids

2.1.7.1 Expression vectors for mammalian cells (General expression vectors and reporter plasmids)

A) General expression vectors

Name/ vector type	Coding Insert	Reference/ Source
pXM	pXM backbone expression vector containing adenoviral promoter	Moriggl et al., 1996
pXM-STAT5A	mouse STAT5A	Moriggl et al., 1996
pMX IRES GFP	pMX backbone IRES-GFP based bicistronic retroviral vector	Onishi et al., 1996
pMX STAT5A IRES GFP	mouse STAT5A	Onishi et al., 1996
pMX STAT5A 1*6 IRES GFP	constitutive active mouse STAT5A	Onishi et al., 1996
pMX dnSTAT5A IRES GFP	dominant negative mouse STAT5A Y705F	Döll, F., PhD thesis
pMX STAT5A Δ750 IRES GFP	dominant negative mouse STAT5A with deletion of transcription activation domain	Döll, F., PhD thesis

B) Luciferase reporter plasmids

Name	Clone DNA Insert	GenBank Acc.	Nucleotide position
pGL3 basic	Luciferase reporter backbone vector is enhancer-less and promoter -less	Promega	

pGL3 ba c153 L_10	c153 (743bp) wild type	NT_039260.6	25716483-25715741
pGL3 ba c153(336)_WT_6F	c153 (339bp) forward orientation wild type		25715922-25716261
pGL3 ba c153(336)_WT_3R	c153 (339bp) reverse orientation wild type		25716261-25715922
pGL3 ba c153(336)_M1_35R	c153 (339bp) reverse orientation with mutation in STAT site I		25716261-25715922
pGL3 ba c153(336)_M2_48R	c153 (339bp) reverse orientation with mutation in STAT site II		25716261-25715922
pGL3 ba c153(336)_M1+M2_55R	c153 (339bp) reverse orientation with mutation in STAT site I and II		25716261-25715922
pGL3 ba c153(336)_M1-3_11R	c153 (339bp) reverse orientation with mutation in STAT site I, II and III		25716261-25715922
pGL3 ba c153(336)_M1-2+4_16R	c153 (339bp) reverse orientation with mutation in STAT site I, II and IV		25716261-25715922
pGL3 ba c153(336)_M1-4_20R	c153 (339bp) reverse orientation with mutation in STAT site I, II, III and IV		25716261-25715922
pGL3 ba c194_11F	c194 (793bp) forward orientation wild type	NT_162143.2	19410300-19409508
pGL3 ba c194_29R	c194 (793bp) reverse orientation wild type		19409508-19410300
pGL3 ba c236(293)_5F	c236 (293bp) forward orientation wild type	NT_165760.1	7795553-7795845
pGL3 ba c236(293)_6R	c236 (293bp) reverse orientation wild type		7795845-7795553
pGL3 ba c236(293)_M1_1R	c236 (293bp) reverse orientation with mutation in STAT site I		7795845-7795553
pGL3 ba c236(293)_M2_7R	c236 (293bp) reverse orientation with mutation in STAT site II		7795845-7795553
pGL3 ba c236(293)_M1+M2_22R	c236 (293bp) reverse orientation with mutation in STAT site I and II		7795845-7795553
pGL3 ba c831 L_6	c831 (740bp) wild type	NT_039590.6	15820938-15821677
pGL3 ba c831 L_6_M1_21	c831 (740bp) with mutation in the STAT5 site		15820938-15821677

2.1.8 Bacteria and bacterial culture media

To prepare plasmid (p) DNA for subsequent molecular DNA applications, the *Escherichia coli* host strain DH5α was used. The bacterial cultures were grown in standard Luria Bertani (LB) medium (10 g/l Tryptone, 10 g/l NaCl, 5 g/l yeast extract). The LB culture broth was made up in deionized water and autoclaved. Ampicillin was used as a selection agent, since all pDNA used in this work carried the ampicillin resistance gene. Ampicillin was added at a final working concentration of 100 µg/ml.

***E.coli* host strain**	**Genotype**
DH5α	*end*A1 *hsd*R17 (rk^- mk^-) *sup*E44 *thi* $^-\lambda^-$ *rec*A1 *gyr*A96 *rel*A1 80dlac^{q}ZD M15

2.1.9 Tissue culture reagents and cell lines

2.1.9.1 Tissue culture reagents

	Company
DMEM (Dulbeccos minimal Essential Medium), 4,5 g/l D-Glucose	BioWhittacker, Verviers (Belgium)
RPMI 1640 (Roosevelt Park memorial Institute Medium)), 4,5 g/l D-Glucose	BioWhittacker, Verviers (Belgium)
L-Glutamine	PAA, Pasching (Austria)
Penicillin/Streptomycin	PAA, Pasching (Austria)
Fetal Calf Serum (FCS)	Invitrogen, Eggenstein
Phosphate Buffered Saline (PBS) (1x)	PAA, Pasching (Austria)
Trypsin-EDTA (1x)	Invitrogen, Eggenstein
Isotone II	Becton Dickinson, Heidelberg
Prolactin from sheep pituitary (1,000 I.U.)	Sigma-Aldrich, Taufkirchen
Insulin	Sigma-Aldrich, Taufkirchen
Dexamethasone	Sigma-Aldrich, Taufkirchen
Dimethylsulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen

The tissue culture media is supplemented with 10% (v/v) FCS, 2 mM L-Glutamine, 50 U/ml penicillin and 50 µg/ml Streptomycin unless stated otherwise. For complement inactivation the FCS is heat inactivated for 30 minutes (min) at 56°C.

2.1.9.2 Cell lines

Cell line	**Cell type**	**ATCC number/ source**	**Culture media**
HeLa TA	human cervical adenocarcinoma cell line, with Tet-transactivator	CCL-2	DMEM
HC11	mouse mammary epithelial cell line	Ball et al., 1988	RPMI 1640 Growth medium: 10 ng/ml EGF, 5 µg/ml Insulin Starve Medium: 2% (v/v) FCS, 5 µg/ml Insulin

2.1.10 Online Bioinformatic resources and tools

PCR primer design:

- Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)

by Rozen and skaletsky (2000)

Normalization of real-time PCR expression data:

- geNorm (http://medgen.ugent.be/~jvdesomp/genorm/)

Database DNA similarity search software:

- BLAST (http://www.ncbi.nlm.nih.gov/BLAST/)

- BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat)

DNA bioinformatic Databases:

- NCBI (http://www.ncbi.nlm.nih.gov/)
- UCSC Genome Browser (http://genome.ucsc.edu/)
- Ensembl Genome Browser (http://www.ensembl.org/index.html)

Transcription factor Motif search software:

- TESS:Transcription Element Search System (http://www.cbil.upenn.edu/cgi-bin/tess/tess)
- Genomatrix/ MatInspector (http://www.genomatix.de/products/MatInspector/)

Other software:

- SOURCE (http://source.stanford.edu/cgi-bin/source/sourceSearch)

GeneReports linked to UCSC Browser

- Dragon Genome Explorer Promoter Search and Gene Regions (http://research.i2r.a-star.edu.sg/promoter/); Dragon TSS Desert Masker ver. 2.0 and Dragon Promoter Finder ver.1.5

Online ChIP Assay protocols:

- http://genomecenter.ucdavis.edu/farnham/farnham/protocol.html

Farnham lab protocols for Chromatin Immunoprecipitation (ChIP) protocol for Tissues and Cloning and ChIP-CpG island microarray binding analysis.

UCSC Genome Bioinformatics Browser sequence alignments:

The genomic assemblies used in the alignment of each species by the UCSC Genome Bioinformatics Browser is as follows: mouse, (Feb 2006, mm8); rat (Nov 2004,rn4); rabbit (May 2005, oryCun1); human (Feb 2006, hg18); chimp (Jan 2006, panTro2); dog (May2005, canFam2); cow (Mar 2005, bosTau); armadillo (May 2005, dasNov1); elephant (May2005, loxAfr1); tenrec (Jul 2005, echTel1); opossum (Jan 2006,monDom4); chicken (Feb 2004, galGal2).

2.2 Methods

2.2.1 General molecular methods for cloning and working with nucleic acids

2.2.1.1 Production of competent cells

The preparation of competent bacteria (*E. coli* strain DH5α) was performed after the protocol adapted from Hanahan (Hanahan, 1983). Competence of *E. coli* can be induced by treatment with calcium chloride before DNA addition. The calcium destabilizes the cell membrane and allows for a calcium phosphate-DNA complex to adhere to the cell surface. The DNA is then readily taken up when the cells are exposed to a brief heat shock of 42°C. Briefly, a bacterial culture was grown while shaking at 37°C in about 150 ml LB broth until an OD_{600} of 0.3-0.4 has been reached. Cells are then centrifuged (2,000 x g, 10 min, 4°C) and the cell pellet is gently resuspended in 50 ml ice-cold solution I (100 mM $RbCl_2$, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8). The cells are then incubated on ice for 2 hours followed by centrifugation (2,000 x g, 10 min, 4°C). The cell pellet is gently resuspended in 4 ml solution II (10 mM MOPS, 10 mM $RbCl_2$, 75 mM $CaCl_2$, 15% (v/v) glycerol, pH 7.0) and incubated on ice for additional 15 minutes. 100 µl of competent cells are transferred into sterile microcentrifuge tubes and frozen in liquid nitrogen. Competent cells are stored at
-80°C.

2.2.1.2 Transformation of bacteria

Competent bacteria are thawed on ice. Plasmid DNA (50 ng to 1 µg) is added to 50-100 µl competent cells, mixed gently and incubated on ice for at least 20 min.
Cells are subjected to a brief heat shock for 90 seconds (sec) at 42°C to facilitate DNA uptake and then incubated on ice for 2 minutes. 400 µl of LB media is added and the cells incubated at 37°C on a shaker for 1 hour. The bacteria are then spread on appropriate antibiotic-containing plates to select for bacteria carrying the desired plasmid DNA with the resistance marker. The plates are incubated over night at 37°C to allow colonies to grow for further processing.

2.2.1.3 Isolation of plasmid DNA (A Mini preparation procedure)

The mini preparation procedure of plasmid DNA is based on the principle of alkaline-detergent lysis method (Birnboim, 1983) and followed by a butanol precipitation of the DNA.

Briefly, 1.4 ml of a bacterial culture grown over night is transferred into a microcentrifuge tube and pelleted by centrifugation (6,000 x g, for 1 min). The bacterial pellet is first resuspended in P1 buffer (50 mM Tris•Cl pH 8.0, 10 mM EDTA, 100 µg/ml RNase A) and then in 100 µl of P2 buffer (200 mM NaOH, 1 % (w/v) SDS) to lyse the cells followed by vigorous inversion of the tubes. After a 5 minutes incubation time 75 µl of P3 buffer (3 M potassium acetate, pH 5.5) is added and the tubes are inverted again several times. This is followed by another 5 minutes incubation time upon which the tubes are centrifuged (20,000 x g, 10 min). The cleared bacterial lysate is then transferred to new tubes and 1.5 volumes 6 M Guanidinisothiocyanate (GITC) and 2 volumes of butanol are added. The tubes are inverted several times and then centrifuged (20,000 x g, 10 min) to pellet the precipitated DNA. The DNA pellet is washed with 70% ethanol, air dried and resuspended in 50 µl water. In general the yield of plasmid DNA was 1-2 µg/ml of bacterial culture. The buffers P1, P2 and P3 used in this protocol were part of the QIAGEN® Plasmid kits (Qiagen, Hilden).

2.2.1.4 Isolation of plasmid DNA (A Maxi preparation procedure)

For the isolation of pure, transfection grade plasmid DNA the QIAGEN® Plasmid Maxi kit (Qiagen, Hilden) was used. The growth of bacterial cultures and purification of plasmid DNA was performed according to the manufacturers specifications. The purified plasmid DNA was resuspended in water or TE buffer (pH 8.0) and stored at –20°C. The DNA concentration was determined by UV spectrophotometry.

2.2.1.5 Measurement of DNA concentration

The DNA concentration can be determined by measuring the absorbance at 260 nm (A_{260}) in a conventional spectrophotometer using a quartz cuvette. One unit of absorbance at 260 nm wavelength corresponds to 50 µg/ml double stranded (ds) DNA. The total amount of DNA is calculated as follows:

Measured (A_{260}) x dilution factor of sample x 50 = µg ds DNA / ml

When only small sample amounts of DNA are available the DNA concentration can be measured directly using only 1 µl of sample DNA with the NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington USA).

2.2.1.6 Restriction endonuclease digestion of DNA

Restriction endonucleases recognize, bind and cleave DNA at specific target sequences. As a general rule one unit of restriction endonuclease digests 1 µg of substrate DNA in 1 hour.
Plasmid DNA (0.5-5 µg) was digested with suitable restriction enzymes using the digestion conditions (appropriate choice of buffers, enzyme unit concentration, optimal restriction reaction temperatures and length of digestion time) suggested by the manufacturers.

2.2.1.8 5'-dephosphorylation of DNA

To prevent the self-ligation of the linearized vector DNA during ligation the 5' phosphate from the vector DNA was removed with calf intestinal phosphatase (CIP) (NEB, Frankfurt/Main). For every 1 µg of linearized vector DNA 0.5 Units of CIP was added and incubated for 1 hour at 37°C.

2.2.1.7 Ligation of DNA

T4 DNA ligase (NEB, Frankfurt/M) was used for cloning of restriction fragments and joining linkers to blunt ended DNA according to manufacturers suggestions. The T4 DNA ligase catalyzes the formation of a phosphodiester bond between the 5' phosphate and the 3' hydroxyl end of two juxtaposed double stranded DNA molecules. Briefly, for cloning purposes approximately 25-50 ng of linearized and 5'-dephosphorylated plasmid DNA vector was used at a vector:insert ratio of 1:1 (as determined on analytical agarose gels). Ligation reactions were performed with at least 4.0 Weiss units of T4 DNA ligase in a final reaction volume of 10 µl and incubated overnight in a water bath set at 16°C.

2.2.1.9 Fill in of 5'-overhangs to form blunt ends in ds DNA

The DNA polymerase I, large (Klenow) fragment from *E.coli* was used to create blunt ends on double stranded DNA and thus allow for effective ligation to other ds DNA with blunt ends. For example shearing of chromatin results in overhangs which need to be blunt ended before ligation to linker or vector DNA.
This DNA polymerase I contains a 5'→3' polymerase activity for filling in 5' overhangs as well as a 3'→5' exonuclease activity that removes 3' overhangs to form blunt ends. The reaction for the DNA polymerase I, large (Klenow) fragment (NEB, Frankfurt /Main) was set up according to manufacturers instructions.

2.2.1.10 Annealing of single stranded DNA oligonucleotides

To obtain the unidirectional double-stranded oligonucleotide linker for generating ChIP chromatin amplicons the two oligos JW102 and JW103 were annealed by mixing in a 1:1 ratio 6.7 µl of each oligonucleotide (10 µM) in 86.6 µl water as described (Oberley M. and P. Farnham Lab Protocol; Section II, Generating Chromatin Amplicons; Day 1). The oligos were boiled for 5 min in a water bath and allow to slowly cool down to room temperature. The unidirectional double-stranded oligonucleotides are then stored at –20°C.

2.2.1.11 Agarose gel electrophoresis of DNA

Based on the size of the DNA fragments to be separated agarose gels of 1 to 2% (w/v) were made up with 1x TBE buffer. The same 1x TBE electrophoresis buffer was used to run the agarose gel. To visualize the DNA samples 0.5 µg/ml ethidium bromide was added prior to poring and running the agarose gel. Ethidium bromide is a fluorescent dye that intercalates between the base pairs of double stranded DNA molecules and allows for the visualization of the ethidium bromide-DNA complex under UV light. The DNA samples are mixed with 1 volume of 10x gel loading buffer (65% (w/v) sucrose, 10 mM Tris-HCl (pH 7.5), 10 mM EDTA, 0.2% (w/v) bromophenol blue) and loaded into the appropriate wells of the agarose gel. The 1 kb plus DNA ladder (Invitrogen, Eggenstein) served as molecular weight marker. Separation of DNA fragments was achieved by running the gels at 60 to 120 Volts for the appropriate time depending on the size of DNA to be separated and the concentration of the agarose gel.

2.2.1.12 Isolation of DNA from agarose gels

DNA fragments that had been separated by standard agarose gel electrophoresis where extracted and cleaned up using the QIAquick® Gel Extraction kit (Qiagen, Hilden) according to manufacturers instructions.

2.2.1.13 Polymerase chain reaction (PCR)

A standard polymerase chain reaction (PCR) allows for the selective amplification of specific DNA regions. To amplify specific DNA regions from mouse genomic DNA for reporter gene assays the TripleMaster® PCR mix (Eppendorf AG, Hamburg) was used since it contains a mixture of three different proteins (Taq DNA polymerase, 3'-5' proofreading enzyme and polymerase enhancing factor) that aid in a high fidelity PCR through GC-rich DNA templates. The PCR reaction mix and the PCR program was set up as outlined in the manual of the TripleMaster® PCR mix.

2.2.1.14 Isolation of RNA from eukaryotic cell lines and tissues

Isolation of total RNA was done using the peqGOLD TriFast™ (Peqlab, Erlangen) and the RNeasy Mini Kit (Qiagen, Hilden) according to manufacturers instructions.
Briefly, for cells lines, the peqGOLD TriFast™ (Peqlab, Erlangen) solution was used, since it allowed storage at –80°C right after harvesting the cells. For each ml TriFast solution used, 200 µl chloroform was added to the sample. The samples are vortexed and then phenol chloroform phases are separated by a 45 min centrifugation at 12,000 x g. The top aqueous phase containing the RNA was transferred to new tubes and mixed 1:1 with 100% Ethanol and then purified using the RNeasy mini kit columns.
For tissue samples, approximately 30 mg of mammary gland tissue was homogenized still frozen in RLT Buffer (Qiagen, Hilden) containing ß-Mercaptoethanol (1 ß-ME:100 RLT buffer) for 1-2 minute at setting 5 using the Ultra-Turrax T25 basic (IKA Werke, Staufen) disperser. The RNA was then purified using the RNeasy mini columns and the RNA was eluted from the columns with DEPC treated water. The RNA was then immediately reverse transcribed or stored at –80°C.

2.2.1.15 Measurement of RNA concentration

The RNA concentration can be determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer using a quartz cuvette. One unit of absorbance at 260 nm wavelength corresponds to 40 µg/ml RNA. The total amount of RNA is calculated as follows:

Measured (A_{260}) x dilution factor of sample x 40 = µg RNA / ml

Alternatively RNA concentration can be measured directly using only 1 µl of sample RNA with the NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington USA).

2.2.1.16 Reverse transcription of RNA (RT-PCR)

For the first strand cDNA synthesis isolated template RNA was reverse transcribed using the Omniscript® Reverse Transcription kit from Qiagen (Hilden). The reaction set up was performed as outlined in the manufacturers instructions. Briefly, to reverse transcribe 2 µg of total RNA the following reaction components were mixed in a final volume of 20 µl: 1x RT buffer, dNTP mix (final concentration 0.5 mM each dNTP), 200 ng hexamer primer $pd(N)_6$ (Amerscham Pharmacia Biotech, Upsala, Sweden), 20 units RNase Inhibitor (RNaseOut, Invitrogen, Karlsruhe) and 4 units of Omniscript Reverse Transcriptase. The reverse transcription reaction was incubated for 1 hour at 37°C followed by heat inactivation of the reverse transcriptase for 5 min at 95°C. In subsequent quantitative real-time PCR reactions 1-3 µl of this cDNA was then used.

2.2.1.17 Quantitative real-time PCR (i-Cycler MyiQ Single Color Real-time PCR Detection System)

For quantitative real-time PCR the i-Cycler PCR machine from Bio-Rad (München) was used for every qPCR reaction, 0.5-3 µl of template (DNA or cDNA) together with 0.5 µl of each specific primer (10 pmol/µl) and 12,5 µl of ABsolute™ QPCR SYBR® Green Mix (Abgene, Epsom, UK) were mixed in a final reaction volume of 25 µl. The reactions were set up in 96 well ThermoFast® PCR plates and sealed with optically clear adhesive sheets (Abgene, Epsom, UK).

For amplification the following PCR program was used: 95°C for 15 minutes (activate DNA polymerase in the ABsolute™ QPCR SYBR® Green Mix), 95°C for 30 seconds, 58°C for 30 sec (annealing) and 72°C for 30 sec (elongation) for maximal 45 cycles.

All PCR reactions were performed in triplicates to compensate for experimental errors. All measurements are presented in form of threshold cycle (C_T) values as determined by the BIO-RAD MyiQ™ software.

2.2.1.18 Analysis of the quantitative Real-Time PCR Data

A) qRT-PCR data analysis for reverse transcript RT-PCR:

The threshold cycle (C_T) data obtained from qRT-PCR on the i-Cycler PCR machine were analyzed as described in the ΔC_T-Method (Bio-Rad, München).

The cycle threshold values obtained for the regulated gene of interest (GOI) and the stably expressed housekeeping gene (HKG) are used to calculate their relative expression levels in different samples (example uninduced versus induced samples).

In general relative expression levels is calculated as described below:

Relative expression = $2^{-(S\Delta CT-C\Delta CT)}$

where:

ΔC_T is the difference in the C_T values between the GOI and the HKG as calculated by (GOI C_T)-(HKG C_T).

$S\Delta C_T$ -$C\Delta C_T$ is the difference between the sample ΔC_T and the control ΔC_T.

The control sample (for example a non-induced sample point) acts as a fixed reference point and here the relative expression level has been arbitrarily set to 1.

B) Calculations for the relative expression levels of putative STAT5 library target genes:

To take into account overall experimental variation within the qRT-PCR experiments, each individual gene of interest (GOI) was normalized against the geometric average of three HKG's. This was done using the 'example-calculations' Excel file provided at the 'geNorm' web page (http://medgen.ugent.be/~jvdesomp/genorm/). This excel file contains example calculations for normalization using more than one reference (HKG) gene. For the results presented here the HKG used as references were 18S RNA, YWHAZ and cyclophilinB). The file was also modified to accommodate the analysis of the 4 different sample conditions and the normalized expression levels of the GOI was displayed in a diagram after rescaling to the geometric mean on a linear scale.

C) qRT-PCR data analysis for ChiP:

The threshold cycle (C_T) data obtained form the ChIP assays on the i-Cycler PCR machine were analyzed by a modified ΔC_T-Method as described by Rascle and Lees (2003).

All data is displayed as % of Input DNA that represents the % of DNA occupancy. The samples were all run in triplicates and normalized to the Input average.

% Input DNA = $2^{\wedge(CT\ Input-\ CT\ IP)}$ x ratio x 100

Ratio for normalization (ratio):

INPUT ratio =
(µl of Input DNA used per PCR reaction / µl final volume Input was resuspended in after DNA purification) x µl of pre-cleared lysate saved for Input

IP ratio =
(µl of IP DNA used per PCR reaction / µl final volume IP was resuspended in after DNA purification) x µl of pre-cleared lysate used per ChIP IP

INPUT ratio value /IP ratio value = ratio for normalization.

A fold difference of at least 3 fold is considered the minimum to be significant, since less than 3 fold difference could be due to a 1 C_T shift (Edelstein et al., 2005).

2.2.2 Tissue culture methods

2.2.2.1 General methods for tissue culture

All cell culture work was performed under sterile laminar flow hoods. The cell lines are cultured in an incubator at 37°C with a relative humidity of 95% and a CO_2 concentration of 5%. Cells are cultured either on cell culture dishes or in flasks.

For induction of cells with lactogenic hormones the following final concentrations are used:

Prolactin (PRL)	5 µg/ml
Dexamethasone	0.1 µM
Insulin	5 µg/ml

The mouse mammary epithelial cell line HC11 was grown to confluency in RPMI 1640 medium supplemented with 10% heat inactivated fetal bovine serum, 1% penicillin-streptomycin and 5 µg/ml insulin (Sigma), 10 ng/ml EGF (Sigma). Before hormone induction HC11 cells were washed 1x with PBS and kept in EGF free RPMI medium containing 2% FCS and and 5 µg/ml insulin (Welte et al., 1994). HC11 cells were induced with 5 µg/ml PRL (31 U/mg, Sigma), and 0.1 µM dexamethasone (Sigma) or left untreated for indicated time points.

2.2.2.2 Passaging of cells

Adherent cells are washed one time with PBS (1x) and then incubated with Trypsin/EDTA for 5 to 10 minutes at 37°C. Once detached, the cells are suspended in the corresponding media containing FCS and seeded out onto dishes at the desired dilution.

2.2.2.3 Cryogenic preservation, storage and thawing of mammalian cells

A near confluent dish of adherent cells is first detached with trypsin. The cells are suspended in 10 ml media and pelleted by cetrifugation (400 x g, 5 min). The supernatant is removed and the cell pellets is resuspended in a cryoprotective medium (10% (v/v) Dimethylsulfoxide (DMSO), 90% (v/v) FCS) and dispensed at 1 ml aliquots into several cryogenic vials. The vials are slowly frozen at -20°C for 24 hours and then transferred into a -80°C freezer or for permanent storage into the vapor phase (-140°C) of a liquid nitrogen freezer.
The cells are thawed quickly with gentle agitation in a 37°C water bath and immediately resuspended in pre-warmed medium to dilute the DMSO. The cells are pelleted (100 x g, 5 min) and then resuspended in fresh medium and seeded out in a suitable culture vessel.

2.2.2.4 Determination of cell numbers with the COULTER Counter

To determine the concentration of cells in suspension, the cells get diluted 1:20 with Isotone II and the number of particles is measured with the Coulter counter (Beckman Coulter, Krefeld). Only particles between 5-18 µm in size get counted since living cells are found in this size fraction.

2.2.2.5 Transient transfections of eukaryotic cells with Superfect reagent.

HeLa TA cells were transfected using the Superfect transfection reagent (Qiagen, Hilden) according to the manufacturers instructions.
In the case of hormone stimulation for a reporter gene assay the cells were induced one day after transient transfection and cultivated for an additional 16 hours.

2.2.3 Molecular Biology techniques

2.2.3.1 Cell lysates preparation for reporter gene assays

For reporter gene assays the cells are cultured in 12 well plates. Adherent cells are washed 2x with ice-cold PBS (1x) buffer and scraped off the surface of the well. Cells from each well are transferred with 1 ml of ice-cold PBS buffer into a 1.5 ml Eppendorf tube and centrifuged (1,000 x g, 5 min). Each cell pellet is then lysed with 50 µl of Lysis buffer (25 mM Glycylglycin pH 7.8, 8 mM $MgSO_4$, 1 mM EDTA, 1 mM DTT, 15% (v/v) Glycerol, 1% (v/v) Triton X-100). After 10 minutes incubation on ice, the cell lysates are centrifuged (16,000 x g, 10 min). The supernatant is then transferred into new tubes and can be either used directly or stored at –20°C until further use.

2.2.3.2 Determination of luciferase and ß-galactosidase activity

To determine the luciferase activity 10 µl of each cell lysate is pipetted into a separate well of a white 96-well plate and is analyzed in a Luminometer (MicroLumat LB96 P, EG&G Berthold, Bad Wilbad) after addition of 50 µl of Luciferin reaction buffer (20 mM Tricine/NaOH, pH 7.8, 1.02 mM $[MgCO_3]_4 \bullet Mg[OH]_2 \bullet H_2O$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33 mM DTT, 270 µM Coenzyme A, 470 µM luciferin, 530 µM ATP).
The luciferase catalyzes the oxidation of the luciferin substrate in the presence of ATP, molecular oxygen and magnesium ions. After the addition of the luciferin the emitted light from this reaction is measured in the luminometer for a total of 30 seconds at a wavelength of 562 nm.
For the determination of the ß-galactosidase activity, 3.3 µl of the corresponding cell lysate and 33.3 µl of ß-galactosidase reaction buffer (1 mM $MgCl_2$, 100 mM Phosphate buffer [$NaPO_4$], pH 8.0, 1% (v/v) Galacton (Tropix, Bedford)) is pipetted into a new white 96-well plate. The plate is sealed with aluminum foil to protect against light and incubated for 45 min under agitation at room temperature (RT).

The light emission is measured in the Luminometer under the same condition as the luciferase activity after addition of the 50 µl of enhancer solution (200 mM NaOH, 10% (v/v) Emerald chemiluminescent amplifier (Tropix, Bedfort). The measured ß-galactosidase activity is used to normalize the luciferase activity.

The normalized Luciferase value for a given sample is obtained by multiplying the measured luciferase value with the quotient of the average ß- galactosidase value (sum of all ß-gal measurements in that data set) and the measured corresponding ß-galactosidase value. At least two independent assays were performed and analyzed.

2.2.3.3 Preparation of mammary gland tissue from mice

Balb/c mice were bought from the Charles River Company (Sulzfeld) and maintained at the in-house Georg-Speyer-Haus Animal facility.

Mammary gland tissue was harvested from adult virgin mice (at least 5 weeks old), 10-18.5 days pregnant mice, 5 days lactating mice and mice that had weaned their pups for at least a weak (Involuting).

In the mouse, the mammary glands form extensive subcutaneous sheets of tissue that extend from the cervical to the inguinal regions as five ventrolateral pairs.

Mammary gland tissues from glands 2, 3, 4 and 5 were isolated from female Balb/c mice as described (Ball, 1998; Pullan and Streuli, 1996). The thoracic glands 2 and 3 and the inguinal glands 4 and 5 were chosen, since they were most easily removed also from virgin mice. Mammary glands were placed on ice immediately after harvesting.

2.2.3.4 Standard Chromatin Immunoprecipitation (ChIP) assay

A) ChIP assay using mammary gland tissue:

ChIP assays using mammary gland tissue were performed in part as described by the Farnham lab protocol (section 2.1.10) and as described below.

Mammary gland tissue was harvested from female Balb/c mice (section 2.2.3.3).

For mammary gland tissue from virgin mice and mice having undergone involution about (10-12 mice) 8-11 gram tissue are required, while for mammary gland tissue from pregnant and lactating mice (1-2 mice) only 2-3 gram of mammary gland tissue are required for a ChIP assay. The differences in the number of mice and amount of mammary gland tissue harvested was adjusted to yield approximately similar size of cell pellets after tissue homogenization. Mammary gland tissue from virgin mice provided in general the lowest yield in terms of cell pellet volume and dictated the amounts finally

used in the immunoprecipitations between the different developmental stages. After harvesting, the mammary gland tissue is chopped into approximately equal thick pieces with a scalpel, to ensure comparable formaldehyde cross-linking conditions between the different stages. The tissue is transferred to a 50 ml falcon tube and washed two times with PBS (1x). Formaldehyde (Merk) is then added to a final concentration of 1% and rotated at room temperature for 20 min. The cross-linking reaction is stopped by adding glycine to a final concentration of 0.125 M. Rotation of the samples is continued at room temperature for 5 min. The samples are then centrifuged at low speed and washed two times with ice cold 1x PBS. The tissue is then chopped up with a scalpel into a very fine paste. To disaggregate the tissue even further, small pea size amounts of paste are added together with about 1 ml ice cold 1x PBS to a loose fitting dounce homogenizer 2 cm^3 (B. Braun, Melsungen) and disaggregated. The resulting solution is strained through a 40 µm nylon cell strainer (Becton Dickinson Labware) using ice cold 1x PBS.

The samples are centrifuged at low speed to pellet the cells and to decant the supernatant. The cell pellets are then washed extensively for at least three times with ice-cold 1x PBS. If a top reddish layer has formed on the cell pellet it needs to be removed. Before proceeding with the lysis, the cell pellets between the different mammary gland stages are adjusted to equal size. (About 50 µl of cell pellet volume in a 1.5 ml Eppendorf tube can be used for about 7 immunoprecipitations).

Cell pellets are then resuspended in 1 or 2 ml of lysis buffer (1% (w/v) SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1, phosphatase Inhibitor cocktails I and II (Sigma), protease inhibitor Complete (Roche)). The final volume of the cell lysis buffer is adjusted so that there are no cell clumps. The samples are incubated on ice for 20 minutes and then sonicated three times for 15 seconds each with a 3 mm tip using a Branson sonifier W-250 (Heinemann, Gmünd) at setting 5.

Supernatants were then recovered by centrifugation at 16,000 x g for 10 min at 4°C. Supernatants from different conditions were all diluted equally 3 to 10 times in dilution buffer (1% (v/v) Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1) and subjected to one round of immunoclearing for 2 hr at 4°C with 2 µg sheared salmon sperm DNA, 2.5 µg preimmune serum and 20 µl of protein A-agarose or A/G plus agarose (of 50% slurry). Fifty microliters of precleared lysate was kept as the input for real-time PCR. Equal volumes (1.5 ml) of precleared lysate was used in each immunoprecipitation (IP) with the antibodies for every condition. Immunoprecipitation was performed overnight with specific antibodies, then 2 µg sheared salmon sperm DNA and 20 µl of protein A or A/G agarose (of 50% slurry) were further added for 1-2 hr at 4°C. The immunoprecipitates (IPs) were washed sequentially for 5 min each in TSE I (0.1% (w/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl), TSE II (0.1% (w/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl) and buffer III

(0.25 M LiCl, 1% (v/v) NP-40, 1% (w/v) deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1). Agarose beads precipitated were then washed three times with TE buffer and eluted three times with 1% (w/v) SDS, 0.1 M $NaHCO_3$. The IP eluates were pooled and together with the input were heated at 65°C overnight to reverse the formaldehyde cross-linking. Supernatants were then incubated for 1 hr at 45°C with proteinase K (40 µg each) and RNase (0.8 µg) and DNA was purified using PCR Spin columns (Qiagen). DNA from immunoprecipitations was resuspended in 100 µl of sterile water and DNA from input was resuspended in 300 µl of sterile water. Quantitative real-time PCR (qRT-PCR) was performed with 1-3 µl of DNA and performed in triplicates. Immunoprecipitation data were normalized to input DNA ($2^{\wedge(CT\ Input-\ CT\ IP)}$ x ratio x 100) and amounts of DNA recovered in the IPs were expressed as percentage of input DNA, as described in the work of Rascle et al. (2003). The ChIP results presented were performed on chromatin from at least two independent ChIP experiments with similar trends.

B) ChIP assay using a mammary epithelial cell line:

ChIP assays for the mammary epithelial cell line HC11 were performed as described in Litterst et al. (2003). Briefly, HC11 cells were grown to high confluency and maintained for 2 days in EGF-free RPMI 1640 medium containing 2% (v/v) fetal calf serum and 5 µg/ml insulin. The cells were stimulated with prolactin (5 µg/ml) and dexamethasone (0.1 µM) for the different time points indicated and cross-linked with 1% (v/v) formaldehyde under agitation for 10 minutes at room temperature. During this time the plates were gently agitated on a rocker. The cross-linking reaction was stopped by the addition of glycine to a final concentration of 0.125 M and agitated for an additional 5 minutes. The HC11 cells were then rinsed with ice-cold PBS and harvested by scraping them of the plate. The cells were collected by centrifugation, washed 1x in PBS and then resuspended in 1 ml of lysis Buffer. The cell lysis and subsequent steps in the ChIP assay were performed as described above in section A. About one 10 cm^2 culture dish of confluent HC11 cells (approximately 13 million cells) was used for about three immunoprecipitations.

2.2.3.5 Construction of a library containing STAT5A-binding targets using the ChIP method

The strategy was to create a library containing putative STAT5A-binding targets immunoprecipitated in ChIP assay from mammary gland tissue (Fig. 2.1).
The standard ChIP protocol was performed using a STAT5A antibody to capture STAT5A associated chromatin, followed by an adaptation of the ChIP cloning protocol from Peggy Farnham's Lab, to create a library with these captured DNA fragments. Using an antibody against STAT5A a standard ChIP assay was performed on chromatin obtained from mammary gland tissue of a 5 day lactating mouse. A standard quantitative RT-PCR was performed using the IgG and STAT5A generated immunoprecipitations (IPs) with primers corresponding to the known positive control to verify that the ChIP assay was successful before proceeding further. The positive control ß-casein promoter primes should show clear enrichment for each individual STAT5A IP over IgG.
Since immunoprecipitations from tissue yield very low amount of DNA a total of 7 individual STAT5A IPs were pooled. In addition, to make certain that enough DNA is available, a ligation-mediated PCR technique (LMPCR) to create an amplicon of the pooled STAT5A IP as outlined by protocols from P. Farnham Lab (http:// genomecenter .ucdavis.edu/ farnham/ farnham/ protocol.html; ChIP-CpG island microarray binding analysis) was used. With LMPCR chromatin from a STAT5A IPs can be amplified that accurately represents the starting DNA population. The LMPCR method involves the blunt ending of the immunoprecipitated chromatin, ligation of a unidirectional double-stranded oligonucleotide linker and finally PCR amplification of the resultant DNA population. The linker-ligated chromatin from the STAT5A IP was then PCR amplified one time for 15 cycles (as described in Oberley M. and P. Farnham lab protocol; section II, Generating Chromatin Amplicons; day 1 and day 2: steps 4 to 7). The cloning of the STAT5 chromatin library into the pCR 2.1 vector was done as described in the TA Cloning Kit from Invitrogen. From the 30 µl STAT5A amplicon (18 ng/µl) 1 µl was used for the ligation reaction with the pCR 2.1 vector. 2 µl of the ligation reaction was then used to transform competent TOP10F' *E. coli* cells. 90 µl of the total 300 µl of transformed cells was plated out on LB plates containing X-Gal/IPTG. This allows for a blue/white screening for the selection of white colonies that contain vectors with inserts. The left over 210 µl of transformed cells were stored at –80°C in 15% (v/v) glycerol. Therefore the STAT5A library presented here represents a sub-fraction of the total STAT5A library. All possible white colonies (containing inserts) were picked from the plates and plasmid mini preps were made to isolated the plasmid DNA. The restriction analysis was performed with EcoRI and subjected to an agarose gel electrophoresis run to determine insert fragment size. Clones containing inserts around 500 bp and larger were sequenced using M13

Reverse primer. Larger fragments were chosen, since they are less likely to contain non-specific DNA according to the advice in the lab cloning protocol of P. Farnham.

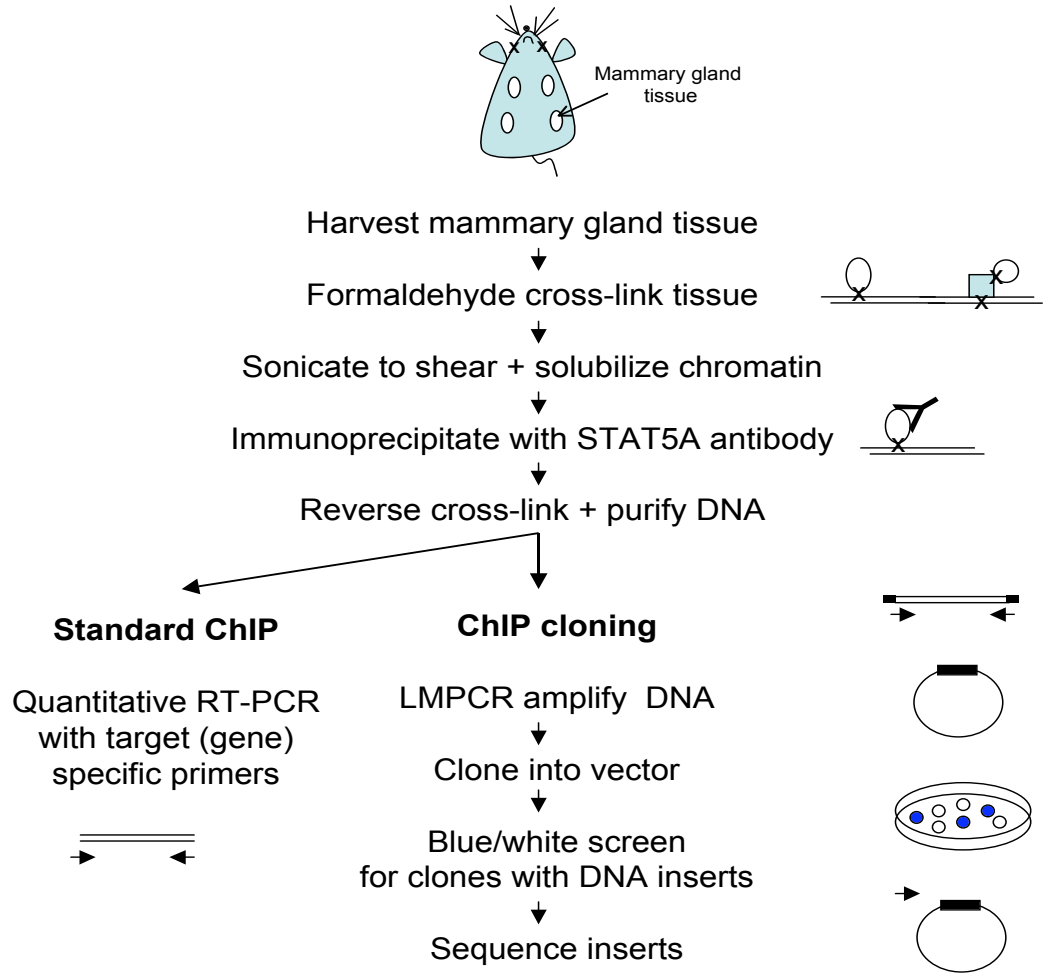


Figure 2.1 Schematic diagram of the workflow used for the cloning ChIP procedure.
For both the standard and cloning procedure the initial ChIP assay steps are similar. The mammary gland tissue is harvested from a female mouse and cross-linked using formaldehyde. This step cross-links any protein-protein and protein-DNA interactions. Sonication then shears and solubilizes the chromatin. Immunoprecipitation with an antibody against STAT5A allows for the capture of DNA that was associated with the antibodies target. The formaldehyde cross-linking is reversed and the DNA purified. A quantitative RT-PCR with primers that amplify a known STAT5A target (such as ß-casein promoter) completes the standard ChIP. To create a library containing novel DNA fragments associated with the factor of interest the ChIP cloning procedure was followed. The purified DNA was LMPCR amplified and then cloned into the plasmid vector pCR 2.1, which allows for a blue/white screening. The screening allows for the selection of white colonies containing inserts. Select clones containing inserts were then sequenced.

3. Results

3.1 Factor recruitment and changes in histone modifications at the endogenous ß-casein and WAP gene regions during mammary gland differentiation

The purpose of the mammary glands is to nourish young with milk and undergoes repeated cycles of growth, differentiation and regression with each pregnancy. Since the main events of differentiation in the mammary gland take place during the adult stages of pregnancy and lactation, the mammary gland is used as a model to study differentiation in adult mammals. Differentiation in the mammary gland is characterized by the activation of milk proteins such as ß-casein and whey acidic protein (WAP) in the mammary epithelial cells. The differentiation-specific expression of these genes requires the ordered interaction of different hormone-regulated transcription factors, the recruitment of co-factors, histone modifications and the remodeling of chromatin in the promoter region. In tissue this complex interplay has not yet been investigated well, therefore it is important to look at this process in mammary gland tissue at different developmental stages.
The hormone prolactin (PRL) is essential for the growth and functional differentiation of mammary gland tissue and for the transcriptional activation of milk proteins (Topper and Freeman, 1980). The transcription factor STAT5 mediates the PRL induced expression of the milk proteins. When both STAT5A and STAT5B are absent, neither ß-casein nor WAP are expressed (Miyoshi et al., 2001). In the absence of STAT5A alone, the expression of the milk proteins is already severely reduced (Liu et al., 1997). Later studies, using conditional STAT5 mutant mice, have further substantiated that STAT5 is intimately involved in the regulation of mammary epithelial cell proliferation, differentiation and survival (Cui et al., 2004).

The aim of this study was to investigate factor recruitment and changes in histone modifications that associate to the endogenous chromatin of two milk proteins (ß-casein and WAP) during different stages of mammary gland differentiation. For this purpose the Chromatin immunoprecipitation (ChIP) assay has been chosen, since this method can be used to evaluate the *in vivo* presence of proteins that occur on endogenous target chromatin. Mammary gland tissue from female Balb/c mice in the following four development stages: adult virgin (V), pregnant (P), lactation (L) and involution (I) were examined.

First the expression of the two target genes to be studied was verified. In order to quantify the abundance of ß-casein and WAP mRNA in mammary gland during the four different developmental stages quantitative real-time PCR (qRT-PCR) was performed. Total RNA isolated from mammary gland tissue was reverse transcribed into cDNA using hexamer primers. The ß-casein and WAP-cDNA was then amplified with sequence specific primers in triplicate qRT-PCR reactions. The levels of ß-casein and WAP mRNA were normalized to the abundance of 18S ribosomal RNA.

A clear induction for both the ß-casein mRNA and WAP mRNA was detected during pregnancy (Fig. 3.1A and B). During lactation the analysis revealed an even stronger expression level compared to pregnancy for both ß-casein and WAP. In comparison, nearly no expression for both genes could be detected in tissue from adult virgin mice, which was similar to involution, where only minimal levels were detected for ß-casein.

These results confirm the differentiation-specific regulated expression of the two milk protein genes during pregnancy and lactation. These genes are therefore suitable for the analysis of histone modification during mammary gland differentiation.

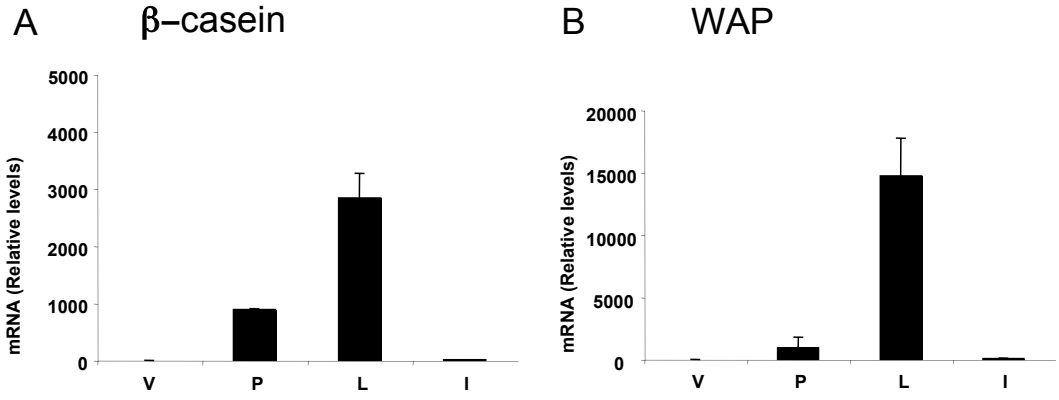


Figure 3.1 Confirmation of ß-casein and WAP mRNA expression in mammary gland tissue.
RNA samples were prepared from mammary gland tissue of Balb/c mice during different stages of mammary gland development (virgin (V), pregnancy (P), lactation (L) and involution (I)). The expression levels of ß-casein (A) and WAP (B) transcripts relative to 18S ribosomal RNA were determined by quantitative real-time PCR.

In order to determine transcription factor binding and the distribution of histone modifications multiple primer sets covering specific regions across the ß-casein and WAP genes were designed. These primers were used in qRT-PCR to analyze the immunoprecipitated (IP) DNA of the ChIP assays. For the ß-casein gene multiple primers sets were designed spanning the region possessing a predicted enhancer region (-6kb) (Kabotyanski et al., 2006; Rijnkels et al., 2003), the 5' region (–3.5kb), the promoter region with two STAT recognition sites (-0,1kb), the transcribed region (transcribed) (+1,7kb), and the 3' region (+7.2kb) (Fig. 3.2A). For the WAP gene the primer were used to cover the 5' region (-3,0kb), the promoter region possessing the

STAT recognition site (-0,5kb), the transcribed region (+1,2kb), and the 3' region (+4,2kb) (Fig. 3.3A). These primers allow for the analysis of transcription factor binding to different regions of the genes (e.g. promoter and transcribed region), but also to scan for changes in the distribution of histone modifications in the far up- (e.g. enhancer) and downstream regions of these genes.

3.1.1 Both STAT5a and STAT5b bind to the enhancer and promoters of the ß-casein and WAP gene in mammary gland tissue

To investigate direct STAT5 DNA binding *in vivo*, ChIP assays were performed in tissue from mice at different stages of mouse mammary gland development. The phenotype of conventional STAT5a-knockout mice demonstrated that STAT5a is the main mediator of mammary gland development (Liu et al., 1997) ChIP was also performed for STAT5B, since both phosphorylation activated STAT5A and STAT5B are able to bind the same DNA binding motifs, the gamma-interferon activation site (GAS). Latent STAT5A and STAT5B proteins are found in the mammary tissue of virgin, pregnant, lactating and involuting mice at comparable levels, but in the phosphorylation activated state, the highest levels of both STAT5s is observed during pregnancy and throughout lactation (Liu et al., 1996). In addition, early *in vitro* experiments using electrophoretic mobility shift assays (EMSA) had demonstrated that during lactation both STATA and STATB display DNA-binding activity as hetero- and homodimers (Liu et al., 1996). Therefore, antibodies specific for either STAT5A or STAT5B were used to immunoprecipitate cross-linked chromatin. Purified genomic DNA was then subjected to qRT-PCR with primers amplifying the different regions of the ß-casein and WAP genes described above (Fig. 3.2A, Fig. 3.3A). Both STAT5A and STAT5B were recruited to the STAT5 binding sites of the endogenous ß-casein and WAP promoter regions in lactating and in pregnant mouse mammary gland tissue (Fig. 3.2B and C, Fig. 3.3B and C). In addition, binding of both STAT5A and STAT5B was detected at the distal enhancer of the ß-casein (-6kb) in pregnant and lactating mice. In comparison, no binding of either STAT5A or STAT5B was observed at the 5´, transcribed region or 3´ region of the ß-casein or the WAP gene. Interestingly, analysis of the ß-casein upstream sequence predict a classical STAT5 binding at -3525 to -3507 nucleotide 5´ from the transcription start site (TSS) of the ß-casein gene, which is covered by the 5' primers. However, no binding to this region could be detected in ChIP. This may suggest that this STAT5 site is not accessible to STAT binding or that the antibodies used failed to recognize STAT5 due to epitope masking. These data suggests that in mammary gland tissue both STAT5A and STAT5B are recruited to the STAT5

recognition elements in the promoters of the endogenous ß-casein and WAP gene during pregnancy and lactation.

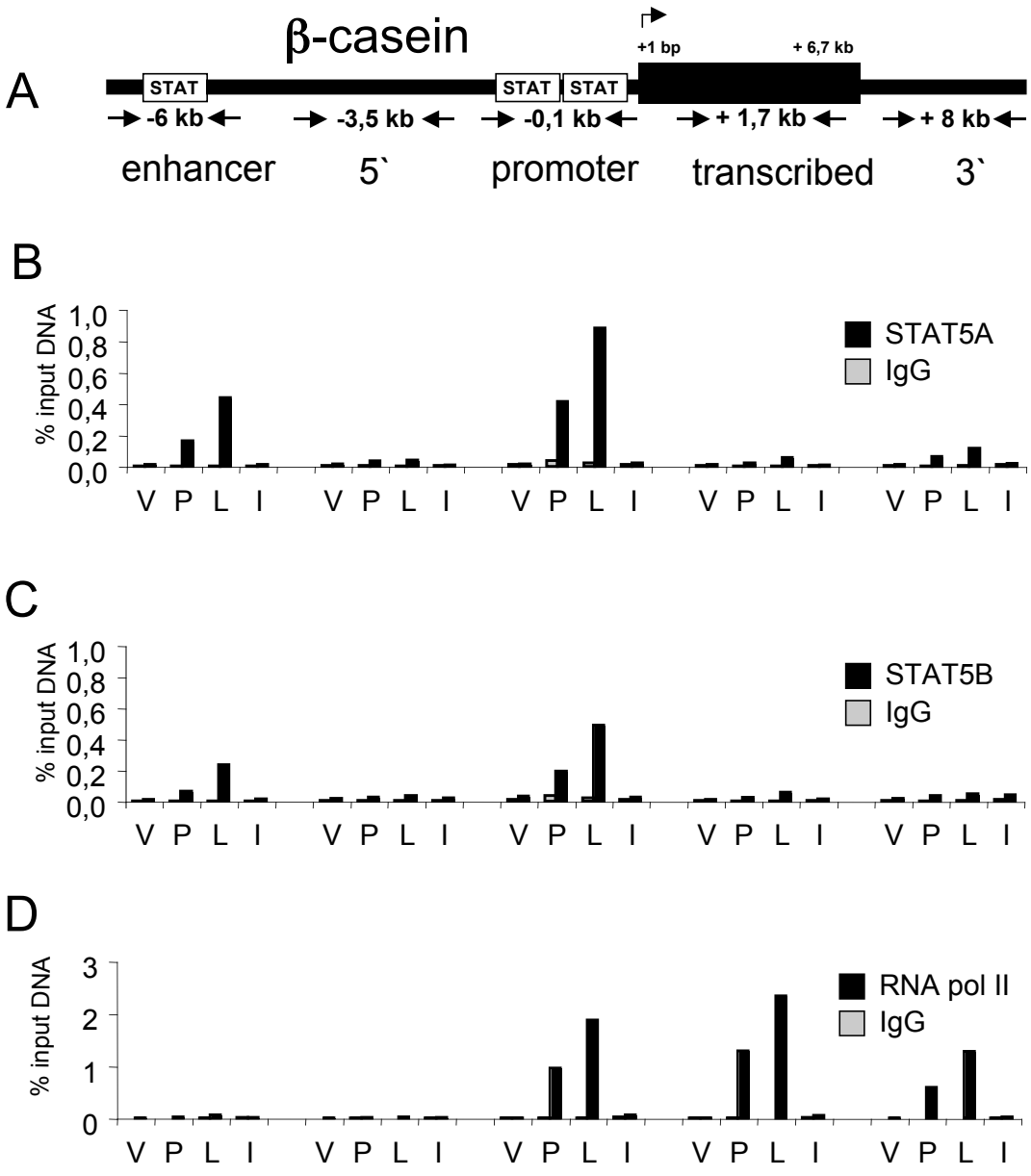


Figure 3.2 Recruitment of the transcription factors STAT5A, STAT5B and RNA polymerase II to the endogenous ß-casein gene in mammary gland tissue.
A. Schematic structure of the murine ß-casein gene.
Square boxes represent two STAT binding sites, together with their positions relative to the transcription initiation site (black +1 arrow). The relative locations of the primers covering the enhancer, 5', promoter, transcribed and 3' regions of the gene used in evaluating transcription factor recruitment and histone modifications are depicted by black arrows.
B,C and D. Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific against STAT5A (STAT5A), STAT5B (STAT5B), phosphorylated CTD RNA polymerase II (RNA pol II) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the different regions of the ß-casein gene depicted in A. The immunoprecipitation data were normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA.

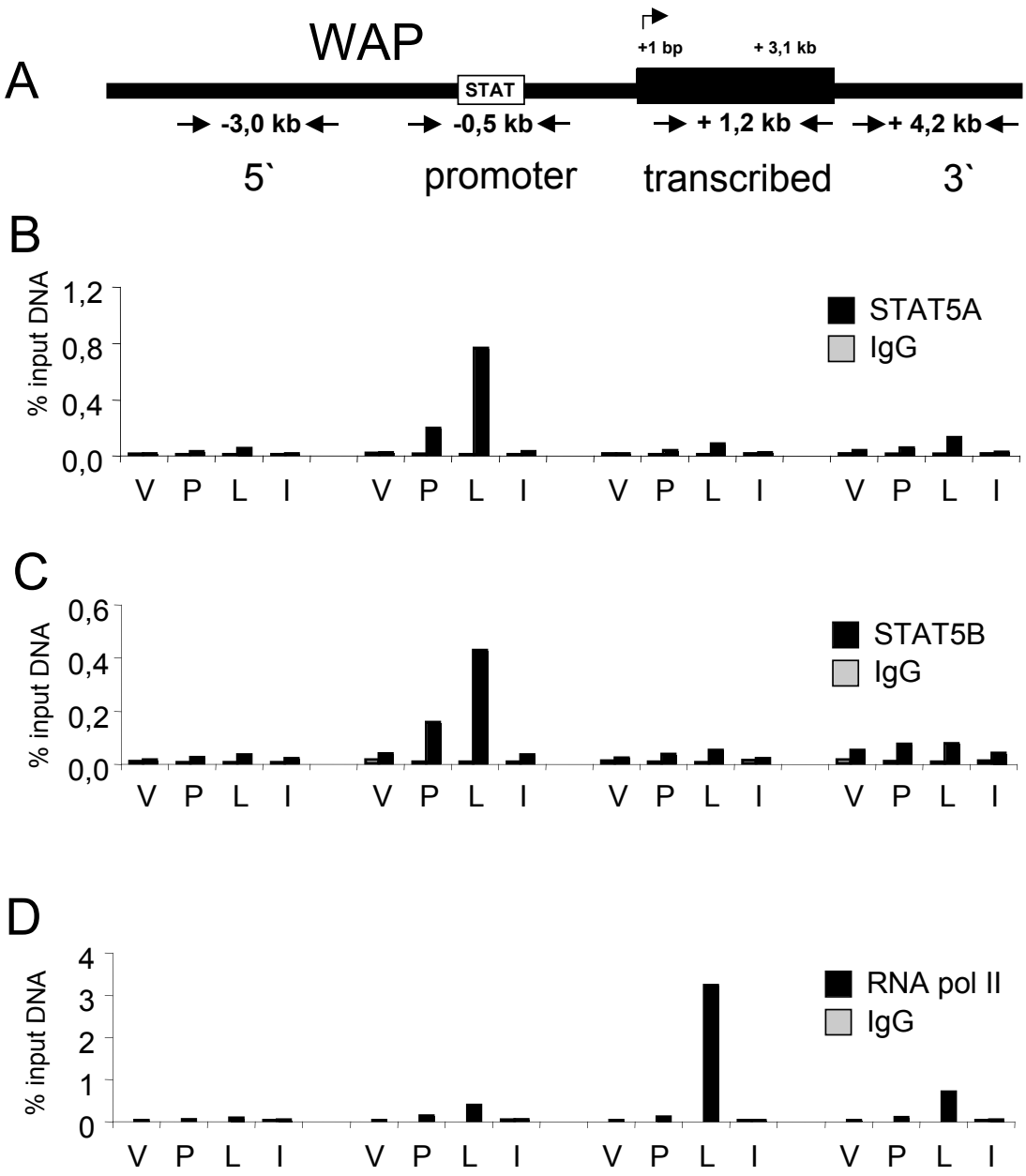


Figure 3.3 Recruitment of the transcription factors STAT5A, STAT5B and RNA polymerase II to the endogenous WAP promoter in mammary gland tissue.
A. Schematic structure of the murine WAP gene.
The square box represents one STAT binding site relative to the transcription initiation site (black +1 arrow). The relative locations of the primers covering the, 5', promoter, transcribed and 3' regions of the gene used in evaluating transcription factor recruitment and histone modifications are depicted by black arrows.
B,C and D. Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific against STAT5A (STAT5A), STAT5B (STAT5B), phosphorylated CTD RNA polymerase II (RNA pol II) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the different regions of the WAP gene depicted in A. The immunoprecipitation data were normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA.

3.1.2 RNA polymerase II is recruited to the promoter, transcribed region and 3´region of the ß-casein and WAP genes during pregnancy and lactation

In Eukaryotes, transcription of protein-coding genes is performed by RNA polymerase II. Being the major component of the basic transcription machinery, its presence indicates transcriptional activity.

To examine if the RNA polymerase II is recruited to the milk protein genes ChIP assays were performed using an antibody specific against the phosphorylated C-terminal domain (p-CTD) of RNA polymerase II. The p-CTD form of RNA polymerase II is considered to be a transcriptional active form of RNA polymerase II. The results from the ChIP assays show that p-CTD RNA polymerase II is associated most strongly with the STAT5 binding sites containing promoter region and the transcribed region of both the ß-casein and the WAP genes in pregnant and lactating mammary gland tissue (Fig. 3.2D and 3.3D). This is in agreement with the confirmed expression of ß-casein and WAP in mammary gland tissue at these differentiation stages (Fig. 3.1). The RNA polymerase II is also found to be associated, although to a lesser extent, at the 3' region of both the ß-casein and WAP genes in the mammary gland from pregnant and lactating mice (Fig. 3.2D and 3.3D). In comparison to the other regions, no such strong recruitment of the p-CTD RNA polymerase II to the distal enhancer region of ß-casein was detected. These results show co-localization of the transcriptional active RNA polymerase II to the endogenous ß-casein and WAP genes, during pregnancy and lactation when these two genes are expressed.

3.1.3 Distribution of histone modifications during different mammary gland stages

The N-terminal tails of histones H3 and H4 can be modified by various post-translational modifications such as acetylation and methylation (Fischle et al., 2003). The specific histone modifications can be spread over large stretches of chromatin or be found concentrated at specific regions of genes. Specific N-terminal tail histone modifications have been associated with actively transcribed euchromatin or silent heterochromatin. To determine what histone modifications occur on the ß-casein and WAP genes in the various mammary gland stages, antibodies specific for modified histone N-terminal tails were used in ChIP assay together with multiple primer sets covering the various regions.

3.1.4 Increased histone acetylation at the ß-casein and WAP genes during pregnancy and lactation

Antibodies specific for acetylated histone H3 (Ac-H3) showed an increase of acetylation at the promoter and transcribed region of the ß-casein gene in pregnant mice (Fig. 3.4B). Maximal acetylation levels were observed in mammary gland tissue from lactating mice. At this differentiation stage, enhanced acetylation was also observed on the enhancer and the 5´region of the ß-casein gene. The increased histone acetylation indicates a more open chromatin structure that is associated with transcriptional active chromatin. In comparison to the other regions the lowest levels of histone H3 acetylation was found at the 3' region of the ß-casein gene. Interestingly, a slight increase in histone H3 acetylation was still observed on the transcribed region of the ß-casein gene during involution, when the expression of this gene has subsided, suggesting a possibly still open chromatin structure. For WAP, there was an increase of histone H3 acetylation (Ac-H3) during lactation at the 5', promoter and the transcribed region (Fig. 3.5B). Antibodies specific for acetylated histone H4 (Ac-H4) showed a pronounced aceylation of this core histone during lactation at the 5´region of both the ß-casein and WAP gene (Fig. 3.4C and 3.5C). An increase of histone H4 acetylation could also be detected in tissue from pregnant and lactating mice at the (enhancer for ß-casein), 5', promoter, and transcribed region of both genes (Fig. 3.4C and 3.5C). Interestingly, levels of histone H4 (Ac-H4) acetylation during lactation are more enriched at the 5' regions of both genes.

In tissue from pregnant and lactating mice there is a clear increased acetylation of histones H3 and H4 indicating a more open chromatin structure at the regions investigated. This enhanced acetylation of these core histones correlates nicely with the transcriptional activity of both the ß-casein and WAP genes during pregnancy and lactation.

The observed enrichment pattern of acetylated histone H4 (Ac-H4), especially at the 5' regions of both ß-casein and WAP was reconfirmed with another antibody specific against Tetra-acetylated histone H4 (data not shown). This antibody against tetra-acetylated histone H4 recognizes the N-terminal tail of Histone H4 that is acetylated specifically at lysines (K) 5, 8, 12 and 16. Therefore the individual lysines of Histone H4 were further investigated to see if any differences in the pattern and distribution could be detected.

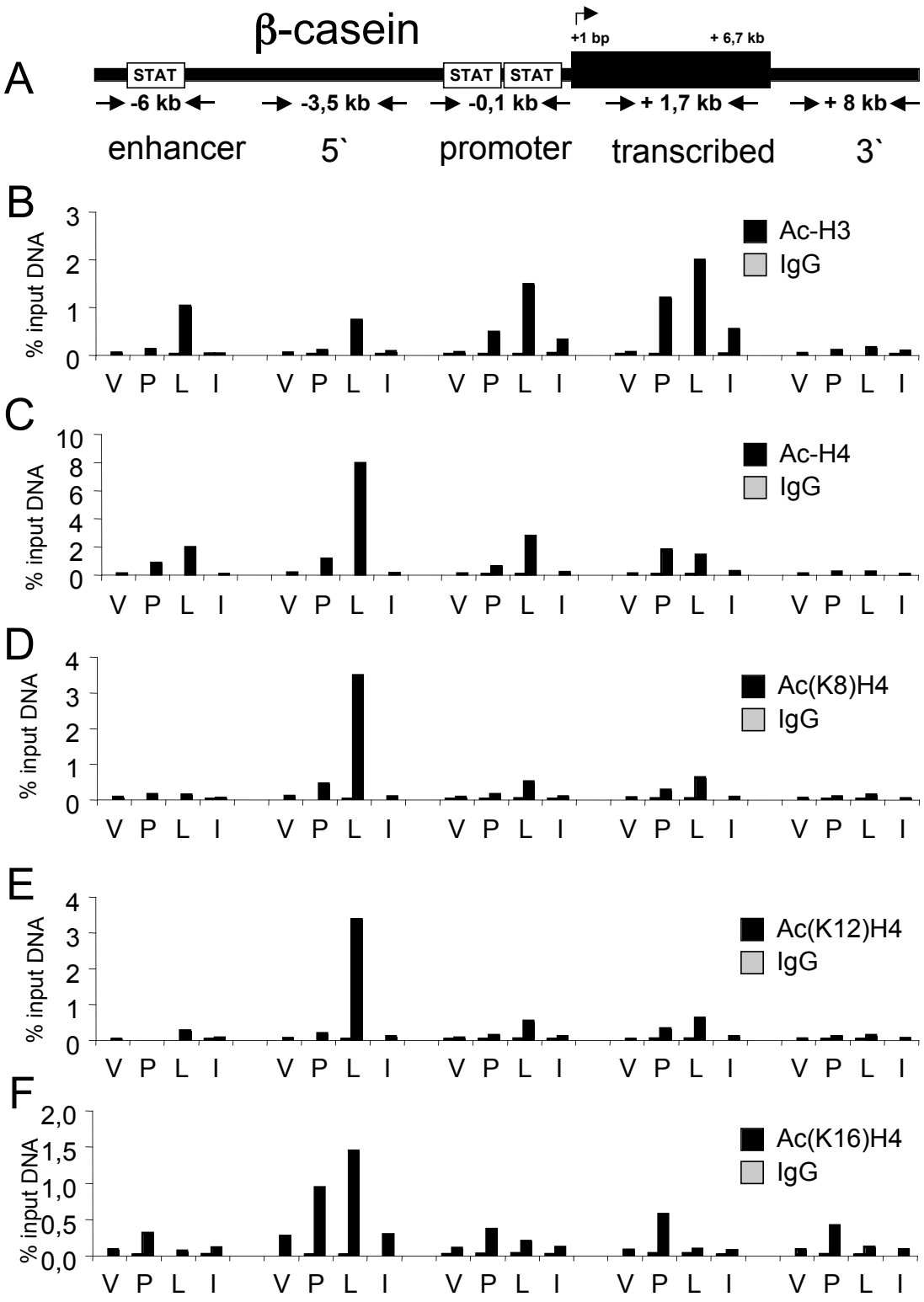


Figure 3.4 Distribution of histone lysine (K) acetylation across the endogenous ß-casein gene in mammary gland tissue.

(A) Schematic representation of the ß-casein gene.

The relative locations of the primers used in evaluating the histone modifications in different regions of the ß-casein gene are depicted by black arrows.

(B) Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for acetylated histone H3 (Ac-H3), acetylated histone H4 (Ac-H4), acetylated lysine 8 of histone H4 (Ac(K8)H4), acetylated lysine 12 of histone H4 (Ac(K12)H4), acetylated lysine 16 of histone H4 (Ac(K16)H4) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions of the ß-casein gene depicted in (A).

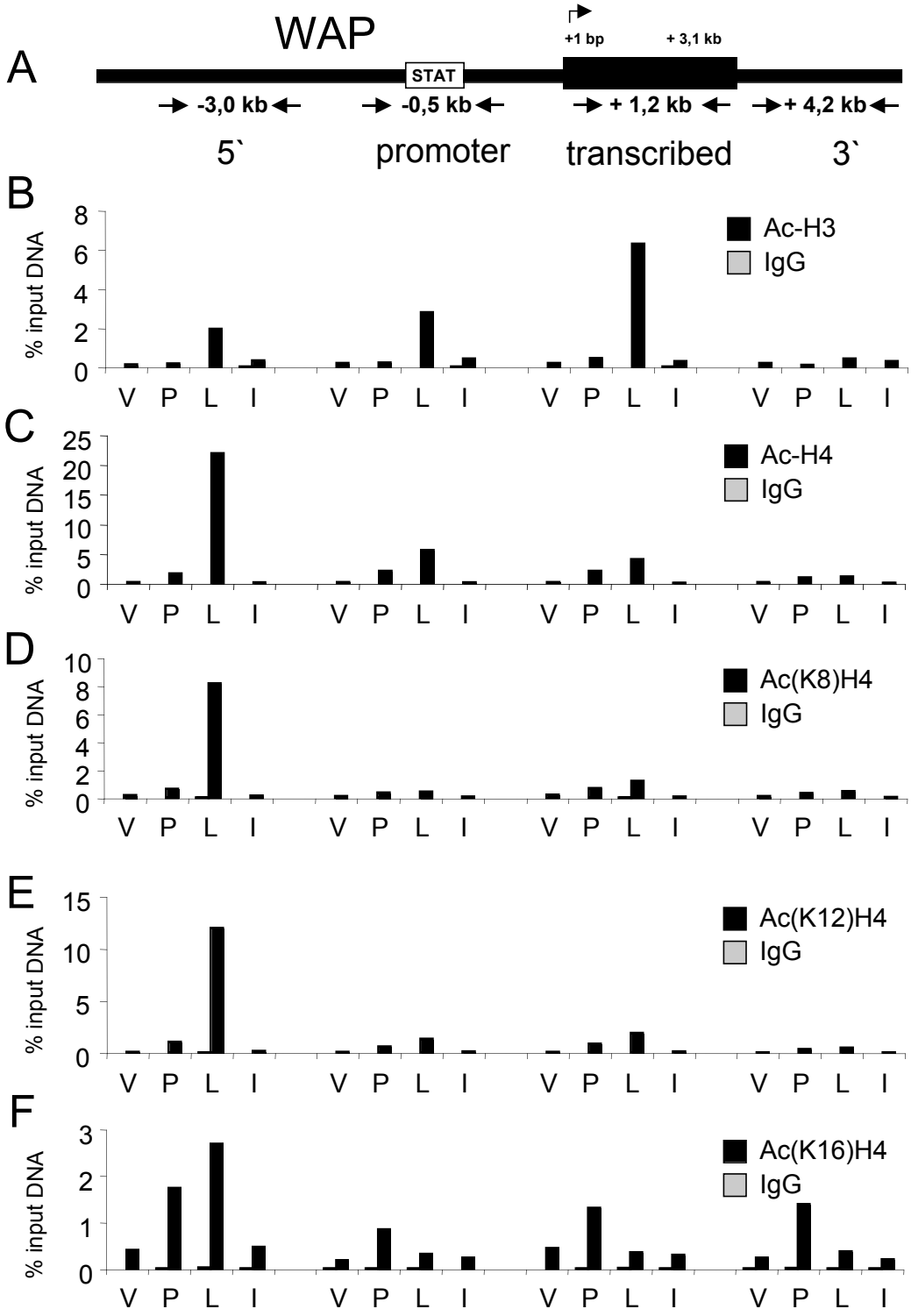


Figure 3.5 Distribution of histone lysine (K) acetylation across the endogenous WAP gene in mammary gland tissue.
(A) Schematic representation of the ß-casein gene. The relative locations of the primers used in evaluating the histone modifications in different regions of the ß-casein gene are depicted by black arrows.
(B) Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for acetylated histone H3 (Ac-H3), acetylated histone H4 (Ac-H4), acetylated lysine 8 of histone H4 (Ac(K8)H4), acetylated lysine 12 of histone H4 (Ac(K12)H4), acetylated lysine 16 of histone H4 (Ac(K16)H4) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions of the WAP gene depicted in (A).

3.1.5 Local histone acetylation of specific lysine residues occurs on the ß-casein and WAP genes during mammary gland differentiation

To determine individual lysines that are acetylated on the N-terminal tails of histone H4, antibodies specific against acetylated lysine 8 histone H4 (Ac(K8)H4), acetylated lysine 12 histone H4 (Ac(K12)H4) and acetylated lysine 16 histone H4 (Ac(K16)H4) were used in ChIP (Fig. 3.4D, E, F and 3.5D, E, F). The pattern of individual acetylated lysines on histone H4 partly mirror the pattern observed for the general acetylated histone H4 (Ac-H4), with a more pronounced enrichment of acetylated lysine 8 and 12 found specifically during lactation at the 5´ region of both the ß-casein and the WAP gene (Fig. 3.4 and 3.5). In comparison, the least enrichment for this modification was observed in the enhancer and 3' region of the ß-casein gene. This indicates that there is a localized acetylation of histone H4 lysines at the 5´ region of both the ß-casein and the WAP gene in mammary gland tissue from lactating mice.
In the ChIP assay using the antibody specific to histone H4 acetylated on lysine 16, there was a preferential enrichment found in the mammary gland tissue from pregnant and lactating mice (Fig. 3.4F and 3.5F). To note, is that there is already a general higher background level of lysine 16 acetylation observed with this antibody in tissue from virgin mice and mice having undergone involution. When taking into account that in quantitative RT-PCR a minimum of a 3 fold enrichment is required to consider a result significantly enriched (Edelstein et al., 2005), this modification appears only enriched at the 5' region during lactation.
Overall, the general enrichment of histone H3 and histone H4 acetylation in the upstream and transcribed region, together with the specific histone H4 lysine acetylation at the 5´ region of both ß-casein and WAP, indicate a open chromatin structure that is associated with transcriptional active chromatin during pregnancy and lactation.

3.1.6 Changes in histone methylation at the ß-casein and WAP genes during mammary gland differentiation

In addition to acetylation, methylation of specific lysine residues on the N-terminal tail of histones has a strong correlation with transcriptional active or inactive chromatin. The histone H3 N-terminal tail can be methylated on various lysine residues. The methylation of lysine 4 (K4) of histone H3 has been associated with transcriptional permissive and active genes. Specifically, di-methylation of (K4) histone H3 is found in

both active and inactive euchromatic genes whereas tri-methylation is found on active euchromatic genes (Santos-Rosa et al., 2002).

In contrast, methylation of lysine 9 or 27 (K9 or K27) of histone H3 or lysine 20 (K20) of histone H4 have been associated with transcriptional repressed genes (Martin and Zhang, 2005; Lachner et al., 2003; Nishioka et al., 2002; Rice and Allis, 2001).

In order to determine what role histone methylation plays in the regulation of the ß-casein and WAP genes during mammary gland differentiation, antibodies against specific methylated histone tails residues were used in ChIP assays. Antibodies specific for di-methylated histone H3 (2M(K4)H3) show an increase of di-methylated (K4) histone H3 in tissue from pregnant and lactating mice at the enhancer (for ß-casein), 5', promoter, transcribed region and 3' region of the ß-casein gene (Fig. 3.6B). Similar results were obtained for the WAP gene (Fig. 3.7B). However, the observed di-methylated (K4) H3 was less elevated during pregnancy and lactation at the 3' region when compared to the other regions of the WAP gene (Fig. 3.7B). In addition, the levels of di-methylated (K4) H3 during involution decrease again for WAP, while levels often remain elevated in the promoter and transcribed region of ß-casein gene (compare Fig. 3.6B and 3.7B). This was observed in four independent ChIP assays. Assuming that the ß-casein gene is silent at a transcriptional level during involution, this indicates that the di-methyl (K4) H3 mark may persist for considerable time after transcription has subsided. The di-methyl (K4) H3 mark has been described as a modification that may act as a post-activation signal that serves to mark chromatin as having recently been transcriptionally active (Kouskouti and Talianidis, 2005).

3.1.7 The presence of tri-methylated (K4) histone H3 correlates with the active transcriptional state of ß-casein and WAP gene during pregnancy and lactation

Antibodies specific for tri-methylated histone H3 (3M(K4)H3) showed enrichment of tri-methylated (K4) histone H3 in tissue from pregnant and in lactating mice for both the ß-casein and WAP genes (Fig. 3.6C and 3.7C). The enrichment of tri-methylated (K4) histone H3 was observed in the transcribed region of both the ß-casein and WAP genes and often, but not always in the promoter region of the ß-casein gene. This mark, associated with active euchromatic genes, correlates well with the active state of the ß-casein and the WAP gene in mammary gland tissue from pregnant and lactating mice.

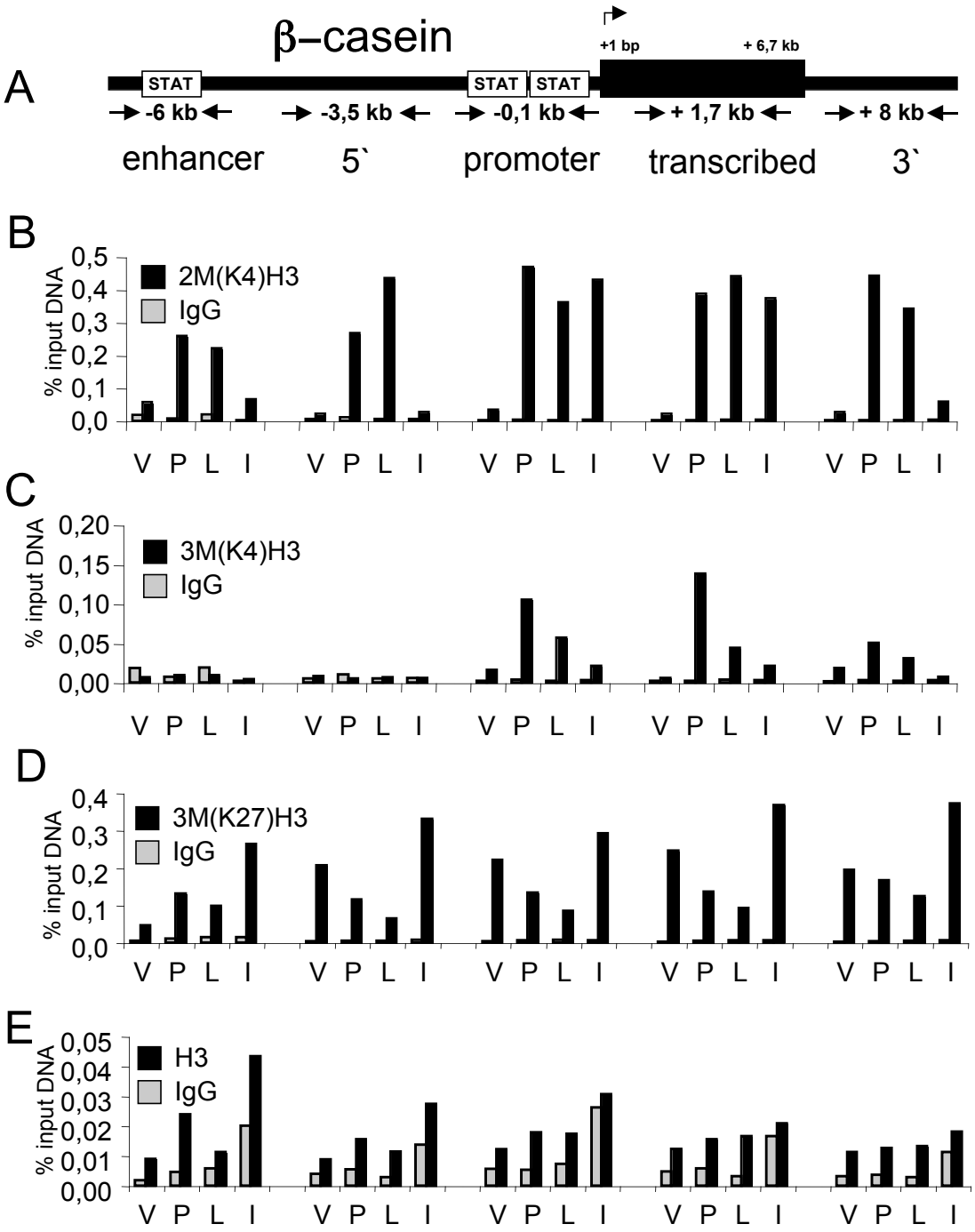


Figure 3.6 Distribution of histone H3 lysine methylation across the endogenous ß-casein gene in mammary gland tissue.
(A) Schematic representation of the ß-casein gene. The relative locations of the primers used in evaluating the histone modifications in different regions of the ß-casein gene are depicted by black arrows.
(B) Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for di-methylated (K4) histone H3 (2M(K4) H3), tri-methylated (K4) histone H3 (3M(K4) H3), tri-methylated (K27) histone H3 (3M(K27) H3), unmodified N-terminal histone H3 (H3) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions of the ß-casein gene depicted in (A).

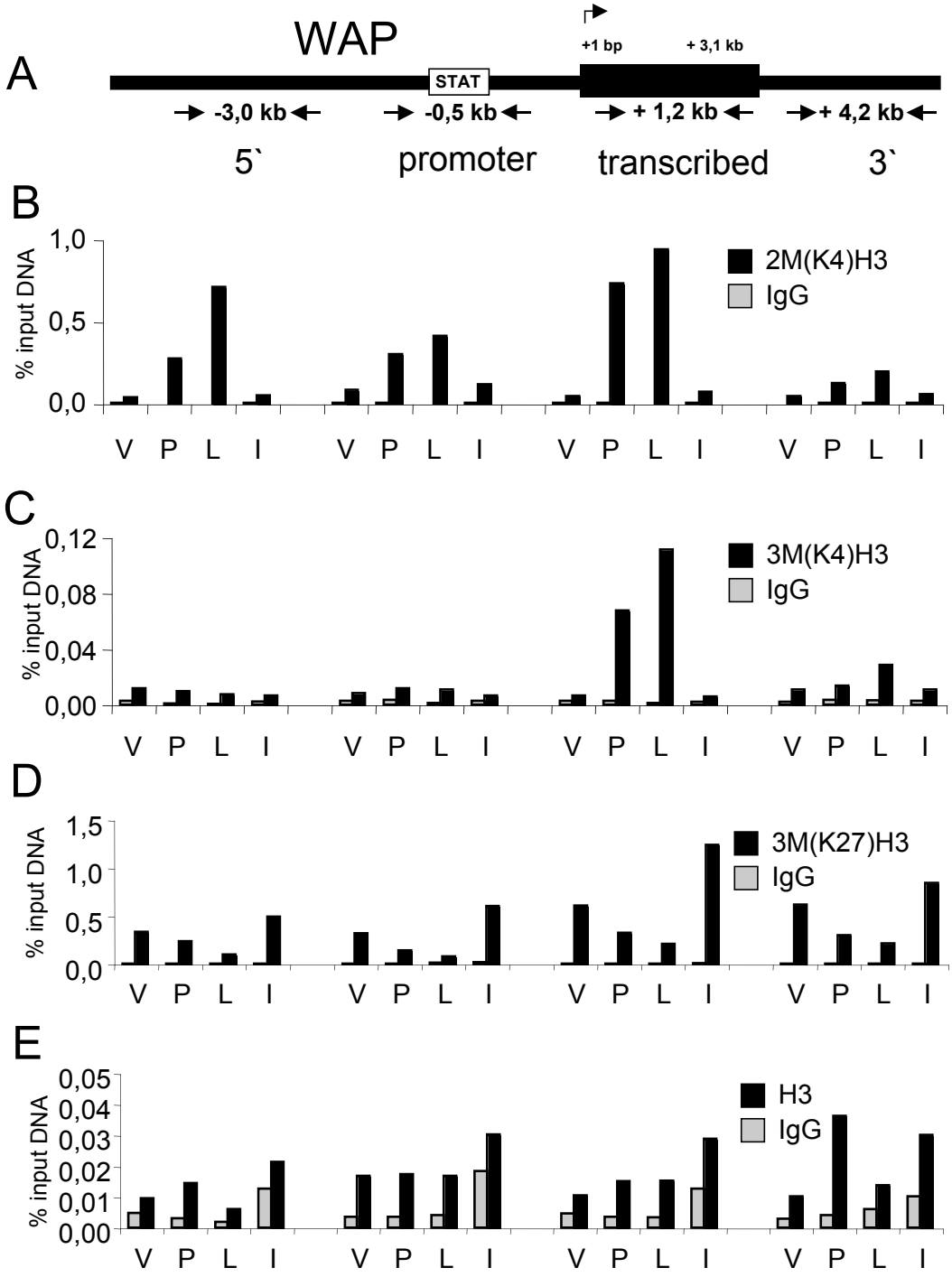


Figure 3.7 Distribution of histone H3 lysine methylation across the endogenous WAP gene in mammary gland tissue.
(A) Schematic representation of the WAP gene. The relative locations of the primers used in evaluating the histone modifications in different regions of the WAP gene are depicted by black arrows.
(B) Mammary gland tissues were harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for di-methylated (K4) histone H3 (2M(K4) H3), tri-methylated (K4) histone H3 (3M(K4) H3), tri-methylated (K27) histone H3 (3M(K27) H3), unmodified N-terminal histone H3 (H3) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions of the WAP gene depicted in (A).

3.1.8 Tri-methylated lysine 27 of histone H3 is associated with the ß-casein and the WAP gene in mammary gland tissue

Apart from the histone methylation of lysines that correlates with chromatin that is transcriptional active, there are three histone modifications, namely methylation of lysine 9 and 27 of histone H3 and methylated lysine 20 of histone H4, that have been described to be associated with transcriptional repressed chromatin.
To examine if histone modifications that correlates with transcriptional repressed chromatin, may also be associated with the target genes in a differentiation specific manner, ChIP assays were performed with antibodies specific to tri-methylated (K9)H3 or tri-methylated (K20)H4. The results did not show significant enrichment of tri-methylated (K9)H3 or tri-methylated (K20)H4 at either genes in the different stages of mammary gland tissue investigated (data not shown). A faint enrichment was observed in ChIP assays using an polyclonal antibody recognizing di-methylated (K9)H3 at the ß-casein promoter (data not shown). It is therefore possible that some di-methylation of lysine 9 on histone H3 is present. In comparison the ChIP analysis performed with the antibody against tri-methylated (K27) H3 show significant enrichment over control IgG (Fig. 3.6D and 3.7D). This modification could be detected at all stages of mammary gland differentiation in both ß-casein and WAP genes and it was not restricted to a specific region. Slightly elevated levels of tri-methylated lysine 27 on histone H3 were found in tissue from virgin and involuting mice mainly in the transcribed region of ß-casein in two independent ChIP assays. This may correlate with the expected transcriptional silence of the two milk proteins in mammary gland tissue from virgin and involuting mice.

3.1.9 Histone H3 abundance

To gain information on the general abundance of histone H3 in the different regions investigated, a ChIP analysis was performed using an antibody against the core histone H3. ChIP assays performed with this antibody gave only low background levels close to the control IgG (Fig 3.6E and 3.7E). Since this antibody detects the unmodified N-terminal tail of Histone H3, these results imply that in the regions investigated the majority of histone H3 N-terminal tails might be modified and therefore may not be recognized by this antibody. Additional, ChIP assays were performed with an antibody recognizing the C-terminal domain of histone H3. Contrary to the above antibody, the results gained with this antibody showed an enrichment in tissue from lactating or some times pregnant mice (data not shown). When following the general notion that

within transcribed genes there is a displacement or loss of histones (Morillon et al., 2006), these results are contrary to what was expected. It is not clear either of these results reflect the actual abundance of histone H3 in the mammary gland at the regions investigated. If the latter antibody possibly recognizes a specific variant of the C-terminal domain of histone H3 that is prevalent during lactation, the data may not reflect the actual abundance of histone H3.

3.1.10 Histone modifications and transcription factor binding at STAT5 responsive genes in mammary epithelial cell lines

To ensure that the histone modifications observed in the mammary gland tissue arise from chromatin modifications of epithelial cells and are not partly due to other cell types found in the tissue, histone modifications were analyzed in a pure mammary epithelial cell line. The HC11 cell line can be used as a model system to study mammary epithelial cell differentiation *in vitro* (Ball et al., 1988). Confluent grown HC11 cells can be differentiated by removal of EGF and addition of the lactogenic hormones prolactin, insulin and glucocorticoids leading to the expression ß-casein (Welte et al., 1994). Because ß-casein transcription has been shown to be dependent on the synergistic action of both STAT5 and the glucocorticoid receptor (Wyszomierski et al., 1999; Lechner et al., 1997; Stoecklin et al., 1997; Doppler et al., 1989), a three day pretreatment of HC11 cells with the glucocorticoid hormone analog, dexamethasone was performed. The capability of the mammary epithelial cell line HC11 to express ß-casein in response to lactogenic hormone stimulation after a three day dexamethasone pretreatment was verified by using real-time PCR.
The values were normalized to the abundance of 18S ribosomal RNA. As a control the expression of another STAT5 target gene namely SOCS3, which act as a negative feed back regulator of STAT5 was monitored as well (Tonko-Geymayer et al., 2002).
Induction of the ß-casein mRNA was first detected 0.5 hour after stimulation with lactogenic hormones. A gradual rise in the induction of ß-casein mRNA could be observed over the next 4 hours of lactogenic hormone treatment (Fig. 3.8A left panel). In contrast to ß-casein, the SOCS3 induction is transient in HC11 cells, with high mRNA levels reached already after 0.5 hours of induction (Fig. 3.8A, right panel). This is in accordance to what has been described in others studies (Tonko-Geymayer et al., 2002).

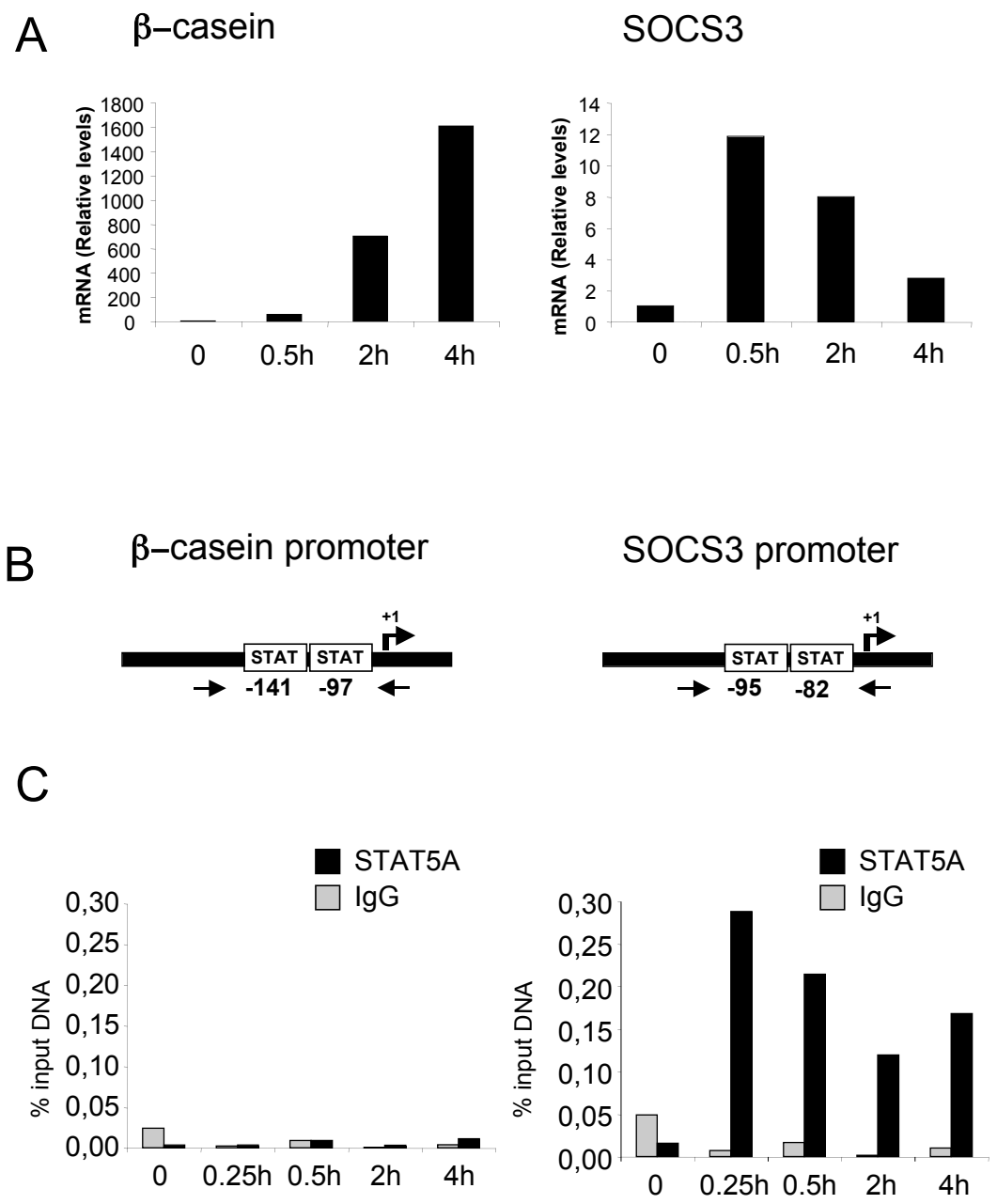


Figure 3.8 Confirmation of ß-casein and SOCS3 mRNA expression and recruitment of STAT5 to the target gene promoters in mammary epithelial cells .
(A) RNA samples were prepared from HC11 cells that were pretreated with dexamethasone for 3 days and then induced with lactogenic hormones PRL (5 µg/ml), insulin (5 µg/ml) and dexamethasone (0,1 µM) for the different hour (h) time points indicated. The expression levels of ß-casein and SOCS3 transcripts relative to 18S ribosomal RNA were determined by real-time PCR.
(B) Schematic structure of the murine ß-casein and SOCS3 promoter. Square boxes represent two STAT binding sites, together with their bp positions relative to the transcription initiation site (black +1 arrow). The relative locations of the primers covering the STAT5 recognition elements are marked by black arrows.
(C) After 3 day dexamethasone pretreatment HC11 cells were treated with lactogenic hormones PRL (5 µg/ml), insulin (5 µg/ml) and dexamethasone (0,1 µM) for the different time points indicated. Cells were harvested and analyzed by ChIP with an antibody specific for STAT5A and unrelated IgG (IgG). Antibody-coprecipitated DNA samples were analyzed by real-time PCR with primers flanking the region of the STAT5 recognition elements of the specified promoters depicted in (B).

Next the binding of STAT5A to the endogenous ß-casein and the SOCS3 promoter was analyzed. For ChIP analysis primers were used that flank the two STAT5 response elements in the ß-casein and SOCS3 promoter (Fig. 3.8B).
ChIP assays were performed with differentiation competent HC11 cells pre-treated with dexamethasone for 3 days and then stimulated with lactogenic hormones over various time points. No significant binding of STAT5A to the ß-casein promoter at any time point of induction was observed, although STAT5A was found to be inducible and specifically bound to the endogenous SOCS3 promoter (Fig. 3.8C). This data indicates, that the ß-casein promoter in the mammary epithelial cell line HC11 might not be accessible for STAT5 binding and even induction of differentiation by the lactogenic hormones may not be efficient enough to open the chromatin structure. This is in contrast to the results observed for the SOCS3 promoter, where activation of STAT5 by prolactin induction is sufficient to recruit STAT5A to the promoter.

3.1.11 Tri-methylated lysine 27 of histone H3 is also associated with the ß-casein gene promoter in the mammary epithelial cell line HC11

To determine if the tri-methylation of lysine 27 on histone H3 observed in the mammary gland tissue is part of the repressed chromatin state of the ß-casein promoter in the HC11 mammary epithelial cell line, ChIP analysis was performed with the antibody against 3M(K27)H3. ChIP assays were performed with differentiation competent HC11 cells stimulated with lactogenic hormones over prolonged time points. For this assay the three day dexamethasone pretreatment was omitted, since it has been described to have a slight repressive effect on SOCS3 mRNA expression (Tonko-Geymayer et al., 2002). A significant association of 3M(K27) histone H3 could be observed on the ß-casein promoter of the HC11 cells (Fig. 3.9A, left panel). No significant change in the level of lysine 27 tri-methylation was detected on the ß-casein promoter region even after prolonged induction of up to 68 hours with lactogenic hormones. In comparison to the ß-casein promoter only an insignificant association of histone H3 tri-methylated lysine 27 to the SOCS3 promoter was observed at all time points of induction (Fig. 3.9A, right panel). The high levels of histone H3 lysine 27 tri-methylation found at the ß-casein promoter indicates that there are a significant number of HC11 cells that display this histone mark. This is the same mark that has also been observed in the mammary gland tissue and is considered a mark to be associated with repressed chromatin (Fig. 3.6D). In comparison, the SOCS3 promoter was not found to be in a repressed state considering the levels of tri-methylated (K27) histone H3.

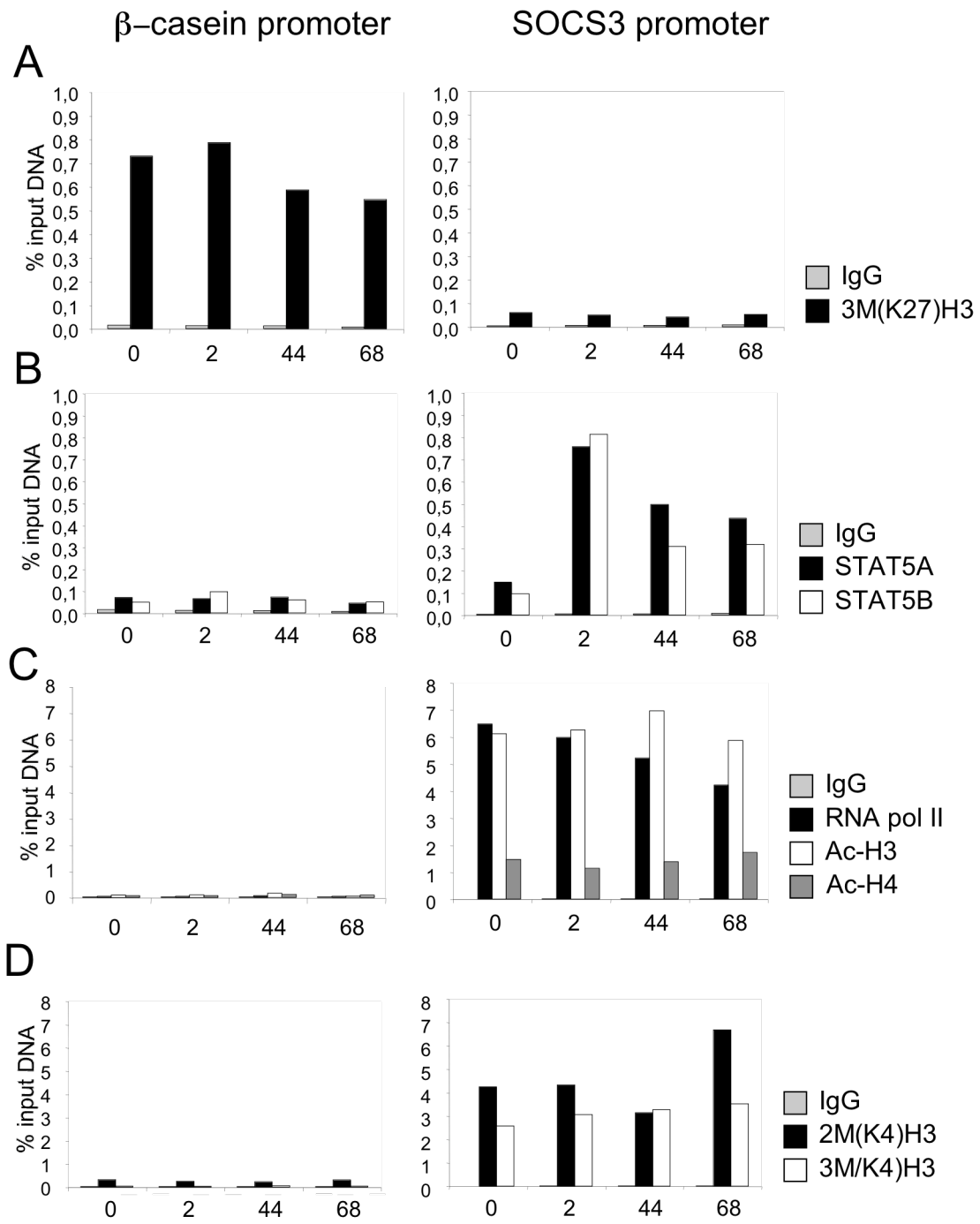


Figure 3.9 Histone acetylation, methylation and recruitment of STAT5 and RNA polymerase to the ß-casein and SOCS3 promoters in mammary epithelial cells.
HC11 cells were treated with lactogenic hormones PRL (5 µg/ml), insulin (5 µg/ml) and dexamethasone (0,1 µM) for the different hour time points indicated. Cells were harvested and analyzed by ChIP with antibodies specific for tri-methylated lysine 27 of histone H3 (3M(K27)H3), STAT5A and STAT5B, phosphorylated CTD RNA polymerase II (RNA pol II), acetylated histone H3 (Ac-H3) and H4 (Ac-H4), di-methylated lysine 4 of histone H3 (2M(K4)H3), tri-methylated lysine 4 of histone H3 (3M(K4)H3), and unrelated IgG (IgG). Antibody-coprecipitated DNA samples were analyzed by real-time PCR with primers flanking the region of the STAT5 recognition elements of the specified promoters.

3.1.12 Both STAT5A and STAT5B are recruited to the SOCS3 promoter of differentiation competent HC11 cells

To determine if not only STAT5A, but also STAT5B can be recruited to the endogenous ß-casein and SOCS3 promoters, antibodies specific for STAT5A and STAT5B were used in the ChIP.
Both STAT5A and STAT5B were found to be inducible and specifically bound to the endogenous SOCS3 promoter in response to prolactin and dexamethasone treatment in HC11 cells (Fig. 3.9B, right panel). In comparison, no significant recruitment of either STAT5A or STAT5B was observed for the STAT5 RE of the ß-casein distal enhancer (data not shown) and promoter (Fig. 3.9B, left panel). This remained so even after lactogenic hormone treatment for up to 68 hours by which time ß-casein is detected by RT-PCR (data not shown). These results are in accordance with the former data (Fig. 3.8C).

3.1.13 RNA polymerase II and histone modifications are found at the SOCS3 promoter, but not at the ß-casein promoter

Additional, ChIP assays were performed to clarify if RNA polymerase II or any histone modifications that are commonly associated with transcriptional active genes could be detected at the ß-casein or SOCS3 promoter. ChIP assays performed with the p-CTD RNA polymerase II antibody showed association of RNA polymerase II to the SOCS3 promoter during all time points, while no recruitment of RNA polymerase II was observed on the ß-casein promoter (Fig. 3.9C). To note, is that a significant level of p-CTD RNA polymerase II is already associated with the SOCS3 promoter in unstimulated cells (Fig. 3.9C, left panel). This suggests that RNA polymerse II has already been recruited to the SOCS3 promoter independent of prolactin induction and that the SOCS3 gene may already be transcriptional active at a basal level (Fig. 3.9C).
ChIP assays performed with anti-acetylated Histone H3 antibody showed elevated levels of acetylated histone H3 at the SOCS3 promoter during all time points of induction, yet no significant levels of acetylated histone H3 were observed on the ß-casein promoter (Fig. 3.9C). While for acetylated H4 the levels remained overall low (Fig. 3.9C). Since increased histone acetylation is associated with open and transcriptional active chromatin, the elevated histone H3 acetylation suggests an open chromatin structure for the SOCS3 promoter.
ChIP analysis with 2M(K4)H3 and 3M(K4)H3 antibodies show enrichment of 2M(K4)H3 and 3M(K4)H3 at the SOCS3 promoter during all time points, while hardly any

significant levels of these histone modifications were observed on the ß-casein promoter (Fig. 3.9D). The 3M(K4)H3 methylation found in active euchromatic genes correlates well with the transcriptional active SOCS3 gene, while for ß-casein this histone modification was not observed at any significant elevated levels.
Overall, recruitment of the transcription factor STAT5 and the histone modifications that correlate with transcriptional active genes could be observed at the SOCS3 promoter, but at a comparative level not at the ß-casein promoter in HC11 cells stimulated with lactogenic hormones.

3.2 Identification of novel STAT5A-binding sites in mammary gland tissue from a lactating mouse using ChIP

The goal of this part of the study was to use mammary gland tissue from a lactating mouse to identify new differentiation specific STAT5 targets. Mammary epithelial cell differentiation has come to full terms during lactation. STAT5 is expressed during all stages of mammary gland development and is found at high levels in active phosphorylated form during early lactation (day 5) in mammary gland tissue (Liu et al., 1996). During this time the two milk proteins and STAT5 target genes ß-casein and WAP, which are considered markers for mammary epithelial cell differentiation, are expressed. Differentiation of mammary gland is regulated by transcription factor STAT5, as knockout mice show defects in the development of epithelial mammary cells. It has been demonstrated that STAT5 is essential for the proliferation, differentiation and survival of mammary epithelium during pregnancy and lactation using transgenic mice in which the STAT5 gene could be conditionally inactivated during different stages of mammary gland development (Cui et al., 2004). Thus, lactating mammary gland tissue is a good model system to identify new differentiation specific STAT5 targets.

3.2.1 Verification of STAT5A binding to known differentiation specific STAT5 target genes

STAT5 binding sites are known and defined in the ß-casein and WAP genes (Li and Rosen, 1994; Happ and Groner, 1993). The TTC(N3)GAA) consensus sequence is characterized as a STAT5 binding site (Soldaini et al., 2000). In order to determine that the approach to identifying differentiation specific STAT5 targets is valid, the binding of both STAT5A and STAT5B to the ß-casein and WAP promoters had to be confirmed.

Primers were designed to amplify the STAT5 regulatory regions and used in a ChIP assay. Mammary gland tissue was taken from the 4 different developmental stages of the mammary gland, namely from virgin, pregnant, lactating and involution Balb/c mouse tissue. The tissue was cross-linked with formaldehyde and the chromatin was harvested. This was followed by an immunoprecipitation using antibodies specific against STAT5A and STAT5B. As a nonspecific control polyclonal rabbit IgG was used. Both STAT5A and STAT5B bind to the ß-casein and WAP promoters during pregnancy and lactation, but not in mammary gland tissue from virgin and involuting mice (Fig. 3.10A and B).

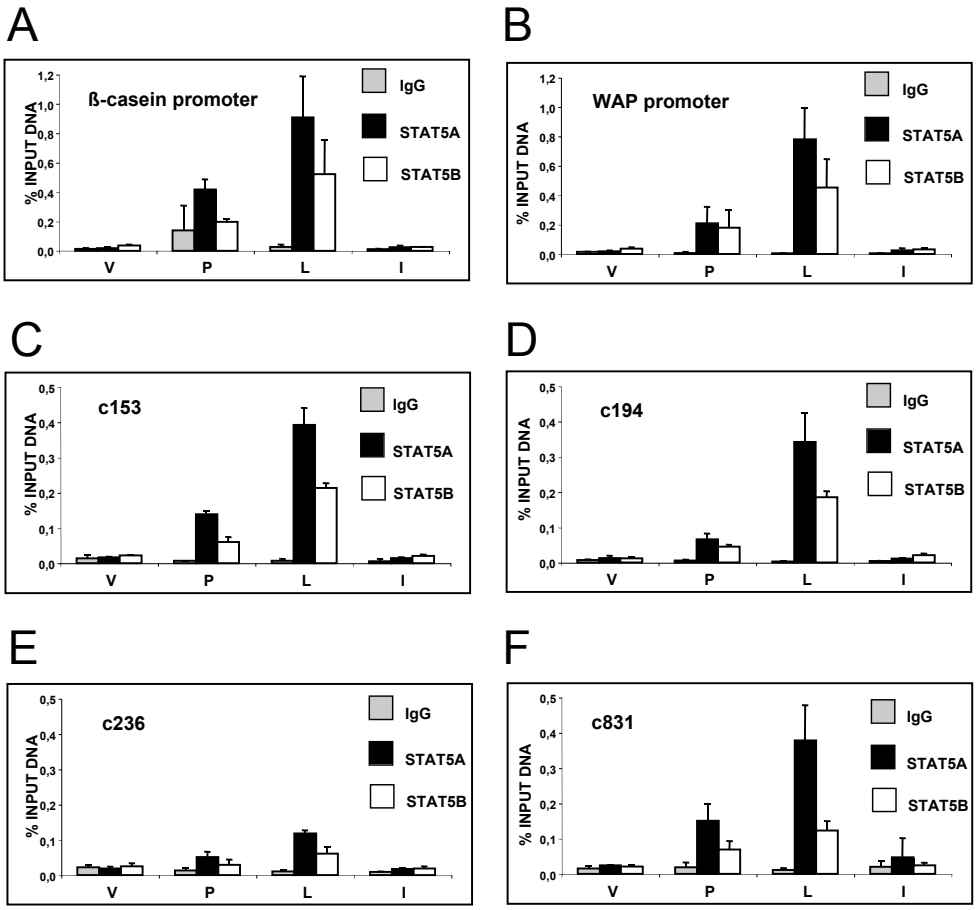


Figure 3.10 Recruitment of STAT5A and STAT5B to the novel STAT5 targets in mammary gland tissue.
Mammary gland tissue was harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for STAT5A and STAT5B and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions flanking the STAT5-binding sites of the two milk protein promoters (ß-casein and WAP) and the novel STAT5 motif-containing clones (c153, c194, c236, c831). The quantitative real-time PCR was performed in triplicates. The immunoprecipitation data was normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA. The data depicted are representative of the trend observed in at least two independent ChIP assays.

No ß-casein or WAP promoter was detected when the non-specific rabbit IgG was used for immunoprecipitation, showing that the immunoprecipitation is specific for

STAT5A and STAT5B. More importantly, this shows that STAT5A and STAT5B bind the promoters specifically during the developmental stages that are associated with mammary epithelial cell differentiation. Therefore, it should be feasible to isolate new differentiation specific STAT5 targets using mammary gland tissue from a lactating mouse.

3.2.2 Approach used for ChIP based cloning and the identification of novel STAT5-binding sites in cloned DNA regions

The Chromatin immunoprecipitation is the ideal method to obtain data on the DNA binding activity of transcription factors *in vivo*. The ChIP method has the potential for the isolation of all the DNA regions that are associated with the transcription factor at the time of cross-linking. The immunoprecipitatied ChIP products can be ligated into a vector, which allows for screening and selection of clones containing appropriate sized inserts. These inserted DNA regions can then be sequenced and their location identified. Therefore, the ChIP assay was used to isolate novel STAT5 binding targets from mouse mammary gland tissue of a lactating mouse. A mixture of computational and experimental approaches were used to analyze the novel STAT5 motif containing sequences obtained from this STAT5A library (Fig. 3.11). The genomic location of these sequences in the mouse genome, as well as the transcription factor (STAT5) recognition motifs within these regions were identified using a computational approach. The *in vivo* binding of STAT5 to the novel STAT5A target sites was re-confirmed experimentally by ChIP. The comparative genomic analysis, whereby the mouse STAT5 containing sequences are aligned to other vertebrate species, provided further information to the degree of evolutionary conservation of confirmed STAT5 motif containing sequences. Additional experimental approaches such as reporter gene assays were used to provide insight into the functionality of the confirmed STAT5 sites. RNA transcripts located in proximity were also analyzed for potential differentiation specific regulation that parallels STAT5 binding during mammary gland development.
The analysis for select novel STAT5 motif containing clones from this ChIP created STAT5A library are presented in the subsequent sections.

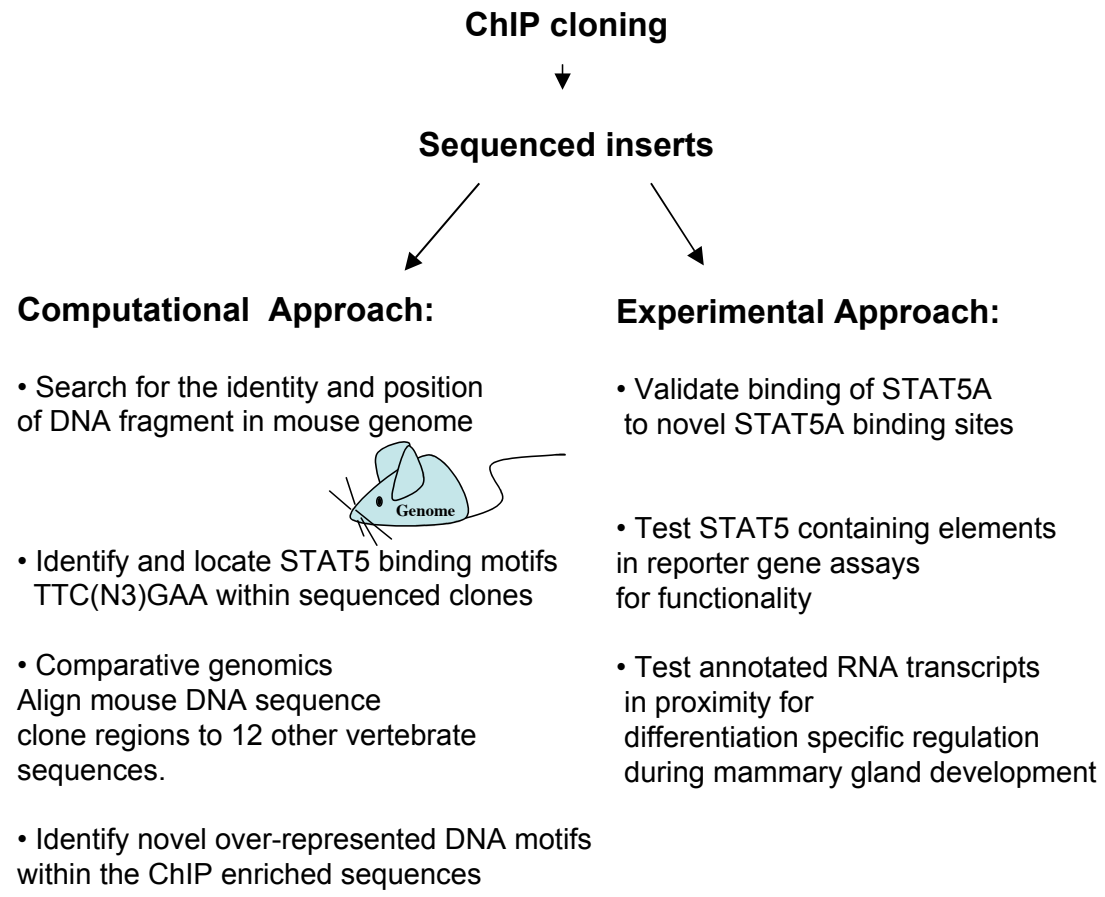


Figure 3.11 Computational and experimental approaches used to analyze the DNA sequences of the STAT5A ChIP-enriched library clones. The inserted DNA fragments from the STAT5A ChIP-enriched library clones are sequenced. To gain more information on the individual clones they are analyzed both by computational and experimental approaches.

The enriched, STAT5-bound chromatin obtained from mammary gland tissue from a lactating mouse was cloned into a bacterial vector to produce a STAT5A chromatin library. Bacteria were transformed and blue/white screening performed to select white colonies containing plasmids with inserts. 967 white colonies were picked and plasmid minipreps were made and restriction digests performed to select only those plasmids containing inserts larger than 400 bp. This was done according to the ChIP cloning protocol from the Peggy Farnham lab, because clones with larger insert size are less likely to contain non-specific DNA. Approximately 350 clones with inserts of more than 400 bp were sequenced. Any clones that failed in sequencing were not resubmitted.

As the first cloned insert sequences became available these were aligned against the mouse genome using the NCBI BLAST program to determine their location within the

genome. In addition, the available sequence of each clone was searched manually for STAT5 and STAT recognition motifs. STAT5, like other members of the STAT family, can bind to the broad consensus sequence TT(N5)AA with the optimal sequence being TTC(N3)GAA for STAT5 (Ehret et al., 2001; Soldaini et al., 2000).

The STAT5, STAT and STAT-like sites were defined as follows; STAT5: TTC(N3)GAA, STAT: TT(N5)AA and STAT-like: (TT(N1-2,4,6)AA). Priority was given to those clones with one or more classic STAT5 sites and fragments with STAT sites nearest to known genes as determined by the NCBI BLAST search. 18 clones were originally selected, that contained STAT5 or STAT-like sites identified in the sequenced portions of the clones (Table 2). Primers were designed to these STAT5 sites for 11 clones (c76, c90, c98, c153, c177, c192, c194, c236, c376, c831 and c878) and for 7 clones with STAT/ STAT-like sites (c180, c234, c240, c241, c372, c836 and c854) (Table 2).

Of these initial 18 clones selected, only 4 (c153, c194, c236 and c831) that contained STAT5 sites were successfully immunoprecipitated in ChIP (See next section 3.2.3). Since the clones had been sequenced just with reverse primers, additional insert sequences, which might contain potential STAT5 sites would have been missed. Therefore, once the sequencing of all submitted clones was completed (yielding sequences for a final 294 clones) the search for STAT5-binding sites within the cloned regions was repeated, but this time extending the region searched an additional 1000 bp upstream and downstream of the original cloned fragments. This extension of the clone regions was done by aligning the individual clone sequences to the corresponding region found in the mouse genome using an automated BLAST search set to a minimum sequence identity of 95% (performed by Dr. Bulashevska from the DKFZ, Heidelberg). The BLAST aligned location points in the mouse genome were extended by 1000 bp upstream and downstream for each individual clone sequence alignment. These extended sequence were then extracted, thus created a new data set for analysis. This data set containing all the extended regions was searched for STAT5 binding sites using the MatInspector Bioinformatics software from Genomatrix (performed by Dr. Bulashevska from the DKFZ, Heidelberg). The STAT5 binding site motif was stringently defined by TTC(N3)GAA. Preliminary analysis of the Bioinformatics output from MatInspector showed that of the extracted clone regions searched approximately 25% had STAT5 binding motifs. Some of these clones could be found multiple times at different locations in the genome and were thus excluded from further analysis, since they appeared mostly not to be unique and often composed of DNA with low complexity. Some of these clones belong to repetitive sequence elements for example c863 aligns to a long interspersed nuclear element (LINE). Designing useful ChIP primers for these regions containing repetitive sequences is therefore challenging.

Table 2 Location of select clones containing STAT and STAT-like sequence motifs relative to known genes.
List displaying the initial 18 clones selected for verification of STAT5 binding by ChIP assay. The location of the clones is indicated in relation to the nearest genes as determined by NCBI BLAST search within the mouse genome. The sequence of the individual STAT5, STAT and STAT-like motifs found for each clone are shown.

Clone	STAT5 motifs	Nearest gene(s)	location	Chr.	STAT5 motif sequences		STAT like motif
153	2 STAT	Hypothetial protein LOC76455	5' -110kb	4	TTCtatGAA	AAGgctCTT	
		Zinc finger, DHHC domain containing 22	3'-238kb				
194	2 STAT5	Hypothetical protein LOC320478	5' -493kb	3	TTCttgGAA	TTCaaaGAA	
		Hypothetical protein XP_61988	3'-404kb				
236	2 STAT5	Hypothetical protein XP_905854	5' -58kb	5	AAGgggCTT	TTCcaaGAA	
		Hypothetical protein LOC75977	3'-1kb				
831	1 STAT5	Zinc finger, SWIM domain containing 6	in gene	13	TTCcatGAA		
76	1 STAT5	Platelet-derived growth factor	5'-105kb	3	TTCtcgGAA		
		RIKEN cDNA A830029E22 gene	3'-481kb				
90	2 STAT5	Branched chain ketoacid dehydrogenase E1	5'-66kb	9	AAGtggCTT	TTCcctGAA	TTTaaaGAA
	1 STAT like	Similar to nucleolar protein GU2	3'-783kb				
98	1 STAT5	Somatostatin receptor 4	5'-2kb	2	TTCaaaGAA		
		Thrombomodulin	3'-6kb				
177	1 STAT5	MEGF10 protein		18	AAGtgtCTT		
180	2 STAT like	Chondoitin sulfate proteoglycan 4	5'-1,6kb	9			AAGtcattCTT
		Hypothetical protein LOC235406	3'-6kb				AAGcagcCTT
192	1 STAT5	RAP quanine nucleotide exchange factor 2	in gene	3	TTCatgGAA		AAGcctTTT
	2 STAT like						TTCcaggGAA
234	2 STAT like	Similar to core histone macro-H2A.2	5'-235kb	8			TTTtagGAA
		Similar to retrovirus-related Pol Polyprotein LINE-1	3'-542kb				TTTgttGAA
240	3 STAT like	Histamine H4 receptor	in gene	18			AAGgctcATT
							TTCtatgGAA
							TTTtctGAA
241	3 STAT like	Trans-acting transcription factor 4 (SP4)	in gene	12			AAGtgTTT
							AAGcatcATT
							AAGgCTT
372	2 STAT like	START domain containing 10	in gene	7			AAggtcTTT
							AAGatcTTT
376	1 STAT5	Rho GTPase activating protein 28	in gene	17	AAGtgaCTT		
836	1 STAT like	Similar to cat ye syndrom critical region protein 2	in gene	6			AAGtcCTTCctGAA
854	1 STAT like	Similar to phosphoribosyl pyrophosphate synthetas	5'-42kb	2			AAGagaaCTT
		Similar to proteosome subunit beta type 3	3'-271bp				
878	2 STAT	Hypothetical protein LOC71947	in gene	11	AAGaagCTT	TTCccaGAA	

To note is that most of the initial STAT-like motifs containing clones, found in the original 18 clone selection list (Table 2) turned up again in this MatInspector based STAT5 motif search with additional new STAT5 recognition motifs in the surrounding 1000 bp. As in the case of c241, which only had STAT-like motifs in the original clone sequence, a new STAT5 site was found located approximately 500 bp away from the site the original primers amplified. This may explain why it did not enrich significantly in the reconfirmation IPs, when taking into account that the average sheared fragment size is about 500 bp.

From the extended data set searched by the MatInspector program for STAT5 binding sites, 7 new candidate clones containing STAT5 sites were selected and tested in ChIP assays for STAT5 binding. The selection of these additional new clones was based on having at least two or more STAT5 motifs preferentially in close proximity to each other and located near known genes (Table 3, lists all confirmed clones including the confirmed clones identified from MatInspector extended data set, c372 and c721 not shown).

Table 3 Location of the 9 confirmed clones relative to known genes.
The STAT5-binding sites were analyzed using the UCSC genome Bioinformatics browser BLAT search (http://www.genome.ucsc.edu/cgi-bin/mmBLAT) for the nearest gene. The locations of the clones in relation to known genes or RNA transcripts were determined by using the accompanying EMBL transcript identifications displayed in the browser.

Clone	Nearest gene	Accession number	STAT5 motif location	Ensembl transcript ID
c20	24b2/STAC2 (SH3 and cysteine rich domain 2)	AJ608761/ NM_146028	about 11,6 kb 5'	ENSMUST00000043983
c34	Arhgef3 (Rho guanine nucleotide exchange factor (GEF) 3)	NM_027871	within Intron 2-3	ENSMUST00000049206
c73	Inpp4b (Inositol polyphosphate-4-phosphatase, type II)	NM_001024617	within last Intron 22-23	ENSMUST00000042529
c153	Nfib-010, EST (Nuclear factor I/B)	Aw106080	Intron 1-2 of EST	OTTMUST00000021179 (vega transcript ID)
c194	Unknown EST	BE373654	about 57 kb 3' of EST	NA
c236	Unknown clone, EST EST EST	BC050254/ BU610001 AK017143 AK042126	Intron 1-2 of EST Intron1-2 Intron 3-4 of EST	ENSMUST00003833694 ENSMUSESTG00003735796 ENSMUSESTG00003735796
c398	Rbm27 (RNA-binding motif protein 27)	AY461716	5' upstream + Intron1-2	ENSMUST00000091920
c831	Zswim6 (Zinc finger, SWIM domain containing 6)	AK122528	Intron1-2 (101 kb)	ENSMUST00000052377
c882	Igsf3 (Immunoglobulin superfamily, member 3)	NM_207205	Intron1-2	ENSMUST00000043983

These clones had initially not been identified as STAT5 containing sites, since the STAT5 sequences are located outside the original sequenced clone fragments. This is exemplified when comparing the nucleotide positions of the sequenced clone fragments for clones c20, c34, c73, c398 and c882 to the position of the annotated STAT5 sites identified in vicinity by MatInspector using the extended sequence dataset (Table 4, c372 and c721 not shown).

Overall, searching the immediate surrounding sequence at the locations of the original clone fragments by Bioinformatic means, appears to be an alternative to the costly sequencing of the insert sequences from both directions.

Table 4 Nucleotide and chromosome positions of the STAT5-binding sites for the 9 clones confirmed as STAT5 binding targets.
The 9 clones and the sequences of the STAT5 binding sites are listed with their nucleotide and chromosome positions. The sequences of the listed clones including +/- 1000 bp of surrounding sequence were searched for STAT5-binding motifs as defined by TTC(N3)GAA using the transcription factor binding site search program Matinspector from Genomatrix (Version 6.2, Okt 2006). A second STAT5 motif, which conforms to the defined STAT5 motif sequence was not identified by the Matinspector™ within clone c236 and is therefore not listed.

clone	Clone chromosome:position	STAT5 sequence (Bold)	STAT5 Chromosome:position
c20	chr11:97881047-97881402	cttg**ttcaaggaa**agctgg	chr11:97881577-97881585
		ttag**ttcttagaa**ttttta	chr11:97881373-97881681
c34	chr14:26177070-26177494	catt**ttctgggaa**gctaag	chr14:26177047-26177055
		caag**ttccaagaa**gtcagt	chr14:26178089-26178079
		ctct**ttctctgaa**ggtctg	chr14:26178512-26178523
c73	chr8:85009850-85010306	agat**ttccaggaa**cacttt	chr8:85008967-85008975
		agag**ttcaaagaa**tctctc	chr8:85009522-850095300
		gggc**ttctgagaa**agcctt	chr8:85009781-85009789
		taag**ttctctgaa**ctcaca	chr8:85009883-85009891
		tgca**ttctgtgaa**acatca	chr8:85010470-85010478
c153	chr4:82038357-82038644	aggt**ttctatgaa**aagccg	chr4:82038470-82038478
c194	chr3:35362366-35362759	tcag**ttcttggaa**acagag	chr3:35362472-35362480
		cata**ttcaaagaa**cctgct	chr3:35362693-35362701
c236	chr5:22914227-22914652	tcta**ttcttggaa**aagaga	chr5:22914295-22914304
c398	chr18:42401306-42401715	gagt**ttccaggaa**aaaaac	chr18:42400981-42400989
		ggag**ttcactgaa**ctctcc	chr18:42401157-42401165
		ccgg**ttctttgaa**gagagc	chr18:42402641-42402649
c831	chr13:108947025-108947483	gagg**ttcatggaa**ctggcg	chr13:108947329-108947337
c882	chr3:101536707-101537020	catc**ttcagagaa**ggcaga	chr3:101537632-101537640
		acag**ttcttggaa**ttaaca	chr3:101537904-101537912

3.2.3 Confirmation of STAT5A and STAT5B binding to novel target sequences

To make sure that the STAT binding sites identified are real STAT5 targets, primers that flank these putative binding sites were designed and tested in ChIP assays during the different stages of mouse mammary gland development.

ChIP assays were performed with antibodies against either STAT5A or STAT5B, because STAT5A and STAT5B can bind to the same response elements. A total of 25 (18 old + 7 new) different clones were analyzed for STAT5 recruitment using ChIP. In mouse mammary gland tissue from lactating and pregnant mice, both STAT5A and STAT5B were recruited to the putative STAT5 binding sites of the following

endogenous clone regions: c153, c194, c236, c831 and c20, c34, c73, c398, and c882 (Fig. 3.10C-F and Fig. 3.12A-E).

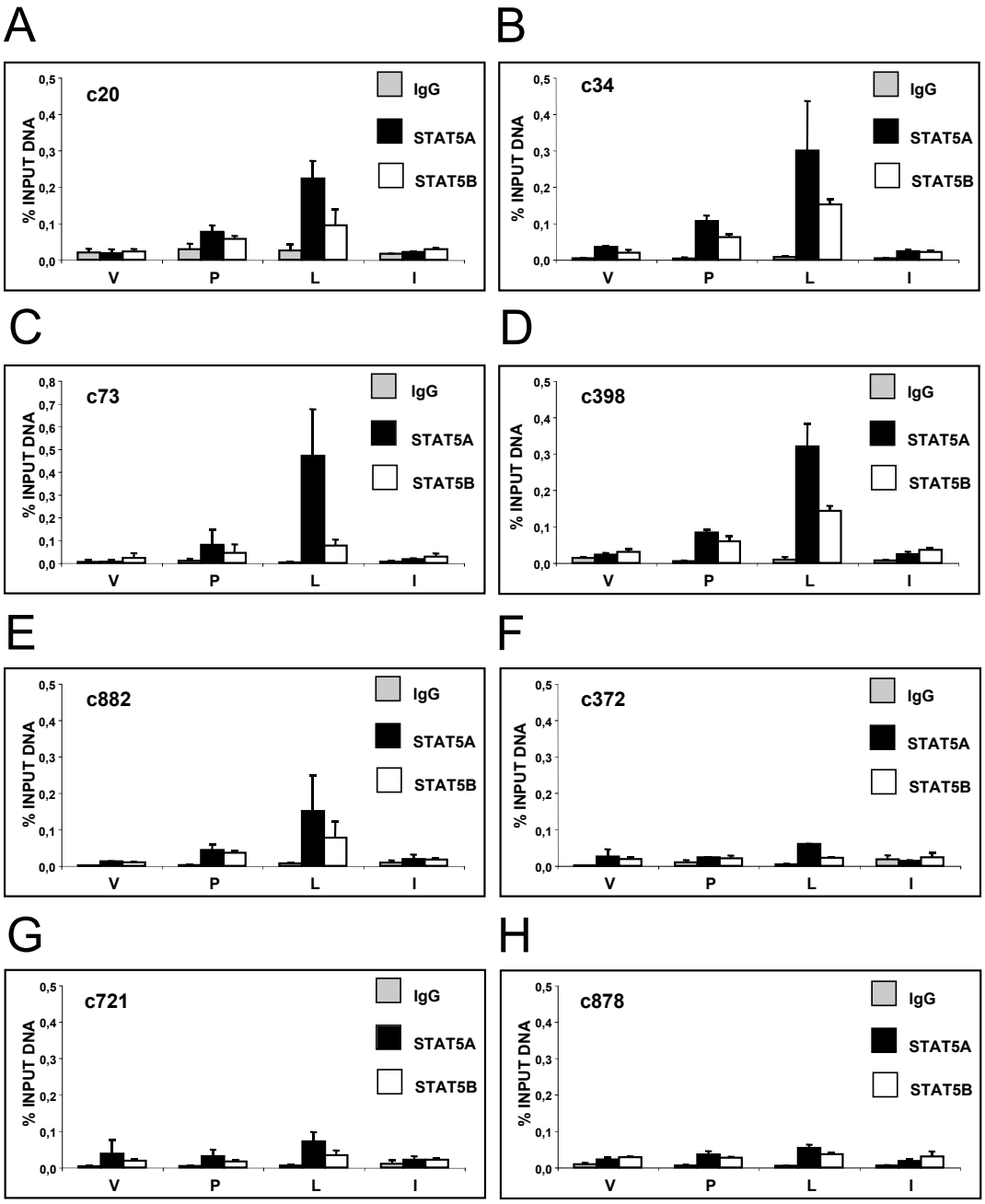


Figure 3.12 Recruitment of STAT5A and STAT5B to the novel STAT5 target sequences in mammary gland tissue.
Mammary gland tissue was harvested from virgin (V), pregnant (P), lactating (L) and involuting (I) mice and analyzed by ChIP with antibodies specific for STAT5A and STAT5B and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions flanking the STAT5-binding sites of additional novel STAT5 motif-containing clones (c20, c34, c73, c398, c882, c372, c721, c878). The real-time PCR was performed in triplicates. The immunoprecipitation data was normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA. The data depicted is representative of the trend observed in at least two independent ChIP assays.

In comparison, no STAT5 recruitment was observed during the stage of involution or in the mammary gland tissue from virgin mice. The ChIP assay was specific for STAT5A and STAT5B, since only very low levels of the clone regions were captured in the IP with the non-specific control IgG antibody. These results show that both STAT5A and STAT5B can bind 9 novel STAT5 target regions specifically during the mammary gland developmental stages of pregnancy and lactation. The recruitment of STAT5B mirrors the trend observed for STAT5A. Clone regions like c372, c721 and c878 are example of regions that represent borderline cases in terms of STAT5 recruitment (Fig. 3.10F-H). The overall values of bound DNA (% input DNA) are low in comparison to the other clone regions, but specific STAT5 recruitment can still be observed when compared to non-specific recruitment of IgG during lactation. These clone regions might represent regions that are bound by STAT5 with lower affinity.

Overall, the results show the isolation of 9 unique/novel STAT5 targets, which recruit both STAT5A and STAT5B proteins. This recruitment is seen mainly during pregnancy and lactation when mammary epithelial cells undergo proliferation and differentiation. Therefore the ChIP assay was used successfully to isolate new differentiation specific STAT5 targets from lactating mammary gland mouse tissue. And these independent ChIP assays show that the DNA sequences isolated represent true STAT5 binding sites.

The individual STAT5-binding sites for the 9 clones are listed (Table 3). Most of the clones, except c153 and c831 contain at least two or more classic STAT5 binding motifs within 1000 bp from the original cloned fragment.

To have a more controlled environment, where the contribution of individual signals can be dissected, ChIP assays for STAT5 recruitment to the novel STAT binding regions were tested in the HC11 cells line, which is a representative model for mammary epithelial cell differentiation. Interestingly, like in the case of ß-casein, none of the 9 newly cloned STAT5 target regions showed STAT5 recruitment in confluent HC11 cells induced with lactogenic hormones over various time points (Fig. 3.13, compare A with C-H and no data shown for c20, c398 and c882). On the other hand, the positive control SOCS3 faithfully showed STAT5 recruitment in the same ChIP experiments (Fig. 3.13B). These results show that in HC11 cells induced with lactogenic hormones, no STAT5 recruitment was observed to the ß-casein or any of the novel STAT5-binding regions when compared to the SOCS3 promoter. This might indicate that possibly additional signals are required before STAT5 recruitment to these sites in this cell line can be observed.

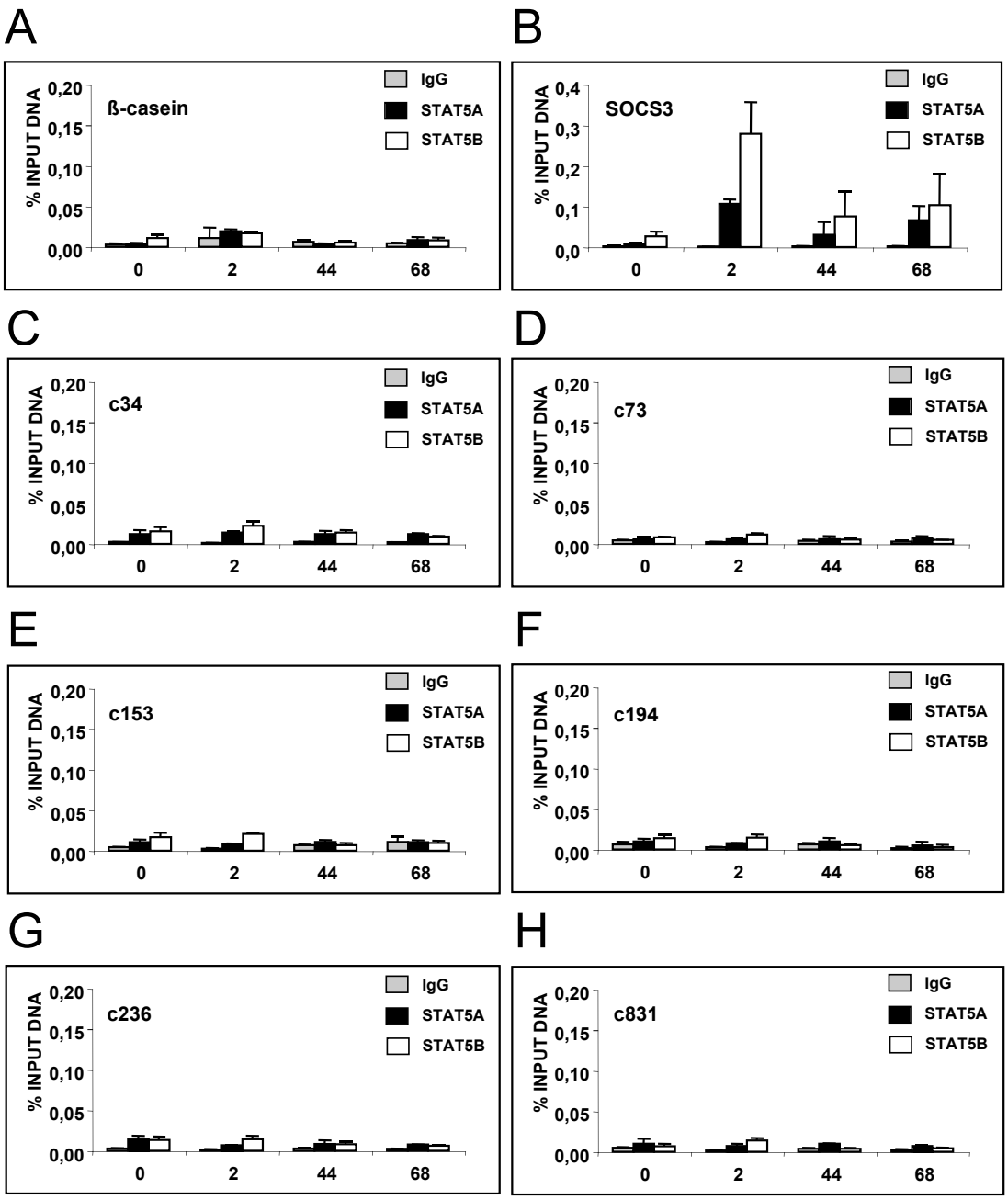


Figure 3.13 Recruitment of STAT5A and STAT5B to the novel STAT5 target sequences in mammary epithelial cells.
After a 3 day pre-treatment with dexamethasone, confluent HC11 cells were treated with lactogenic hormones prolactin (5 µg/ml), insulin (5 µg/ml) and dexamethasone (0,1 µM) for the different time points (hours) indicated. The cells were harvested and analyzed by ChIP with antibodies specific for STAT5A and STAT5B and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions flanking the STAT5-binding sites. Results for two known STAT5-binding genes (A) ß-casein and (B) SOCS3, as well as six novel STAT5 motif-containing clones (C-H) c34, c73, c153, c236, c194 and c831 are depicted. The quantitative real-time PRC was performed in triplicates. The immunoprecipitation data was normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA. The data depicted is representative of the trend observed in at least two independent ChIP assays.

3.2.4 Investigation into RNA polymerase II recruitment and histone 2M(K4)H3 modifications at four of the novel STAT5 binding sites

Specific N-terminal tail histone modifications have been associated with actively transcribed euchromatin or silent heterochromatin. In general di-methylation of lysine 4 of histone H3 correlates with a permissive state of euchromatin in which genes can be either active or potentially active (Santos-Rosa et al., 2002). To assess if the STAT5-binding regions are located in such a euchromatin environment and if changes between the two mammary gland developmental stages (virgin and lactation) can be observed, ChIP assays were performed with antibodies against di-methyl(K4)H3. In addition, ChIP was performed with an antibody against the phosphorylated C-terminal domain of RNA polymerase II, since the RNA polymerase II is the major component of the pre-initiation complex and its association with a STAT5-binding region would indicate transcriptional activity. This analysis was performed on the first four clones c153, c194, c236 and c831 initially identified and verified to recruit STAT5. The results for c153 show that there is a strong increase in RNA polymerase II recruitment to the region during lactation (Fig. 3.14A). In the virgin tissue there are elevated levels of 2M(K4)H3 already present, which increase 3-fold during lactation (Fig. 3.14.B). These results suggest that RNA polymerase II recruitment is differentiation stage specific and the elevated presence of 2M(K4)H3 in the virgin tissue indicates that the chromatin of this c153 region is permissive to gene transcriptional activity. For clone c194 there is only a minimal association of RNA polymerase II compared to IgG and barely a 2-fold difference between virgin and lactating tissue (Fig. 3.14C). In comparison to c153 this is only a poor association of RNA pol II to this region (Fig. 3.14, compare A and C). The situation is similar for the histone modification, only low levels of 2M(K4)H3 are detected in comparison and there is barely a 3-fold difference between the virgin and the tissue from lactating mice (Fig. 3.14D). These results indicate that c194 contains less 2M(K4)H3 and there is less RNA polymerase II recruitment than found for any of the other three clone regions. The clone 236 displays already strong recruitment of RNA polymerase II in virgin tissue, with a 2-fold increase during lactation (Fig. 3.14E). This indicated that RNA polymerase is already associated with the region and the high levels of immunoprecipitated 2M(K4)H3 (Fig. 3.14F) suggest that the region is located in euchromatin. For c831 the levels remain low and there appears to be hardly any difference in RNA polymerase II recruitment between the two stages (Fig. 3.14G). The levels of detected 2M(K4)H3 remain high with hardly any variance between the two stages (Fig. 3.14H). This indicates that the c831 region is located in a euchromatin environment just as the other two STAT5-binding regions c153 and c236.

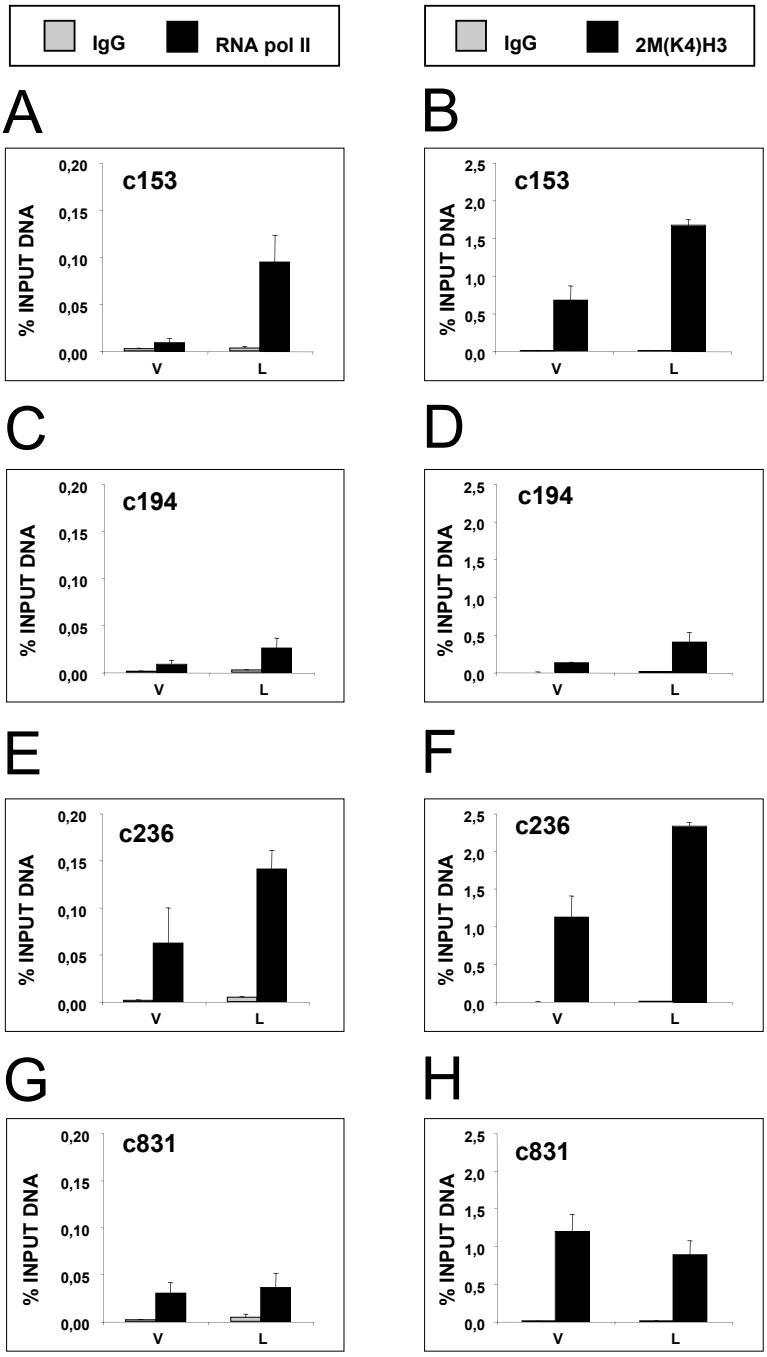


Figure 3.14 Recruitment of RNA polymerase II and 2M(K4)H3 histone modification on four STAT5-binding target sequences.
Mammary gland tissue was harvested from virgin (V) and lactating (L) mice and analyzed by ChIP with antibodies specific for phosphorylated CTD of RNA polymerase II (RNA pol II) and di-methylated (K4) histone H3 (2M(K4)H3) and unrelated IgG antibody (IgG). Antibody co-precipitated DNA samples were analyzed by real-time PCR with primers amplifying the regions flanking the STAT5-binding sites of the following novel STAT5 motif-containing clones (c153, c194, c236, c831). The quantitative real-time PCR was performed in triplicates. The immunoprecipitation data was normalized to input DNA and amounts of DNA recovered expressed as a percentage of input DNA. The data depicted is representative of the trend observed in at least one independent ChIP assays.

3.2.5 Comparative sequence analysis of the novel STAT5-binding sites between different vertebrate species

Differentiation specific STAT5-binding has been confirmed in mammary gland tissue for nine individual clone regions *in vivo*. To gain more insight into a possible functional role of the STAT5-binding sites, a computational approach in addition to an *in vitro* functional assay was chosen. Because sequences containing functionally important regulatory DNA sequences are often conserved during evolution, (Thomas et al., 2003; Loots et al., 2000) a multiple sequence alignment of the STAT5-binding site sequences was performed, to determine how evolutionary conserved these STAT sites are among the genomes of different species. Using the UCSC Genome Bioinformatics browser, 12 different vertebrates sequences were aligned to the mouse genome sequences. In this section the genomic location of the mouse STAT5-binding sites and their homology to other species will be described for the following nine clones: c20, c34, c73, c153, c194, c236, c398, c831, c882.

3.2.5.1 The second STAT5 motif of the c20 clone shows homology between the mouse and rat sequence

Within 175 bp from the sequenced c20 clone fragment there are two STAT5-binding sites (Table 3 and Fig. 3.15A). Both STAT5 site I (TTCAAGGAA) and STAT5 site II (TTCTTAGAA) have the classic STAT5-binding motif TTC(N3)GAA. The two STAT5-binding sites are separated from each other by 88 bp. For the STAT5 site I none of the species that had sequences aligned retained a STAT-binding motif. For the second STAT5 site, the same STAT5 motif was found conserved in the rat, but not in the other species that had sequences available to be aligned. This indicates that there is sequence homology between the two rodents (mouse and rat) at least for the second STAT5-binding site. This sequence homology might be linked due to its close proximity to a highly conserved region. As seen in the top histogram, right adjacent (approximately 15 bp) to the STAT5-binding site II is a region of approximately 24 bp with high sequence conservation between the mouse, rat, human, chimp, rhesus monkey, dog and cow (Fig. 3.15A). For this region an alignment was not possible for all species.

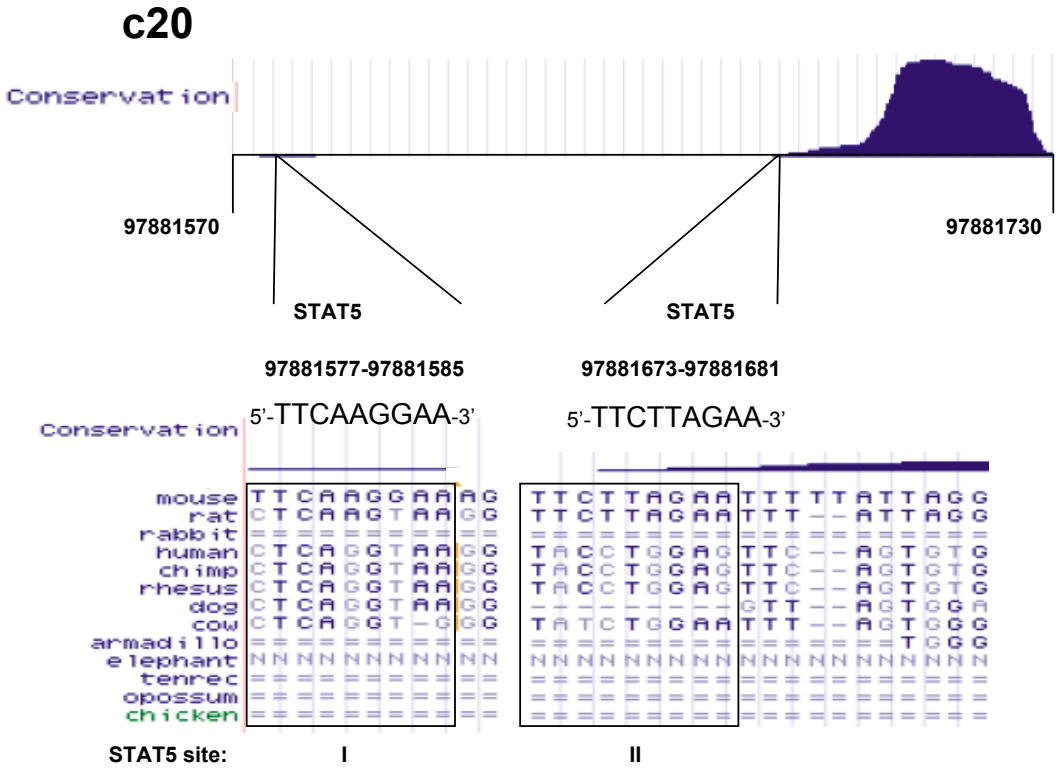


Figure 3.15A The second STAT5 motif of the c20 clone region shows homology between the mouse and rat sequence.
The upper panel shows a histogram of a 160 bp region containing the two STAT5-binding sites associated with c20 where increased conservation is displayed by increased height. The nucleotide positions of both the region and STAT5-binding sites are given and are based on the UCSC genome browser genomic assembly (Feb 2006, mm8) in mouse Chromosome 11. The locations of the STAT5-binding sites within the 160 bp region are indicated by lines converging to a point. The middle panel shows the individual mouse STAT5-binding motifs and their nucleotide locations as given in the alignment. The lower panel shows the exact sequence identified in the sequence alignment using a screen capture of the UCSC genome Bioinformatics browser, and depicts the evolutionary conservation between the mouse genome and 12 other vertebrate species. The STAT5-binding sites are boxed.

3.2.5.2 The c34 clone is located within the second intron of the Arhgef3 gene and all three STAT5 motifs show evolutionary conservation in at least 5 different mammals

The three STAT5-binding sites from clone 34 are located in the second intron of the Rho guanine nucleotide exchange factor (GEF)3 (Arhgef3) gene (Ensemble-Exon view). The second intron is about 97 kb in size and the STAT5 sites are located in a 1,475 bp large segment that lies 84,577 bp downstream and 11,150 bp upstream of exon 2 and 3, respectively (Fig. 3.15B). For this region no alignment was available for the chicken and opossum.

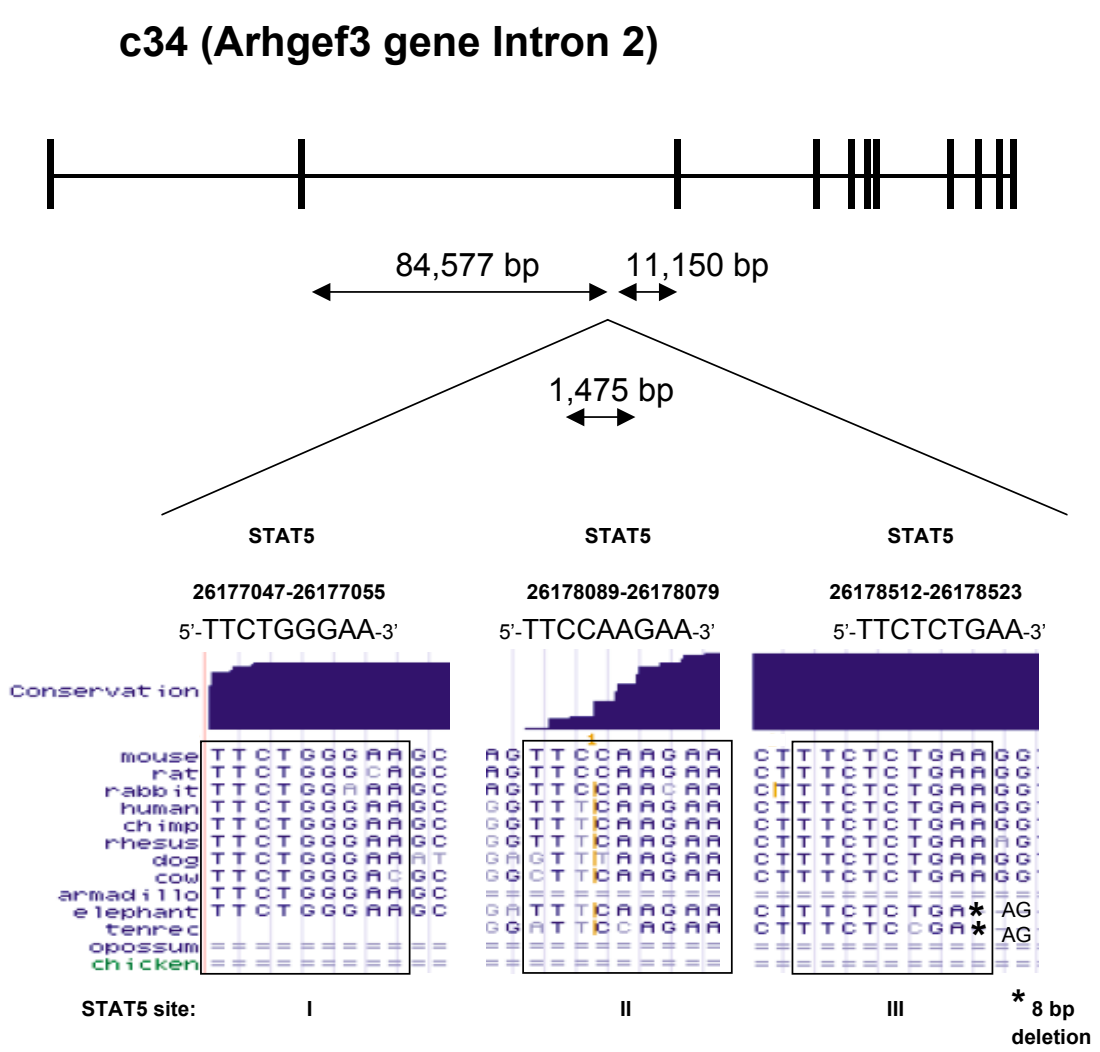


Figure 3.15B The c34 clone region is located in the second intron of the Arhgef3 gene and has three highly conserved STAT5-binding sites.
The upper panel depicts the schematic representation of the Arhgef3 gene with the approximate locations of the exons (vertical lines) and the introns (horizontal lines) (not drawn to scale). The location of the STAT5-binding sites are indicated below. The distance of the STAT5-binding sites to the second and third exon is noted below the gene structure. The lower panel shows the mouse STAT5-binding sites and their base pair location based on the UCSC genome browser genomic assembly (Feb 2006, mm8) in mouse Chromosome 14. The lower panel shows the exact sequences identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. Increased conservation is displayed by increased height in the histogram. The STAT5-binding sites are boxed.

The STAT5-binding site I (TTCGGGGAA) was exactly conserved in the 3 primates (human, chimp and rhesus), the dog, the armadillo and the elephant (Fig. 3.15B) lower panel). There is one nucleotide difference between the mouse motif and the rat, cow and rabbit sequences, and only in the rabbit does this single nucleotide change still conform to the STAT motif (TT(N5)AA). About 1,034 bp downstream of the STAT5 site I is the second STAT5 site II (TTCCAAGAA). The STAT5-binding site II is exactly

conserved in the 5 mammals, namely the rat, human, chimp, rhesus monkey, and elephant (Fig. 3.15B) lower middle panel). In the rabbit, dog, cow and tenrec there is a single nucleotide change in the STAT5 motif and again only the rabbit can retain the STAT motif. Another 444 bp downstream of STAT5 site II is a third STAT5 site (TTCTCTGAA). Here there is 100 % conservation between the mouse and 7 other mammals (rat, rabbit, human, chimp, rhesus monkey, dog and cow (Fig. 3.15B) lower right panel). Both the tenrec (TTCTCTGAA) and elephant (TTCTCCGAA) also maintain a STAT5 motif in their sequences. Due to a 8 bp deletion, right at the 3 prime end of the STAT motif in the sequence of both the tenrec and elephant, the alignment program placed an empty space holder when displaying the 8 bp sequence of the other species. The subsequent nucleotides in the tenrec and elephant sequence are shown right after the star symbol representing the 8 bp deletion. Thus, the STAT5-binding site III conforms to the STAT5 motif for 9 different mammals, indicating a high evolutionary conservation within the mammals investigated for this STAT5 site. The evolutionary conservation observed for all three STAT5-binding sites in at least 5 different mammalian species indicates that these STAT5 sites might be important DNA-binding elements.

3.2.5.3 The c73 clone is located within the last intron of the Inpp4b gene and three out of the five STAT5-binding sites show homology between the mouse and rat sequence

The five STAT5-binding sites from clone c73 are located in the last intron of the Inpp4b (inositol polyphosphate-4-phosphatase, type II) gene (Ensemble-Exon view). The last intron is approximately 49 kb in size and the STAT5 sites are located in a 956 bp large segment that lies 41,939 bp downstream and 6,485 bp upstream of exon 22 and 23, respectively (Fig. 3.15C). A comparative sequence alignment for these mouse STAT5 sites was observed for the rat only. The mouse STAT5-site I (TTCCAGGAA) and V (TTCTGTGAA) show no homology to the rat in that the STAT motif was not maintained due to 1-2 nucleotide changes. On the other hand, the mouse STAT5 sites II (TTCAAAGAA), III (TTCTGAGAA) and IV (TTCTGTGAA) show exact homology to the rat sequences. This suggests that there is sequence homology between the two rodents (mouse and rat) for these three STAT5 sites.

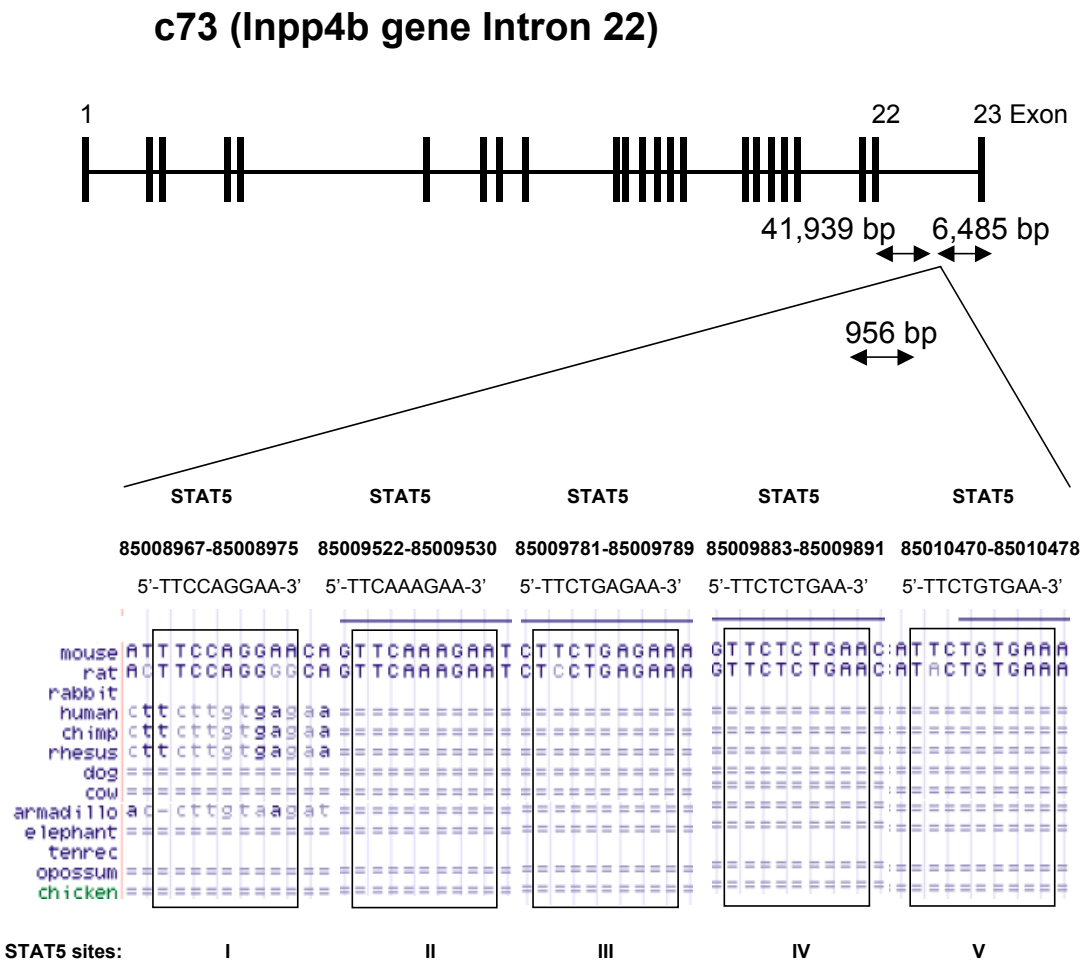


Figure 3.15C The c73 clone region is located in the last intron of the Inpp4b gene and contains five STAT5-binding sites of which three show homology to the rat sequence.
The upper panel depicts the schematic representation of the Inpp4b gene with the approximate locations of the exons (vertical lines) and the introns (horizontal lines) (not drawn to scale). The locations of the STAT5-binding sites are indicated below. The distance of the STAT5-binding sites to exon 22 and the last exon is noted below the gene structure. The lower panel shows the five mouse STAT5-binding sites and their base pair location based on the UCSC genome browser genomic assembly (Feb 2006, mm8) in mouse Chromosome 8. The lower panel shows the exact sequences identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. Increased conservation is displayed by increased height in the histogram. The STAT5-binding sites are boxed.

3.2.5.4 The 339bp region of c153 clone is evolutionarily conserved and contains one STAT5 consensus binding site and 3 STAT-like sites

Within the 743 bp region cloned for c153 there is a 339 bp region that is conserved by 80% between mouse and human genome. As shown in the histogram most of the 339 bp region of c153 appears to be well conserved within all the 12 vertebrate species except for the chicken for which no alignment was possible (Fig. 3.15D).

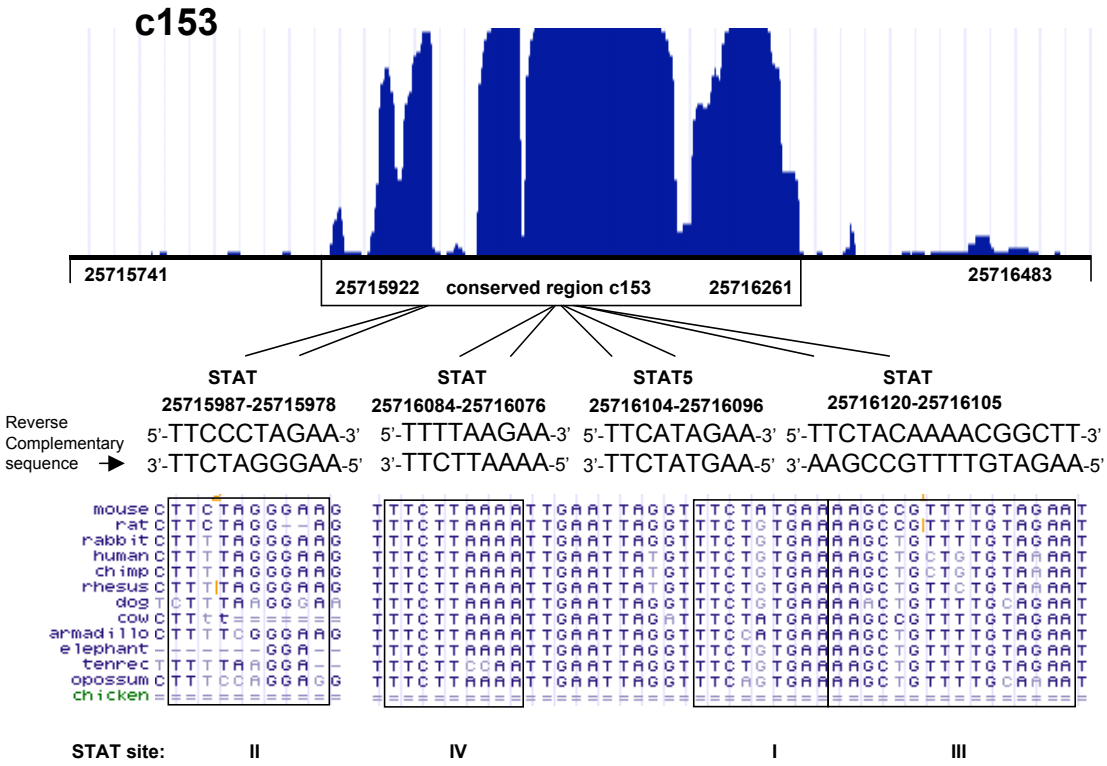


Figure 3.15D The 339 bp region of the c153 clone is highly conserved and contains one STAT5 and 3 STAT-binding sites.
The upper panel shows a histogram of the 743 bp region of c153 where increased conservation is displayed by increased height. The genomic location of the c153 region is marked below the histogram as well as the 339 bp region of c153 that is highly conserved between species. The nucleotide locations are based on the NCBI reference NT_039260.6/Mm4_39300_36 chromosome 4. The location of the various STAT-binding sites, found within the conserved 339 bp region of c153, is indicated below by lines. The middle panel shows the individual mouse STAT-binding sites and their reverse complementary sequences given in the alignment as well as their nucleotide locations. The lower panel shows the exact sequence identified in a multiple alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. The STAT-binding sites are boxed and numbered according to their initial identification in the sequence search.

This suggests that the conserved region might be under selective constraint. Within this highly conserved region are 4 putative STAT-binding sites. The putative STAT-binding site designated as STAT5 is conserved for all mammals displayed, and

conforms perfectly to the STAT5-binding site motif TTC(N3)GAA. The sequence of the motif is the same between the mouse and cow, while in the armadillo and opossum there is a slight variation within the central nucleotides of the motif. All other species display the same exact sequence for this STAT5 motif as seen in the rat. The high evolutionary conservation of the STAT5-binding motif between the different mammalian species investigated indicates that this STAT5 site is important. The mouse STAT site III (5'-TTCTACAAAACGGCTT-'3) right adjacent to the STAT5-binding site is comprised of a head to head STAT-binding motif (TT(N5)AA) in all the mammals except the primates. About 11 bp upstream of the STAT5-binding site is a STAT-binding site IV with a TT(N5)AA motif that is identical between the species, except for the tenrec. Finally there is another STAT-like site II (TTC(N4)GAA) approximately 89 bp upstream of the STAT5-binding site in the mouse. Here there are gap(s) in the available rat, tenrec, elephant and cow sequences, while the rabbit and primates exhibit the same nucleotide sequence (TTTTAGGGAA) which can be interpreted variably as a TT(N3-5)GAA motif. The armadillo matches with one nucleotide difference the primate sequence and the dog maintains a TT(N4)GAA while the opossum does not conform to the TT(N3-5)AA motif. Overall, this site shows less conservation between the investigated mammals than the other STAT-binding sites. This suggests that it is evolutionary less conserved and possibly not as important as the other sites.

3.2.5.5 The c194 region contains two putative STAT5-binding sites which show homology between the mouse and the rat genome

The c194 clone contains two classic STAT5-binding sites and two STAT-like sites (Fig. 3.15E). Both STAT5 site I (TTCTTGGAA) and STAT5 site II (TTCAAAGAA) have the classic STAT5-binding motif TTC(N3)GAA. The STAT5 site I is separated by 8 bp from a STAT-like site III (TTGCCTGGAA), which can be interpreted as a TT(N5)GAA motif. An alignment was obtained for this mouse sequence solely to that of the rat and only the sequence motif of the STAT-like site III was maintained. For the STAT5 site I a difference in the first T nucleotide of the rat sequence eliminated the classic STAT5-binding motif. The second classic STAT5 site was conserved in sequence with the rat, but not with any of the primates or the dog sequence. The mouse STAT-like site IV site (TTCAAAAGAA) was also not conserved within any of the other species aligned. The sequence conservation of the c194 STAT5 and STAT-like binding sites was weak between the mouse and rat, pointing to a poor evolutionary conservation of these sites.

c194

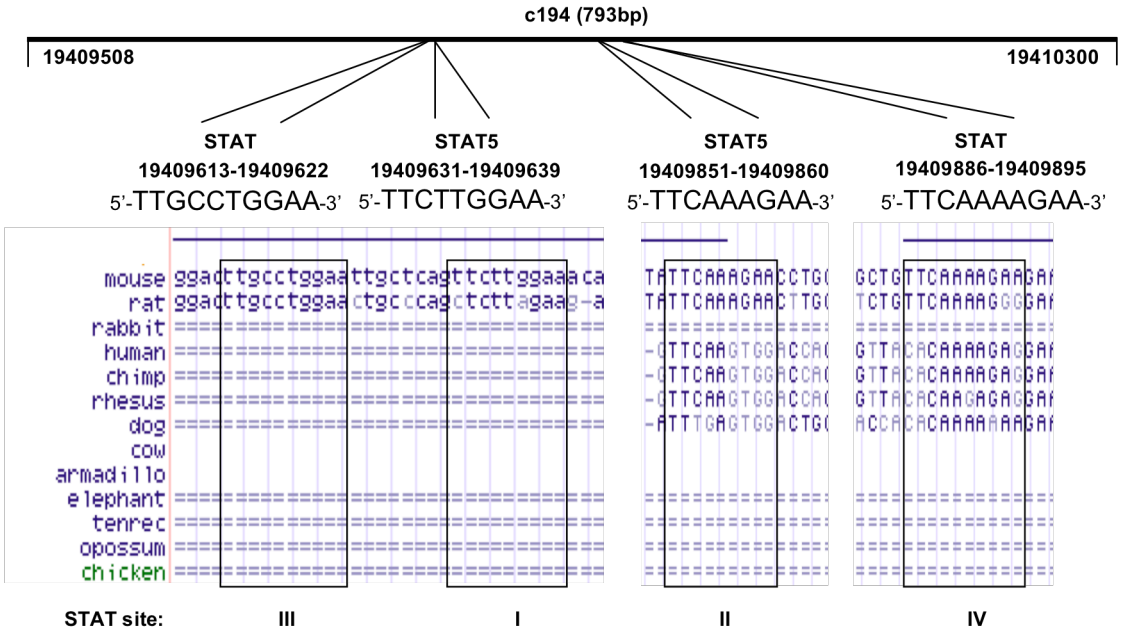


Figure 3.15E The c194 clone region contains two STAT5 and two STAT-binding sites of which two STAT motifs show homology to the rat sequence.
The genomic location of the 793 bp region of c194 is shown (horizontal line). The location is based on the NCBI reference NT_162143.2/Mm3_159849_36 chromosome 3. The positions of the various STAT-binding sites within the 793 bp region of c194 are indicated below. The lower panel shows the individual mouse STAT-binding sites and their base pair locations. The panel shows the exact sequence identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation of the STAT sites between the mouse and 12 other vertebrate species. The STAT-binding sites are boxed and numbered according to their initial identification in the sequence search.

3.2.5.6 The c236 region contains two putative STAT5-binding sites which are conserved between mouse and rat genome

The c236 region contains two STAT5-binding sites that both have the classic STAT5-binding motif TTC(N3)GAA. As seen in the histogram (Fig. 3.15F) there are multiple stretches within the 293 bp c236 region that show a higher conservation than the other regions. The two STAT5-binding sites are approximately 104 bp apart, with the STAT5-binding site II (TTCCAAGAA) being located in a section with higher homology than the STAT5 binding site I (AAGGGGCTT). The sequence of the mouse STAT5-binding site II is exactly conserved within the rat, rabbit and elephant and also in the cow with the exception of one nucleotide difference within the center of the motif. The primates retain a STAT-like binding TT(N4)GAA motif, even though there is a deletional gap

between the two adenines where in the aligned sequences of the other species there are found four adenines. The alignment program defined a gap and decided to place the second A at the fourth position. The sequence of the STAT5-binding site I (AAGCCCCTT) is a classic STAT5 TTC(N3)GAA motif and is exactly conserved between the mouse and the rat. All 3 primates, the cow and the elephant are more variable in the sequence, but still maintain the STAT motif TT(N5)AA. Taken together, the STAT5-binding site II has a higher conservation between the species for the classic STAT5 TTC(N3)GAA motif, suggesting that this binding site is more relevant for the c236 region than the STAT5-binding site I.

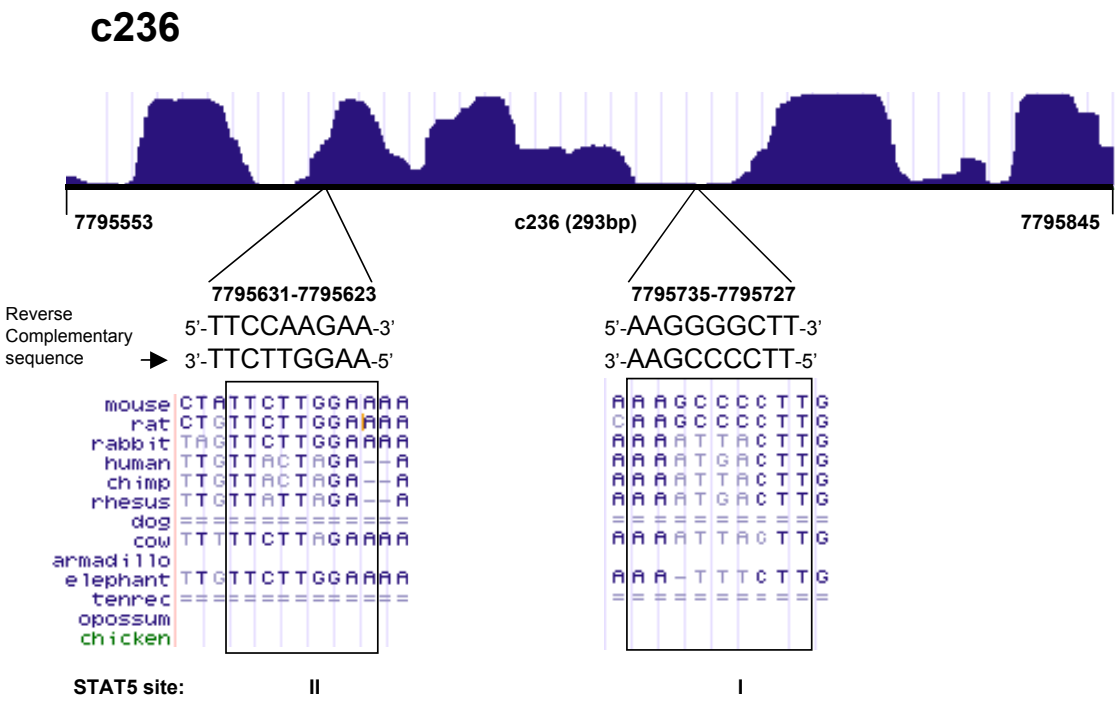


Figure 3.15F The c236 clone region contains two STAT5-binding sites that are conserved in rodents and retain STAT motif homology in other mammalian species.
The upper panel shows a histogram of the 293 bp region of c236 where increased conservation is displayed by increased height. The genomic location of the c236 region is marked below the histogram and is based on the NCBI reference NT_165760.1/Mm5_163274_36 chromosome 5. The locations of the two STAT5-binding sites within the 293 bp region of c236 are indicated below. The lower panel shows the individual mouse STAT5-binding sites and their base pair locations as well as the reverse complementary sequence as given in the alignment. The panel shows the exact sequence identified in the sequence alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. The STAT-binding sites are boxed and numbered according to their initial identification in the sequence search.

3.2.5.7 The c398 clone is located in the promoter and within the first intron of the RBm27 gene and one of the three STAT5-binding sites shows evolutionary conservation in 9 different mammals

The first and second STAT5 binding sites are located –370 bp and -197 bp upstream from the first exon of the RNA-binding motif protein 27 (RBm27) gene, respectively (Ensemble-Exon view) (Fig. 3.15G). The third STAT5-binding site III (TTCTTTGAA) is located in the first intron of the RBm27 gene (Ensemble-Exon view). The first intron is about 12,5 kb in size and the STAT5 site III is located 1,075 bp downstream and 11,431 bp upstream of exon 2 (Fig. 3.15G).

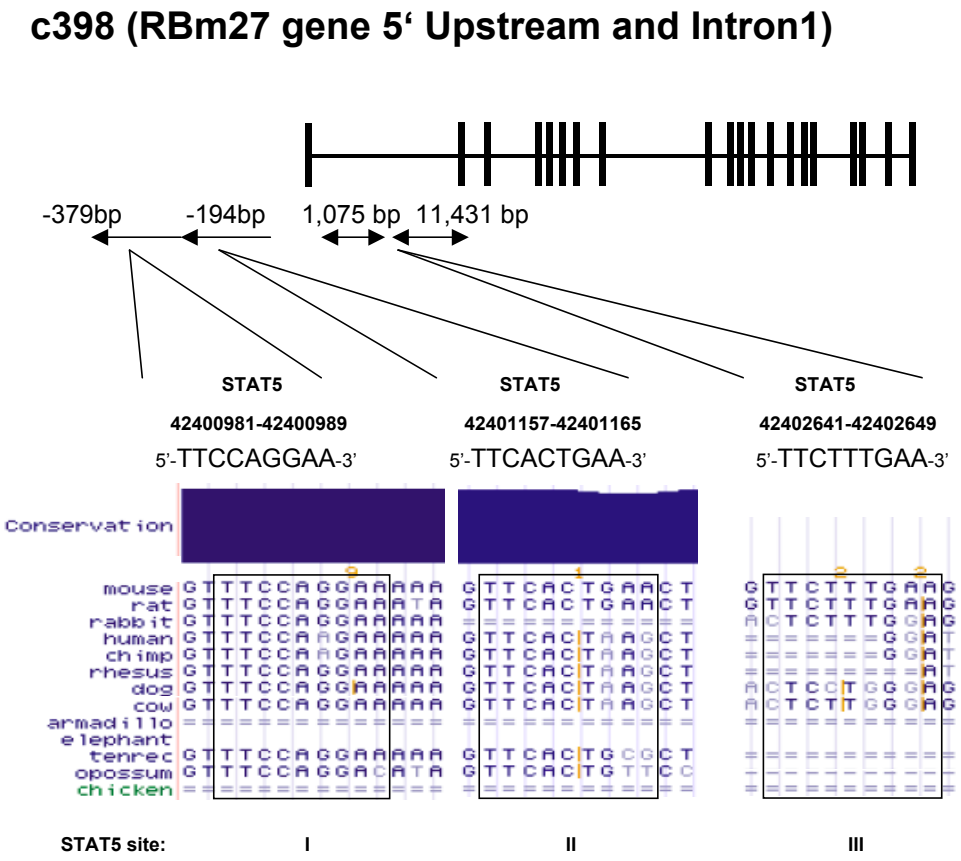


Figure 3.15G The c398 clone region is located in the 5’ upstream and first intron of the RBm27 gene and two out of the three STAT5-binding sites show high homology to other mammalian species.
The upper panel depicts the schematic representation of the RBm27 gene with the approximate locations of the exons (vertical lines) and the introns (horizontal lines) (not drawn to scale). The location of the STAT5-binding sites is indicated below. The distance of the STAT5-binding sites to the first exon and second exon is noted below the gene structure. The lower panel shows the three mouse STAT5-binding sites and their base pair location based on the UCSC genome browser genomic assembly (Feb 2006, mm8) in mouse Chromosome 18. The lower panel shows the exact sequences identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. Increased conservation is displayed by increased height in the histogram. The STAT5-binding sites are boxed.

The STAT5-binding site I (TTCCAGGAA) motif is 100% conserved between the mouse and the rat, rabbit, rhesus monkey, dog, cow and tenrec (Fig. 3.15G, lower left panel). There is only one nucleotide difference between the mouse STAT5 motif and the human, chimp and opossum sequences. In the human and chimp sequence (TTCCAAGAA) this single nucleotide change still conform to the STAT5 motif (TTC(N3)GAA). But in the opossum sequence the STAT motif (TT(N5)AA) is not maintained. No alignment was available for the chicken, armadillo and elephant. This evolutionary conservation, observed for the STAT5-binding sites I in at least 9 different mammalian species, indicates that this STAT5 site might be an important DNA-binding element found in the 5 prime upstream region –379 bp from the first exon of the RBm27 gene. The mouse STAT5-binding site II (TTCACTGAA) motif is also found in a region with higher sequence homology. With exception of the rat though, all other species aligned did not retain the STAT motif, due to double nucleotide changes in the sequence. At least the mouse STAT5 sites II (TTCACTGAA), and III (TTCTTTGAA) show exact homology to the rat sequences. This indicates that there is at least sequence homology between the two rodents (mouse and rat) for all three STAT5 sites.

3.2.5.8 The STAT5 binding site from the c831 clone is found within the first intron of the Zswim6 gene and is evolutionary conserved in 6 different mammals

The STAT5A binding site from the clone c831 is located within the 101 kb intron of the Zinc finger, SWIM domain containing 6 (Zswim6) gene (Ensemble- Exon view) and lies 62 kb downstream and 38 kb upstream of exon 1 and 2, respectively. No alignment was available for the chicken, opossum and rabbit and only partial alignment was available for the tenrec. The STAT5-binding site (TTCATGGAA) was 100% conserved between the mouse and the following 6 species: rat, human, chimp, rhesus, dog and elephant (Fig. 3.15H). There is only one and two nucleotide difference in the cow and armadillo sequence respectively, but these changes still conform to the basic STAT-binding site core consensus TT(N5)AA (Soldaini et al., 2000). The evolutionary conservation of this STAT5-binding site found in 6 different mammalian species indicates that this STAT-binding site may be an important DNA-binding element. Interesting to note is that the STAT5 site is 13 bp adjacent to a region of high homology, which may warrant further investigation.

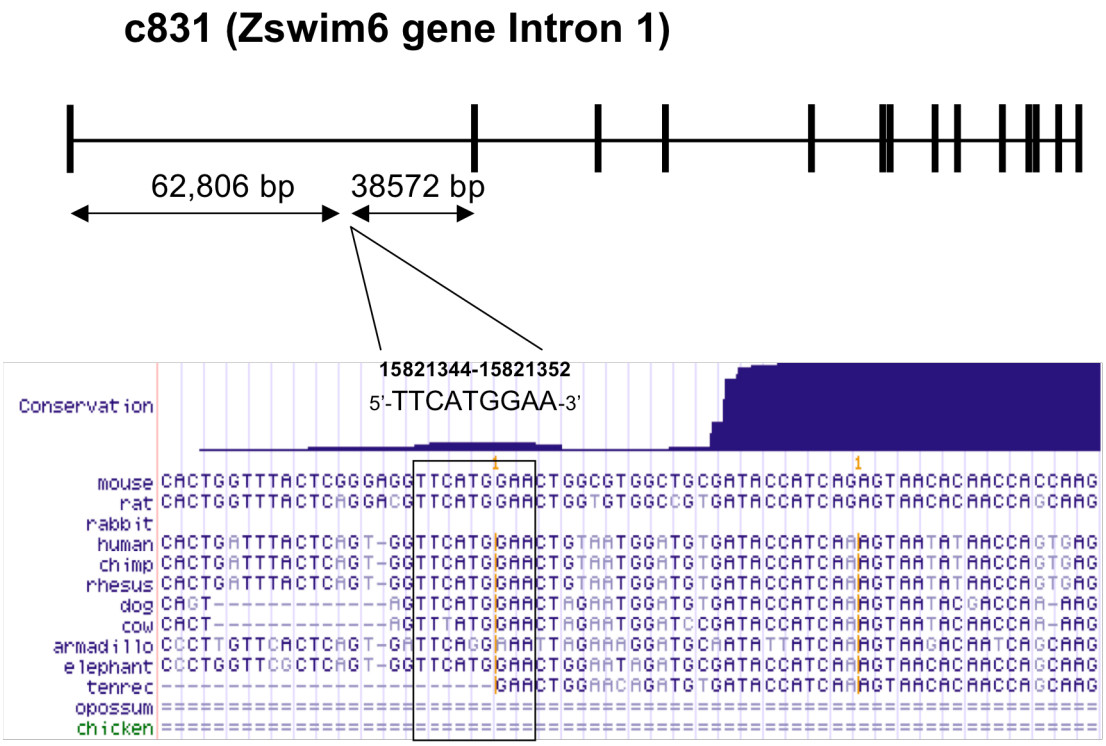


Figure 3.15H The c831 clone region is located in the first intron of the Zswim gene and contains a STAT5-binding site which displays high homology to other mammalian species.
The upper panel depicts the schematic representation of the Zswim6 gene with the approximate locations of the exons (vertical lines) and the introns (horizontal lines) (not drawn to scale). The location of the STAT5-binding site is indicated below. The distance of the STAT5-binding site to the first and second exon is noted below the gene structure. The mouse STAT5-binding site and its base pair location are based on the NCBI reference NT_039590.6/Mm3_39630_36 chromosome 13. The lower panel shows the exact sequence identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. Increased conservation is displayed by increased height in the histogram. The STAT5-binding site is boxed.

3.2.5.9 The c882 clone is located within the second intron of the Igsf3 gene and the two STAT5-binding sites show homology between the mouse and rat sequence

Two STAT5-binding sites from clone 882 are located in the second intron of the Immunoglobulin superfamily, member 3 (Igsf3) gene (Ensemble-Exon view). The second intron is about 46,7 kb in size and the STAT5 sites are located in a 280 bp large segment that lies 30,078 bp downstream and 16,599 bp upstream of exon 2 and 3, respectively (Fig. 3.15I). No alignment was available for the chicken, opossum, tenrec and elephant. The STAT5-binding site I (TTCAGAGAA) motif was exactly conserved between the mouse and the rat and rabbit (Fig. 3.15I lower left panel). The aligned sequences from the other species did not conform to the STAT motif due to

one or more nucleotide changes in the motif sequences. About 280 bp downstream of the STAT5 site I is the second STAT5 site II (TTCTTGGAA). For the second STAT5 site II the exact same STAT5 motif was found conserved in the rat, but not in any of the other species that had sequences available to be aligned (Fig. 3.15l lower right panel). Taken together these two STAT5-binding sites show homology for the STAT5 motif at least between the two rodents (mouse and rat).

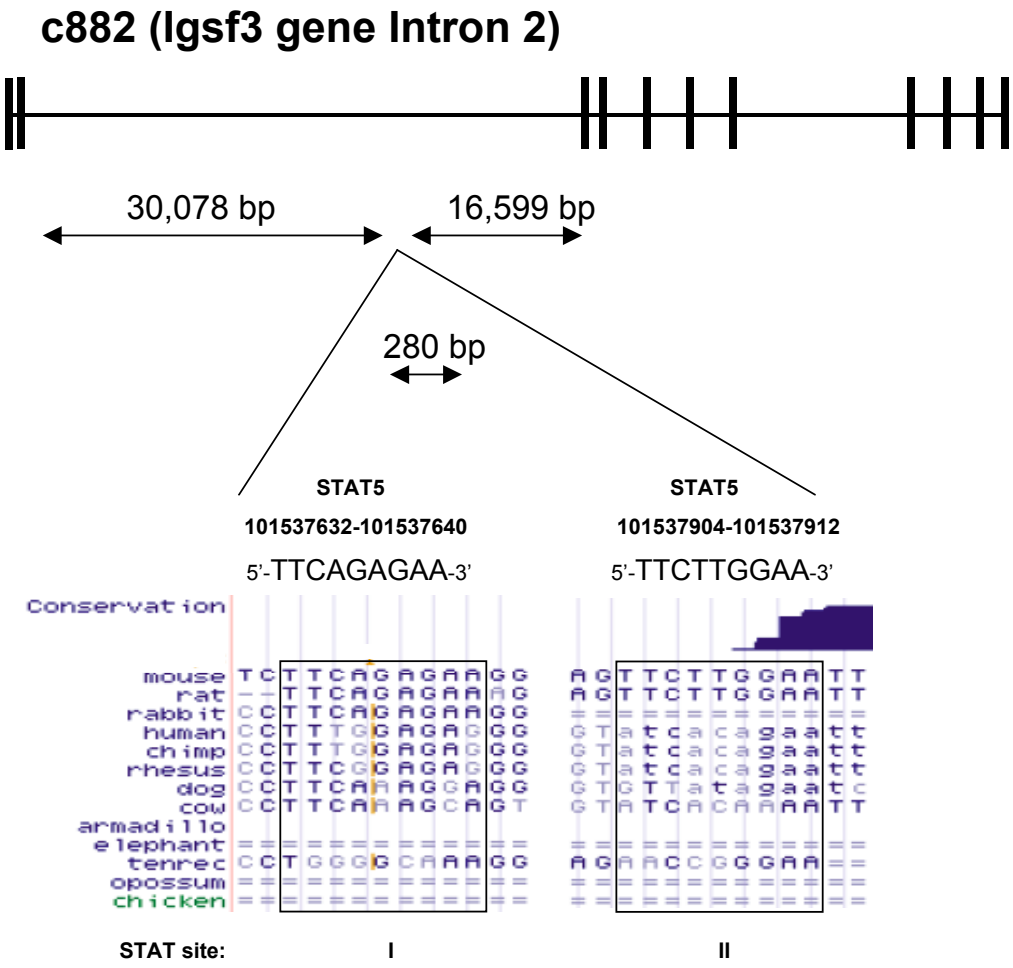


Figure 3.15l The c882 clone region is located in intron 2 of the Igsf3 gene and contains two STAT5-binding sites that show homology to the rat sequence.
The upper panel depicts the schematic representation of the Igsf3 gene with the approximate locations of the exons (vertical lines) and the introns (horizontal lines) (not drawn to scale). The locations of the STAT5-binding sites are indicated below. The distance of the STAT5-binding sites to the second and third exon is noted below the gene structure. The lower panel shows the two mouse STAT5-binding sites and their base pair location based on the UCSC genome browser genomic assembly (Feb 2006, mm8) in mouse Chromosome 3. The lower panel shows the exact sequences identified in the vertebrate alignment using a screen capture of the UCSC genome browser, and shows the evolutionary conservation between the mouse and 12 other vertebrate species. Increased conservation is displayed by increased height in the histogram. The STAT5-binding sites are boxed.

Overall, no alignment of the 9 clone regions was observed to the chicken sequence, which is also a vertebrate but not a mammal. The mouse c153 region aligned to all other 11 mammalian species tested, while other mouse clone regions such as c73 barely aligned to the highly related rat sequence. The computational analysis has provided information on the exact positions of the clone regions and their STAT motifs within the mouse genome. The comparative genomics approach, has allowed for the comparison of the sequences from the mouse clone regions to the genomic sequences of other species. This information provides hints as to how much of the sequences are in common and possibly functional important, but does not replace the need for experimental verification of the potential STAT5 regulatory elements in the clone regions.

3.2.6 Investigation into the potential function of the novel STAT5-binding sites using reporter gene assays

The sequence alignments made between the mouse STAT5-binding site motifs of the nine clones and the genomes of different vertebrate species, showed a diverse range in sequence homology. Some sites displayed sequence homology only to the rat or no homology at all, while other STAT5 motifs had high sequence homologies to different mammalian species. STAT5-binding motifs with high homology between species might be functionally important DNA-binding elements. Since STAT5 is a transcription factor involved in regulating transcriptional activity of target genes, an *in vitro* reporter gene assay was chosen to investigate the functional role of the STAT5-binding sites. The first four clones (c153, c194, c236 and c831) that had been initially identified and reconfirmed to bind STAT5 in ChIP were chosen for further analysis. For these four clones the STAT5-binding site containing regions were inserted into basic luciferase reporter gene constructs to determine if the STAT5-binding sites in these regions can support STAT5 mediated transcriptional activity.

3.2.6.1 The region of c153 confers transcriptional activity that depends on the STAT binding sites

To show that the c153 region is functional, the 743 base pair region of c153 was inserted into the pGL3 basic luciferase reporter gene (Fig. 3.16A).

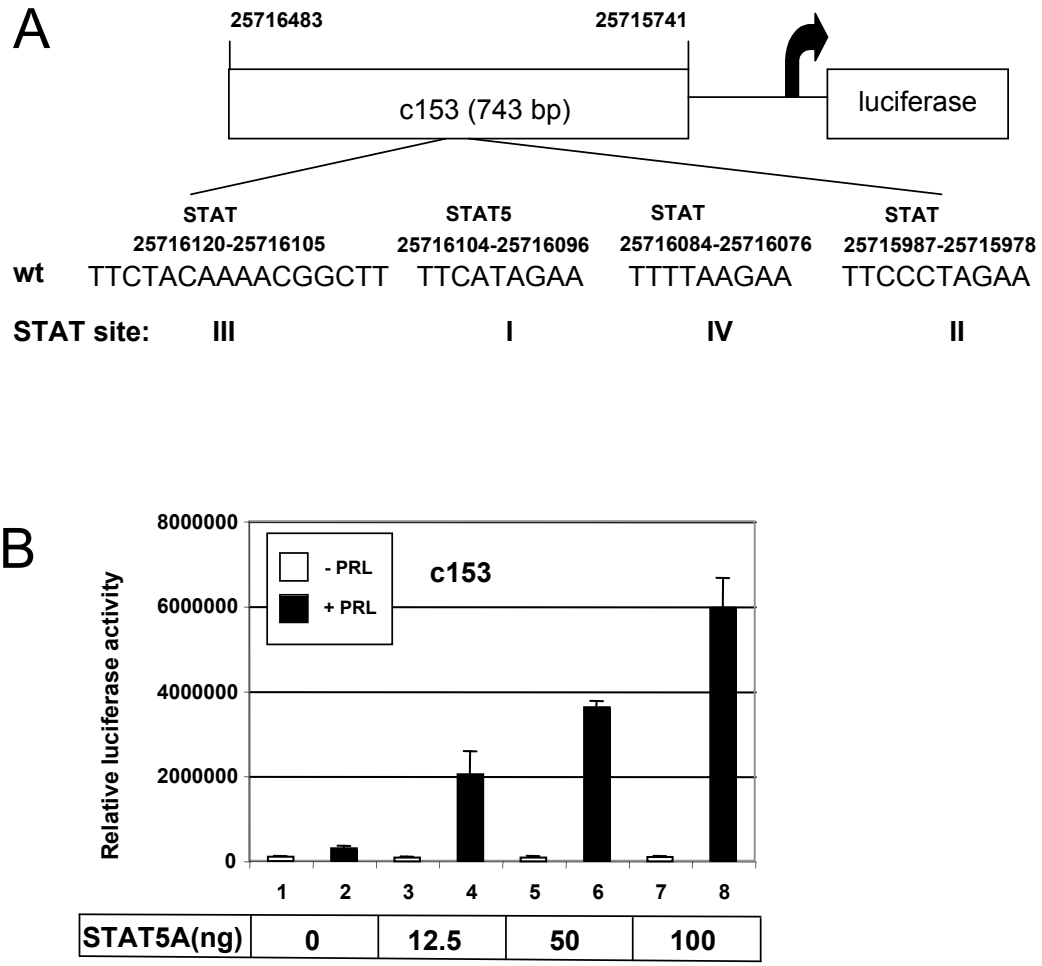


Figure 3.16 The c153 region is transcriptionally responsive to STAT5.
(A) Schematic representation of the luciferase reporter construct containing the longer 743 bp c153 region.. The nucleotide positions for the genomic location of the c153 region and of the STAT-binding sites are indicated and based on the NCBI reference NT_039260.6/Mm4_39300_36 chromosome 4. The nucleotide sequence and location in the mouse genome of the wild type (wt) STAT5 and STAT-like binding sites are indicated. The STAT-binding sites are numbered according to their initial identification in the sequence search.
(B) HeLa TA cells were transiently transfected in duplicate samples with the 743 bp c153 region luciferase reporter construct (c153 L_10) (1.25 µg) together with either an empty control vector (pMX) or with increasing amounts of wt STAT5A expression constructs (12.5 ng, 50 ng and 100 ng). The HeLa TA cells were co-transfected with the prolactin receptor (12.5 ng) and the SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection half of the duplicate samples were treated for 16 hours with prolactin (5 µg/ml) (dark bars) or left un-treated (white bars). After cell lysis, transcriptional activity of the luciferase reporter gene was measured and normalized to ß-galactosidase activity. Each luciferase assay was repeated two times.

Duplicate samples of HeLa TA cells were transiently transfected with the c153 region reporter construct, prolactin receptor, SV-40-*Lac*Z and increasing amounts of the wild type (wt) STAT5A expression construct. For each sample the corresponding sister well was treated with prolactin or left untreated and the luciferase activity was determined and normalized to ß-galactosidase. There is only a minimal increase in transcriptional

activity seen upon prolactin treatment for the c153 region reporter construct when the cells are not co-transfected with exogenous STAT5A (Fig. 3.16B, lanes 1 and 2). This almost negligible increase in luciferase activity may be due to the endogenous STAT5A. Increasing transcriptional activity by the c153 region reporter construct was observed with corresponding increases in the amounts of exogenous STAT5A (Fig. 3.16B, lanes 4, 6 and 8). Samples not induced with prolactin remained unchanged at a basal level (Fig. 3.16B, lanes 3, 5 and 7). These results indicate that the 743bp c153 region is transcriptionally responsive to STAT5A in a concentration dependent manner.

To prove if the reporter activity is directly due to STAT5A, dominant negative and constitutive active STAT5A mutants were tested for their effect on the c153 region transcriptional activity (Fig. 3.17).

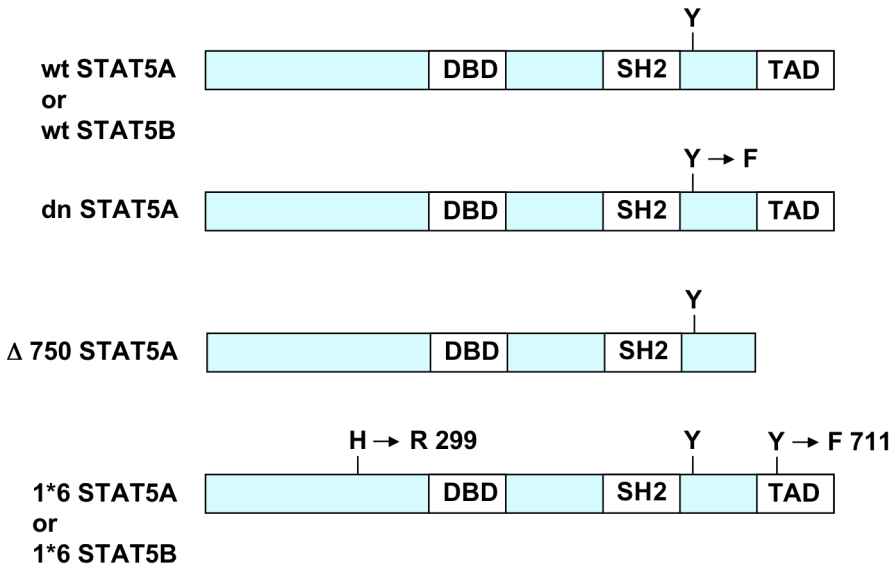


Figure 3.17 Schematic representation of wild type and mutant STAT5 expression constructs.
The wild type (wt) STAT5A or STAT5B with the major functional domains are displayed; the DNA-binding domain (DBD), the SH2 domain (src homology domain), the tyrosine (Y) residue, that is phosphorylated during activation and the C-terminal transactivation domain (TAD). In the dominant negative STAT5A mutant the essential tyrosine residue that is important for the posphorylated dimerization of two STAT proteins is mutated to phenylalanine (F). In the delta 750 STAT5A mutant the C-terminal transactivation domain is deleted. The constitutive active 1*6 STAT5A mutant has two point mutations, one upstream of the DNA-binding domain (histidine (H) 299 to arginine (R)) and one in the C-terminal transactivation domain (serine (S) 711 to phenylalanine (F)) (Onishi et al., 1998). The same mutations are present in 1*6 STAT5B mutant, since the amino acids surrounding the mutations are conserved in both STAT5A and STAT5B.

The c153 region reporter construct was transfected into HeLa TA cells with the different mutant STAT5 expression constructs (Fig. 3.18B). Addition of the control expression vector pMX resulted in no change in transcription activity when induced with prolactin, while addition of wt STAT5A resulted in a clear 16-fold induction of luciferase activity (Fig. 3.18B, lanes 2 and 4).

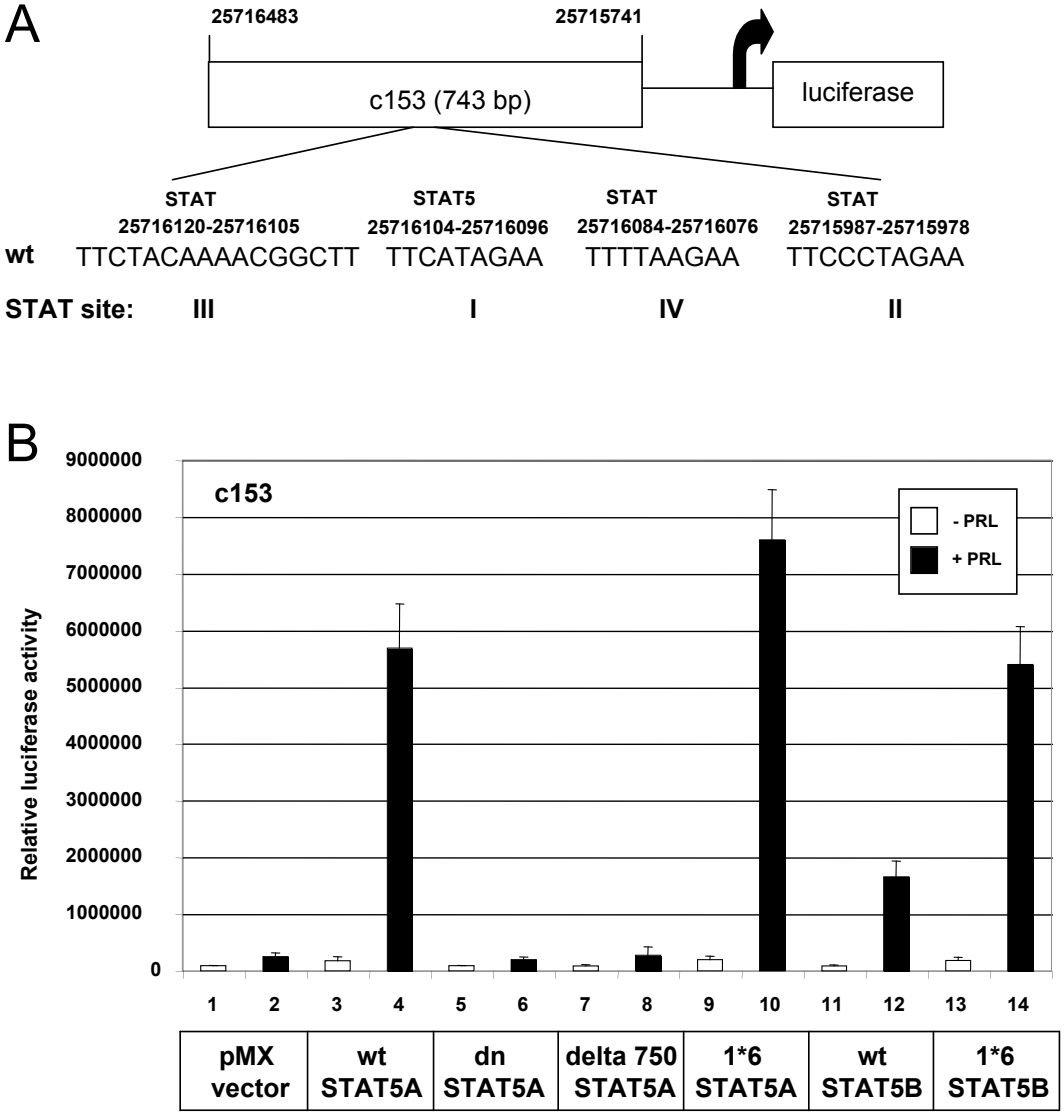


Figure 3.18 Effects of STAT5 mutants on the transcriptional activity of the c153 region.
(A) Schematic representation of the luciferase reporter construct containing the 743 bp c153 region. The nucleotide positions for the genomic location of the c153 region and of the STAT-binding sites are indicated and based on the NCBI reference NT_039260.6/Mm4_39300_36 chromosome 4. The nucleotide sequence and location in the mouse genome of the wild type (wt) STAT5 and STAT-like binding sites are indicated. The STAT-binding sites are numbered according to their initial identification in the sequence search
(B) HeLa TA cells were transiently transfected in duplicate samples with the wild type c153 region (c153) (1.25 µg) together with either an empty control vector (pMX) (100 ng) or with one of the following STAT5 expression constructs (100 ng) (Fig. 3.8), wild type STAT5A (wt STAT5A), dominant negative STAT5A (dn STAT5A), dominant negative delta 750 STAT5A (delta 750 STAT5A), constitutively active 1*6 STAT5A (1*6 STAT5A), wild type STAT5B (wt STAT5A), or constitutively active 1*6 STAT5B (1*6 STAT5B). The cells were co-transfected with the prolactin receptor (12.5 ng) and the SV-40_*LacZ* expression construct (12.5 ng). 24 hours after transient transfection the cells were either treated with prolactin (dark bars) or left un-treated (white bars) for 16 hours. After cell lysis the transcriptional activity of the luciferase reporter gene was measured and normalized to ß-galactosidase activity. Each luciferase assay was repeated at least two times.

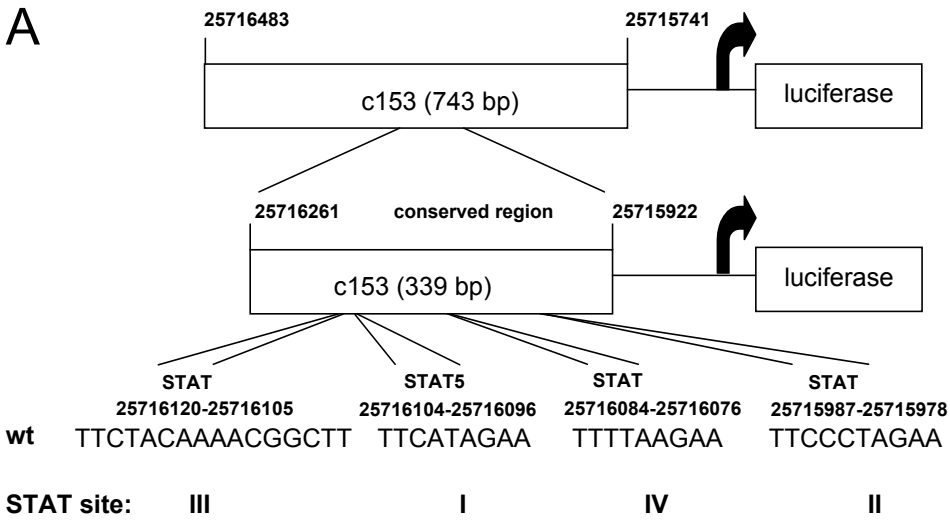
No luciferase activity of the c153 region reporter construct could be induced with prolactin when either the dominant negative mutant dnSTAT5A or the Δ750 STAT5A mutants was co-transfected (Fig. 3.18B, lanes 6 and 8). Since the dnSTAT5A and the Δ750 STAT5A mutant is defective in activating transcription, the results suggest that a functional STAT5A must be present to get a prolactin induced luciferase activity by the c153 region reporter construct. Co-transfection with the constitutive active 1*6 STAT5A expression construct resulted in a slightly higher than wt STAT5A luciferase activity (compare lanes 4 and 10). Contrary to the expectation, the constitutive active 1*6 STAT5A construct was not able to efficiently induce luciferase activity without prolactin induction (lane 9 and 10).

Previous ChIP experiments showed that STAT5A as well as STAT5B are recruited to the c153 region in mammary gland tissue from lactating mice, *in vivo*. To determine if STAT5B could also activate the reporter activity of the c153 region reporter construct, HeLa TA cells were transfected either with wt STAT5B or the constitutive active 1*6 STAT5B mutant expression constructs together with the c153 region reporter construct. Co-transfection with the wt STAT5B resulted in a 5-fold increase in luciferase activity and a 15-fold increase in luciferase activity with the constitutive active 1*6 STAT5B construct when compared to the prolactin induced response of the control expression vector (Fig. 3.18B, compare lanes 4, 12 and 14). These results indicate that not only STAT5A but also STAT5B can mediate the prolactin induced luciferase activity of the c153 region.

Overall, the data suggests that the c153 region is transcriptionally responsive to both STAT5A and STAT5B

3.2.6.2 The highly conserved 339 bp region of c153 confers reporter activity

Within the 743 bp region of c153 is a highly conserved 339 bp region which has 80% homology to the sequence found in the human genome (as described in detail in section 3.2.5.4, Fig. 3.15D). This highly conserved 339 bp region of c153 contains also the putative STAT sites. To focus on this evolutionary conserved region without the possible interference from other transcription factor binding sites found within the non-conserved sequence of the 743 bp region, the shorter 339 bp region was sub-cloned into the pGL3 basic luciferase reporter construct (Fig. 3.19A). A comparison to the 743 bp construct showed that the shorter 339 bp conserved region still retains half of the luciferase activity observed for the longer construct (Fig. 3.19B, compare lane 4 and 6). Nonetheless, the shorter reporter construct is still clearly prolactin inducible when compared to the pGL3 basic backbone construct (lanes 2 and 6).



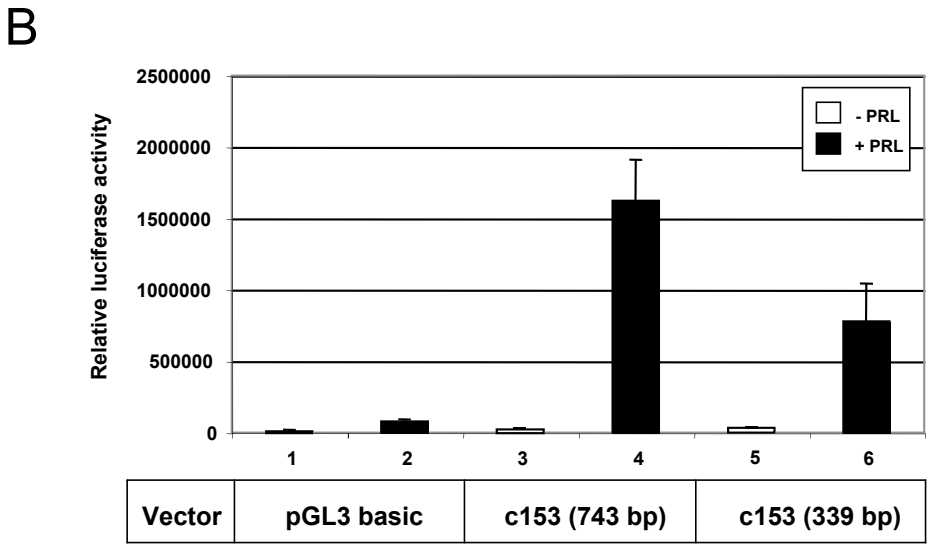


Figure 3.19 The shorter 339 bp evolutionary conserved region of c153 remains transcriptionally responsive to STAT5.
(A) Schematic representation of the luciferase reporter construct containing the longer 743 bp c153 region and the evolutionary conserved 339 bp c153 region. The nucleotide positions for the genomic location of the c153 region and of the STAT-binding sites are based on the NCBI reference NT_039260.6/Mm4_39300_36 chromosome 4. The nucleotide sequence and location in the mouse genome of the wild type (wt) STAT5 and STAT-like binding sites are indicated. The STAT-binding sites are numbered according to their initial identification in the sequence search.
(B) Duplicate samples of HeLa TA cells were transfected with empty luciferase reporter vector (pGL3 basic) (1.25 µg), or the reporter construct containing the long 743 bp c153 region (c153 (743 bp)) (1.25 µg), or the evolutionary conserved 339 bp c153 region (c153 (339 bp)) (1.25 µg). HeLa TA cells were co-transfected with wt STAT5A expression construct (100 ng), prolactin receptor (12.5 ng) and SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection half of the cells in sister wells were treated for 16 hours with prolactin (5 µg/ml) (dark bars) or left un-treated (white bars). After cell lysis the transcriptional activity of the luciferase reporter gene was measured and normalized to ß-galactosidase activity. Each luciferase assay was repeated at least two times.

Because the c153 region is not in the close vicinity of a known gene or its promoter, the c153 region could be potentially an enhancer, which might act in an orientation-independent way. Enhancers are *cis*-acting regulatory DNA sequences capable of increasing transcription independent of their distance or orientation from the linked gene (Blackwood and Kadonaga, 1998). Therefore both orientations of the 339 bp conserved region were tested in the luciferase assay for reporter activity (Fig. 3.20B). Both orientations were able to activate the reporter gene activity, with one orientation is more effective. The c153 region in reverse orientation was chosen for further experiments, since the overall luciferase activity was more intense than observed for the forward orientation (Fig. 3.20B, lanes 4 and 8) and is in the same orientation within the pGL3 luciferase reporter construct as the larger 743 bp c153 region reporter construct shown in previous experiments. In the genome the c153 region is located approximately 62 kb from the 5 prime side of nuclear factor I/B (NFI/B). The reverse oriented plasmid harbors the c153 region in correct orientation and upstream position to the reporter gene as it would be oriented towards the NFI/B gene in the mouse genome.

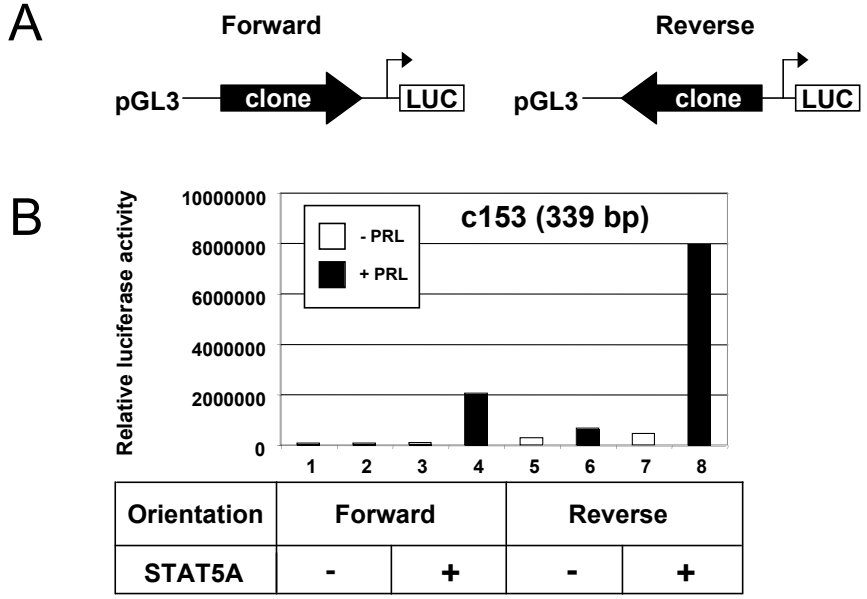


Figure 3.20 Differences in reporter activity were observed depending on the orientation of the conserved 339 bp c153 region.
(A) Schematic representation of the pGL3 basic luciferase (LUC) reporter construct depicting the arbitrary forward and reverse orientation of the inserted clone regions.
(B) Duplicate samples of HeLa TA cells were transfected with the pGL3 luciferase reporter construct containing either the forward or reverse oriented 339 bp c153 region (1.25 µg) together with either the control empty expression vector or the wild type STAT5A expression construct (100 ng). 24 hours after the transient transfection half of the samples were treated with prolactin (5 µg/ml)(dark bars) for 16 hours or left untreated (light bars). The transcriptional activity of the luciferase reporter gene was measured after cell lysis and normalized to ß-galactosidase activity. The luciferase assay was performed one time.

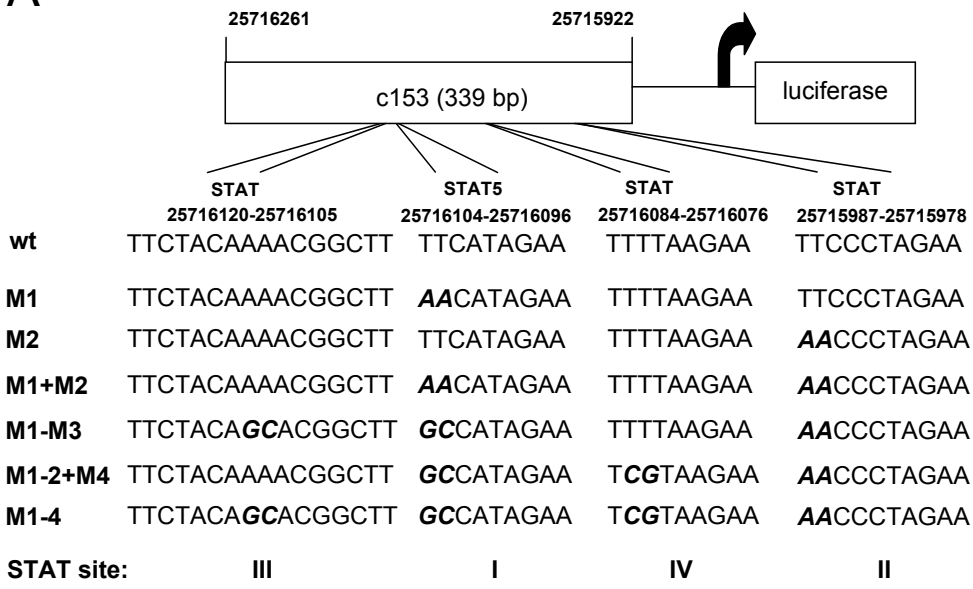
3.2.6.3 The STAT-binding sites of the c153 region are involved in the transcriptional activity

To determine if the STAT5 and STAT-like binding site found within the conserved c153 region are functional, mutations of these sites were performed and tested in luciferase assays. The first two nucleotides in the STAT5 site were mutated from wt TT to AA (Fig. 3.21A). Later these nucleotides were mutated instead to GC to remove any influence of an artificial created binding site (Mef2) by the initial AA mutation. There is a 10-fold increase in luciferase activity of the wt c153 reporter construct over the control pGL3 reporter construct when induced with prolactin (Fig. 3.21B, lanes 2 and 4). The double AA mutation in the particular STAT5-binding site I (M1) nearly results in a loss of half of the luciferase activity (Fig. 3.21B, lanes 4 and 6). This suggests that this STAT5-binding site found in the c153 conserved region confers part of the transcriptional activity, since the mutation of this site reduces this activity by half. Since half the luciferase activity remains the other putative STAT sites were also mutated. Mutation of the designated STAT site II (M2) did not result in a significant change of luciferase activity when compared to wt (Fig. 3.21B, compare lane 4 and 8). Mutation of both the STAT5 site I and STAT site II (M1+M2), resulted in a reduced luciferase activity comparable to the M1 mutant (Fig. 3.21B, compare lane 4, 6 and 10). Taken together this result would indicate that in comparison to the STAT5-binding site I, the STAT site II is not required for the observed luciferase activity, since its mutation does not reduce the activity, neither alone or in combination with the STAT5 site M1 mutation.

To find out if the other STAT-like binding sites are also involved in conferring reporter activity to the conserved c153 region additional mutations were introduced into the sequences of the remaining STAT-like binding sites III and IV. Additional mutations on the initial double mutant (M1+M2) of the STAT site M3 or M4 result in additional reduction of the luciferase activity (lanes 11-14). Mutation of all 4 sites (M1-4) together results in a significant reduction of luciferase activity in comparison to the wt reporter construct (compare lanes 4 and 16). There is barely a 2-fold increase of luciferase activity remaining when compared to the pGL3 basic control vector without insert (lane 2 and 16). These results suggest that the STAT-binding sites found within the conserved c153 region confer transcriptional activity and that when all 4 putative STAT-binding sites are mutated this reduces the activity significantly. It appears that the most important STAT sites are STAT site I, III and IV, while STAT site II seems less involved.

Overall the data suggests that the prolactin induced transcriptional activity observed for the c153 region is directly due to STAT5 and depends mostly on the STAT-binding sites.

A



wt	TTCTACAAAACGGCTT	TTCATAGAA	TTTTAAGAA	TTCCCTAGAA
M1	TTCTACAAAACGGCTT	***AA***CATAGAA	TTTTAAGAA	TTCCCTAGAA
M2	TTCTACAAAACGGCTT	TTCATAGAA	TTTTAAGAA	***AA***CCCTAGAA
M1+M2	TTCTACAAAACGGCTT	***AA***CATAGAA	TTTTAAGAA	***AA***CCCTAGAA
M1-M3	TTCTACA***GC***ACGGCTT	***GC***CATAGAA	TTTTAAGAA	***AA***CCCTAGAA
M1-2+M4	TTCTACAAAACGGCTT	***GC***CATAGAA	T***CG***TAAGAA	***AA***CCCTAGAA
M1-4	TTCTACA***GC***ACGGCTT	***GC***CATAGAA	T***CG***TAAGAA	***AA***CCCTAGAA
STAT site:	**III**	**I**	**IV**	**II**

B

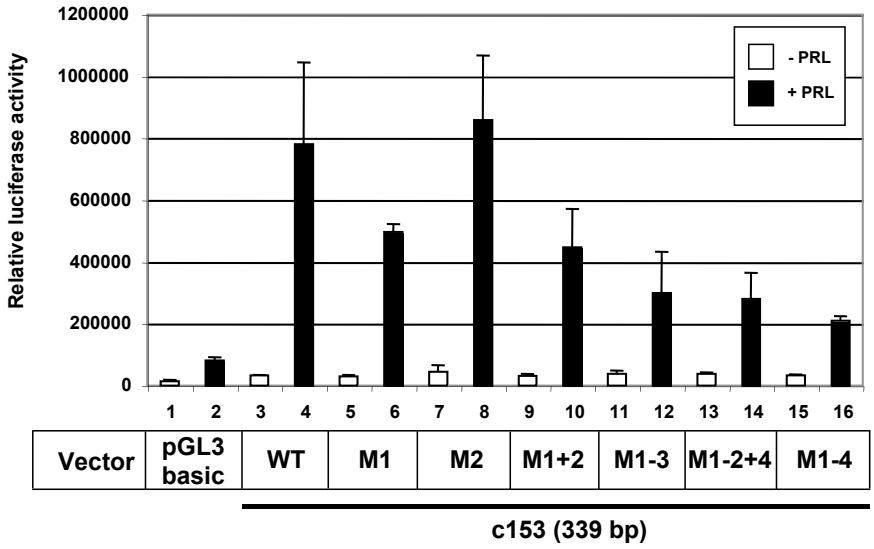


Figure 3.21 Mutation of the STAT5 and STAT-binding site within the evolutionary conserved c153 region reduces the transcriptional activity.

(A) Schematic representation of the luciferase reporter construct containing the 339 bp evolutionary conserved c153 region. The nucleotide positions for the genomic location of the STAT-binding sites based on the NCBI reference NT_039260.6/Mm4_39300_36 chromosome 4 are given.

The wild type (wt) STAT-binding sites are indicated, as well as the nucleotides mutated in each of the various mutant (M) constructs. The c153 region reporter constructs with mutations in the STAT5 consensus site I (M1), or mutation in the STAT site II (M2) or mutations in both the STAT5 consensus site I and the STAT site II (M1+M2) are depicted. Subsequent mutations were added to the double mutant (M1+M2), creating additional mutations in the STAT site III (M1-M3), mutations in the STAT site IV (M1-2+M4), or mutations in all four putative STAT sites I, II, III, and IV (M1-4). The STAT-binding sites are numbered according to their initial identification in the sequence search.

(B) Duplicate samples of HeLa TA cells were transfected with empty reporter vector (pGL3 basic) (1.25 µg), or with the reporter construct containing the wild type 339 bp conserved c153 region (WT) (1.25 µg), or the 339 bp conserved c153 region with mutations in one or more of the four putative STAT-binding sites

(1.25 µg). The cells were co-transfected with the wt STAT5A expression construct (100 ng), the prolactin receptor (12.5 ng) and the SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection half of the sister wells were either treated with prolactin (5 µg/ml) for 16 hours (dark bars) or left un-treated (white bars). The transcriptional activity of the luciferase reporter gene was measured after cell lysis and normalized to ß-galactosidase activity. The luciferase assays were performed at least two times.

3.2.6.4 The region of c194 displays unreliable reporter gene activity in response to exogenous STAT5

The c194 clone contains two classic STAT5-binding sites and two STAT-like sites within a stretch of about 793 bp (Fig. 3.15E). Comparative sequence analysis has shown homology to the rat sequence for the two STAT sites, STAT5 site II and STAT like site III (Fig. 3.15E). The c194 region is not in the close vicinity of any known mouse gene. Because the potential function of the STAT site-containing region is not clear, both orientations of the 793 bp c194 region were tested in the luciferase assay for reporter activity (Fig. 3.22A). HeLa TA cells were transiently transfected with wt STAT5A expression construct and either orientation of the c194 region reporter construct. The c194 region in reverse orientation displayed a higher luciferase activity than observed for the forward orientation (Fig. 3.22B, lane 4 and 8). In addition, both orientations were tested in an initial luciferase assay for their response to increasing amounts of exogenous STAT5A (data not shown). Here, the forward orientation was prolactin inducible, but remained at similar levels of activity regardless of the increase in STAT5A concentration and for the reverse c194 region reporter construct the levels of luciferase activity varied. This hints that the luciferase activity of c194 region is STAT5A responsive, but may not be concentration dependent. Due to the combination of low homology found in the STAT-binding sites in regard to other species and the unreliable induction of transcriptional activity in response to STAT5A, further analysis of this clone has been delayed.

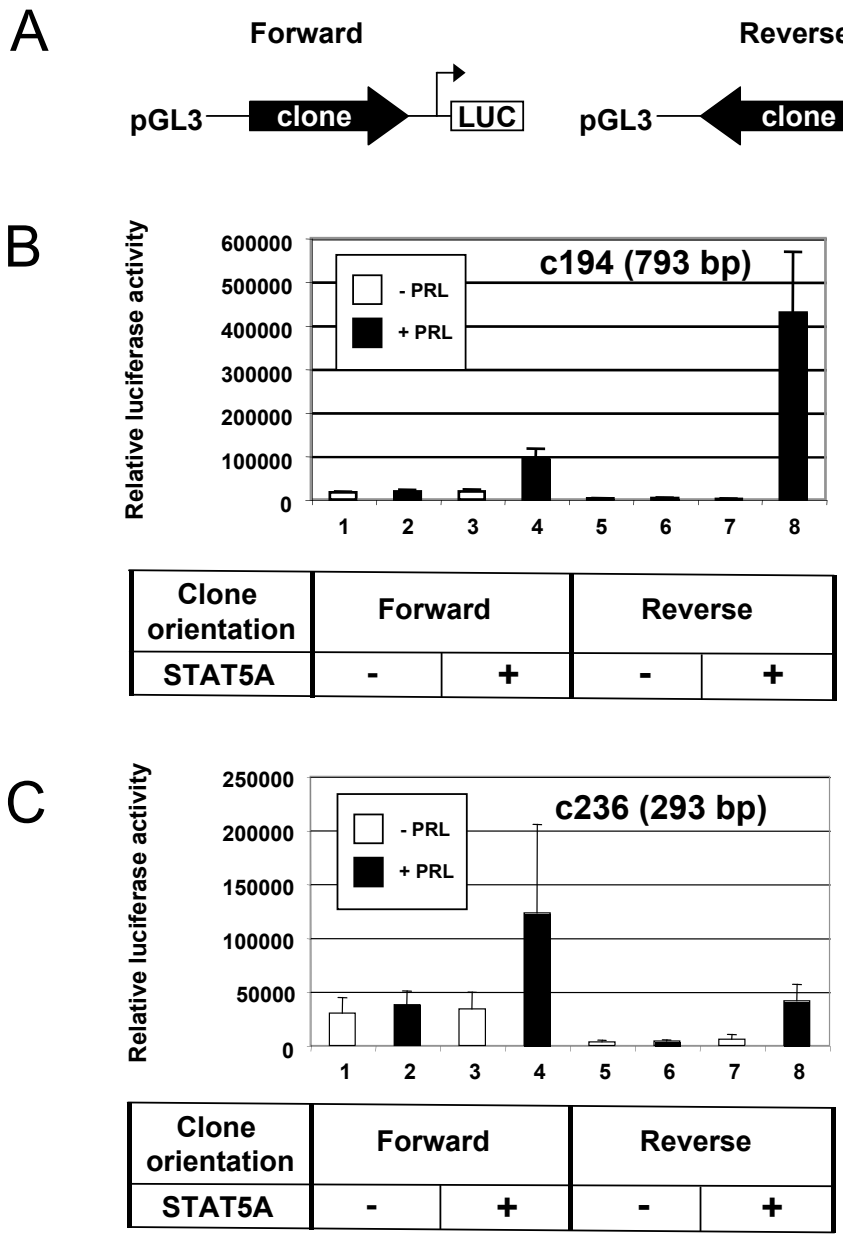


Figure 3.22 Differences in reporter activity were observed depending on the orientation of the cloned regions for c194 and c236.
(A) Schematic representation of the pGL3 basic luciferase (LUC) reporter construct depicting the arbitrary forward and reverse orientation of the inserted clone regions.
(B and C) The cloned regions, c194 (793 bp) or alternatively c236 (293 bp) (1.25 µg), oriented in either the forward or reverse orientation within the pGL3 reporter constructs, were transfected in HeLa TA cells together with either the control empty expression vector or the wild type STAT5A expression construct (each 100 ng). 24 hours after the transient transfection half of the duplicate samples were treated with prolactin (5 µg/ml)(dark bars) for 16 hours or left untreated (light bars). The transcriptional activity of the luciferase reporter gene was measured after cell lysis and normalized to ß-galactosidase activity. The luciferase assays were performed two times.

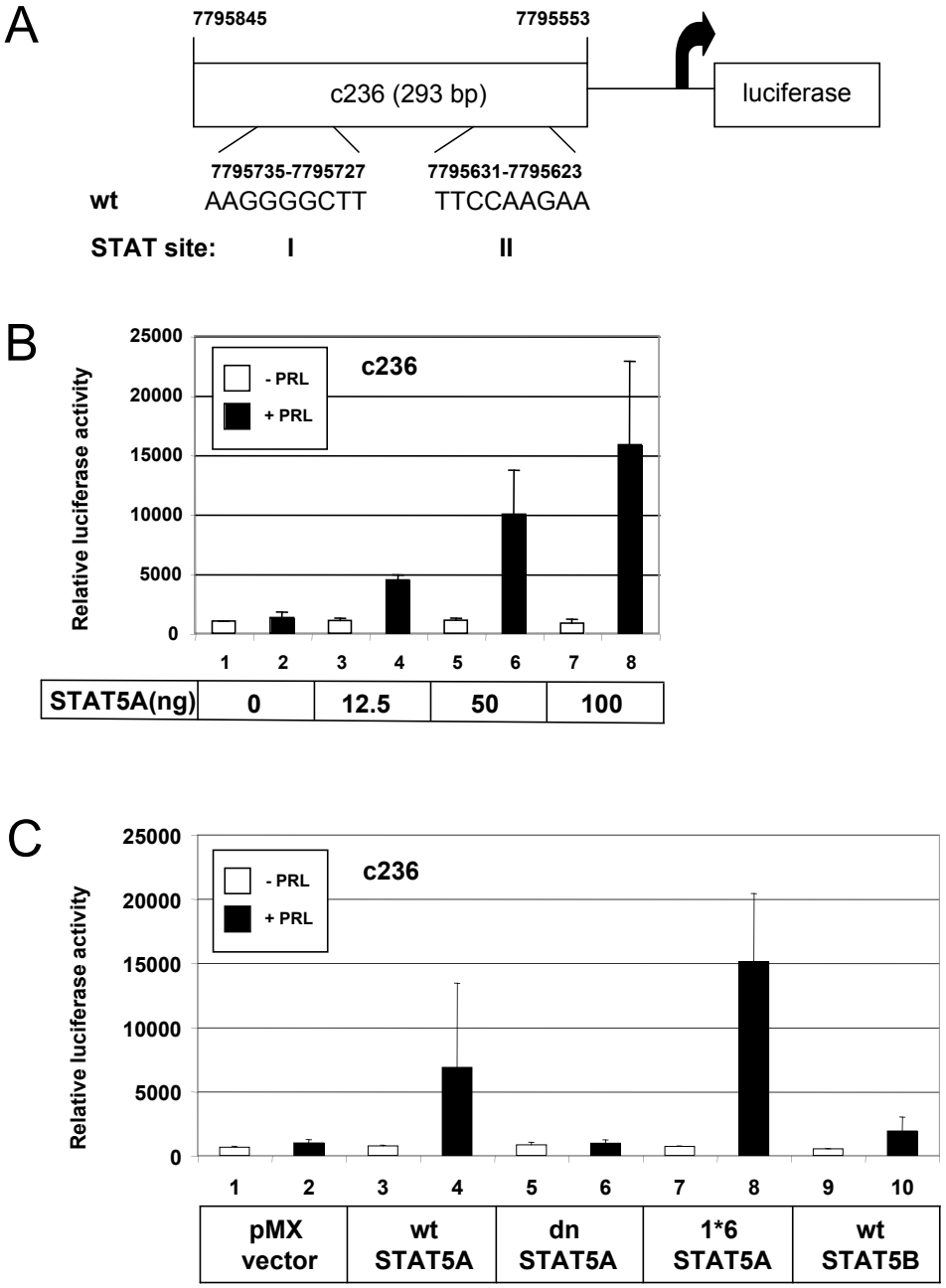


Figure 3.23 The c236 region is responsive to different concentrations of STAT5A and mutants of STAT5 affect its transcriptional activity.

(A) Schematic of the pGL3 basic reporter construct containing the 293 bp c236 region. The indicated nucleotide positions for the genomic location of the c236 region and of the two STAT-binding sites are based on the NCBI reference NT_165760.1/Mn5_163274_36 chromosome 5. The two wild type (wt) STAT-binding sites are depicted.

(B) HeLa TA cells were transfected with the c236 (293 bp) region reporter construct (1.25 µg) together with either an empty control vector (pMX) or with increasing amounts of the wt STAT5A expression construct (12.5 ng, 50 ng and 100 ng). HeLa TA cells in duplicate samples were also co-transfected with the prolactin receptor (12.5 ng) and the SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection the cells were either treated with prolactin (5 µg/ml) (dark bars) or left un-treated (white bars) for 16 hours. After cell lysis the transcriptional activity of the luciferase reporter gene was measured and normalized to ß-galactosidase activity. Each luciferase assay was repeated two times.

(C) The c236 region reporter construct (c236) (1.25 µg) together with either an empty control vector (pMX) (100ng) or with one of the following STAT5 expression constructs (100ng): wild type STAT5A (wt STAT5A), dominant negative STAT5A (dn STAT5A), constitutively active 1*6 STAT5A (1*6 STAT5A), or wild type STAT5B (wt STAT5A) (Fig. 3.8) was transiently transfected into HeLa TA cells as described above (B).

3.2.6.5 The region of c236 confers transcriptional activity that depends on the more conserved STAT5-binding site

To show that the c236 region is functional, the 293 bp fragment was inserted into the pGL3 basic luciferase reporter construct (Fig. 3.23A) and was tested in transient transfection assays with or without wt STAT5A expression construct. Both orientations of the 293 bp conserved c236 region were tested in the luciferase assay for reporter activity (Fig. 3.22A and C). The forward orientation showed only a 2-fold increase in luciferase activity in response to PRL treatment (lanes 3 and 4), while the reverse orientation responded to PRL induction with a 4-fold increase in luciferase activity (lanes 7 and 8). The c236 region in reverse orientation was chosen for further experiments, even though the overall luciferase activity was less intense than for the forward orientation, because the background luciferase activity was lower and the results more consistently reproducible (Fig. 3.22C).

To examine if the c236 region is responsive to different amounts of STAT5, reporter activity was tested with increasing concentrations of wt STAT5A expression constructs (Fig. 3.22B). Increasing amounts of exogenous STAT5A result in a corresponding increase in luciferase activity of the c236 region reporter construct (lanes 2, 4, 6 and 8). Samples not induced with prolactin remained unchanged at a basal level (lanes 1, 3, 5 and 7). These results indicate that the 293 bp fragment of the c236 region is transcriptional responsive to STAT5A in a concentration dependent manner.

To find out if this reporter activity is due directly to STAT5A, a dominant negative and a constitutive active STAT5A mutant was tested for their effect on the c236 region transcriptional activity. The c236 region reporter construct was transfected into HeLa TA cells with the different mutant STAT5 expression constructs (Fig. 3.23C). Addition of the control expression vector pMX resulted in no change of the transcription activity when induced with prolactin, while addition of wt STAT5A resulted in an increase of luciferase activity (lanes 1-4). No luciferase activity of the c236 region reporter construct could be induced with prolactin when the dominant negative dnSTAT5A mutant was co-transfected (lanes 5 and 6). Since the dnSTAT5A mutant is defective in activating transcription, the results suggest that a functional STAT5A must be present to induce luciferase activity of the c236 region reporter construct in response to prolactin. Co-transfection with the constitutive active 1*6 STAT5A expression construct resulted in a higher than wt STAT5A luciferase activity (compare lanes 4 and 8). As shown for the other reporter constructs, the constitutive active 1*6 STAT5A construct was not able to efficiently induce luciferase activity without prolactin induction.

ChIP experiments have shown recruitment of STAT5A as well as a weaker recruitment of STAT5B to the c236 region in mammary gland tissue from lactating mice *in vivo*

(Fig. 3.10). Therefore, STAT5B was tested for its ability to mediate transcriptional activity of this region in the reporter gene assay. Co-transfection with the wt STAT5B resulted in an almost negligible increase in luciferase activity when compared to the prolactin induced response of the control expression vector (Fig. 3.23C, lanes 2 and 10). The c236 region continually displays a weaker luciferase activity than the other clone regions (c153 or c831) tested (compare Fig. 3.23B with Fig. 3.16B and Fig. 3.25B). This weaker reporter activity matches also the lower amount of STAT5A recruitment (% Input DNA) to the c236 region in the ChIP assays (Fig. 3.10E). Nevertheless, the data indicates that the c236 region is transcriptionally responsive to STAT5A. The involvement of STAT5B remains unclear, due to the weak activity observed.

3.2.6.6 The evolutionary more conserved STAT5-binding sites in the c236 region is necessary for transcriptional activity

To determine the contribution of each the two STAT5 binding sites towards the transcriptional activity observed for the c236 region, each site was mutated and then tested in luciferase assay. The last two nucleotides in each STAT5 binding site were mutated from the wild type AA to TT (Fig. 3.24A). For the unmodified c236 reporter construct there is a 2,5-fold increase in luciferase activity upon prolactin stimulation when compared to the pGL3 basic control reporter construct (Fig. 3.24B, lanes 2 and 4). Mutation of the less conserved STAT5 binding site I (M1) resulted in a surprising, but reproducible increase of prolactin induced luciferase activity (lane 6). A search with TESS^{TM} (Transcription Element Search System), a web tool for predicting transcription factor binding sites in DNA sequences, shows that the mutation of the STAT5 M1 site does not create any additional known transcription factor binding site, but instead results in the loss of an overlapping NF-1 (-like proteins) binding site. The loss of this site may result in the increased luciferase activity observed for the M1 mutation, but further study would be needed to elucidate the actual cause. Mutation of the more conserved STAT5-binding site II (M2) resulted in a strong reduction of prolactin induced luciferase activity (Fig. 3.24B, lane 8).

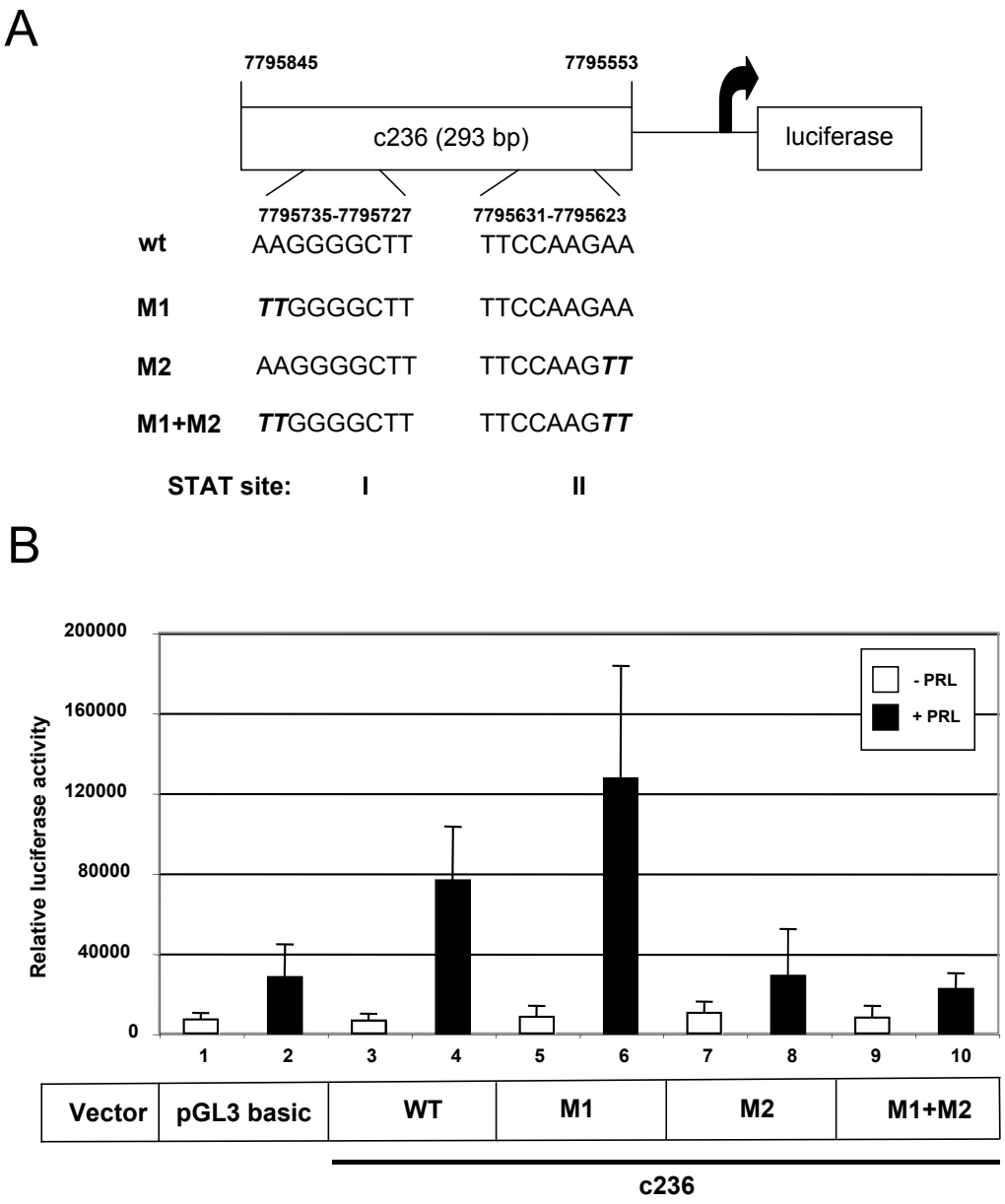


Figure 3.24 Mutation of the more conserved STAT5-binding site within the c236 region strongly reduces its transcriptional activity.
(A) The nucleotide positions for the genomic location of the 293 bp c236 region and of the two STAT-binding sites are based on the NCBI reference NT_165760.1/Mn5_163274_36 chromosome 5. The two wild type (wt) STAT-binding sites are indicated, as well as the nucleotides mutated in each of the mutant (M) constructs.
(B) Duplicate samples of HeLa TA cells were transiently transfected either with the empty control reporter vector (pGL3 basic) (1.25 µg), with the reporter construct containing the wild type c236 region (c236 (293 bp)) (1.25 µg), or the c236 region with mutations in one or both of its STAT5 binding sites (1.25 µg). The c236 region reporter constructs with either mutations in the STAT5-binding site I (M1), mutations in the STAT5-binding site II (M2) or mutations in both STAT5-binding sites (M1+M2) are depicted. The cells were also co-transfected with the wt STAT5A expression construct (100 ng), prolactin receptor (12.5 ng) and SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection half of the sister wells were either treated with prolactin (5µg/ml) for 16 hours (dark bars) or left un-treated (white bars). The Luciferase activity was determined and normalized to ß-galactosidase. Each luciferase assay was repeated two times.

Mutation of both STAT5-binding sites (M1+M2) together resulted in a clearly reduced luciferase activity as seen also for the M2 mutation (compare lanes 8 and 10). In the double mutant the original increase in activity caused by the M1 mutation had no influence on the loss of activity caused by the M2 mutation. Therefore the M1 mutation is not able to offset or rescue the loss of luciferase activity caused by the M2 mutation. This suggests that the conserved STAT5-binding site II of the c236 region confers the observed transcriptional activity and mutation of this site strongly reduces this activity.

3.2.6.7 The Zswim6 intronic region (c831) confers transcriptional responsiveness

To show that the STAT5 consensus site found within the Zswim6 intron is functional, a 740 base pair region containing the STAT5-binding site was inserted upstream of the pGL3 basic luciferase reporter gene (Fig. 3.25A). The orientation of the mouse Zswim6 intronic region corresponds to the natural position found within the Zswim6 gene (Fig. 3.15H). Therefore the other orientation was not tested. To determine if the c831 region is responsive to STAT5A, luciferase reporter gene assay was performed in transiently transfected HeLa TA cells with the Zswim6 region reporter construct and increasing amounts of wild type STAT5A expression vector (Fig. 3.25B).

Without co-transfected exogenous STAT5A, no significant change in transcriptional activity was observed for the Zswim6 reporter construct induced with prolactin (lanes 1 and 2). This suggests that endogenous levels of STAT5 in HeLa TA cells may not be sufficient to induce transcriptional activation of the Zswim6 intronic region (c831) reporter construct. When induced with prolactin, increasing amounts of exogenous STAT5A expression vector result in a corresponding gradual increase of transcriptional activity by the Zswim6 reporter construct, while untreated samples remain unchanged at a basal level (Fig. 3.25B, lanes 3-8). This shows that the Zswim6 intronic region is transcriptionally responsive to STAT5A in a concentration dependent manner.

To determine if this reporter activity is specifically due to STAT5, dominant negative and constitutive active forms of STAT5 were also tested (Fig. 3.17 and Fig. 3.25C). Equal amounts of these different forms of STAT5 were individually transfected in combination with the Zswim6 reporter construct into HeLa TA cells. The control pMX backbone vector did not exceed basal levels whereas the addition of wild type STAT5A resulted in a 6-fold induction upon prolactin treatment (Fig. 3.25C, lanes 1-4).

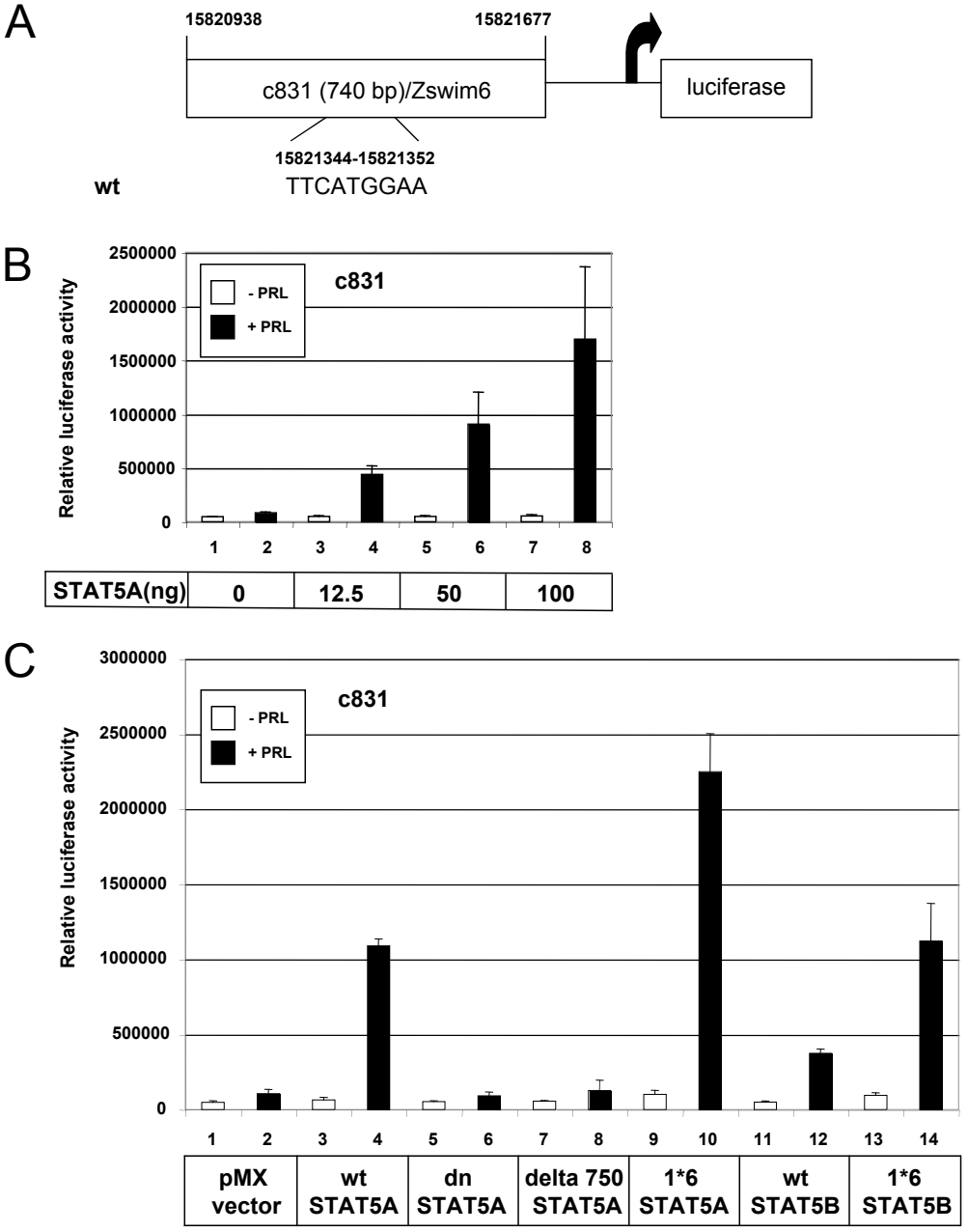


Figure 3.25 The Zswim6 intronic region (c831) is transcriptional responsive to STAT5 and STAT5 mutants affect this transcriptional activity.

(A) Schematic representation of the reporter construct containing the Zswim6 intronic (c831) region. The indicated nucleotide positions for the genomic location of the 740 base pair Zswim6 intronic region and of the STAT5-binding site are given based on the NCBI reference NT_039590.6/Mm3_39630_36 chromosome 13. The wild type (wt) STAT5 binding site is shown.

(B) Duplicate samples of HeLa TA cells were transiently transfected with the wild type Zswim6 intronic region luciferase reporter construct (c831 (740 bp)) (1.25 µg) together with either an empty control vector pMX or with increasing amounts of wt STAT5A expression constructs (100 ng). The cells were co-transfected with the prolactin receptor (12.5 ng) and the SV-40_*LacZ* expression construct (12.5 ng). 24 hours after transient transfection the cells were either treated with prolactin (5 µg/ml) (dark bars) or left un-treated (white bars) for 16 hours. After cell lysis the transcriptional activity of the luciferase reporter gene was measured and normalized to ß-galactosidase activity.

(C) The wild type Zswim6 intronic region (c831) (1.25 µg) together with either an empty control vector (pMX vector) (100ng) or with one of the following STAT5 expression constructs (100ng) (Fig. 3.8): wild type STAT5A (wt STAT5A), dominant negative STAT5A (dn STAT5A), dominant negative delta 750 STAT5A (delta 750 STAT5A), constitutively active 1*6 STAT5A (1*6 STAT5A), wild type STAT5B (wt STAT5A), constitutively active 1*6 STAT5B (1*6 STAT5B), was transiently transfected into HeLa TA cells as described above (B).

This transcriptional activity of the Zswim6 reporter construct was lost and remained at basal levels when either the dominant negative mutant dnSTAT5A or the Δ750 STAT5A mutant was co-transfected (lanes 5-8). This data indicates that functional STAT5 must be present to obtain a prolactin induced transcriptional activation of the Zswim6 intronic region reporter construct. The constitutive active 1*6 STAT5A mutant results in a doubling of the wt STAT5A response (compare lanes 4 and 10). Contrary to the expectation this mutant did not act in a classical constitutive active manner because prolactin stimulation for efficient transcriptional activation of the reporter construct was still required (lanes 9 and 10).

Since ChIP experiments showed that STAT5B was also recruited to the c831/Zswim6 intronic STAT5 site *in vivo* (Fig. 3.10F), the question arises if STAT5B can also activate the reporter activity of the Zswim6 intronic region construct. Therefore HeLa TA cells were transfected with either wild type STAT5B or the constitutive active 1*6 STAT5B mutant together with the Zswim6 reporter construct (Fig. 3.25C). The Zswim6 reporter construct responded in comparison to the control vector with a 4-fold increase in luciferase activity when co-transfected with wild type STAT5B and a 6-fold increase in luciferase activity when co-transfected with the constitutive active 1*6 STAT5B mutant (compare lanes 2, 12 and 14). Although weaker than STAT5A the results suggests that STAT5B can also directly activate the reporter activity of the Zswim6 intronic region in response to prolactin stimulation. Overall the data shows that the Zswim6 intronic region is transcriptionally responsive to STAT5A and STAT5B.

3.2.6.8 The consensus STAT5-binding site in the Zswim6 intronic region confers reporter gene activity

To determine the contribution of the STAT5-binding site to the transcriptional activity of the Zswim6 intronic region the STAT5A-binding site was mutated. The first two nucleotides in the STAT5 site were mutated from the wild type TT to AA (Fig. 3.26A) and tested in luciferase assays to determine its effect on the transcriptional activity of the Zswim6 reporter construct (Fig. 3.26B). In comparison to the pGL3 basic control reporter construct the wild type construct, resulted in a 12-fold increase in luciferase activity upon prolactin stimulation (lanes1-4). Mutation of the STAT5-binding site resulted in a clear 4-fold reduction of prolactin induced luciferase activity (Fig. 3.26B, compare lanes 4 and 6). This suggests that the STAT5-binding site of Zswim6 intronic region confers transcriptional activity and mutation of this site strongly reduces this activity.

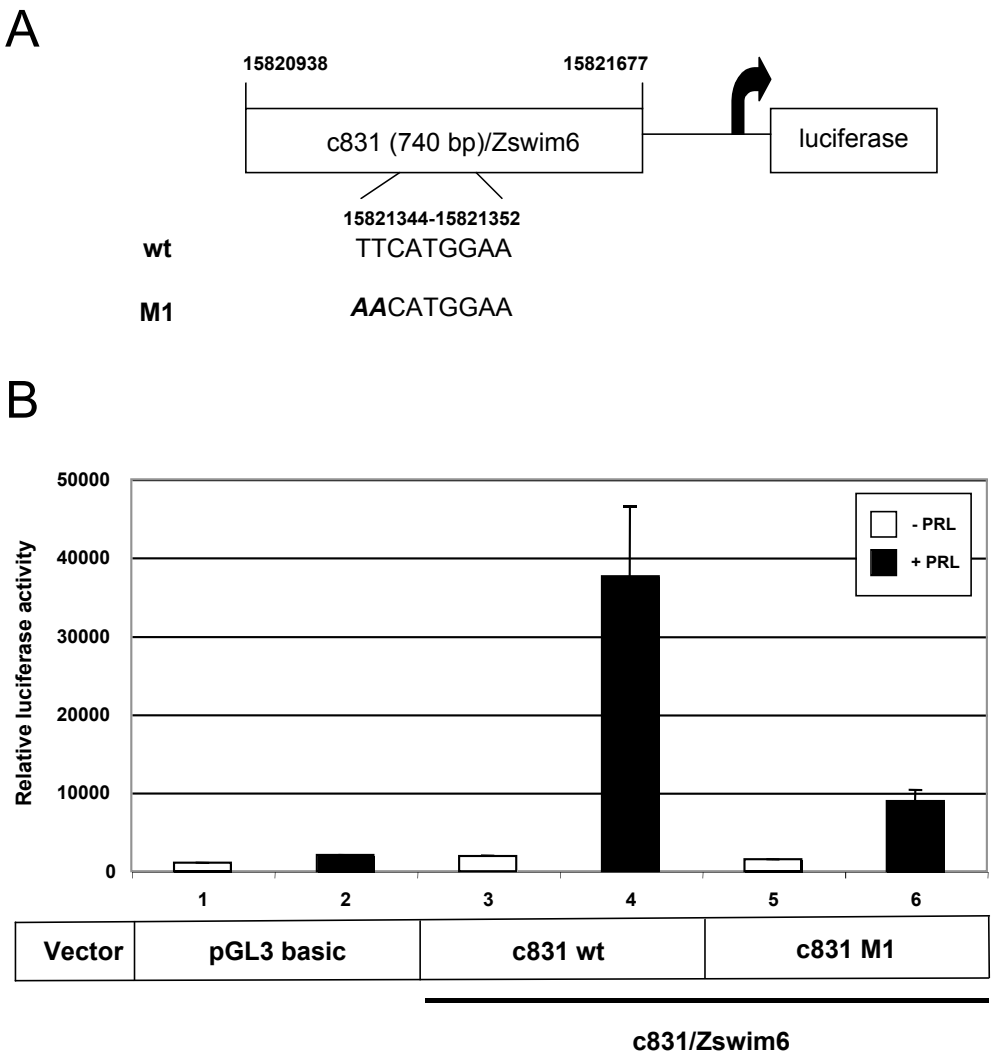


Figure 3.26 Mutation of the STAT5-binding site within the Zswim6 intronic region reduces its transcriptional activity.
(A) Schematic representation of the luciferase reporter construct containing the Zswim6 intronic region. The nucleotide positions for the genomic location of the 740 base pair Zswim6 intronic region and of the STAT5-binding site are based on the NCBI reference NT_039590.6/Mm3_39630_36 chromosome 13. The wild type (wt) STAT5 binding site is indicated, as well as the nucleotides mutated in the mutant (M1) construct
(B) Duplicate samples of HeLa TA cells were transfected either with empty vector (pGL3 basic) (1.25 µg), with the luciferase construct containing the wild type Zswim6 intronic region (c831 wt) (1.25 µg), or the Zswim6 intronic region with mutations in the STAT5 consensus site (c831 M1) (1.25 µg). The HeLa cells were also co-transfected with the wt STAT5A expression construct (100 ng), the prolactin receptor (12.5 ng) and the SV-40_*Lac*Z expression construct (12.5 ng). 24 hours after transient transfection half the sister wells were either treated with prolactin (5µg/ml) for 16 hours (dark bars) or left un-treated (white bars). The luciferase activity was determined and normalized to un-treated for each construct. Each luciferase assay was repeated two times.

Taken together, the previous reporter gene analysis shows that the investigated cloned regions can initiate transcription. The reporter gene activity is mediated by STAT5 and depends mainly on STAT5-binding motifs within the cloned regions (c153, c236 and c831) analyzed. This suggests that these genomic elements may be involved in the transcriptional regulation of endogenous genes. Therefore, the expression of transcripts associated with the cloned regions was analyzed for changes in expression during the different mammary gland stages.

3.2.7 Analysis of RNA transcripts found at or near the cloned STAT5-binding site regions

The regulation of the genome is more complex than initially thought. Many genes have multiple promoters within which there are multiple transcription start sites (TSS) (Sandelin et al., 2007; Birney et al., 2007; Carninci et al., 2006, 2005). The differential, often tissue specific usage of the promoters and TSS and the expression of coding and non-coding RNA's result in further layers of complexity.
For the cloned STAT5-binding regions there are multiple possibilities of transcripts that could potentially be regulated. These transcripts vary between the different available computational databases. To determine if the association of STAT5 with the novel binding sites has functional consequences, the levels of the nearest known RNA transcripts were investigated for the following three clones c153, c236 and c831 (Table 3). The transcript regulation in the mammary gland during the four differentiation stages of pregnancy, lactation, involution and in virgin mice was determined by quantitative real-time PCR.

3.2.7.1 The expression of known STAT5 target genes and housekeeping genes during different mammary gland developmental stages

The obtained data was normalized to the geometric average (Vandesompele et al., 2002) of three housekeeping genes (HKGs), S18 ribosomal RNA (18SRNA) transcript, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ) and cyclophilinB within the same samples. This was done to ensure that any variation within the HKG during the mammary gland lactation cycle is compensated. The general variance found for each individual HKG was maximal about 2.5-fold between the different stages (Fig. 3.27A, B and C). The variance, seen between each individual HKG normalized against the geometric average of all 3 HKG, was defined as an indicator of experimental fluctuation. Any variation observed that is

below 3-fold will not to be considered significant. However, one has to keep in mind that with this stringent method of normalization very low changes in RNA expression may not be detected. The relative RNA levels found in the virgin mammary gland stage is designated in these experiments as the undifferentiated control and is set arbitrarily to 1. As positive controls, the three known STAT5 target genes ß-casein, WAP and SOCS3 were tested for their mRNA expression levels in the four different mammary gland developmental stages (Fig. 3.27D-F).

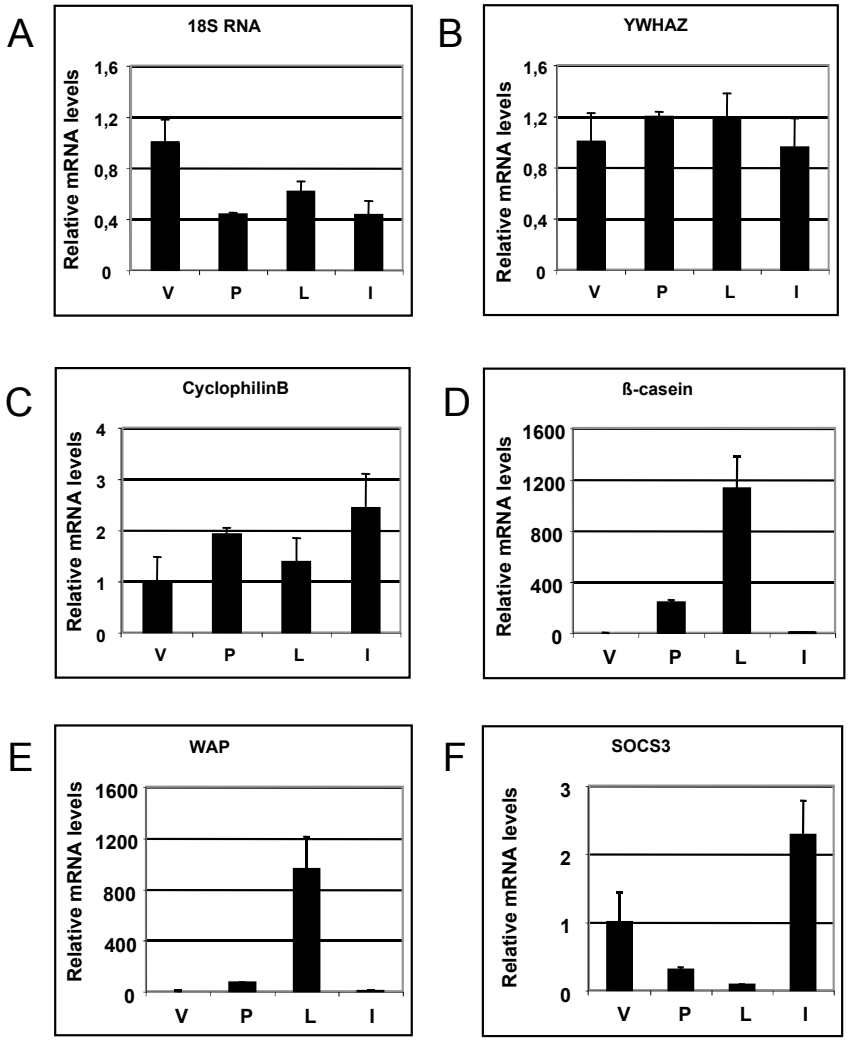


Figure 3.27 Expression analysis of housekeeping genes and known STAT5 target genes during mammary gland development.
Transcript expression levels were determined by quantitative real-time PCR in samples derived from Balb/c mouse mammary gland tissue of four different developmental stages. The developmental stages are indicated at the horizontal axis of each panel. Designation of stages: V, adult non-pregnant female (Virgin); P, approximately 10-15 days post conceptum (pregnancy); L, day 5 of lactation (lactation); I, adult females that weaned from 4 to 6 weeks old pups (involution). 2 µg of total RNA prepared from homogenized mammary gland tissue was reverse transcribed into cDNA and then the relative transcript amounts were quantified in triplicates by real-time PCR using the specified primers. The relative RNA level for each transcript was normalized against the geometric average of the following 3 housekeeping genes.
(A, B and C) The three housekeeping genes 18S RNA, YWHAZ and cyclophilinB were used as controls for data normalization.
(D, E and F) The expression of the known STAT5 target genes ß-casein, WAP and SOCS3 are shown as positive controls. The results shown are representative of the trends observed in two independent experiments.

For the two milk proteins, ß-casein and WAP, the analysis revealed the expected elevated levels during pregnancy and lactation. At the Involution stage the expression levels approached the minimal detected levels found in virgin mice (Fig. 3.27D and E). In contrast to the milk proteins, SOCS3 displayed a different expression pattern (Fig. 3.27D, E and F). For SOCS3, the expression levels were elevated in virgin mice and even higher during involution when compared to pregnancy and lactation. These results are consistent with a detailed study on the regulation of SOCS3 transcripts during the different stages of mouse mammary gland development (Tonko-Geymayer et al., 2002). To note is that the overall relative RNA level for SOCS3 is significantly lower than for the positive controls ß-casein and WAP. This may be the result of a weaker expression or a more rapid turnover of the SOCS3 transcripts in comparison to the milk protein transcripts.

3.2.7.2 The expression of NFI/B transcript correlates with the *in vivo* association of STAT5A to the conserved c153 region

The conserved c153 region containing the STAT-binding sites is not located at a classic promoter or within a known gene (Table 3). A search program used to identify any possible transcription start sites (TSS) in vicinity of the c153 region did not identify any TSS in the sequence 2000 bp upstream or downstream of the c153 region (Dragon program for TTS search). A good indicator of promoters is also the presence of nearby CpG islands (Bajic et al., 2004). The nearest CpG island that could be identified is at the 5 prime end of NFI/B gene and according to UCSC browser miRBase search there are also no annotated microRNAs in close vicinity (UCSC bioinformatics browser). Therefore, further investigation was focused on already annotated transcripts found near the c153 region.

Initial searches with the NCBI databases to identify already annotated transcripts in the region were performed. Based on NCBI BLAST searches within the mouse, the c153 region is located 110 kb 5 prime of a hypothetical protein Loc76455 (NCBI reference assembly NT_039260.6/Mm4_39300_36), which is a shorter transcript associated with the NFI/B gene (Table 2). The alternative Celera reference assembly (NW_001030747.1/Mm4_118890_36) mapped the c153 region 62 kb from the 5' side of Nuclear Factor I/B locus. The results from the analysis of the mouse genome database match approximately the 65 kb distance that the conserved human c153 region is found located 5' from the human NFI/B gene (reference assembly NT_008413.17/Hs9_8570).

Using the UCSC bioinformatics browser, subsequent searches revealed an expressed sequence tag (EST) named AW106080, which spans the conserved c153 region (Fig. 3.28A). This AW106080 transcript, was originally identified in a Sugano mouse embryo mewa cDNA library and is comprised of two transcripts. EST AW106080 aligns to Chr4 (81969552-82176886) on the UCSC genome browser. The transcript part 1 is located at Chr4 (82176886-82176746) and transcript part 2 Chr4 (81970000-81969552). The c153 region is found between the two transcripts at Chr4 (82038634-82038298); with STAT5 site TTCATAGAA located at Chr4 (82038478-82038469). Thus the AW106080 EST sequence flanks the c153 site, but there is still a 138 kb separation between cDNA part 1 and 68 kb from cDNA part 2. In GenBank the accession AW106080 is a 5' EST that is linked to the cDNA clone (ID IMAGE:2225331) and is associated with the putative identity of a mouse mRNA NFI-B2 isoform. Within the Ensembl database, AW106080.1 is listed as supporting evidence for the Vega (Vertebrate genome annotation database) annotated transcript Nfib-010, with genomic location at Chr4 (81969552-82176981). This transcript is also classified as a product belonging to the NFI/B gene. Here the c153 region and STAT5-binding site is located within the first intron of the NFib-010 transcript. Apart from the mouse AW106080.1 transcript there is also an additional comparative human mRNA (Homo AK131233) transcript, which also flanks the c153 region with exon 1 and 2 (Fig. 3.28A). This human mRNA is located at Chr9 (14077860-14388982) within the human genome and was found in a cDNA library from amygdala brain tissue. AK131233 is considered very similar to nuclear factor 1/B according to NCBI nucleotide sequence database. Taken together, the information indicates that the conserved c153 Stat5-binding site is in approximate 62 kb proximity of NFI/B, and is spanned (or within the putative first intron) by a NFI/B associated transcript. This is the case both in the mouse and human, supporting the idea that this sequence maybe important and possibly involved in NFI/B regulation. Since NFI/B is the closest known gene found near the conserved c153 region/STAT5-binding site, the question arose if NFI/B is differentially regulated during the different mammary gland developmental stages investigated.

To determine if the different transcripts are expressed in mammary gland tissue RT-PCR primers were designed for two different transcripts (NFI/B and AW106080) to test the expression levels in the four mammary gland developmental stages. For NFI/B a 4- to 7-fold increase in relative RNA expression was observed during pregnancy and lactation when compared to virgin mice (Fig. 3.28B). An even stronger 12-fold increase in NFI/B relative RNA levels was observed during the stage of pregnancy and lactation when compared to Involution (Fig. 3.28B).

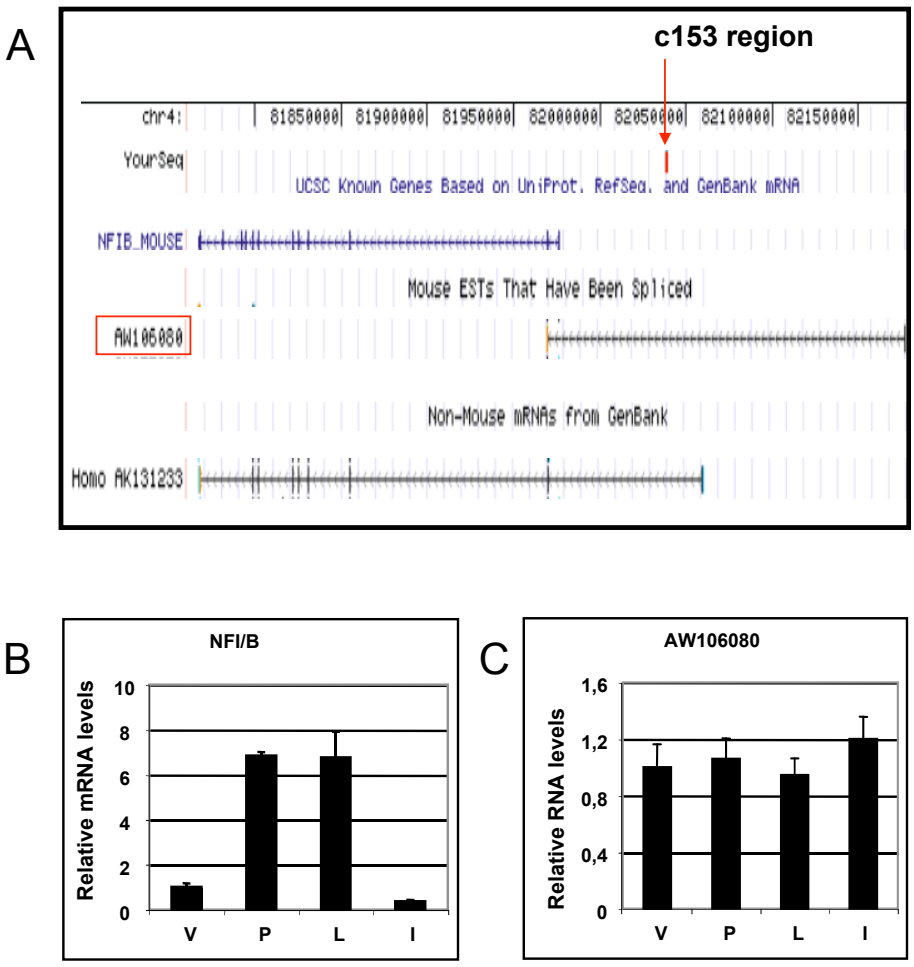


Figure 3.28 Expression analysis of transcripts associated with the c153 region.
(A) Excerpt from UCSC Genome Bioinformatics Browser (BLAT search details) window showing different transcripts spanning or in close proximity to the c153 region (location marked by vertical arrow).
(B) Relative expression level of the mouse mRNA NFI/B.
(C) Relative expression level of the mouse EST transcript AW106080.
Transcript expression levels were determined by quantitative real-time PCR in samples derived from Balb/c mouse mammary gland tissue of four different developmental stages. The developmental stages are indicated at the horizontal axis of each panel. Designation of stages: V, adult nonpregnant female (Virgin); P, approximately 10-15 days post conceptum (pregnancy); L, day 5 of lactation (lactation); I, adult females that weaned from 4 to 6 weeks old pups (involution). 2 µg of total RNA prepared from homogenized mammary gland tissue was reverse transcribed into cDNA and then the relative transcript amounts were quantified in triplicates by real-time PCR using the specified primers. The relative RNA level for each transcript was normalized against the geometric average of the 3 housekeeping genes (18S RNA, YWHAZ and CyclophilinB). The results shown are representative of the trends observed in two independent experiments.

This data shows that NFI/B relative mRNA levels increase in a similar pattern, as do the milk proteins during the differentiation specific stages of mammary gland development. The expression of NFI/B correlates nicely with the *in vivo* association of STAT5A to the conserved c153 region during the same mammary gland development stages, pregnancy and lactation. Taken together with the functional *in vitro* data using

reporter gene assays the results are suggestive that the conserved c153 region via STAT5 may be involved in regulating NFI/B.

For the AW106080 transcript no more than a 2.5-fold variance between the different conditions was observed (Fig. 3.28C). This indicates, that there is no more variation in AW106080 transcript expression between the different developmental stages than has been observed for the HKGs. In addition the AW106080 transcript appears to be present already in virgin mice and in the subsequent pregnancy, lactation and involution stages tested. The cycle threshold (Ct) values reached in these quantitative RT-PCR experiments, during the exponential phase with the AW106080 primers, are comparable in level to those found for the housekeeping gene YWHAZ (Ct 19-21). These results suggest that the AW106080 transcript is present, but appears not to be regulated during differentiation.

3.2.7.3 Multiple putative transcripts are found associated with the c236 region

Using the UCSC browser multiple mRNAs and ESTs were located that span or are in close proximity to the c236 region/STAT5 binding site. The following two mouse mRNAs Ak017143 and Ak042126 from GenBank and the 4 mouse ESTs BU610001, BY716854, BB236757 and BB638939 span the c236 region (Fig. 3.29A). The mRNA Ak042126 and the two ESTs BB236757 and BB638939 appear to group together in a similar region that is distinct to the region (exons) covered by the mRNA Ak017143 and the EST BU610001. In addition, there are two transcripts, one mRNA (AK013579) and one EST (CA979556), that are in close proximity (130-500 bp) to, but do not span the c236 region. The mouse mRNA, Ak017143 was identified in a cDNA library made from pregnant mouse ovary and uterus tissue. The product has not been classified yet. Primers designed against mRNAs Ak017143 failed in the quality control, but primers of the EST BU610001, which spans a similar region to mRNAs Ak017143, passed quality control. In quantitative RT-PCR these primers showed high Ct values (31-33) suggesting that there is maybe very little cDNA template available in the samples. Also there is no consistent trend discernable between the different experiments (Fig. 3.29B). Taken together, this indicates that hardly any template cDNA of BU610001 is available in the samples, making it difficult to ascertain difference between the stages. Due to the different types of annotated transcripts found around the c236 region additional primers need to be designed and tested, that also cover these different mRNA and ESTs. Especially, candidates like the mRNA Ak042126 need to still be covered, since the comparative expression data from different tissues indicate their elevated presence

in mammary gland tissue (S.O.U.R.C.E search, BC050254, EST expression information).

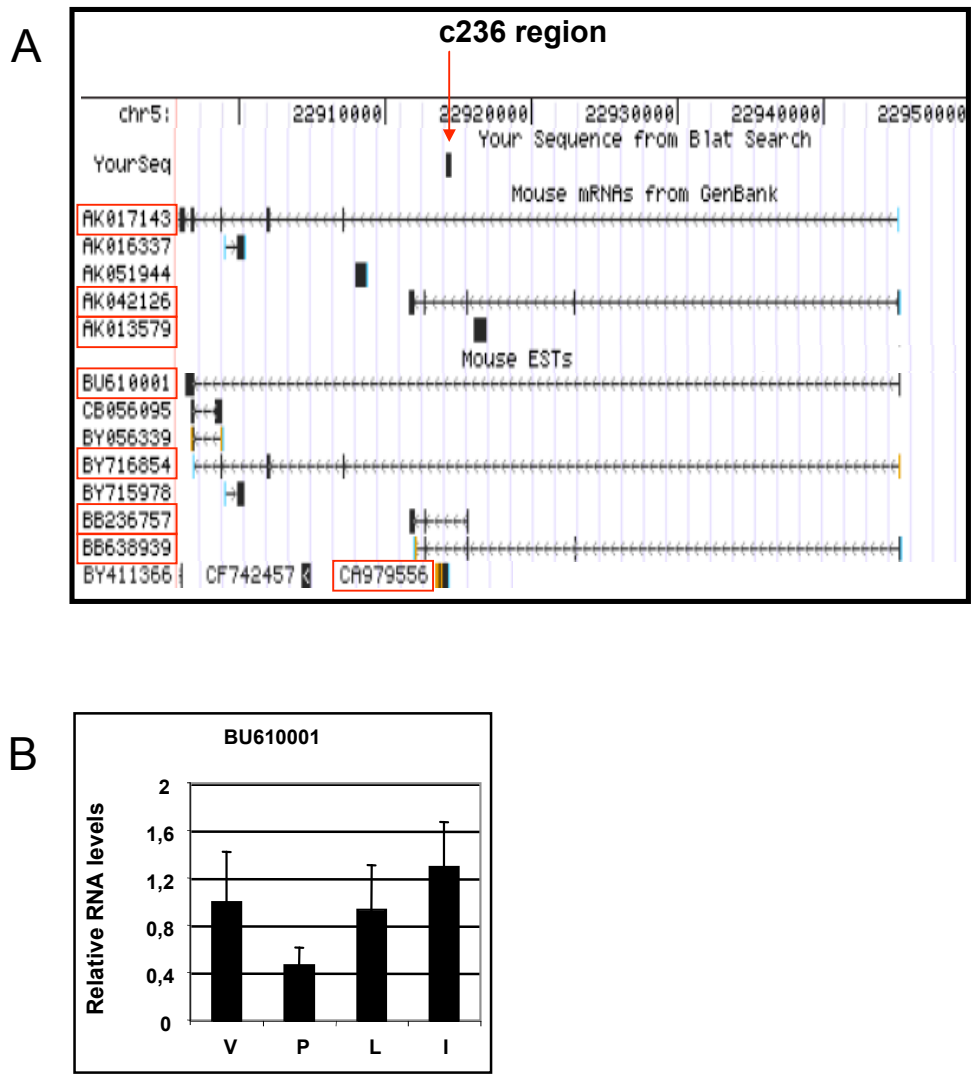


Figure 3.29 Expression analysis of transcripts associated with the c236 region.
(A) Excerpt from UCSC Genome Bioinformatics Browser (BLAT search details) window showing different transcripts spanning or in close proximity to the c236 region (location marked by vertical arrow).
(B) Relative expression level of the mouse EST transcript BU610001.
Transcript expression levels were determined by quantitative real-time PCR in samples derived from Balb/c mouse mammary gland tissue of four different developmental stages. The developmental stages are indicated at the horizontal axis of each panel. Designation of stages: V, adult nonpregnant female (Virgin); P, approximately 10-15 days post conceptum (pregnancy); L, day 5 of lactation (lactation); I, adult females that weaned from 4 to 6 weeks old pups (involution). 2 µg of total RNA prepared from homogenized mammary gland tissue was reverse transcribed into cDNA and then the relative transcript amounts were quantified in triplicates by real-time PCR using the specified primers. The relative RNA level for each transcript was normalized against the geometric average of the 3 housekeeping genes (18S RNA, YWHAZ and CyclophilinB). The results shown are representative of the trends observed in two independent experiments

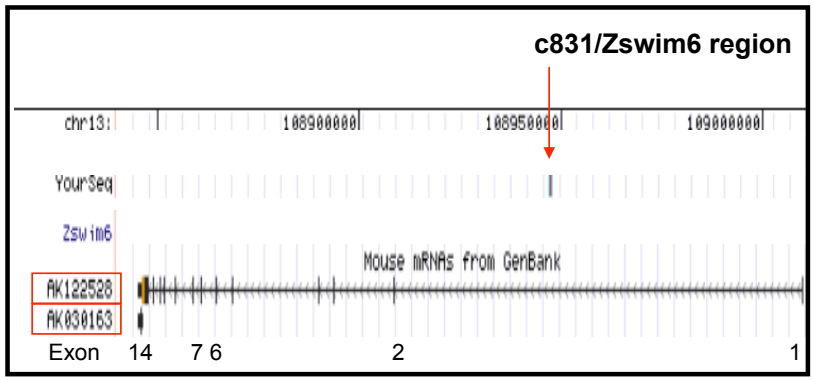
3.2.7.4 Multiple putative transcripts are found associated with c831/Zswim6

The clone c831 has a conserved STAT5-binding site that is located in the first intron of the Zswim6 gene. This first intron has a size of over 100 kb and the STAT5-binding site is located 62 kb downstream from the first exon of Zswim6. An example for functional regulatory STAT5-binding sites located within the first intron, about 200 kb away from the first exon, of a gene it regulates have been described in the literature for the NCAM2 gene (Nelson et al., 2006). Based on the data that STAT5 binds *in vivo* during lactation and pregnancy to the c831 STAT5-binding site region (Fig. 3.10F) and also confers transcriptional activity in reporter assays (Fig. 3.26B), the question arose if the association of STAT5 with this novel binding site has a functional consequence on the mRNA level of the Zswim6 gene during mammary gland differentiation.

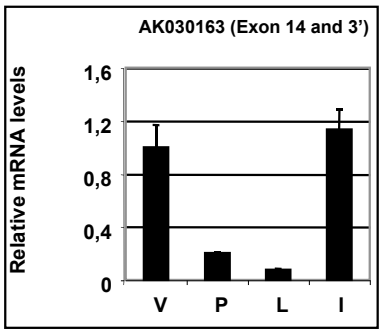
Primers were obtained based on the Zswim6 linked mRNA AK030163 from Primer Bank™. AK030163 mRNA has been identified in a cDNA library obtained from adult mouse testis. The primers cover exon 14 and the 3' end of the Zswim6 gene. In quantitative RT-PCR experiments the relative RNA levels of Zswim6 (AK030163) were always at least 4.5-fold higher in virgin mice and during involution compared to the pregnant and lactating stages. (Fig. 3.30B). The involution stage retained a strong increase over the pregnancy and lactation stages. These results indicate that the transcript (AK030163) is more strongly expressed during involution and in virgin mice and suggest that the expression levels might be downregulated in pregnant and lactating mammary gland tissue. Interesting to note is that transcript (AK030163) displays a similar trend in the expression pattern as found for SOCS3 (compare Fig. 3.30B and Fig. 3.27F). Also, like SOCS3 the overall relative RNA levels for Zswim6 are significantly lower than for the positive controls ß-casein and WAP (compare Fig. 3.30B-D and Fig. 3.27D-F). This may also be due, as proposed for SOCS3 in the study, to a weaker induction and or a more rapid turnover of the Zswim6 transcripts.

Because the transcript from AK030163 does not span the c831 region found in the first intron of Zswim6, two additional primer sets were designed based on the AK122528 transcript that spans the c831 region (Fig. 3.30A). The primers designated (XM3583115) and (AK122528) cover the exon 1 to 2 and exon 6 to 7 boundary of Zswim6, respectively. For XM3583115 the Ct values (30-36) were in a high range compared to the HKGs, suggesting that there is very little cDNA template available in the samples. But interesting is, that in the involution stage the relative RNA levels are decreased 3- to 4-fold in comparison to pregnancy and lactation (Fig. 3.30D).

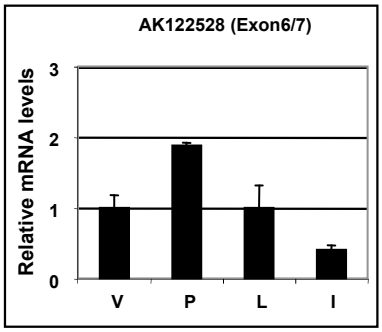
A



B



C



D

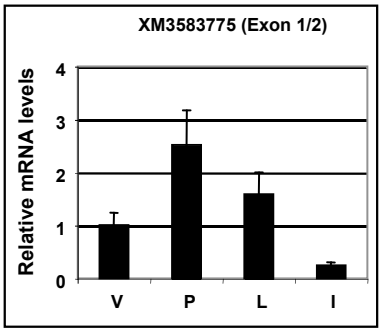


Figure 3.30 Expression analysis of transcripts associated with the c831 region.
(A) Excerpt from UCSC Genome Bioinformatics Browser (BLAT search details) window showing different transcripts spanning or in close proximity to the c831 region (location marked by vertical arrow).
(B-D) The relative expression level of the mouse Zwim6 mRNA: AK030163, AK122528 and XM3583775, respectively.Transcript expression levels were determined by quantitative real-time PCR in samples derived from Balb/c mouse mammary gland tissue of four different developmental stages. The developmental stages are indicated at the horizontal axis of each panel. Designation of stages: V, adult nonpregnant female (Virgin); P, approximately 10-15 days post conceptum (pregnancy); L, day 5 of lactation (lactation); I, adult females that weaned from 4 to 6 weeks old pups (involution). 2 µg of total RNA prepared from homogenized mammary gland tissue was reverse transcribed into cDNA and then the relative transcript amounts were quantified in triplicates by real-time PCR using the specified primers. The relative RNA level for each transcript was normalized against the geometric average of the 3 housekeeping genes (18S RNA, YWHAZ and CyclophilinB). The results shown are representative of the trends observed in two independent experiments.

For transcript AK122528 the Ct ranges between 25 and 30. Here there is no more than a maximal 3-fold change between the involution (lowest) and pregnancy (highest) stage (Fig. 3.30C). But the pattern is similar to the data obtained for the XM3583115 primers. These results contradict the patterns found by the Zswim6 primers

(AK030163). A detailed investigation into the possible Zswim6 mRNA isoforms/splice variants that are possibly expressed during mammary gland differentiation needs to be performed, to get any indications of potential regulatory function of STAT5 in this region.

3.2.8 Identification of novel over-represented DNA motifs within the STAT5A-ChIP enriched sequences

A major step in transcriptional regulation is the binding of the transcription factor to the target DNA in a sequence-specific manner.

Because the enriched DNA from the ChIP assay is at least 100 bp or larger this does not allow for the direct identification of the shorter transcription factor binding sites within the larger enriched DNA sequences. Known and defined binding sites can be identified, but variations from an already known binding site or novel binding sites would remain unnoticed within the ChIP-enriched sequences.

Often the binding specificities of some transcription factors is less pronounces and involve very short stretches of nucleotides (about 4-8 bp) that also may contain uninformative positions. Binding affinities of transcription factors are encoded as position weight matricies (PWM), where these matricies specify the sequence motif by allotting a base distribution for each of its positions.

The search of all the STAT5 ChIP-enriched sequences for the consensus GAS motif TTCNNNGAA using TESS allowed for the identification of 34 individual clone sequences (Table 4) containing exactly this motif. In this search of the non-redundant sequence data set 14% (34/236) of the sequences contained this stringent defined STAT5 consensus motif. Therefore, an analysis with Trawler (Ettwiller et al., 2007) was performed by Dr. Mirana Ramialison (Wittbrodt Group, Developmental biology Unit, EMBL in Heidelberg) to identify any over-represented transcription factor binding motifs within the STAT5A ChIP-enriched sequences.

To clearly identify over-represented DNA motifs within the ChIP-enriched sequences the sequences have to be compared to a background data set. This background data set was specifically built taking into account the distribution of the STAT5-ChIP enriched sequences in the mouse genome. Using a BLAST search alignment the genomic location of each clone sequence was mapped and its position in regard to surrounding annotated genes was determined. From the list of 236 clones that had been unambiguously mapped to the mouse genome, those that still contained highly repetitive sequences (example satellite DNA sequences) were removed. And any clone sequences that overlapped with each other were combined into one sequence contig.

Table 4 List of 34 clones containing the STAT5 sequence motif from the TESS analysis of the non-redundant ChIP-enriched sequence data set.
The list displays the 34 clones that contain STAT5-binding motifs in their sequence that conform exactly to the TTCNNNGAA motif as determined by using a TESS string search of the non-redundant 236 sequence data set. The total length (bp) of the DNA sequence searched is listed for each clone or contig. The individual STAT5 motifs found for each clone are shown and the location (bp) of the STAT5 motifs within each clone sequence are given.

Name	Length	STAT5 sites	Location
25	386	TTCAGGGAA	232
37	393	TTCT AAGAA	127
73	467	TTCAGAGAA	426
76	430	TTCTCGGAA	284
80	422	TTCAGAGAA/ TTCAGGGAA	141/ 339
90	394	TTCCCTGAA	293
98	429	TTCAAAGAA	336
153	282	TTCTATGAA	114
177	329	TTCAGGGAA	2
194	394	TTCTTGGAA/ TTCAAAGAA	107/ 328
236	383	TTCCAAGAA	351
237	432	TTCAGAGAA	242
243	371	TTCCAAGAA	356
270	410	TTCTGGGAA	245
315	427	TTCTCAGAA/ TTCTCCGAA	260/ 281
320	420	TTCCGAGAA	169
469	401	TTCAGAGAA	133
591	386	TTCTTCGAA	275
594	366	TTCTACGAA	172
600	318	TTCTCTGAA	279
611	160	TTCCAGGAA	137
612	442	TTCCTGGAA	222
719	425	TTCAGAGAA	328
831	463	TTCCATGAA	147
838	463	TTCAAAGAA/ TTCTATGAA	252/ 327
878	434	TTCCCAGAA	394
888	374	TTCATGGAA	93
901	464	TTCCCTGAA	402
906	453	TTCCTGGAA	434
946	460	TTCGTAGAA	162
Contig 2	422	TTCCAAGAA	107
Contig 3	1293	TTCTTAGAA	1027
Contig 6	206	TTCCAGGAA	79
Contig 9	467	TTCCTGGAA	36

This final dataset composed of a total of 198 sequences contains only non-redundant sequences that mapped to the mouse genome and non-repetitive DNA sequences (Appendix 9.1). The highly repetitive sequences had to be excluded for the subsequent Trawler analysis, since motif elements within the repetitive sequences would dominate over all other over-represented motifs in the sequences. Interesting to note is that almost 48% (96/198) of the clone sequences are located in the intron regions of annotated transcripts (Fig. 3.31A).

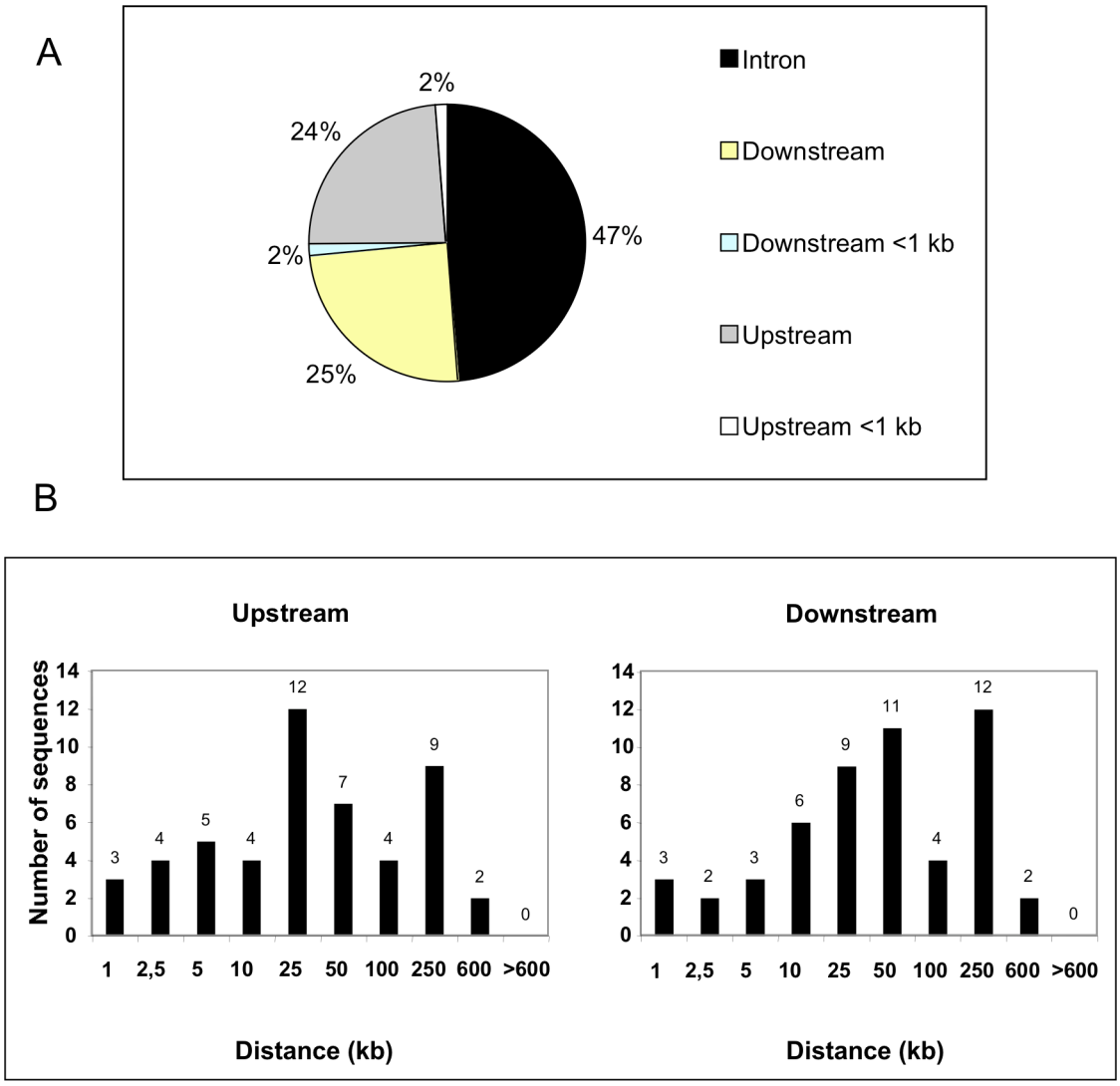


Figure 3.31 Distribution of the STAT5 ChIP-enriched fragments relative to the nearest annotated gene.
(A) The chart depicts in percentage the upstream or downstream location of the STAT5A ChIP-enriched fragments relative to annotated genes. For this purpose only the non-redundant and non-repetitive 198 sequence data set mapped by Dr. Ramialison using the Ensembl database was used. Sequences within genes were assigned to the intron category even if part of the clone sequence overlapped with an exon region of the gene.
(B) Distribution of the distance (kilobase pairs (kb)) the clone sequences are found located upstream or downstream from the nearest annotated gene.

The other half (102/198) is found distributed equally upstream or downstream from the nearest annotated gene. The distance of the clone sequences from the nearest gene varied. To note is that only about 2,9% (3/102) of the sequences are found located within 1 kb upstream or downstream from the nearest gene. The rest displayed a variable distributed over the next 600 kb distance from the nearest gene, with distances of around 25-50 kb and 100-250 kb being most abundant (Fig. 3.31B).

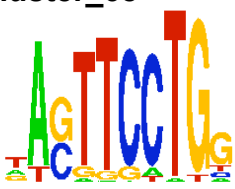
According to the distribution found for the sequences, an appropriate background data set was built composed of randomly picked sequences with similar genomic background.

The following parameters were used in Trawler Analysis (http://ani.embl.de/trawler/result/job_33704). The minimum number of times the motif in the sample sequences is observed was set to 10 occurrences. The minimum length of a motif had to be 9 bp. This was done since the consensus GAS STAT5 binding site is 9 bp long (TTCNNNGAA). Only a minimum number of 2 mismatches in the motif were allowed.

Overall, sixteen matrices in 151 clone sequences were found in comparison to the background to be significantly over-represented as determined by highest z-score values. The z-scores represent the standard deviation of the occurrence of the motif in the sample when compared to the random background dataset. Only those motifs with a significant high z-score are displayed by Trawler (Ettwiller et al., 2007). The over-represented matrices are classified into families and each family is subdivided into clusters depending on position weight matrices similarity. The family motif 68 received the overall highest z-score (z-score =11) and is subdivided into 5 cluster motifs (cluster 68, 79, 78, 80 and 83) (Fig. 3.32). The newly identified over-represented motifs from the ChIP-enriched sequences identified were also compared by Trawler with motifs already described for binding specificity of known transcription factor collected in database such as Transfac (Matys et al., 2003) and Jaspar (Vlieghe et al., 2006). For the motif family_68 the cluster_78 is clearly annotated as STAT-binding site, even though there are only 3-4 (GGAA) well-defined bases. This STAT resembling motif from cluster_78 though does not overlap with the classic GAS STAT5 sites (TTCNNNGAA) initially identified in, for example, the clone c153 sequence, but lie adjacent to these sites (Fig. 3.33A). Therefore the consensus STAT5 GAS binding site motif is not found over-represented as already shown by the results from the previous TESS search. In the motif family_68 the cluster_79 is similar to the c-REL binding site and the cluster_80 is annotated as bHLH and ETS-class binding site. The motif family_68_cluster_68 and cluster_83 do not resemble any motifs in the transcription factor databases and are thus novel over-represented motifs. Of the motif family_68 the cluster_68 is the most prominent cluster. This motif family_68_cluster_68 is found significantly over-represented and is found also in the ChIP re-confirmed clones c153 and c831 adjacent to the classic GAS STAT5-binding motif identified in these sequences (Fig. 3.33A-B).

Matrix family_68

cluster_68

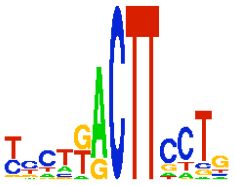


NAGTTCCTGG

Transfac and Jaspar hits

No similar matrices found.

cluster_79



TNCTKACTTCCTG

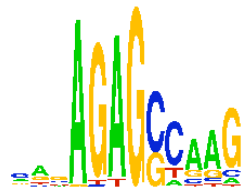
Name	Source	ID	Divergence	Consensus
	JASPAR	MA0023	0.602	GGGGWTTTCC
c-REL	JASPAR	MA0101	0.685	NGGGNTTTCC
CREL	TRANSFAC	M00053	0.686	NGGGNTTTCC
CP2	TRANSFAC	M00072	0.709	GCNCNANCCAG

cluster_78

AGCCAGGAANN

Name	Source	ID	Divergence	Consensus
STAT	TRANSFAC	M00223	0.694	TTCCCGGAA
	JASPAR	MA0081	0.697	AGMGGAA
ELK1	TRANSFAC	M00007	0.898	NNNACMGGAAGTNCNN

cluster_80



NNANAGAGCCAAG

Name	Source	ID	Divergence	Consensus
bHLH			0.657	NNCNNNTG
ETS_class			0.825	RRRGGARR

cluster_83

GCTCTCTGC

No similar matrices found.

Figure 3.32 Over-represented motifs are found in the STAT5A ChIP-enriched sequences. (A) Excerpt adapted from the Trawler output (http://ani.embl.de/trawler_job 33704) performed Dr. Ramialison using the non-redundant and non-repetitive 198 sequence data set. On the left panel are shown the sequence logos representing the position weight matrices found for the 5 individual clusters (68, 79, 78, 80 and 83) comprising the family_68 matrix. The size of the characters represents the relative frequency of the corresponding bases. The family_68 matrix achieved the highest z-score 11. Next to each individual cluster matrices are reported any motifs found to display similarities to known transcription factors binding sites annotated in the Transfac and Jaspar databases. The name of the transcription factor is listed next to its ID, which allows for retrieving information on the transcription factor in the relevant source databases. The degree of difference between the transcription factor consensus motif and the cluster matrices are evaluated by their degree of divergence.

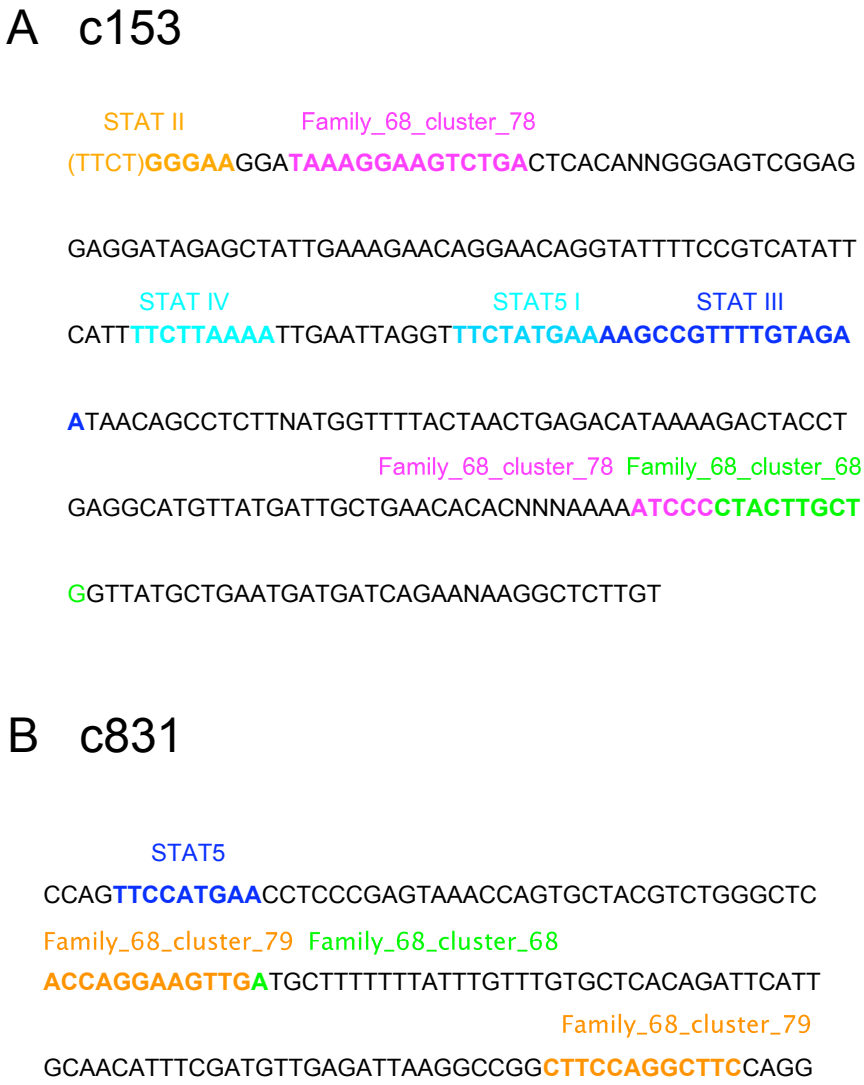


Figure 3.33 Diagram depicting STAT sites and over-represented motifs in two example clones.
The sequences in the clone c153 (A) and c831 (B) that show similarity to a specific designated family motif cluster are in bold. STAT- and STAT5-binding sites are also marked in bold in the sequences. Note not all motif family clusters found by Trawler analysis in the representative clones are depicted.

Overall, several over-represented motifs were identified using Trawler. Those identified matrices that do not resemble motifs from the transcription factor binding site databases are novel. Other motifs identified resemble the binding sites of certain transcription factor motifs. The motif family_68_cluster_78 was defined by Trawler to resemble the STAT-binding motif as defined in the Transfac database. This supports that this STAT5A ChIP enriched DNA library contains a significant over-representation of a motif that resembles the known STAT-binding motif.

4. Discussion

4.1 STAT5, RNA polymerase II recruitment and histone modification changes at different regions of the ß-casein and WAP genes

The mammary gland is an organ that develops mainly after birth in the adult animal. The mouse mammary gland tissue proliferates and differentiates in response to systemic hormones during puberty, pregnancy and lactation. This process repeats itself with each subsequent pregnancy. One of the main signaling pathways involved in the proliferation and differentiation of the mammary gland is the prolactin/Jak2/STAT5 signaling pathway. This signaling pathway is crucial for the functional differentiation of the mammary gland, as has been exemplified by the phenotypes observed upon loss of these factors in genetically engineered mice (Wagner et al., 2004; Cui et al., 2004; Shillingford et al., 2002; Miyoshi et al., 2001). Germline deletion of STAT5a results in impaired mammary gland development and lactogenesis (Liu et al., 1997). STAT5 deficient mammary gland epithelial cells are unable to form epithelial alveoli and express certain milk protein genes (Miyoshi et al., 2001). Conditional deletions of STAT5 from the mammary epithelial cells during pregnancy have demonstrated that STAT5 is essential not only for the functional differentiation, but also for the maintenance and survival of the mammary epithelial cells (Cui et al., 2004).

This work describes the recruitment of STAT5A, STAT5B and RNA polymerase II and the pattern of post-translational histone modifications at two milk protein genes throughout different mammary gland development stages. For this purpose the milk protein genes ß-casein and WAP were chosen, since these are known STAT5 target genes and specifically expressed in differentiated epithelial cells of mammary gland tissue. Expression of the two milk proteins ß-casein and WAP was detected in mammary gland tissue from pregnant and lactating mice, but not in the tissue from virgin mice and mice having undergone involution (Fig. 3.1). Therefore tissue from the latter two mammary gland developmental stages (virgin and involution) was included for comparative purposes.

This study has not only included the promoter regions containing the STAT5 regulatory elements of ß-casein and WAP, but also regions far upstream from the transcription start sites TSS (5' region), the transcribed region in the main body of the gene and regions downstream from the end of each gene transcript (3' region). This approach allows for a more complete picture of any changes in histone modifications occurring across the general regions of each gene.

STAT5A and STAT5B are both highly related proteins that can bind the same regulatory elements. The phenotypes of STAT5a and STAT5b deficient mice have shown that the two have redundant and non-redundant functions, suggesting that some of their transcriptional targets may be identical and others not. In Stat5a deficient mice the expression of WAP is almost completely lost, while some ß-casein can still be detected (Liu et al., 1997). This indicates differences, in the ability of the two highly related STAT5 transcription factors in regulating WAP expression. A way, differences may be propagates is through the differential recruitment of factors via the STAT5 C-terminal transactivation domain (TAD). This TAD domain is the most variable domain between STAT5A and STAT5B and is believed to contribution to the transcriptional regulation of the target gene by recruitment of cofactors (Liu et al., 1995). In addition, differences can be enhanced by differential expression of STAT5A or STAT5B and different temporal kinetics in the DNA –binding by the two STAT5s to regulatory elements.

To analyze this, ChIP assays and specific antibodies were used to investigate the *in vivo* recruitment of factors to the endogenous target genes in mammary gland tissue during different stages of development.

4.1.1 Both STAT5A and STAT5B are found associated at the endogenous promoter regions of ß-casein and WAP in mammary gland tissue

With specific antibodies against STAT5A and STAT5B this work shows the *in vivo* recruitment of both STAT5A and STAT5B to the endogenous promoter regulatory region of ß-casein and WAP in tissue from pregnant and lactating mice (Fig. 3.2 and Fig. 3.3). Binding of both STAT5A and STAT5B is also observed on the ß-casein putative enhancer region, containing the STAT5 binding site, in tissue from pregnant and lactating mice. This STAT5 binding to the promoters of the two milk genes correlates with their mRNA expression during pregnancy and lactation (Fig. 3.1A and B). In comparison, no STAT5 recruitment was observed in these regions (enhancer and promoter) in tissue from mature virgin mice or mice having undergone involution. These results clearly show that both STAT5A and STAT5B are bound to the endogenous target genes during pregnancy and lactation, the mammary gland developmental stages, where the major proliferation and functional differentiation of mammary epithelial cells has taken place. Although other studies have observed STAT5 binding in tissue culture cell systems using mammary epithelial cells to ß-casein and gamma-casein (Xu *et al*., 2007; Kabotyanski et al., 2006) this is the first description of *in vivo* binding of STAT5A and STAT5B to the endogenous ß-casein and

WAP promoter in mammary gland tissue. WAP is a direct STAT5 target gene that is largely confined to secretory alveolar cells (Miyoshi et al., 2001; Liu et al., 1997). The expression of WAP begins in the latter half of pregnancy and its expression is considered as a terminal differentiation marker for mammary epithelial cells. The cell type and differentiation stage specific control conferred by the WAP promoter region has been utilized extensively in transgenic mice to target and control expression specifically in differentiating mammary epithelial cells. Therefore, it is relevant to investigate STAT5 *in vivo* recruitment to the endogenous promoter in mammary gland tissue. Interesting, is that both STAT5A and STAT5B are found recruited to the endogenous WAP promoter, since in the STAT5A deficient mouse WAP expression is nearly lost (Teglund et al., 1998) and STAT5B appears not able to completely compensate for STAT5A. No difference was observed between the binding of either STAT5A and STAT5B to the two milk protein regulatory elements in the promoters during the investigated stages of pregnancy and lactation. This demonstrates that the endogenous STAT5 regulatory regions are accessible and consequently bound by both STAT5A and STAT5B in tissue during both stages. These observation are in accordance with the original studies in mammary gland tissue. Both STAT5A and STAT5B are found expressed and activated during pregnancy and lactation and both display DNA-binding activity as hetero- and homo-dimers during lactation as determined by *in vitro* EMSA assays (Liu et al., 1996). In other cell types, like the murine Ba/f3 cells line, interleukin-3 induced STAT5A and STAT5B binding to the same regulatory region has also been observed. Here though, it could be shown that the binding to some targets occurred with different kinetics (Nelson et al., 2004).
Therefore, it could be feasible that STAT5A or STAT5B might demonstrate different DNA-binding kinetics sometime during other time points of the 21 day pregnancy and subsequent lactation period. One complication for detecting different kinetics in DNA-binding on target regulatory elements in mammary gland tissue is that within the mammary gland tissue the epithelial cells are not all perfectly synchronized.

4.1.2 p-CTD RNA polymerase II is recruited to the promoter, transcribed and 3` region of ß-casein and WAP in mammary gland tissue

In eukaryotes the RNA polymerase II is the most important RNA polymerase involved in the transcription of protein-coding genes. RNA polymerase II requires the assistance of general transcription factors to initiate transcription by formation of a pre-initiation complex on the core promoter of protein-coding genes (Kornberg, 2007). On the carboxy (C)-terminus of the largest RNA pol II subunit, are located heptapeptide

(YSPTSPS) repeats referred to as the CTD (c-terminal domain) (reviewed in Hirose and Ohkuma, 2007). The CTD of RNA polymerase II is phosphorylated at serine 2 and 5, where phosphorylation of serine 5 occurs already during transcriptional initiation and at serine 2 during elongation. The phosphorylation of the CTD of RNA polymerase II is essential for transcription.

Results from this work here, demonstrate that the p-CTD RNA polymerase II is associated with the promoter, transcribed region and too a lesser extent to the 3' region of both ß-casein and WAP during lactation (Fig. 3.2D and Fig. 3.3D). As expected, these results show the localization of a transcriptional active p-CTD RNA polymerase II across the body of the endogenous ß-casein and WAP genes, when the mRNA of both genes is expressed. In comparison, to the other ß-casein regions investigated that show bound p-CTD polymerase II, no enrichment of p-CTD polymerase II was observed on the putative enhancer region (-6 kb) or 5' region in mammary gland tissue from lactating mice. This was unexpected in regard to the enhancer region, since observations made in the mammary epithelial HC11 cell system, showed a 3-fold phosphorylated RNA polymerase II accumulation at almost equal levels at both the ß-casein promoter and enhancer regions upon prolactin and glucocorticoid induction (Kabotyanski et al., 2006). The authors speculated that possibly there is a looping mechanism involved where the enhancer and promoter come into close proximity by looping out the intervening DNA. These results hint at differences between the HC11 cell line system and the mammary gland tissue. In tissue during lactation, difference in p-CTD polymerase II load between the enhancer and the promoter regions of ß-casein suggest that the p-CTD polymerase II on the promoter might not be in close proximity to the putative enhancer, since then the enhancer region would be expected to show a similar high enrichment in the ChIP assays. This might indicate that there is no close contact between the putative enhancer and the ß-casein promoter, precluding a looping model between the two regions. Future chromatin conformation capture assays may be able to determine if a physical interaction actually occurs between the putative enhancer and TSS of ß-casein (promoter region).

4.1.3 Changes in histone modifications across the different regions of the ß-casein and WAP gene in mammary gland tissue

The chromatin has to be accessible for transcription by RNA polymerase II to occur successfully. One mechanism involved in the dynamic modulation of chromatin

structure is the post-translational modification of the N-terminal tails of histones. The histone tails are subject to many post-translational modifications such as acetylation, methylation and phosphorylation (reviewed in Kouzarides, 2007). Acetylation of histone H3 and H4 as well as the di- and tri-methylation of lysine 4 of Histone H4 are associated with active transcription, whereas methylation of lysine 27 of Histone H3 is found at inactive genes or regions of heterochromatin. Two main mechanisms on how post-translational histone modification function have been postulated (Li et al., 2007; Clayton et al., 2006). Histone modifications such as acetylation are thought to shift the overall net charge of the nucleosomes, thus that there is a change in the DNA–histone interaction on the nucleosome leading to a relaxation or loosening up of the chromatin structure. In addition, the histone modifications can serve as potential docking sites for specific effector proteins and are believed to be read by proteins that are involved in chromatin remodelling and or in the process of transcription. Post-translational histone modifications have been found in distinct localized patterns within the genes and their surrounding upstream and downstream regions. The locations of these histone posttranslational modifications are regulated and are involved in the process of transcription (Li et al., 2007).

Therefore, histone H3 and H4 lysine acetylation at distinct regions of the ß-casein and WAP gene, namely at the enhancer (-6 kb only for the TSS of the ß-casein gene), 5´region (3-3.5 kb upstream of the genes TSS), the promoter region, the transcribed region and 3´region (about 1 kb downstream from the 3' end of the genes mRNA transcript) were investigated.

4.1.4 Histone H4 lysine acetylation is localized to the 5' region of both the ß-casein and WAP gene during lactation

Early studies had shown, that acetylated histones are associated with active genes and DNase I-sensitive chromatin domains (reviewed in Kouzarides, 1999). And the increased acetylation of histones has been associated with a less compact and more accessible chromatin structure (Berger, 1999). Genome wide mapping of histone modification patterns to gene expression patterns has positively connected histone hyperacetylation with actively transcribed chromatin (Kurdistani et al., 2004; Schubeler et al., 2000). By using antibodies recognizing the general acetlyated lysine residues on the N-terminal tails of histones H3 and H4, increased histone acetylation in the 5', promoter and transcribed regions of both the ß-casein and WAP genes during pregnancy and lactation was shown (Fig. 3.4 B, C and Fig. 3.5 B, C). This suggests a

more relaxed and open chromatin structure in these regions during the differentiation specific expression of these milk protein genes.

Regions of localized histone hyper-acetylation have been described at other very distal *cis*-acting elements from activated genes. For example localized H3 hyper-acetylation has been shown at the locus control region (LCR) during gene activation in the human ß-globin locus (Schubeler et al., 2000). Most interesting is, that the results presented here clearly demonstrate a pronounced localized histone H4 hyper-acetylation found at the distal 5' regions of the ß-casein (-3,5kb) and WAP (-3.0kb) genes during pregnancy and lactation (Fig. 3.4C and 3.5 C). This 5' region specific histone hyper-acetylation has been observed also in other genes from different studies. A similar localization effect of histone H4 hyper-acetylation has been described in the extreme 5' region (-2,8kb) of the E-selectin gene during TNFalpha induced expression in endothelial cells (Edelstein et al., 2005).

Also for another milk protein alpha s1-casein, histone H4 hyper-acetylation of the upstream distal enhancer (-3.4 kb) has been demonstrated using ChIP assay in primary rabbit mammary cells (Jolivet et al., 2005). The extracellular matrix (ECM) and not PRL was shown to maintain a high amount of histone H4 acetylation upstream of the alpha s1-casein gene in the distal prolactin- and ECM-sensitive enhancer. The study by Jolivet et al. (2005) provides evidence that the ECM regulates the PRL/STAT5 dependent expression of the alpha s1-casein gene in primary rabbit mammary cells by local modification of the chromatin structure. It is therefore tempting to speculate that the localized histone H4 hyper-acetylation, observed in the 5' regions of the mouse ß-casein and WAP genes during pregnancy and lactation in tissue, may be under similar regulation by the ECM as seen for the rabbit alpha s1-casein gene.

The N-terminal tails of the core histones can be acetylated at multiple individual lysine residues. The antibody against tetra-acetylated histone H4 recognizes the N-terminal tail of Histone H4 that is acetylated specifically at lysines 5, 8, 12 and 16. Acetylation of these lysines is conserved between different species such as in mice and yeast and the acetylation of these lysines has shown effects on the transcriptional potential in yeast (Dion et al., 2005). Individual point mutations in H4 lysine, 5, 8, and 12 barely showed any effect on the expression of genes in yeast, but when these mutations are combined a cumulative change in the expression profile of specific genes was observed (Dion et al., 2005). This suggests a possible redundancy in function for H4 lysine 5, 8, and 12. An exception is the histone H4 lysine 16, since point mutations in this individual lysine 16 of H4 resulted in a change of gene expression (Dion et al., 2005). In yeast it has been described that this modification has a functional role in maintaining transcriptional permissive DNA domains by disrupting the formation of higher-order chromatin structure (Shogren-knaak et al., 2006; Dion et al. 2005).

Therefore in this work individual acetylated lysines of Histone H4 were further investigated.

ChIP assays using site-specific anti-acetyl antibodies recognizing histone H4 lysine K8, K12 and K16 resulted in a distinct enrichment of the 5´region in the case of ß-casein and the WAP gene from mammary gland tissue of lactating mice (Fig. 3.4 and 3.5 D, E and F) This suggests that in the differentiated epithelial cells during lactation the ß-casein and WAP genes have nucleosomes at this 5' region whose histone H4 tails have been acetylated on the lysine K8 and K12 more strongly than in the other regions investigated.

For the acetylation of histone H4 lysine 16 a smoother distribution was observed across the different regions analyzed, with only a slight enrichment observed in the 5´region of the ß-casein and the WAP gene during lactation and pregnancy. The relative change in this modification during the distinct differentiation stages was minimal in comparison to the other histone H4 acetylated lysines analyzed. Nonetheless at the regions investigated, the presence of this mark suggests a de-condensed chromatin structure, since acetylated lysine 16 of histone H4 is involved in the prevention of the formation of higher order chromatin structure and is implicated in the maintenance of euchromatin (Shogren-knaak et al., 2006).

Apart from suggesting an open chromatin structure at the 5' region, the specific acetylated lysines on the histone H4 tails may serve to recruit bromodomain (factors) containing chromatin-regulating proteins. For example acetylated H4-K8 histone tails can mediate the recruitment of SWI/SNF complexes (Agalioti et al., 2002) that can in turn remodel nucleosomes. A recent study shows the recruitment of the bromodomain containing BRG-1 protein to the promoters of beta-casein and gamma-casein gene and that the SWI/SNF-dependent chromatin remodelling is required for transcription of these mammary specific genes (Xu et al., 2007). Attempts to detect BRG-1 by ChIP assay in mammary gland tissue resulted in variable recruitment during different mammary gland stages, which could not be resolved (data not shown).

The localized histone H4 hyper-acetylation observed in the extreme 5' regions of the ß-casein and WAP genes correlate well with the activation of the milk proteins during lactation. It seems likely to speculate that the changes in localized H4 acetylation observed at the 5' regions during lactation are due to the specific recruitment of histone acetyltransferase (HAT) activity containing proteins by differentiation-specific transcriptional activators. CBP is a histone acetyltransferase that can acetylate K8 of histone H4. The CBP-acetylated H4 K8 has been described to allow for BRG1 recruitment to gene promoters via BRG1s bromodomain (Agalioti et al., 2002). HATs, like CBP and p300 are involved as co-activators in STAT5 mediated gene regulation (Pfitzner et al., 1998). In addition, another study has shown that histone deacetylase

activity is also required for transcriptional activation by STAT5, through recruitment of HDAC1 (Rascle et al., 2003). The chromatin structure and its accessibility is influenced by a dynamic process of acetylation and deacetylation of histone N-terminal tails by HAT and HDAC activity containing factors. Therefore, the recruitment of CBP and HDAC1 to the ß-casein and WAP gene in mammary gland tissue was tested (data not shown). However, no reproducible binding of these histone modifying enzymes could be detected in the mammary gland tissue in ChIP experiments. These results imply that the enhanced acetylation observed at the 5' region during differentiation is a modification that might not require the continuous action or presence of HATs and HDACs. However, it may be possible, that the association of CBP or HDAC1 might be to transient to be detected in these experiments. The ChIP assay can capture a snapshot picture of what is happening at that moment in time as an average of the whole in the different cell types of the mammary gland tissue. Thus short lived, dynamic processes in tissue samples may not be detected as well as in a controlled induced and synchronized cell population. Alternatively, other HATs or HDACs not tested for might be responsible for the acetylation or deacetylation of histone H4 lysines in the 5' region. For example in budding yeast complexes containing members of the MYST HAT family like Sas2 can also acetylate nucleosomal histone H4 lysine 5, 8,12 and K16 (Fukuda et al., 2006). Further studies using high-resolution tiling arrays to narrow down the exact region of localized H4 lysine acetylation within the 5' region might help to further delineate its involvement in the differentiation-specific milk gene transcription.

4.1.5 Histone H3 lysine methylation changes across the different regions of both milk protein genes during lactation

Methylation of the lysine residues on the N-terminal tails of histones correlates with transcriptional active or repressed chromatin. Specifically, lysine 4 of histone H3 can be di or tri-methylated. Di-methylated lysine 4 of histone H3 is a mark associated mainly with transcriptional permissive or active chromatin, while tri-methylation is associated with actively transcribed genes (Santos-Rosa et al., 2002). Tri-methylation of H3K4 was detected in the promoter, transcribed region and to a lower extent in the 3´region of the ß-casein gene during pregnancy and lactation (Fig. 3.6C). No histone H3K4 tri-methylation modification could be detected in the enhancer and 5´region of the ß-casein gene. For WAP tri-methylation of H3K4 was observed mainly in the transcribed region (Fig. 3.7C). These results are consistent with the findings of Schneider et al. (2004) that in higher eukaryotes tri-methylation of H3K4 is strongly and

preferentially associated with the transcribed regions of active genes. In contrast to tri-methylation of H3K4, the di-methylation of this residue resist the inactivation of the ß-casein gene during involution. In comparison, using the same chromatin immunoprecipitated samples, this remnant di-methylation of H3K4 was not observed for the WAP gene during involution (compare Fig. 3.6B and 3.7B). The persistence of the di-methylation of H3K4 mark at the promoter and transcribed region of the ß-casein gene may represent a genomic mark of recent transcriptional activity as has been described in an inducible system for the GAL10 gene in yeast (Ng et al., 2003). Persistent hyper-methylation of H3K4 found within the transcribed region some time after transcriptional inactivation of the gene has been implicated to provide a molecular memory mark of recent transcriptional activity (Kouskouti and Talianides, 2005; Ng et al., 2003).

In the mammary gland tissue investigated a persistence of the tri-methylation mark was not observed. This may be because in this system more dynamic histone changes, which can be more readily observe in an inducible system with synchronized cell populations, can not be pinpointed so precisely. The H3K4 di-methylation mark was observed in all analyzed regions of the ß-casein gene during pregnancy and lactation. The question arises if the H3K4 di-methylation mark may be continuously distributed over the whole casein locus, especially since the expression of casein genes are co-ordinately regulated during pregnancy and lactation (Hobbs et al., 1982). Of the five casein genes comprising the casein locus, ß-casein is located 10 kb and 60 kb from the next neighbouring casein gene (Rijnkels et al., 1997). Continuous H3K4 methylation has been observed over other long stretches of DNA such as for a 17 kb region at the ß-globin locus (Kim and Dean, 2004). A broad distribution of histone H3 lysine 4 di-methylation has also been observed across multiple active genes of the Hox cluster in human and mouse (Bernstein et al., 2005). Future studies using a high resolution mapping may clarify if this is also the case for the regions neighbouring ß-casein and/or the whole casein locus.

The data here shows that methylation of H3K4 is enhanced along the regions of the ß-casein and WAP genes and that this correlates with the active transcription of these genes. The changes observed in H3K4 methylation during mammary gland differentiation at the different regions indicate here also the action of specific histone methylation enzymes. The histone methyltransferases (HMTs) responsible for the di- and tri-methylation of H3-K4 at the ß-casein and WAP genes still need to be determined. HMTs contain SET methyltransferase domains that are highly specific for lysines to be methylated. For example, both SET1 and SET7/9 are HMTs, that specifically can methylate K4 of H3 histones (Wang et al., 2001; Briggs et al., 2001). The mechanisms by which H3 lysine 4 methylation controls transcription is still not

clear (Francis et al. 2005). Studies in eukaryotes show a role for methylation of H3 in the initiation and elongation of transcription by RNA polymerase II (Gerber et al., 2003; Hampsey and Reinberg, 2003). The results of these studies suggest that histone lysine methylation may not only act as a stable epigenetic mark of potential active genes, but also may directly be involved in regulation of gene transcription. For example the cell type specific transcription factor pancreatic duodenal homeobox-1 (Pdx-1) was shown to link H3-methylation and RNA polymerase II elongation to Insulin gene transcription (Francis et al., 2005).

Although, the mammary gland tissue systems did not allow for the detection of fast acting dynamic changes in histone modifications, a good correlation between recruitment of STAT5, RNA-Polymerase II, histone acetylation and histone H3K4 methylation was observed at the differentiation stages when the milk protein genes are expressed in the mammary tissue. While lysine 4 histone H3 methylation is associated with transcriptional active genes, another methylation on the lysine 27 of histone H3 is correlated with inactive genes. The tri-methylation of lysine 27 on histone H3 is correlated with the silencing of homeobox (HOX) genes (Cao et al., 2002) and is associated with X chromosome inactivation (Plath et al., 2003).

Histone H3K27 tri-methylation, but non of the other tri-methylation marks (tri-methyl H3K9 and tri-methyl H4K20) investigated and known to be associated with transcriptional repressed chromatin, was observed at the ß-casein and WAP gene in either mammary gland tissue or the mammary epithelial cell model system used (data not shown). The disability to detect tri-methylated (K9)H3 or tri-methylated (K20)H4 modifications suggests that these histone modifications may not be involved in the transcriptional repression of the two WAP and ß-casein genes investigated or may be due to epitope masking.

Some enrichment was observed in ChIP assays using an polyclonal antibody recognizing di-methylated (K9)H3 on the ß-casein promoter (data not shown). It is therefore possible that di-methylation of lysine 9 of histone H3 may be present. Alternatively, there might be a slight cross-reactivity of the polyclonal antibody with the histone H3 lysine 9 and 27 methylation mark, since the lysine 9 and 27 methylation sites reside within a highly related sequence motif ARKS on the histone H3 tail (Fischle et al., 2003b).

When the milk protein genes are not expressed and active histone marks like H3K4 tri-methylation have decrease or are not detected anymore, a slight increase in H3K27 tri-methylation is found (at the transcribed region) in tissue from the virgin and involution stages (V and I) (compare Fig. 3.6-7C and D). This indicates that mammary gland development is associated with distinct changes in histone lysine methylation at the differentiation specific marker genes WAP and ß-casein.

In this study mammary gland tissue was used, which is composed of epithelial and stroma cells. Only the luminal epithelial cells undergo functional differentiation in pregnancy to produce milk. The general level of histone H3K27 tri-methylation observed during all the different mammary gland differentiation stages might result from simultaneous detection of this modification in other cell types in the tissue like epithelial and stroma (adipocyte and fibroblasts) cells. Therefore, to determine if the changes in tri-methylation of lysine 27 on histone H3 observed in the mammary gland tissue is also part of the repressed ß-casein chromatin of epithelial cells and not only due to other cell types found in the tissue, a pure mammary epithelial cell line was analyzed.

4.1.6 Differences in the recruitment of STAT5, RNA polymerase II and histone modifications at the ß-casein and SOCS3 promoter in the HC11 cell line

The HC11 mammary epithelial cell line is a model system to study mammary epithelial cell differentiation *in vitro*. Confluent grown HC11 cells can be differentiated by removal of EGF and addition of the lactogenic hormones prolactin and glucocorticoid leading to the expression of epithelial differentiation specific markers such as the milk protein ß-casein (Doppler et al., 1989). With stably transfected reporter constructs the HC11 cell line has been used to define the minimal and essential responsive promoter elements in the ß-casein promoter (Doppler et al., 1989). Recent studies also used this model system to analyze the assembly and disassembly of transcription factors and cofactors at the ß-casein promoter in response to prolactin and glucocorticoid by ChIP experiments (Kabotyanski et al., 2006). In addition to the ß-casein milk protein gene, other STAT5 target genes, such as *socs3* and *cis* have been shown regulated by lactogenic hormones in HC11 cells (Tonko-Geymayer et al., 2002).
In HC11 cells, tri-methylation of H3K27 at the ß-casein promoter (Fig. 3.9A) was detected. This indicates that this repression-associated marker also seen in the mammary gland tissue is part of the epithelial cell chromatin and not solely due to other cell types found in the tissue. The presence of the tri-methylated lysine 27 on histone H3 indicates, that the repression of these milk protein genes could be mediated by specific histone methyltransferases (HMTs) during differentiation. Histone lysine methyltransferases complexes containing the Enhancer of zeste Homology 2 (EzH2) can tri-methylate lysine 27 of histone H3 (Cao et al., 2002). Therefore ChIP assays using antibody against Enhancer of zeste were performed in mammary gland tissue, but these did not yield reproducible results (data not shown).

In comparison to the ß-casein promoter, only minimal levels of tri-methylated lysine 27 on histone H3 were detected at the SOCS3 promoter (Fig. 3.9A) indicating that this gene might be regulated by a different kind of mode. Lactogenic hormone treatment induced binding of both STAT5A and STAT5B to the SOCS3 promoter, but no increase in the binding of STAT5A or STAT5B to the ß-casein promoter was observed (Fig. 3.8C and Fig. 3.9B). In addition to the failure to induce STAT5 binding, no recruitment of the RNA polymerase II to the ß-casein promoter was observed. In the case of the SOCS3 the RNA-Polymerase II could already been detected on the promoter without hormone treatment in HC11 cells. The association of the potentially transcriptional active form of the phosphorylated-CTD RNA polymerase II to the SOCS3 promoter suggests, that RNA polymerase II has already been recruited to the SOCS3 promoter independent of prolactin induction and that the SOCS3 gene may already be transcriptional active at a basal level (Fig. 3.9C). It has been demonstrated, in another study that SOCS3 transcripts can be induced by STAT5-independent signaling pathways in HC11 cells (Tonko-Geymayer et al., 2002). In addition, the presence of insulin required in the medium for optimal HC11 culture may influence the SOCS3 basal level, since insulin itself has been shown to induce SOCS3 mRNA expression in skeletal muscle cells (Sadowski et al., 2001). This correlates with the presence of the histone marks generally associated with transcriptional active genes such as acetylated H3 and H4 and di- and tri-methylated H3K4 found on the SOCS3 promoter (Fig. 3.9C, D). Apart from the repressive mark none of the other active histone marks seen on the SOCS3 promoter could be observed on the ß-casein promoter. Even prolonged lactogenic hormone treatment could not induce a significant change in histone H3 or H4 acetylation or histone H3K4 methylation at the ß-casein promoter (Fig. 3.9C and D). This indicates a possible closed chromatin conformation at the ß-casein promoter in the majority of HC11 cells. The lack of STAT5 and RNA polymerase II recruitment to the ß-casein promoter suggest that the ß-casein promoter is not as accessible as the SOCS3 promoter in response to lactogenic hormone treatment. The same differences were also observed when the *wap* and the *cis* genes in HC11 cells were analyzed (data not shown).

Overall, recruitment of the transcription factor STAT5 and the histone modifications that correlate with transcriptional active genes and open chromatin structure could be observed at the SOCS3 promoter. On a comparative level, this was not observed at the ß-casein promoter in HC11 cells stimulated with lactogenic hormones. It may be that additional endocrine and extracellular matrix-derived signals are required to fully open up chromatin structure of the ß-casein promoter. Especially, since the expression of milk protein genes is regulated by a complex interplay of hormones, cell-cell contacts and cell-extracellular matrix interactions (Topper and Freeman, 1980).

Contrary to the cell culture system used, the mammary gland tissue seems to have the optimal conditions to differentiate and to alter the chromatin structure, which allows for the binding of STAT5 and recruitment of RNA polymerase II as determined in the ChIP assays. The success of the ChIP assays using mammary gland tissue from lactating mice, may be due to the fact that differentiation specific milk protein expression levels, such as ß-casein and WAP in cell culture models are generally lower than compared to the *in vivo* expression observed in lactating mammary gland tissue (Clarkson et al., 2006). Initial characterization of the HC11 cell line showed by use of immuno-fluorescence that only 30% of the cells displayed ß-casein expression after five days of induction with lactogenic hormones (Ball et al., 1988). Thus there might not be enough fully differentiated HC11 cells to detect factor binding efficiently by ChIP.

Although ß-casein is detected in the induced HC11 cells this does not necessarily equate with a high number in differentiated HC11 cells. The ChIP assay provides information based at the genomic DNA level, where each cell has only two alleles of ß-casein gene, while the detection of the differentiation specific milk protein RNA can be enhanced by mechanisms regulating RNA stability and is not limited by number per cell. Therefore, a low number of fully differentiated HC11 cells could contribute to the poor detection by the ChIP method of transcription factors and histone modifications at the ß-casein and WAP gene. Even after long-term treatment for 68 hours with lactogenic hormones, efficient binding of STAT5A and RNA-Polymerase II was not observed, nor were comparable changes in the histone modifications found in the mammary gland tissue. The HC11 cell line was initially derived from the mouse mammary epithelial cell line COMMA-1D, which requires in addition to the stimulation with lactogenic hormones the cultivation on extracellular matrix (ECM) to express endogenous ß-casein (Danielson et al., 1984; Schmidhauser et al., 1990). On the other hand, the HC11 cell line is thought not to require growth on ECM like the parental COMMA-1D, since this cell line can produce its own endogenous extracellular matrix components to establish differentiation competence (Chammas et al., 1994). The ECM appears to be an important component in mammary epithelial cell differentiation, since recent data have shown that the ECM controls the histone H4 actylation at the alpha1-casein gene in rabbit primary mammary cells (Jolivet et al., 2005). This indicates that controlling the histone modifications in the milk protein genes might be the level by which differentiation inducing signals like from the ECM and surrounding cells can control the accessibility of the gene and thereby regulate the binding of the transcription factors like STAT5.

4.1.7 Model of the histone modification changes observed at the different milk protein gene regions in mammary gland tissues

Based on the results from the ChIP assay the following model represents a summary of the findings observed at the different regions of the ß-casein gene during the four different mammary gland stages investigated (Fig. 4.1).

In mammary gland tissue from virgin mice (V) and mice having weaned their young (I) no significant enrichment of histone marks associated with active transcription, such as lysine acetylated histone H3 and H4 or lysine 4 methylation of histone H3, was observed at the five different regions of ß-casein gene investigated. No STAT5 or p-CTD RNA polymerase II was detected in any of the regions (enhancer, 5, promoter, transcribed and 3'). Instead tri-methylated lysine 27 histone H3, a mark associated with repressed chromatin, was observed distributed across all regions of ß-casein (Fig. 4.1 V and I).

With pregnancy (P) (day 15) and lactation (L) (day 5) both STAT5A and STAT5B are found bound at the promoter and putative enhancer region of ß-casein containing STAT5 recognition motifs (Fig. 4.1 P+L). This is accompanied by recruitment of the p-CTD RNA polymerase II at the promoter and transcribed region of ß-casein. In addition there is a strong increase in histone H3 and H4 acetylation, with a specific localized enrichment of acetylated histone H4 (lysine 8 and 12) at the 5' region of ß-casein. This enhanced histone acetylation is associated with transcriptional active transcribed genes and indicates an open and accessible chromatin structure. In addition, there is increased tri-methylation of lysine 4 on histone H3 observed in the promoter and transcribed region of ß-casein. The presence of this active associated histone mark further support the notion of active transcription in the region at that stage. Furthermore, there is a strong increase in di-methylation of lysine 4 on histone H3 found spread across all five regions of ß-casein during lactation (Fig. 4.1 P+L). Overall, these histone marks known to be associated with transcriptional active genes together with the recruitment of STAT5 and the major component of an active basal transcription machinery (p-CTD RNA polymerase II), correlate well with the active transcription and expression of ß-casein during lactation. In addition to an open chromatin structure and transcriptional activity, the presence of the observed histone modifications hint at the action of HATs and HMTs involved in the regulation of ß-casein transcription.

Once the young are weaned, the STAT5 and active p-CTD RNA polymerase II as well as the active associated histone marks previously observed during lactation are not observed any more in the regions of ß-casein (Fig. 4.1 compare I and P+L). An exception is the di-methylated lysine 4 histone H3 mark (Fig. 4.1 I).

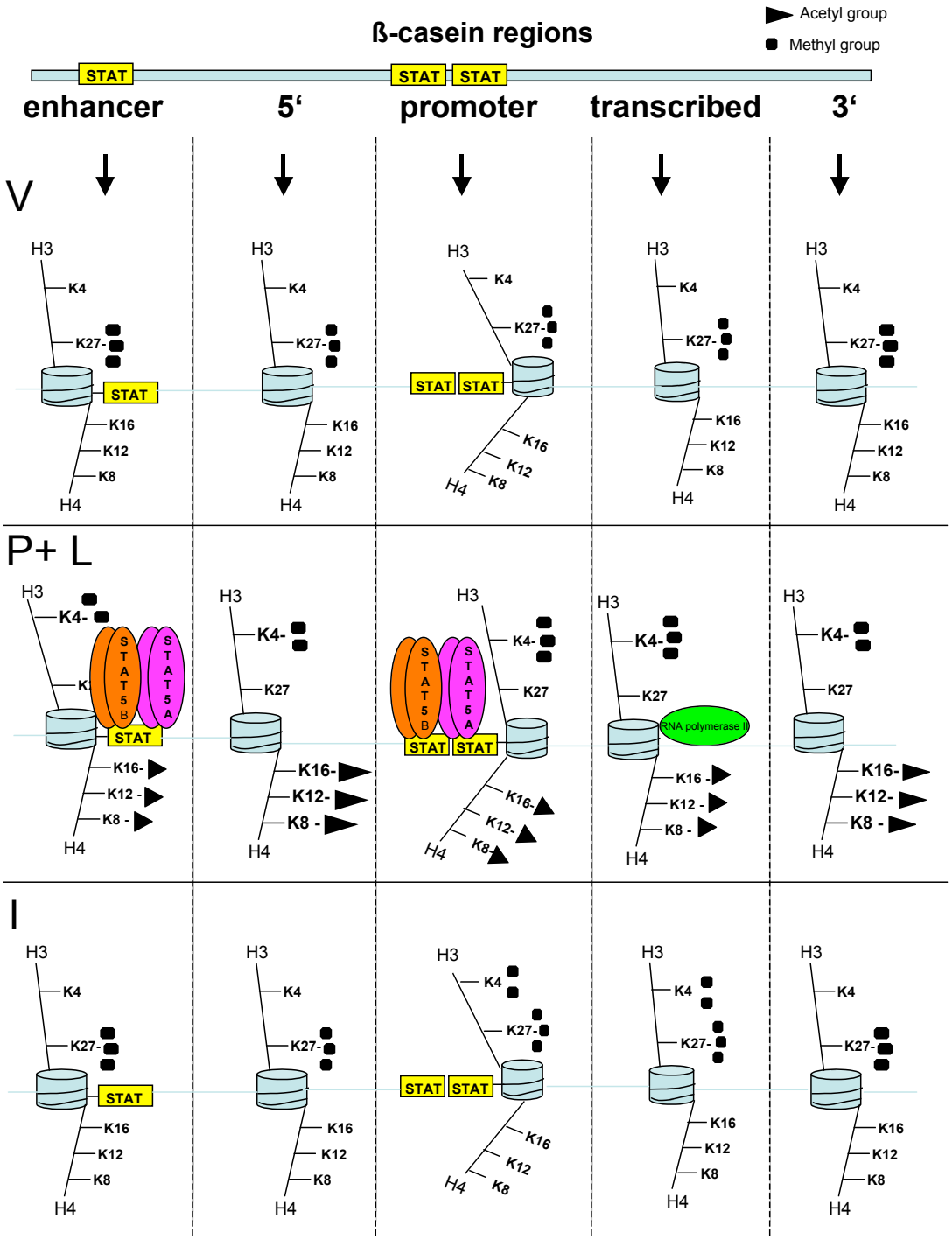


Figure 4.1 Descriptive model of ß-casein regions showing STAT5 and p-CTD RNA polymerase II recruitment and changes in histone modifications during different mammary gland developmental stages.
Schematic representation of the predicted histone modifications and recruitment of STAT5 and RNA polymerase II to the endogenous ß-casein gene regions investigated in mammary gland tissue from virgin (V), pregnant (P), lactating (L) and involuting (I) mice. Depicted are the five different DNA regions (enhancer, 5', promoter, transcribed and 3') of the ß-casein gene investigated. Each region is represented by a picture of a boxed nucleosome with one N-terminal tail depicted for histone H3 and H4. Only the lysines (K) residues that were investigated are shown. Acetylated lysines are indicated with triangles and methylated lysines by rectangles. Two or three rectangles indicate di- or tri-methylation, respectively.

This permissive mark was often found retained on the promoter and transcribed region of ß-casein and may potentially act as a memory mark for a recent transcription event. This is not so in the case of WAP, which in comparison to ß-casein did not retain the elevated levels of this histone mark in any of its regions investigated, hinting at a differential regulation. Meanwhile, the lysine 27 tri-methylation of histone H3, considered a mark associated with transcriptional repressed chromatin, is observed at levels similar to those found in virgin mouse tissue (Fig. 4.1 compare I and V).

In conclusion, these results indicate that the differentiation specific expression of the milk proteins ß-casein and WAP during pregnancy and lactation correlate with distinct covalent histone modifications. During the different mammary gland stages there are distinct quantitative differences between the 5' region, promoter, transcribed region and 3' regions of the ß-casein and WAP genes with regard to STAT5 and RNA polymerase II recruitment and the localization of histone modifications. The regions of the transcriptional active ß-casein and WAP genes (during pregnancy and lactation) are characterized by specific histone hyper-acetylation and lysine 4 histone H3 methylation (active and permissive histone marks). When the genes are inactive, the same regions lose the active chromatin marks, but still retain the repressive histone H3 lysine 27 methylation mark.

4.2 Analysis of the cloned STAT5A ChIP Library

For gene expression analysis, microarrays are used to identify regulated target genes, but this also can lead to the identification of secondary regulated genes. An advantage of the ChIP cloning procedure is the unbiased identification of novel and direct STAT5-binding elements in their natural cellular environment regardless of their genomic location. This approach allows for new insights into STAT5-mediated gene regulation during mammary gland differentiation.

The *in vivo* recruitment of STAT5 to DNA in the intact chromatin environment of mammary gland tissue has not yet been extensively studied. This is essential, since the chromatin environment has an important influence on regulating gene expression. Mammary gland tissue was used for cloning of novel STAT5-binding sites, because sufficient levels of differentiation specific STAT5-binding was observed in the tissue, but not in the mammary epithelial cell line HC11 (Section 4.1.6).

4.2.1 Verification of the STAT5A Library

One aim of this study was to identify new differentiation specific STAT5 regulatory elements in mammary gland tissue from lactating mouse. For this purpose the ChIP cloning method was used to identify novel STAT5 transcription factor binding sites. In this work, *in vivo* binding of STAT5A and STAT5B to the ß-casein and WAP promoter regions containing the STAT5-binding motifs could be confirmed (Fig. 3.10A-B). The specific recruitment of STAT5 to the promoters of these two differentiation specific milk protein genes occurred during the developmental stages of pregnancy and lactation in the mammary gland tissue. These results verify that the conditions for an effective immunocapture of STAT5-bound chromatin from mammary gland tissue has been established and can be used to isolate potential new STAT5 targets.

4.2.2 Isolation and validation of 9 novel STAT5 targets from mammary gland tissue during pregnancy and lactation

Nine out of a total of 25 clones tested from the ChIP library were verified to be unique and novel STAT5 targets, which can recruit both STAT5A and STAT5B proteins (Fig. 3.10C-F and Fig. 3.12A-E). This recruitment is seen mainly during pregnancy and lactation when the epithelial cells in the mammary gland tissue undergo proliferation and differentiation. Therefore, the ChIP assay was used successfully to isolate nine new differentiation specific STAT5 targets from lactating mammary gland mouse tissue. In addition, these independent ChIP assays show that the DNA sequences isolated represent true STAT5-binding sites. A general drawback of the cloning procedure in the ChIP cloning method is, that there is also cloning of non-specific sequences. In addition, not all identified transcription binding sites in the cloned sequences can be validated as true binding sites (Maston et al., 2006).

Approximately 83% of the immunocaptured, cloned and sequenced fragments from the STAT5 mammary gland tissue library could be localized to the mouse genome with significant sequence similarity using the NCBI BLAST search. The remaining 17% (22/128) of sequences that had not aligned to the mouse genome with significant similarity often displayed repeating sequence elements or just had poor quality sequences with lots of undetermined base (N) calls. In future, sequential double immunoprecipitations with the STAT5 antibody during the ChIP assay procedure might reduce the nonspecific sequences and enhance the number of STAT5-chromatin interaction sites identified.

In this thesis, the success rate of 36% (9/25) achieved for verified clones containing STAT5-binding site, that bind STAT5A and STAT5B during pregnancy and lactation, is in accordance with the percent range of verified STAT5-binding sites achieved in a study performed by Nelson et al. (2004). Here two independent STAT5 libraries from hematopoitec Ba/F3 cells, induced with IL-3, were generated, one for STAT5A and STAT5B, respectively. For the STAT5A library 25% (3/12) and for the STAT5B library 35% (6/12) of the STAT5 binding sites were shown as verified by ChIP assay (Nelson et al., 2004). Although, this study elegantly combined a linker-ligation mediated PCR with a subtraction step to remove more of the unspecific sequences this measure did not appear to elevate the overall percent of verified clones that bound STAT5. Therefore the method used in this thesis was satisfactory.

The remaining 16 out of 25 tested clones analyzed in this thesis did not give reproducible and strong signals in ChIP with STAT5A or STAT5B antibodies, even though their sequences contain either STAT5 or STAT-like binding sites. This is in agreement with the findings of another study, that evaluated a genome wide chromatin library of STAT5-binding sites from human breast cancer cells using *in vitro* EMSA and *in vivo* ChIP assays. LeBaron et al. (2005) found for half of their sequences that the presence of a STAT5-binding motif did not assure the binding of STAT5 to the immunocaptured fragment. The reason for this remains unclear. These STAT-motif containing fragments could have been immunoprecipitated non-specifically. Alternatively, these fragments can represent very low affinity STAT5-binding sites that do not display differences in the recruitment level between the different mammary gland developmental stages.

The high background is a problem inherent with the labor intensive ChIP cloning method. Other variations of the CHIP technique are used to analyze transcription factor binding and so identify possible regulatory elements and their target genes. Chromatin immunoprecipitated DNA can be hybridized to DNA microarrays (ChIP-on-chip). These DNA arrays may cover promoter regions, specific regions surrounding TSS or CpG islands. Therefore the ChIP-on-chip experiments are limited to whatever DNA the mircoarray covers. Recently though, arrays are available that cover the entire non-repetitive sequences of the genome. Such genome-wide analysis have been used to identify new DNA regions bound by transcription factors like estrogen receptor alpha (ERalpha) in both cultured cells and in mouse liver tissue (Gao et al., 2008; Carroll et al., 2006 and 2005). In combination with general expression analysis under the same conditions this is a powerful tool to identify transcription factor regulatory elements involved in control of nearby target genes.

4.2.3 Histone modifications and RNA polymerase II recruitment to 4 of the novel STAT5-binding regions

In general active promoters are marked by acetylation of histone H3 and H4 and methylation of histone H3 lysine 4 (reviewed in Barrera and Ren, 2006). Recently, Heintzman et al. (2007) has demonstrated in the human genome, that transcriptional active promoter and enhancers elements are marked by distinct chromatin signatures, such as methylation of histone H3 lysine 4, and that these can be used to predict potential regulatory DNA elements. The presence of a member of the basal transcription machinery and specific histone modifications, can be used as an indicator to determine if a specific DNA region is located in transcriptional active euchromatin. Both the phosphorylated CTD of RNA polymerase II and the histone H3 lysine 4 di-methylation mark, correlate with active and potentially transcriptional active chromatin. Comparison of the different STAT5-binding regions with each other detected different strengths of phosphorylated CTD of RNA polymerase II and histone H3 lysine di-methylation (Fig. 3.14). According to the histone H3 lysine di-methylation levels, all clones appear to be located in regions of potential transcriptional active euchromatin. But only the two clones 153 and 236 showed the most difference between the virgin and lactation stages. This indicates that both RNA polymerase II recruitment and the histone mark are elevated differentiation stage specifically during lactation for at least these two clones. This correlates with the differentiation specific recruitment of STAT5 to these clones during lactation in mammary gland tissue. In contrast to the mammary gland tissue, no recruitment of STAT5 to the clones could be observed in the inducible HC11 epithelial cell culture system (Fig. 3.13.). The same problem had been encountered previously in this work for ß-casein in the same cell system (Section 4.1.6). Here, no STAT5 could be recruited or changes in the histone modifications be observed at the ß-casein promoter (Section 3.1.13). For epithelial cells especially, the cell-cell and cell-matrix interactions are important for proper development and regulation of gene expression (reviewed Lin and Bissell, 1993). Since the STAT5A-binding clones have been isolated originally from lactating mammary gland tissue, the mammary epithelial cell culture system may not be able to provide the same conditions as encountered by the differentiated mammary epithelial cells in tissue from a lactating mouse. Experiments are underway to test the verified clones for STAT5-binding in another mammary epithelial cell line. The EpH4 cell line requires growth on an extracellular matrix (ECM) for expression of ß-casein and may therefore provide necessary signals derived from the ECM to allow for STAT5 recruitment to the cloned regions.

4.2.4 Characterization of the novel STAT5-binding sites

For purpose of discussion the information gained by computational and experimental means for each of the three clones (c153, c236 and c831), that contain novel STAT5-binding sites, will be first discussed individually.

4.2.4.1.1 Clone c153

The data gathered for clone c153 indicated, that the STAT5-binding site-containing distal 5'-flanking region of the NFI/B gene might be a cis-regulatory DNA region. This DNA region potentially mediates STAT5 induced NFI/B gene expression in mice during lactation. Using a comparative genomics approach it was determined, that approximately 80% of the 339 bp region of c153 clone is evolutionarily conserved between the mouse and the human genome. The 339 bp region of c153 contains one consensus STAT5-binding site and 3 STAT-like sites (Fig. 3.15D). The STAT5-binding site I and the immediate surrounding sequence is highly conserved between the mouse and the 10 other mammals aligned. The high degree of evolutionary conservation for the 339 bp region and the STAT5 site indicate, that this site is functionally important. Elevated levels of di-methylated (K4) histone H3 and recruitment of RNA polymerase II mark this region as potentially transcriptional active during lactation (Fig. 3.14B). Investigation into the potential function of the novel STAT5- and STAT-binding sites using reporter gene assays shows that the highly conserved 339 bp region confers reporter gene activity (Fig. 3.19B). The conserved region mediates transcriptional activity that depends on the STAT5 I and the STAT-binding sites IV and III. When mutated in the reporter gene assays the conserved c153 region reporter gene activity declines sharply (Fig. 3.21B). This indicates, that all three STAT sites are contributing to the observed reporter gene activity. The opposite is true for the other STAT-binding site II, which displays the least homology of the four sites and is located in a short stretch of lower homology within the 339 bp region. A mutation of this site appears to have no effect on the STAT5 mediated reporter gene activity. To note is that this STAT site II sequences has a consensus STAT motif with a central spacing of 4 instead of 3 bp, which is the optimal spacing preferred by STAT6 (Schindler et al., 1995). This may explain why mutations of this site had no observable effect on the STAT5 mediated PRL induction reporter assays. Taken together, the data infers that this STAT site II is of less importance for the STAT5 mediated c153 reporter gene activity observed. On the other hand, the STAT5- I and STAT-binding sites IV and III, which display a high degree of evolutionary conservation within the mammals, are responsible for the

STAT5 mediated c153 reporter activity observed. Thus, this work shows that the consensus STAT5-binding site I as well as the STAT-binding sites III and IV display high evolutionary conservation and are important in mediating PRL induced reporter gene activity of the c153 region.

An interesting aspect regarding these three STAT sites is their close proximity to each other. The STAT5-binding site I is 11 bp from STAT site IV and is also right adjacent to the STAT site III, which itself is comprised of two STAT motifs in a head to head orientation (Fig. 3.15D). There is a distance of 7 bp from the consensus STAT5 site I to the outer most of the two head to head STAT motifs, that comprise the STAT site III. Thus, they appear in their layout to resemble two tandemly linked STAT motifs (site I and III). This close spacing of the STAT sites to each other has been described to be important in STAT5A dimer and tetramer formation (Soldaini et al., 2000). STAT5A can bind either as a dimer or preferably as a tetramer on tandemly linked STAT motifs. The optimal spacing distance for STAT5A tetramer-binding between two STAT motifs is 6 to 7 bp in length (Soldaini et al., 2000). Often, a TCC or GAA half STAT gamma interferon activated sequence (GAS) motif is found at the distance of more than 5 bp from a consensus GAS motif (TTCNNNGAA). The STAT5 tetramers can bind to a wide range of tandemly linked consensus and non-consensus motifs (Soldaini et al., 2000). This would be the case, for the spacing between the consensus STAT5 site I and the STAT site III and opens the possibility that tetramer formation at these sites of the c153 region are conceivable.

Multiple STAT5 responsive genes are known to display two tandemly linked STAT motifs (see listed in Soldaini et al., 2000). STAT5 tetramer complexes have been shown to bind for example the Glycosylation-Dependent Cell Adhesion Molecule 1 (GlyCAM1) promoter, interleukin 2 receptor alpha (IL-2Rα) enhancer and the CIS promoter (Hou et al., 2003; Verdier et al., 1998; Meyer et al., 1997). Further investigation is necessary to determine if either or both STAT5 dimers and tetramers bind to this highly conserved group of STAT sites in the c153 region and in what combination this occurs at during the differentiation stages of mammary gland development.

Apart from STAT5A, the *in vivo* binding of STAT5B to the c153 conserved region during pregnancy and lactation has also been demonstrated using ChIP (Fig. 3.10C). The use of both wild type and constitutive active STAT5B in the reporter assays indicate, that STAT5B can also positively drive the c153 region luciferase reporter activity (Fig. 3.18B). This shows that STAT5B, as well as STAT5A, can associate with the STAT-binding sites and influence the transcriptional activity conferred through the STAT5-binding sites located within the c153 region. This is not surprising, since previous *in vitro* experiments have shown that both STAT5A and STAT5B homodimers

have similar DNA-binding preferences (Soldaini et al., 2000). Another study has demonstrated that their ChIP cloned STAT5 targets can also recruited both STAT5A and STAT5B *in vivo* (Nelson et al., 2004). Nelson and colleagues were able to observe different kinetics in the STAT5A or STAT5B recruitment to some targets in this inducible IL-3 Ba/F3 cell culture system. In this thesis no difference in the kinetics of STAT5A and STAT5B recruitment was observed in either the pregnancy and lactation stages of mammary gland development investigated. This may be because a cell tissue culture system, contrary to mammary gland tissue, is made up of a homogeneous population of cells that can be synchronized and induced under controlled conditions.

4.2.4.1.2 STAT5 recruitment to the c153 region correlates with the transcription of the nearest gene NFI/B

Based on the assumption that STAT5 regulates transcription, a search was performed to identify the closest known genes or annotated RNA transcripts. The nearest know gene to the c153 region is the Nuclear factor I (NFI)/B gene. NFI/B belongs to a family of transcription factors that are involved in the expression of mammary specific genes like the milk proteins WAP and ß-lactoglobulin (Li and Rosen, 1995; Watson et al., 1991). The c153 region lies about 62 kb 5 prime to the NFI/B gene. The expression of NFI/B mRNA transcript correlates with the *in vivo* association of STAT5A to the conserved c153 region during the mammary gland development stages of pregnancy and lactation (Fig. 3.28B) and Fig. 3.10C). No significant change in RNA levels was observed for the EST transcript spanning the c153 region. This indicates that the EST transcription might not be influenced on a transcriptional level, by the binding of STAT5 (Fig. 3.28C). Due to the high homology between mouse and human of the STAT5-binding site and the c153 region, it would be interesting to find out if the human NFI/B associated transcript expression also correlates to STAT5-binding in the c153 region. For further evaluation, a complete transcript annotation from lactating mammary gland tissue stage in this area would be required.

The elevated expression of the mouse NFI/B transcript that was detected during lactation is in accordance with other studies. Relative expression levels of the NFI/B gene at different stages of mammary gland development were determined using RNase protection and RT-PCR assays (Mukhopadhyay et al., 2001). Mukhopadhyay et al. found that NFI/B is already expressed in the virgin mammary gland, but that NFI/B displays the highest levels of expression during lactation and decrease again in expression during involution. In addition, the authors showed that the major NFI splice

isoform NFI/B2 in cooperation with STAT5 and GR could activate transcription of the WAP distal promoter in transient reporter gene assays. Previously, both STAT5 and NFI have been implicated to play a role in WAP gene expression in transgenic mice (Li and Rosen, 1995). With this connection in mind, it seems quite possible that during pregnancy and lactation, STAT5 through the STAT5-binding sites within the conserved c153 region is involved in modulating the expression of NFI/B. The 339 bp DNA region from c153, which is located 62 kb 5 prime to the NFI/B gene might act as a distal enhancer of NFI/B. Together then NFI/B and STAT5 could regulate the expression of the milk proteins.

The presence of RNA polymerase II on the c153 DNA region during lactation (Fig. 3.14A) does not preclude the region from being a putative enhancer. For example, RNA polymerase II has been shown associated with enhancers of genes, such as the androgen-responsive prostrate-specific antigen (PSA) gene (Louie et al., 2003; Shang et al., 2002). Another example is the ß-globin gene locus, where the distal enhancers in the locus control region act over large distances on the various globin promoters. Here it has been proposed that the DNA enhancer associated RNA polymerase II either reaches the promoter through a DNA looping or tracking mechanism. Specifically, for the human epsilon-globin gene locus a DNA tracking mechanism has been proposed. (Zhu et al., 2007). Alternatively the 339 bp DNA region from c153 may be part of a promoter of an until now still not annotated transcript for a coding or non-coding RNA transcript. The fact that the 339 bp region induces reporter gene activity already in the promoter-less and enhancer-less pGL3 basic plasmid, suggests that this region is possibly either a promoter or part of a promoter. Searches for transcription start sites (TSS) in vicinity of the c153 region did not result in the identification of any new identifiable TSS (Dragon program for TTS search). This may be due to the search parameters of the program used. Most programs search for CpG islands to determine TSS sites and currently no perfect programs are available to identify all TSS via bioinformatics searches (Sandlin et al., 2007; Birney et al., 2007). In the human genome about half the genes are found near CpG islands and are considered a good predictor for promoters (Bajic et al., 2004). Thus, the presence of a CpG islands has a high predictive value for the identification of this class of promoters. The nearest CpG island though that could be identified is at the 5 prime end of NFI/B gene (UCSC browser).

The cloned chromatin regions that are confirmed as positive *for in vivo* STAT5-binding by ChIP may be involved in the transcriptional control of nearby genes or alternatively of small regulatory RNAs (Cawley et al., 2004). Work done by Cawley et al. show that the un-bias mapping of transcription factor binding sites along the human chromosomes 21 and 22 appears to point to a widespread regulation of not only

protein coding RNAs but also of non-coding RNAs. A transcription factor binding search with the TESS program, showed a proximal sequence element (PSE) site near the STAT5 site I of c153. These PSE sites are often found upstream of TSS from snRNA genes, which can be transcribed by either RNA polymerase II or III (Li et al., 2004). Thus, there could be the possibility that the conserved c153 region and STAT5 is involved in regulating small nuclear RNAs. A UCSC browser miRBase search though, revealed no known annotated microRNAs in close vicinity (UCSC browser) of the clone 153 region.

4.2.4.2 Clone c236

The c236 region contains two STAT5-binding sites in proximity (100 bp) to each other. In terms of a chromatin context, where about 147 bp of DNA wrap around a nucleosome, the distance between the two STAT5-binding sites would bring both in close proximity. Based on the histogram, depicting the degree of conservation between the different mammals for the 292 bp c236 region (Fig. 3.15D), there are many areas of high homology hinting that there may be regulatory potential within this regions of conserved elements. As shown in this work by the use of transient reporter genes assays the region of c236 confers transcriptional activity that depends on STAT5A and the more conserved of the two STAT5-binding site II (Fig. 3.24B). The STAT5 recruitment to the c236 region observed during lactation in vivo and the reporter gene activity was overall weaker than compared to the other two clone regions tested (c153 and c831). This may indicate overall weaker STAT5-binding sites and possibly less importance of STAT5 in the function of driving gene transcriptional activity. Further investigation into the multiple RNA transcripts that span this region is being performed to determine if there are modulations in the transcripts levels that correlate to the *in vivo* binding of STAT5A and STAT5B to the region. Of special interest would be AK017143, a Riken cDNA (5031425E22) found in a library made with tissue from the reproductive organs (uterus and ovaries) of an 11 day pregnant mouse. In this Riken transcript the STAT5-binding site would be located in the first intron of the cDNA (UCSC Browser REFsource). Work is in progress to determine if this cDNA is also found in mammary gland tissue during pregnancy and lactation.

4.2.4.3 Clone c831

The novel STAT5-binding site from the c831 clone is found within the first intron of the Zswim6 (Zinc finger, SWIM domain containing 6) gene and is evolutionary conserved between 6 different mammals (Fig. 3.15H). This evolutionary conservation of the STAT5-binding site indicates that this STAT-binding site may be an important DNA-binding element. Since functional sequences are subject to evolutionary selection, the relatively unchanged (conserved) region, infers that such a region might be functional (reviewed in Miller et al., 2004). Next to the STAT5-binding site about 13 bp away is a 34 bp region with high homology between the different mammalian species investigated. This is a unique feature within this clone and warrants further investigation. Using TESS, a search for potential transcription factor (TF) binding sites was performed within this region for both the human and the mouse sequence. Some transcription factor binding site in common for both human and mouse were GATA, Lim only protein (Lmo2), c-myb, the homeodomain protein engrailed-1 (En-1) and yin yang 1 (YY1). In the human and the other two primates investigated, there is also a TFIID site (TTATATT), which is not found in the corresponding mouse sequence (TT**G**T**G**TT). At the STAT5-binding site in the human there is an overlap with a potential glucocorticoid receptor (GR), progesterone receptor (PR) and Oct-2 site. While in the mouse potential Oct-1, YY1, c-myb and Dioxin-Ah Receptor complex (AhR) binding sites overlap with the STAT5-binding site. Within the mouse conserved 34 bp region is also a PR site. Many of these transcription factors have been observed to act on the transcriptional regulation of STAT5 target genes (reviewed in Litterst et al., 2005). For example, PR has been shown to repress STAT5/GR activation of ß-casein gene (Buser et al., 2007) and YY1 is implicated in the repression of ß-casein basal transcription (Meier and Groner, 1994; Raught et al., 1994). Because TF searches in even short DNA sequences result in a multitude of possible TF-binding sites it remains to be experimentally tested if any of these additional sites are bound by any of these potential transcription factors during mammary gland development.

The finding of a conserved and possibly functional STAT5-binding site within the first intron of the Zswim6 gene is in accordance with other studies. The location of a functional STAT5-binding site within the first intron has been demonstrated in an independent study for the NCAM2 gene (Nelson et al., 2006). In an earlier study with a CHIP library made from IL-3 induced Ba/f3 cells, the analysis revealed that over half (54%) of the STAT5-binding sites were found within the introns of known genes (Nelson et al., 2004). Slightly less than half (23%) of these STAT5-binding sites were located within the first intron. Interesting to note is, that the STAT5 consensus sites within the NCAM2 gene lies within a very large intron, with about 200 kb to the first

exon and 82 kb from the second exon (Nelson et al., 2006). This is similar to the situation of the STAT5-binding site within the first intron of the Zswim6 gene, where the STAT5-binding site is located 62 kb from the first exon and 38 kb from the second exon (Fig. 3.15H). This very large distance from the annotated TSS and how exactly the mechanism would be for STAT5 to influence transcription from within these large introns remains to be elucidated. Potentially there may be an alternative not yet annotated promoter and TSS within Zswim6. It is also not uncommon to find transcription factors binding regions within genes. Genome-wide analysis of ERalpha binding has shown that almost half the binding sites are located within genes and that the binding site distribution can span the whole gene (Gao et al., 2008).

The reporter gene assay shows that the c831 region containing the consensus STAT5-binding site can drive reporter gene activity. Mutation of the STAT5-binding site as well as co-transfection of dominant negative STAT5 expression constructs abrogate this activity and demonstrate that the reporter activity is mainly dependent on the consensus STAT5-binding site (Fig. 3.25B-C and Fig. 3.26B). To determine if the association of STAT5 with its intronic STAT-binding site has a functional consequence on the mRNA level during mammary gland differentiation, primers had been designed to multiple locations of the Zswim6 gene. The result obtained for the Zswim6 primers that cover the last exon 14 suggest that during binding of STAT5 the transcript levels are decreased during pregnancy and lactation when compared to the transcript levels of tissue from the virgin and involution stage (Fig. 3.30B). Taken together this indicates that STAT5 may act as a potential repressor. On the other hand, that would contradict the reporter gene assays. Since the tests are performed using a plasmid where the c831 region is not in its native chromatin environment and in a cell line that may be missing potential transcription factors and cofactors found in the mammary gland during differentiation, may explain differences in gene regulation. There may also be potential interactions with factors located at the highly conserved region adjacent to the STAT5-binding site. The results obtained for the primers covering the exon 1/2 and 6/7 of Zswim6 (Fig. 3.30C-D) are similar to each other in that there is a strong reduction of transcripts during involution in comparison to pregnancy. But these results are contradictory to those for the primers covering transcript AK030163 (Fig. 3.30B). Therefore, a precise picture of the transcripts that get expressed during mammary gland differentiation need to be first determined for Zswim6, before the influence of the potential regulatory function of STAT5-binding to this region can be determined. As with all STAT5-binding sites found so far, one would require correlation with all the transcripts from the nearby area that are differentially regulated during mammary gland development.

4.2.5 Distinct over-represented motifs found in the non-repetitive STAT5 library sequences

Transcription factors establish which sets of genes are expressed in specific tissues in response to certain environmental cues by binding specific DNA target sequences. These DNA target elements are often short in sequence and degenerate. Therefore searching DNA sequences for consensus binding sites by computational means alone, without relevant support from experiments, would yield thousands of biological irrelevant binding sites for that particular transcription factor.

The hormone regulated and mammary specific expression of ß-casein and WAP is in part conferred by composite response elements (Mukhopadhyay et al., 2001; Rosen et al., 1999) and is mediated by the cooperative interaction of STAT5 and other specific transcription factors like GR and C/EBPß that bind these *cis*-regulatory elements. Therefore searching for other motifs that co-occur in the STAT5A immunoprecipitated DNA might identify other TF that co-operate with STAT5A in mediating signaling during lactation.

Surprisingly, the known consensus GAS (gamma interferon activated sequence) motif (TTCNNNGAA) preferentially bound by STAT5, was not found over-represented in the non-repetitive STAT5 library sequence data set, analyzed by Trawler. Instead a motif (family 68_cluster78) that partially resembles the known STAT-binding site was found significantly over-represented in the non-repetitive STAT5A ChIP-cloned sequence dataset analyzed (Fig. 3.32). Here the partial 4 base sequence GGAA of the motif (family 68_cluster78) is found within the palindromic STAT consensus sequence. This implicates that the STAT5A ChIP enriched DNA library contains a significant over-representation of a motif that resembles part of the consensus STAT-binding motif. It might be, that during lactation STAT5A (in complex or cooperation with other factors) might have alternative binding site preference. Other transcription factors have been described which also bind motifs that do not have strong homology to the consensus sequence. For example, using a combined ChIP-CpG island microarray approach, the new target promoters identified for E2F, contain sites that do not resemble the classic E2F consensus motifs previously identified (Weinmann and Farnham, 2002; Weinmann et al., 2002). Differences in stimulus and cell context as well as other factors cooperating with STAT5 might modulate which subset of DNA target sequences are bound.

Interesting is that the cluster_78 motif does not overlap with any of the initial identified STAT sites in the clone 153 already analyzed in the reporter gene assays (Fig. 3.33A-B). Therefore, additional experiments are underway to clarify if this cluster_78 motif can

bind STAT5A and if mutation of this motif has any affect on the STAT5 mediated reporter gene activity.

The motif for cluster_79 resembles part of a c-rel consensus binding motif. C-rel is a member of the NFkappaB transcription factor family and is thus closely linked to the activation of the cell cycle regulator, cyclin D1 by the RANK-NFkappaB pathway in the mammary gland (Hennighausen and Robinson, 2005). Cyclin D1 is also a direct STAT5 target gene and is thus an example of a gene where both the NFkappaB and STAT5 pathways converge to regulate a target gene important in mammary gland differentiation. Although the actual binding of c-rel needs still to be verified to the representative motifs in clone c831, this hints that the data from the library might provide insight into possible cis-regulatory elements where STAT5 and other transcription factors converge in regulating targets during lactation. From the other over-represented motifs it is noteworthy to mention cluster 68-68. This over-represented motif did not match any annotated consensus binding sites in the transcription factor binding site databases searched. But it is found in both verified STAT5-binding clones c153 and c831. The role of the over-represented motifs in the STAT5 mediated transcriptional activity still remains to be investigated.

5. Summary

The mammary gland of mice serves as a model system for studying differentiation in an adult animal. With the beginning of pregnancy the mammary epithelial cells undergo functional differentiation to produce milk for nourishment of the young. The transcription factor STAT5 mediates the cytokine-induced induction of the milk proteins during pregnancy and lactation in response to the lactogenic hormone prolactin. In addition to transcription factors that mediate transcription of their target genes by recruitment of the general transcription machinery to the DNA-regulator regions, specific post-translational modifications on the N-terminal tails of histones also influence expression. These histone modifications can affect chromatin structure, which is a main control barrier to transcription, by directly altering accessibility of the chromatin and by providing binding surfaces for protein complexes that can further modulate chromatin structure and regulate transcription.

In this work N-terminal histone modification marks that associate with open, permissive and repressed chromatin where investigated in different regions of two milk proteins during four stages of mammary gland development. Using the chromatin-immunoprecipitation (ChIP) assays increased acetylation of histone H3 and H4 at the 5' region, promoter and transcribed regions of ß-casein and whey acidic protein (WAP) gene were observed during pregnancy and lactation when these genes are expressed. The presence of these histone marks, which are associated with a relaxed chromatin structure, correlates with the recruitment of STAT5A and STAT5B to the promoter containing regulatory regions as well as the detection of the phosphorylated RNA polymerase II in the transcribed gene region. Both di- and tri-methylation of histone H3 lysine 4, that mark permissive and active chromatin respectively, were enriched in tissue from pregnant and lactating mice. In comparison tri-methylation of histone H3 lysine 27, a mark associated with repressed chromatin, could be observed during all stages of mammary gland tissue investigated, but appears slightly elevated in the tissue from virgin mice when ß-casein and WAP are not expressed. Together these results illustrate that the expression of the two milk proteins genes at distinct stages of mammary gland differentiation correlate with specific changes in histone modifications.

In mammary gland tissue STAT5A is important for the mammary gland epithelial cell differentiation and survival during lactation. Yet many genomic

target regions that STAT5A actually bind and which are involved in regulation of gene expression during lactation still remain unknown. Therefore, the second part of this thesis was focused on the identification of novel STAT5-binding sites that are differentiation specifically bound by STAT5A in mammary gland tissue during lactation.

In summary, the results demonstrate that the ChIP cloning method was employed successfully for the cloning of a STAT5A library and the identification of new STAT5 targets in mammary gland tissue from lactating mice. Nine of the newly identified STAT5-binding targets were verified to differentiation specifically bind STAT5A and STAT5B *in vivo* during the developmental stages of pregnancy and lactation. Even though the selection of the tested clones was biased towards STAT5-binding sites near or at known genes and for multiple STAT5 binding sites, only one out of the nine validated STAT5-binding regions is located in a traditional defined proximal promoter (c398). Except for two STAT5-binding regions (c20 and c194), which are located at least 10 kb from the next annotated known gene, six are located in the intronic regions of annotated mRNA or EST transcripts. Three, out of four verified STAT5-binding regions tested in reporter gene assays for functionality, display the ability to drive reporter gene activity in a STAT5 dependent manner. This transcriptional activity is due to the STAT5-binding sites within the cloned regions as determined by mutational analysis. Of special interest is the clone (c153) that contains one STAT5 and three STAT-like sites within a 339 bp region that is evolutionary conserved by approximately 80% between the mouse and human genome. This STAT5-binding region lies about 62 kb 5 prime of the nuclear factor I/B gene. The expression of the NFI/B mRNA transcript correlates with the *in vivo* association of STAT5A to the conserved c153 region during the mammary gland development stages of lactation. Together, these results suggest that this STAT5-binding site-containing distal 5'-flanking region of the NFI/B gene might be a *cis*-regulatory DNA region that potentially mediates STAT5 induced NFI/B gene expression in mice during lactation.

6. Zusammenfassung (German summary)

Das Milchdrüsegewebe der Maus eignet sich als Modell-System für die Untersuchung von Differenzierungsvorgängen in erwachsenen Tieren. Mit Beginn der Schwangerschaft beginnt die funktionale Differenzierung der Brustepithelzellen, um Milch für die Ernährung der Jungen zu produzieren.

Aktiviert durch das lactogene Hormon Prolaktin, vermittelt der Transkriptionsfaktor STAT5, die Induktion der Milchproteine während der späten Schwangerschaft und Stillzeit.

Nicht nur die Transkriptionsfaktoren, welche die Transkription ihrer Zielgene durch die Rekrutierung der generellen Transkriptionsmaschinerie aktivieren, sondern auch bestimmte post-translationale Modifikationen auf den N-terminalen Schwänzen der Histone, beeinflussen die Genexpression. Diese Histonmodifikationen wirken auf die Chromatinstruktur, welche die Hauptkontrollbarriere zur Transkription darstellt. Dies geschieht durch eine direkte Veränderung der Chromatinzugänglichkeit und durch die Bereitstellung von Bindungsstellen für die Rekrutierung von Proteinkomplexen, welche die Chromatinstruktur weiter modulieren und so die Transkription zusätzlich regulieren.

In dieser Arbeit werden daher N-terminale Histonmodifikationen, welche mit offenem, permissivem und reprimiertem Chromatin assoziiert sind, in den verschiedenen Regionen von zwei Milchproteinen währen der vier Stadien der Brustdrüsenentwicklung untersucht. ß-Casein und Whey Acidic Protein (WAP), sind Milchproteine, die durch Prolaktin aktiviertes STAT5 reguliert werden. Diese Proteine werden am Ende der Schwangerschaft und während der Stillzeit von differenzierten Epithelzellen im Brustdrüsengewebe exprimiert. Mit Hilfe der Chromatin-Immunoprezipitations (ChIP) Methode und Antikörpern, welche definierte Modifikationen erkennen, wurde die erhöhte Acetylierung von Lysinen der Histone H3 und H4 auf den 5’ Regionen, den Promotoren und den transkribierten Regionen der ß-casein und WAP Gene während der Schwangerschaft und Stillzeit nachgewiesen. Im Vergleich zu den anderen untersuchten genomischen Regionen, waren die acetylierten Lysine 8 und 12 von Histon H4 auf der 5’ Region von beiden Milchproteinen während der Schwangerschaft und Stillzeit stark angereichert. Die Anwesenheit dieser Histonmodifikationen, welche mit einer offenen Chromatinstruktur assoziiert werden, korreliert während der Stillzeit mit der Rekrutierung von STAT5A und

STAT5B zu den Promotoren, welche die regulatorischen STAT5-Bindungselemente enthalten. Zur gleichen Zeit ist auch die phosphorylierte Form der RNA Polymerase II auf den transkribierten Genregionen von beiden Genen zu detektieren, was wiederum mit der Expression dieser beiden Gene während der Stillzeit korreliert. Diese Resultate zeigen, dass es möglich ist auch im Gewebe die Rekrutierung von Transkriptionsfaktoren und Histonmodifikationen zu analysieren.

ChIP Untersuchungen mit Antikörpern spezifisch gegen Histon H4 Lysin 16, zeigten eine gleichmäßige Verteilung dieser Modifikation über die verschieden DNA Regionen, mit einer nur schwachen Anreicherung in der 5' Region, während der Schwangerschaft und Stillzeit. Die relative Änderung dieser Modifikation während der verschiedenen Stadien, war im Vergleich zu den anderen analysierten Histon H4 Lysin (8 und 12) Resten, nur sehr gering. Die Acetylierung von Lysin 16 ist in die Prävention einer höher geordneten Chromatinstruktur und der Erhaltung von Euchromatin involviert. Daher impliziert das Auftreten dieser Modifikation auf den untersuchten genomischen Regionen eine dekondensierte Chromatin-Struktur.

Di- und Tri-methylierung von Histon H3 Lysin 4, sind Modifikationen, welche mit permissivem und aktivem Chromatin assoziiert sind. Um die Verteilung auch dieser Modifikationen in den Milchproteingenen während der Stillzeit zu analysieren, wurden Antikörper spezifisch gegen bestimmte methylierte Lysin Reste verwendet.

Während der Schwangerschaft und der Stillzeit, wurde eine Anreicherung von Di-methyliertem Lysin 4 auf Histon H3 an dem Enhancer (für ß-Casein), 5' Region, Promotor, transkribierten Region und der 3' Region von ß-Casein nachgewiesen. Ähnlich sehen die Resultate für WAP aus. In Brustdüsengewebe von Mäusen, die keine Jungtiere mehr säugen, sieht man aber eine deutliche Abschwächung dieser Modifikation in allen Regionen der ß-Casein und WAP Gene. Im Gegensatz zu WAP jedoch, bleibt die Di-methylierung (K4) von Histon H3 in der Promoter und transkribierten Region des ß-Caseingens häufig erhalten. Dies zeigt, dass die Di-Methylierung (K4) von Histon H3 auch eine Zeitlang nach der Stillzeit fortbesteht. Diese Bewahrung der Lysine 4 Methylierung von Histon H3 könnte potenziell als Markierung beziehungsweise Lesezeichen für eine kürzlich noch transkriptionell aktive Chromatinregion dienen. ChIP-Analysen mit Antikörper, spezifisch gegen

Tri-methyliertes Lysin 4 von Histon H3, zeigte eine Anreicherung dieser Modifikation in den transkribierten- und öfters auch in den Promotor-Regionen von ß-Casein und WAP im Brustgewebe von stillenden Mäusen. Diese Histonmodifikation, die mit aktiven Genen im Euchromatin assoziiert ist, korreliert mit der aktiven Transkription beider Milchproteingene während der Stillzeit.

Im Vergleich dazu ist die Tri-methylierung von Histone H3 Lysin 27 mit reprimiertem Chromatin assoziiert und in allen untersuchten Stadien des Brustdüsengewebes nachweisbar. Eine leichte Erhöhung wurde in dem Gewebe von Jungfräulichen Mäusen oder zu einem späteren Zeitpunkt nach der Stillzeit, wenn ß-Casein und WAP nicht exprimiert werden, vorgefunden. Zusammen illustrieren diese Resultate, dass die Expression der zwei Milchproteingene zu den untersuchten Stadien der Brustdrüsendifferenzierung mit spezifischen Veränderungen der Modifikationen der Histone in diesen Genregionen einhergeht.

In Brustdrüsengewebe reguliert STAT5 die Differenzierung und sichert das Überleben der Brustepithelzellen während der Stillzeit. Bis jetzt sind nur wenige genomische Zielregionen, an denen STAT5A tatsächlich binden kann und die in der Regulation der Genexpression während der Stillzeit involviert sind, bekannt. Deshalb befasst sich der zweite Teil dieser Arbeit mit der Identifikation von neuen STAT5-Bindungsstellen, welche differenzierungsspezifisch von STAT5A im Brustdrüsengewebe während der Stillzeit gebunden werden. Hierzu wurde die ChIP-Klonierungsmethode eingesetzt, welche eine genom-weite Identifikation von neuen und direkten STAT5 DNA-Bindungselementen im Genom erlaubt. Neun, der identifizierten Regionen wurden in vivo als differenzierungsspezifische Bindungsstellen von STAT5A und STAT5B verifiziert. Die Bindung von STAT5A und STAT5B wurde nur während der Entwicklungsstadien: Schwangerschaft und Stillzeit, aber nicht im jungfräulichen Stadium oder in Mäusen, die Ihre Jungen abgesetzt hatten, detektiert. Trotz einer Selektion der getesteten Klone, die solche Klone bevorzugte, welche STAT5-Bindungselemente in der Nähe oder in bekannten Genen und mit mehreren STAT5-Bindungselementen haben, ist nur ein Klon (c398) von den neun validierten in einer traditionell definierten proximalen Promoteregion lokalisiert worden. Mit Ausnahme von zwei Klonen (c20 und

c194), welche mindestens 10 kB von dem nächsten annotierten Gen lokalisiert sind, befinden sich die anderen sechs in den intronischen Regionen von annotierten mRNA oder EST Transkripten.

Die Konservierung von Sequenzen zwischen verschiedenen Genomen weist auf eine mögliche funktionelle Bedeutung dieser Regionen hin. Vergleiche der STAT5-bindenden Maus-Sequenzen mit genomischen DNA-Sequenzen von 12 anderen Wirbeltieren zeigten, unterschiedliche Grade der Ähnlichkeit. Manche Sequenzen wiesen eine starke Homologie zu anderen Säugetieren auf, während andere Sequenzen nur Ähnlichkeiten zwischen den zwei Nagetieren Maus und Ratte aufzeigten. Während einige Maus-Sequenzen Ähnlichkeiten zu anderen Säugetiersequenzen aufwiesen, was auf eine engere evolutionäre Beziehung zu einander deutet, zeigte keine der neun STAT5-Bindungssequenzen eine Verwandtschaft zum Huhn Genom.

Die Fähigkeit von vier neu identifizierten STAT5-Bindungsregionen transkriptionelle Aktivierung zu vermitteln wurde in Reportergenexperimenten analysiert. Drei der vier verifizierten STAT5A-Bindungsregionen zeigten eine STAT5 abhängige Reportergenaktivität. Die Transkriptionsaktivität ist dabei abhängig von den STAT5-Bindungsstellen in den klonierten Regionen, wie durch Mutagenese dieser Stellen gezeigt werden konnte. Von besonderem Interesse ist der Klone (c153) welcher ein STAT5 und drei STAT-ähnliche Motive innerhalb einer 339 Bp Region besitzt. Diese ist zu 80% zwischen dem Maus und humanen Genom konserviert und liegt ungefähr 63 kB 5’ von dem Nuclear factor I/B (NFI/B) Gen entfernt. Die Expression des NFI/B mRNA Transkript’s korreliert mit der *in vivo* Assoziation von STAT5A an die konservierte c153 Region, während der Stillzeit. Zusammenfassend, weisen die Resultate darauf hin, dass diese STAT5-Bindungsstellen enthaltende 5’ Region von NFI/B Gen vermutlich ein *cis*-regulatorisches DNA Element darstellt, welches STAT5 induzierte NFI/B Gen Expression in Mäusen während der Stillzeit vermittelt. Insgesamt demonstrieren die Resultate, dass die ChIP-Konierungsmethode erfolgreich für die Klonierung einer Bibliothek von neuen STAT5 DNA-Bindungszielen aus dem Brustdrüsengewebe stillender Mäuse eingesetzt werden konnte.

Eine Analyse der DNA Klonsequenzen der Bibliothek nach dem klassischen Konsensus GAS (gamma interferon activated sequence) STAT5-

Bindungselement (TTCNNNGAA), konnte diese nur in 14% der Klon Sequenzen nachweisen.

Um zu Überprüfen ob STAT5 eventuell alternative Bindungselemente benutzt, beziehungsweise über die Bindung an andere Transkriptionsfaktoren rekrutiert wird, wurde eine Analyse, der DNA Sequenzen der Bibliothek, nach überrepräsentierten Motiven gemacht. Eine solche Analyse erlaubt die Identifikation, auch von nicht bekannten Motiven, anhand Ihrer Häufigkeit. Als Vergleichssequenzen (Sequenzen, die nicht selektioniert wurden) für diese Analyse, dienten zufällige DNA-Sequenzen, mit ähnlichen genomischen Lokalisationen, wie die Klonsequenzen. Mit Hilfe des Bioinformatikprogramms, Trawler, wurden die Sequenzen nach überrepräsentierten Motiven durchsucht. Diese Analyse kann durch die Identifikation von Bindungsstellen anderer Transkriptionsfaktoren auch zur Identifikation von Kooperationspartnern von STAT5 in der DNA-Bindung führen. Ein Teil der gefundenen Motive, weist keine Übereinstimmung mit annotierten Transkriptionsfaktormotiven in den Datenbanken auf. Bei diesen kann es sich somit um neue Bindungsmotive von anderen Transkriptionsfaktoren oder STAT5 handeln. Andere gefundene Motive dagegen ähneln annotierten Transkriptionsfaktorbindungselementen. Insbesondere ein Motiv wurde von Trawler als dem STAT5-Bindungelement ähnlich definiert. Dies weist darauf hin, dass die STAT5A ChIP angereicherte DNA Bibliothek eine signifikante Anzahl von einem Motiv enthält, welches dem Konsensus STAT5-Bindungselement ähnelt. Weitere Analysen müssen klären, ob dieses Motiv tatsächlich ein alternatives Bindungselement darstellt, welches in vivo von STAT5 (alleine oder im Komplex mit anderen Faktoren) im Brustdrüsengewebe von stillenden Mausen gebunden wird und so potenziell STAT5-Zielgene reguliert.

7. Abbreviations

A	Absorbance
Ac	Acetyl
Ahr	Dioxin-Ah receptor complex
Arhgef3	Rho guanine nucleotide exchange factor 3
ATP	Adenosine 5' triphosphate
bp	base pair
bHLH	basic-Helix-Loop-Helix
BLAST	Basic alignment search tool
BLAT	Basic-like alignment tool
C	Carboxyl terminal
CBP	CREB binding protein
ChIP	Chromatin Immunoprecipitation
Chr	Chromosome
CIP	Calf intestinal phosphatase
CIS	Cytokine-inducible SH2-containing domain
C_T	Threshold cycle
CTD	C-terminal domain
DBD	DNA binding domain
DEPC	Diethylpyrocarbonate
DMSO	Dimethylsulfoxide
DNA	Deoxynucleic acid
dNTP	deoxy nucleotide triphosphate
ds	double stranded
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamintetraacetate
EGF	Epidermal growth factor
ELF5	E74-like factor 5
EMSA	Electophoretic mobility shift assay
EN-1	Engrailed 1
EPO	Erythropoietin
ERalpha	Estrogen receptor alpha
ERBB4	Epidermal growth factor receptor (ERBB) family
EST	Expressed sequence tag
Expi	Extracellular proteinase inhibitor
EzH2	Enhancer of Zeste homology 2
FCS	Fetal calf serum
g	Relative gravity
GAS	Gamma-interferon activated sequence
GH	Growth hormone
GITC	Guanidinisothiocyanate
GM-CSF	Granulocyte/macrophage colony stimulating factor
GOI	Gene of interest
GR	Glucocorticoid receptor
GTF	General transcription factor
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HKG	Housekeeping gene
HMT	Histone methyltransferases
HOX	Homeobox
Ig	Immunoglobulin
Igsf3	Immunoglobulin superfamily, member 3
IL	Interleukin
Inpp4b	Inositol polyphosphate-4-phosphatase, type II
I	Involution
IP	Immunoprecipitation
IPTG	Isopropyl-beta-D-thiogalactoside
IRES	Internal ribosomal entry site
JAK	Janus Kinase
kb	kilobase

K	Lysine
L	Lactation
LB	Luria Bertani
LCR	Locus control region
LINE	Long interspersed nuclear element
LMPCR	Ligation mediated PCR
Lmo2	Lim only protein 2
µ	Micro
M	Mol
Me	Methyl
mef2	Monocyte enhancer factor 2
MGF	Mammary gland factor
min	minute
Mm	Mus musculus
mRNA	messenger ribonucleic acid
NCAM 2	Neural adhesion molecule 2
NCBI	National center for Biotechnology Information
NFI/B	Nuclear factor I/B
NH2	amino-terminal
OD	Optical Density
p	Plasmid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
p-CTD	Phosphorylated C-terminal domain
Pdx-1	Pancreatic duodenal homeobox
PIC	Pre-initiation complex
PRL	Prolactin
P	Pregnant
PR	Progesterone receptor
PSE	Proximal sequence element
qRT-PCR	Quantitative real time PCR
RANK-L	Receptor activator of nuclear factor (NF-kappaB)-ligand
RBm27	RNA-binding motif protein 27
RE	Recognition/response element
RNA	Ribonucleic acid
RT-PCR	Real time PCR
SDS	Sodium dodecylsulfate
SH2	Src-homology-2 domain
SOCS	Suppressors of cytokine signaling
STAT	Signal transcducers and activators of transcription
TAD	Transcriptional activation domain
TBP	TATA-box binding protein
TEB	Terminal end buds
TESS	Transcription Element Search System
TF	Transcription factor
TSS	Transcription start site
UCSC	University of California Santa Cruz
UV	Ultra violet
V	Virgin
WAP	Whey acidic acid
wt	wild type
Y	Tyrosine
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
YY1	Yin and yang 1
Zswim6	Zinc finger, SWIM domain containing 6

8. References

Agalioti, T., Chen, G. and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, 111, 381-392.

Auernhammer, C.J., Bousquet, C. and Melmed, S. (1999). Autoregulation of pituitary corticotroph SOCS-3 expression: characterization of the murine SOCS-3 promoter. *Proc Natl Acad Sci U S A*, 96, 6964-6969.

Bajic, V.B., Tan, S.L., Suzuki, Y. and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. *Nat Biotechnol*, 22, 1467-1473.

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W. and Groner, B. (1988). Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *Embo J*, 7, 2089-2095.

Ball, S.M. (1998). The development of the terminal end bud in the prepubertal-pubertal mouse mammary gland. *Anat Rec*, 250, 459-464.

Barrera, L.O. and Ren, B. (2006). The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol*, 18, 291-298.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129, 823-837.

Baus, D. and Pfitzner, E. (2006). Specific function of STAT3, SOCS1, and SOCS3 in the regulation of proliferation and survival of classical Hodgkin lymphoma cells. *Int J Cancer*, 118, 1404-1413.

Berger, S.L. (1999). Gene activation by histone and factor acetyltransferases. *Curr Opin Cell Biol*, 11, 336-341.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. *Nature*, 447, 407-412.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R. et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120, 169-181.

Birnboim, H.C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol*, 100, 243-255.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447, 799-816.

Blackwood, E.M. and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science*, 281, 60-63.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F. and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev*, 15, 3286-3295.

Brisken, C. and Rajaram, R.D. (2006). Alveolar and lactogenic differentiation. *J Mammary Gland Biol Neoplasia*, 11, 239-248.

Brockman, J.L., Schroeder, M.D. and Schuler, L.A. (2002). PRL activates the cyclin D1 promoter via the Jak2/Stat pathway. *Mol Endocrinol*, 16, 774-784.

Bromberg, J.F. (2001). Activation of STAT proteins and growth control. *Bioessays*, 23, 161-169.

Brown, C.E., Lechner, T., Howe, L. and Workman, J.L. (2000). The many HATs of transcription coactivators. *Trends Biochem Sci*, 25, 15-19.

Bry, C., Maass, K., Miyoshi, K., Willecke, K., Ott, T., Robinson, G.W. and Hennighausen, L. (2004). Loss of connexin 26 in mammary epithelium during early but not during late pregnancy results in unscheduled apoptosis and impaired development. *Dev Biol*, 267, 418-429.

Buser, A.C., Gass-Handel, E.K., Wyszomierski, S.L., Doppler, W., Leonhardt, S.A., Schaack, J., Rosen, J.M., Watkin, H., Anderson, S.M. and Edwards, D.P. (2007). Progesterone receptor repression of prolactin/signal transducer and activator of transcription 5-mediated transcription of the beta-casein gene in mammary epithelial cells. *Mol Endocrinol*, 21, 106-125.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298, 1039-1043.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. et al. (2005). The transcriptional landscape of the mammalian genome. *Science*, 309, 1559-1563.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C. et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 38, 626-635.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R. et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122, 33-43.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F. et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat Genet*, 38, 1289-1297.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J. et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, 116, 499-509.

Chammas, R., Taverna, D., Cella, N., Santos, C. and Hynes, N.E. (1994). Laminin and tenascin assembly and expression regulate HC11 mouse mammary cell differentiation. *J Cell Sci*, 107 (Pt 4), 1031-1040.

Clark, D.E., Williams, C.C., Duplessis, T.T., Moring, K.L., Notwick, A.R., Long, W., Lane, W.S., Beuvink, I., Hynes, N.E. and Jones, F.E. (2005). ERBB4/HER4 potentiates STAT5A transcriptional activity by regulating novel STAT5A serine phosphorylation events. *J Biol Chem*, 280, 24175-24180.

Clarkson, R.W., Boland, M.P., Kritikou, E.A., Lee, J.M., Freeman, T.C., Tiffen, P.G. and Watson, C.J. (2006). The genes induced by signal transducer and activators of transcription (STAT)3 and STAT5 in mammary epithelial cells define the roles of these STATs in mammary development. *Mol Endocrinol*, 20, 675-685.

Clayton, A.L., Hazzalin, C.A. and Mahadevan, L.C. (2006). Enhanced histone acetylation and transcription: a dynamic perspective. *Mol Cell*, 23, 289-296.

Copeland, N.G., Gilbert, D.J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J.E., Jr., Mui, A.L., Miyajima, A., Quelle, F.W., Ihle, J.N. et al. (1995). Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics*, 29, 225-228.

Cosma, M.P. (2002). Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell*, 10, 227-236.

Cui, Y., Riedlinger, G., Miyoshi, K., Tang, W., Li, C., Deng, C.X., Robinson, G.W. and Hennighausen, L. (2004). Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. *Mol Cell Biol*, 24, 8037-8047.

Dahmus, M.E. (1995). Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim Biophys Acta*, 1261, 171-182.

Danielson, K.G., Oborn, C.J., Durban, E.M., Butel, J.S. and Medina, D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. *Proc Natl Acad Sci U S A*, 81, 3756-3760.

Dion, M.F., Altschuler, S.J., Wu, L.F. and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci U S A*, 102, 5501-5506.

Doppler, W., Groner, B. and Ball, R.K. (1989). Prolactin and glucocorticoid hormones synergistically induce expression of transfected rat beta-casein gene promoter constructs in a mammary epithelial cell line. *Proc Natl Acad Sci U S A*, 86, 104-108.

Edelstein, L.C., Pan, A. and Collins, T. (2005). Chromatin modification and the endothelial-specific activation of the E-selectin gene. *J Biol Chem*, 280, 11192-11202.

Ehret, G.B., Reichenbach, P., Schindler, U., Horvath, C.M., Fritz, S., Nabholz, M. and Bucher, P. (2001). DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. *J Biol Chem*, 276, 6675-6688.

Ettwiller, L., Paten, B., Ramialison, M., Birney, E. and Wittbrodt, J. (2007). Trawler: de novo regulatory motif discovery pipeline for chromatin immunoprecipitation. *Nat Methods*, 4, 563-565.

Fata, J.E., Kong, Y.Y., Li, J., Sasaki, T., Irie-Sasaki, J., Moorehead, R.A., Elliott, R., Scully, S., Voura, E.B., Lacey, D.L. et al. (2000). The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell*, 103, 41-50.

Fischle, W., Wang, Y. and Allis, C.D. (2003). Histone and chromatin cross-talk. *Curr Opin Cell Biol*, 15, 172-183.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D. and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev*, 17, 1870-1881.

Francis, J., Chakrabarti, S.K., Garmey, J.C. and Mirmira, R.G. (2005). Pdx-1 links histone H3-Lys-4 methylation to RNA polymerase II elongation during activation of insulin transcription. *J Biol Chem*, 280, 36244-36253.

Fukuda, H., Sano, N., Muto, S. and Horikoshi, M. (2006). Simple histone acetylation plays a complex role in the regulation of gene expression. *Brief Funct Genomic Proteomic*, 5, 190-208.

Gao, H., Falt, S., Sandelin, A., Gustafsson, J.A. and Dahlman-Wright, K. (2008). Genome-wide identification of estrogen receptor alpha-binding sites in mouse liver. *Mol Endocrinol*, 22, 10-22.

Gerber, M. and Shilatifard, A. (2003). Transcriptional elongation by RNA polymerase II and histone methylation. *J Biol Chem*, 278, 26303-26306.

Grimley, P.M., Dong, F. and Rui, H. (1999). Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation. *Cytokine Growth Factor Rev*, 10, 131-157.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*, 11, 394-403.

Hampsey, M. and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*, 113, 429-432.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, 166, 557-580.

Happ, B. and Groner, B. (1993). The activated mammary gland specific nuclear factor (MGF) enhances in vitro transcription of the beta-casein gene promoter. *J Steroid Biochem Mol Biol*, 47, 21-30.

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. et al. (2006). Socs2 and elf5 mediate prolactin-induced mammary gland development. *Mol Endocrinol*, 20, 1177-1187.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A. et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39, 311-318.

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

Hennighausen, L. and Robinson, G.W. (2001). Signaling pathways in mammary gland development. *Dev Cell*, 1, 467-475.

Hennighausen, L. and Robinson, G.W. (2005). Information networks in the mammary gland. *Nat Rev Mol Cell Biol*, 6, 715-725.

Hennighausen, L. and Robinson, G.W. (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes Dev*, 22, 711-721.

Hirose, Y. and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev*, 14, 1415-1429.

Hirose, Y. and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem*, 141, 601-608.

Hobbs, A.A., Richards, D.A., Kessler, D.J. and Rosen, J.M. (1982). Complex hormonal regulation of rat casein gene expression. *J Biol Chem*, 257, 3598-3605.

Hoelbl, A., Kovacic, B., Kerenyi, M.A., Simma, O., Warsch, W., Cui, Y., Beug, H., Hennighausen, L., Moriggl, R. and Sexl, V. (2006). Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. *Blood*, 107, 4898-4906.

Hou, Z., Srivastava, S., Mistry, M.J., Herbst, M.P., Bailey, J.P. and Horseman, N.D. (2003). Two tandemly linked interferon-gamma-activated sequence elements in the promoter of glycosylation-dependent cell adhesion molecule 1 gene synergistically respond to prolactin in mouse mammary epithelial cells. *Mol Endocrinol*, 17, 1910-1920.

Humphreys, R.C., Krajewska, M., Krnacik, S., Jaeger, R., Weiher, H., Krajewski, S., Reed, J.C. and Rosen, J.M. (1996). Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development*, 122, 4013-4022.

Jolivet, G., Pantano, T. and Houdebine, L.M. (2005). Regulation by the extracellular matrix (ECM) of prolactin-induced alpha s1-casein gene expression in rabbit primary mammary cells: role of STAT5, C/EBP, and chromatin structure. *J Cell Biochem*, 95, 313-327.

Kabotyanski, E.B., Huetter, M., Xian, W., Rijnkels, M. and Rosen, J.M. (2006). Integration of prolactin and glucocorticoid signaling at the beta-casein promoter and enhancer by ordered recruitment of specific transcription factors and chromatin modifiers. *Mol Endocrinol*, 20, 2355-2368.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. *Cell*, 116, 259-272.

Kiang, D.T., Jin, N., Tu, Z.J. and Lin, H.H. (1997). Upstream genomic sequence of the human connexin26 gene. *Gene*, 199, 165-171.

Kim, A. and Dean, A. (2004). Developmental stage differences in chromatin subdomains of the beta-globin locus. *Proc Natl Acad Sci U S A*, 101, 7028-7033.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature*, 436, 876-880.

Kisseleva, T., Bhattacharya, S., Braunstein, J. and Schindler, C.W. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*, 285, 1-24.

Komarnitsky, P., Cho, E.J. and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, 14, 2452-2460.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, 184, 868-871.

Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A*, 104, 12955-12961.

Kouskouti, A. and Talianidis, I. (2005). Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *Embo J*, 24, 347-357.

Kouzarides, T. (1999). Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev*, 9, 40-48.

Kouzarides, T. (2002). Histone methylation in transcriptional control. *Curr Opin Genet Dev*, 12, 198-209.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128, 693-705.

Kubo, M., Hanada, T. and Yoshimura, A. (2003). Suppressors of cytokine signaling and immunity. *Nat Immunol*, 4, 1169-1176.

Lachner, M., O'Sullivan, R.J. and Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. *J Cell Sci*, 116, 2117-2124.

LeBaron, M.J., Xie, J. and Rui, H. (2005). Evaluation of genome-wide chromatin library of Stat5 binding sites in human breast cancer. *Mol Cancer*, 4, 6.

Lechner, J., Welte, T., Tomasi, J.K., Bruno, P., Cairns, C., Gustafsson, J. and Doppler, W. (1997). Promoter-dependent synergy between glucocorticoid receptor and Stat5 in the activation of beta-casein gene transcription. *J Biol Chem*, 272, 20954-20960.

Levy, D.E. and Darnell, J.E., Jr. (2002). Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol*, 3, 651-662.

Li, S. and Rosen, J.M. (1994). Distal regulatory elements required for rat whey acidic protein gene expression in transgenic mice. *J Biol Chem*, 269, 14235-14243.

Li, S. and Rosen, J.M. (1995). Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. *Mol Cell Biol*, 15, 2063-2070.

Li, C., Harding, G.A., Parise, J., McNamara-Schroeder, K.J. and Stumph, W.E. (2004). Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. *Mol Cell Biol*, 24, 1897-1906.

Li, B., Carey, M. and Workman, J.L. (2007). The role of chromatin during transcription. *Cell*, 128, 707-719.

Lim, C.P. and Cao, X. (2006). Structure, function, and regulation of STAT proteins. *Mol Biosyst*, 2, 536-550.

Lin, C.Q. and Bissell, M.J. (1993). Multi-faceted regulation of cell differentiation by extracellular matrix. *Faseb J*, 7, 737-743.

Litterst, C.M., Kliem, S., Marilley, D. and Pfitzner, E. (2003). NCoA-1/SRC-1 is an essential coactivator of STAT5 that binds to the FDL motif in the alpha-helical region of the STAT5 transactivation domain. *J Biol Chem*, 278, 45340-45351.

Litterst, C.M., Kliem, S., Lodrini, M. and Pfitzner, E. (2005). Coactivators in gene regulation by STAT5. *Vitam Horm*, 70, 359-386.

Liu, X., Robinson, G.W., Gouilleux, F., Groner, B. and Hennighausen, L. (1995). Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc Natl Acad Sci U S A*, 92, 8831-8835.

Liu, X., Robinson, G.W. and Hennighausen, L. (1996). Activation of Stat5a and Stat5b by tyrosine phosphorylation is tightly linked to mammary gland differentiation. *Mol Endocrinol*, 10, 1496-1506.

Liu, X., Robinson, G.W., Wagner, K.U., Garrett, L., Wynshaw-Boris, A. and Hennighausen, L. (1997). Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev*, 11, 179-186.

Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N. and Rando, O.J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol*, 3, e328.

Long, W., Wagner, K.U., Lloyd, K.C., Binart, N., Shillingford, J.M., Hennighausen, L. and Jones, F.E. (2003). Impaired differentiation and lactational failure of Erbb4-deficient mammary glands identify ERBB4 as an obligate mediator of STAT5. *Development*, 130, 5257-5268.

Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M. and Frazer, K.A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, 288, 136-140.

Louie, M.C., Yang, H.Q., Ma, A.H., Xu, W., Zou, J.X., Kung, H.J. and Chen, H.W. (2003). Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. *Proc Natl Acad Sci U S A*, 100, 2226-2230.

Martin, C. and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*, 6, 838-849.

Masso-Welch, P.A., Darcy, K.M., Stangle-Castor, N.C. and Ip, M.M. (2000). A developmental atlas of rat mammary gland histology. *J Mammary Gland Biol Neoplasia*, 5, 165-185.

Maston, G.A., Evans, S.K. and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7, 29-59.

Matsumoto, A., Masuhara, M., Mitsui, K., Yokouchi, M., Ohtsubo, M., Misawa, H., Miyajima, A. and Yoshimura, A. (1997). CIS, a cytokine inducible SH2 protein, is a target of the JAK-STAT5 pathway and modulates STAT5 activation. *Blood*, 89, 3148-3154.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, 31, 374-378.

Meier, V.S. and Groner, B. (1994). The nuclear factor YY1 participates in repression of the beta-casein gene promoter in mammary epithelial cells and is counteracted by mammary gland factor during lactogenic hormone induction. *Mol Cell Biol*, 14, 128-137.

Mertens, C. and Darnell, J.E., Jr. (2007). SnapShot: JAK-STAT signaling. *Cell*, 131, 612.

Meyer, W.K., Reichenbach, P., Schindler, U., Soldaini, E. and Nabholz, M. (1997). Interaction of STAT5 dimers on two low affinity binding sites mediates interleukin 2 (IL-2) stimulation of IL-2 receptor alpha gene transcription. *J Biol Chem*, 272, 31821-31828.

Miller, W., Makova, K.D., Nekrutenko, A. and Hardison, R.C. (2004). Comparative genomics. *Annu Rev Genomics Hum Genet*, 5, 15-56.

Miyoshi, K., Shillingford, J.M., Smith, G.H., Grimm, S.L., Wagner, K.U., Oka, T., Rosen, J.M., Robinson, G.W. and Hennighausen, L. (2001). Signal transducer and activator of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. *J Cell Biol*, 155, 531-542.

Moriggl, R., Gouilleux-Gruart, V., Jahne, R., Berchtold, S., Gartmann, C., Liu, X., Hennighausen, L., Sotiropoulos, A., Groner, B. and Gouilleux, F. (1996). Deletion of the carboxyl-terminal transactivation domain of MGF-Stat5 results in sustained DNA binding and a dominant negative phenotype. *Mol Cell Biol*, 16, 5691-5700.

Morillon, A. (2006). Is histone loss a common feature of DNA metabolism regulation? *Biochem Cell Biol*, 84, 450-462.

Mui, A.L., Wakao, H., O'Farrell, A.M., Harada, N. and Miyajima, A. (1995). Interleukin-3, granulocyte-macrophage colony stimulating factor and interleukin-5 transduce signals through two STAT5 homologs. *Embo J*, 14, 1166-1175.

Mukhopadhyay, S.S., Wyszomierski, S.L., Gronostajski, R.M. and Rosen, J.M. (2001). Differential interactions of specific nuclear factor I isoforms with the glucocorticoid receptor and STAT5 in the cooperative regulation of WAP gene transcription. *Mol Cell Biol*, 21, 6859-6869.

Nelson, E.A., Walker, S.R., Alvarez, J.V. and Frank, D.A. (2004). Isolation of unique STAT5 targets by chromatin immunoprecipitation-based gene identification. *J Biol Chem*, 279, 54724-54730.

Nelson, E.A., Walker, S.R., Li, W., Liu, X.S. and Frank, D.A. (2006). Identification of human STAT5-dependent gene regulatory elements based on interspecies homology. *J Biol Chem*, 281, 26216-26224.

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

Ng, H.H., Robert, F., Young, R.A. and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*, 11, 709-719.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R. et al. (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell*, 9, 1201-1213.

Onishi, M., Nosaka, T., Misawa, K., Mui, A.L., Gorman, D., McMahon, M., Miyajima, A. and Kitamura, T. (1998). Identification and characterization of a constitutively active STAT5 mutant that promotes cell proliferation. *Mol Cell Biol*, 18, 3871-3879.

Ormandy, C.J., Camus, A., Barra, J., Damotte, D., Lucas, B., Buteau, H., Edery, M., Brousse, N., Babinet, C., Binart, N. et al. (1997). Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. *Genes Dev*, 11, 167-178.

Orphanides, G. and Reinberg, D. (2002). A unified theory of gene expression. *Cell*, 108, 439-451.

Pfitzner, E., Jahne, R., Wissler, M., Stoecklin, E. and Groner, B. (1998). p300/CREB-binding protein enhances the prolactin-mediated transcriptional induction through direct interaction with the transactivation domain of Stat5, but does not participate in the Stat5-mediated suppression of the glucocorticoid response. *Mol Endocrinol*, 12, 1582-1593.

Pittius, C.W., Sankaran, L., Topper, Y.J. and Hennighausen, L. (1988). Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. *Mol Endocrinol*, 2, 1027-1032.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B. and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science*, 300, 131-135.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E. et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122, 517-527.

Pullan, S., and Streuli, C. H. (1996) *Epithelial Cell Culture*. Cambridge University Press, Cambridge, UK.

Rascle, A., Johnston, J.A. and Amati, B. (2003). Deacetylase activity is required for recruitment of the basal transcription machinery and transactivation by STAT5. *Mol Cell Biol*, 23, 4162-4173.

Rascle, A. and Lees, E. (2003). Chromatin acetylation and remodeling at the Cis promoter during STAT5-induced transcription. *Nucleic Acids Res*, 31, 6882-6890.

Raught, B., Khursheed, B., Kazansky, A. and Rosen, J. (1994). YY1 represses beta-casein gene expression by preventing the formation of a lactation-associated complex. *Mol Cell Biol*, 14, 1752-1763.

Rice, J.C. and Allis, C.D. (2001). Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr Opin Cell Biol*, 13, 263-273.

Rijnkels, M., Wheeler, D.A., de Boer, H.A. and Pieper, F.R. (1997). Structure and expression of the mouse casein gene locus. *Mamm Genome*, 8, 9-15.

Rijnkels, M., Elnitski, L., Miller, W. and Rosen, J.M. (2003). Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics*, 82, 417-432.

Roh, T.Y., Cuddapah, S., Cui, K. and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci U S A*, 103, 15782-15787.

Rosen, J. M., Wyszomierski, S. and Hadsell, D. (1999). Regulation of milk protein gene expression. *Annu Rev Nutr* 19: 407-36.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386.

Sadowski, C.L., Choi, T.S., Le, M., Wheeler, T.T., Wang, L.H. and Sadowski, H.B. (2001). Insulin Induction of SOCS-2 and SOCS-3 mRNA expression in C2C12 Skeletal Muscle Cells Is Mediated by Stat5*. *J Biol Chem*, 276, 20703-20710.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8, 424-436.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, 419, 407-411.

Schindler, U., Wu, P., Rothe, M., Brasseur, M. and McKnight, S.L. (1995). Components of a Stat recognition code: evidence for two layers of molecular selectivity. *Immunity*, 2, 689-697.

Schindler, C., Levy, D.E. and Decker, T. (2007). JAK-STAT signaling: from interferons to cytokines. *J Biol Chem*, 282, 20059-20063.

Schmidhauser, C., Bissell, M.J., Myers, C.A. and Casperson, G.F. (1990). Extracellular matrix and hormones transcriptionally regulate bovine beta-casein 5' sequences in stably transfected mouse mammary cells. *Proc Natl Acad Sci U S A*, 87, 9118-9122.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C. and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol*, 6, 73-77.

Schones, D.E. and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet*, 9, 179-191.

Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I. and Groudine, M. (2000). Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev*, 14, 940-950.

Shang, Y., Myers, M. and Brown, M. (2002). Formation of the androgen receptor transcription complex. *Mol Cell*, 9, 601-610.

Shillingford, J.M., Miyoshi, K., Robinson, G.W., Grimm, S.L., Rosen, J.M., Neubauer, H., Pfeffer, K. and Hennighausen, L. (2002). Jak2 is an essential tyrosine kinase involved in pregnancy-mediated development of mammary secretory epithelium. *Mol Endocrinol*, 16, 563-570.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R. and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, 311, 844-847.

Smale, S.T. and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu Rev Biochem*, 72, 449-479.

Soldaini, E., John, S., Moro, S., Bollenbacher, J., Schindler, U. and Leonard, W.J. (2000). DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites. *Mol Cell Biol*, 20, 389-401.

Spencer, V.A. and Davie, J.R. (1999). Role of covalent modifications of histones in regulating gene expression. *Gene*, 240, 1-12.

Srivastava, S., Matsuda, M., Hou, Z., Bailey, J.P., Kitazawa, R., Herbst, M.P. and Horseman, N.D. (2003). Receptor activator of NF-kappaB ligand induction via Jak2 and Stat5a in mammary epithelial cells. *J Biol Chem*, 278, 46171-46178.

Stoecklin, E., Wissler, M., Moriggl, R. and Groner, B. (1997). Specific DNA binding of Stat5, but not of glucocorticoid receptor, is required for their functional cooperation in the regulation of gene transcription. *Mol Cell Biol*, 17, 6708-6716.

Strahl, B.D. and Allis, C.D. (2000). The language of covalent histone modifications. *Nature*, 403, 41-45.

Teglund, S., McKay, C., Schuetz, E., van Deursen, J.M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G. and Ihle, J.N. (1998). Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell*, 93, 841-850.

Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C. et al. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 424, 788-793.

Tonko-Geymayer, S., Goupille, O., Tonko, M., Soratroi, C., Yoshimura, A., Streuli, C., Ziemiecki, A., Kofler, R. and Doppler, W. (2002). Regulation and function of the cytokine-inducible SH-2 domain proteins, CIS and SOCS3, in mammary epithelial cells. *Mol Endocrinol*, 16, 1680-1695.

Topper, Y.J. and Freeman, C.S. (1980). Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol Rev*, 60, 1049-1106.

Turner, B.M. (1993). Decoding the nucleosome. *Cell*, 75, 5-8.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. *Bioessays*, 22, 836-845.

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

Van Holde, K.E. (1988) *Chromatin*. New York Springer Press, New York.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 3, RESEARCH0034.

Verdier, F., Rabionet, R., Gouilleux, F., Beisenherz-Huss, C., Varlet, P., Muller, O., Mayeux, P., Lacombe, C., Gisselbrecht, S. and Chretien, S. (1998). A sequence of the CIS gene promoter interacts preferentially with two associated STAT5A dimers: a distinct biochemical difference between STAT5A and STAT5B. *Mol Cell Biol*, 18, 5852-5860.

Vlieghe, D., Sandelin, A., De Bleser, P.J., Vleminckx, K., Wasserman, W.W., van Roy, F. and Lenhard, B. (2006). A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res*, 34, D95-97.

Wagner, K.U., Krempler, A., Triplett, A.A., Qi, Y., George, N.M., Zhu, J. and Rui, H. (2004). Impaired alveologenesis and maintenance of secretory mammary epithelial cells in Jak2 conditional knockout mice. *Mol Cell Biol*, 24, 5510-5520.

Wakao, H., Schmitt-Ney, M. and Groner, B. (1992). Mammary gland-specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89 kDa. *J Biol Chem*, 267, 16365-16370.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P. and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol Cell*, 8, 1207-1217.

Wang, X. and Seed, B. (2003). A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res*, 31, e154.

Watson, C.J., Gordon, K.E., Robertson, M. and Clark, A.J. (1991). Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Res*, 19, 6603-6610.

Watson, C.J. and Burdon, T.G. (1996). Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Rev Reprod*, 1, 1-5.

Weinmann, A.S. and Farnham, P.J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods*, 26, 37-47.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev*, 16, 235-244.

Welte, T., Garimorth, K., Philipp, S. and Doppler, W. (1994). Prolactin-dependent activation of a tyrosine phosphorylated DNA binding factor in mouse mammary epithelial cells. *Mol Endocrinol*, 8, 1091-1102.

West, A.G. and Fraser, P. (2005). Remote control of gene transcription. *Hum Mol Genet*, 14 Spec No 1, R101-111.

Wolffe, A.P. and Hayes, J.J. (1999). Chromatin disruption and modification. *Nucleic Acids Res*, 27, 711-720.

Workman, J.L. and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem*, 67, 545-579.

Woychik, N.A. and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. *Cell*, 108, 453-463.

Wyszomierski, S.L., Yeh, J. and Rosen, J.M. (1999). Glucocorticoid receptor/signal transducer and activator of transcription 5 (STAT5) interactions enhance STAT5 activation by prolonging STAT5 DNA binding and tyrosine phosphorylation. *Mol Endocrinol*, 13, 330-343.

Xu, R., Spencer, V.A. and Bissell, M.J. (2007). Extracellular matrix-regulated gene expression requires cooperation of SWI/SNF and transcription factors. *J Biol Chem*, 282, 14992-14999.

Yao, Z., Cui, Y., Watford, W.T., Bream, J.H., Yamaoka, K., Hissong, B.D., Li, D., Durum, S.K., Jiang, Q., Bhandoola, A. et al. (2006). Stat5a/b are essential for normal lymphoid development and differentiation. *Proc Natl Acad Sci U S A*, 103, 1000-1005.

Zhang, J.G., Farley, A., Nicholson, S.E., Willson, T.A., Zugaro, L.M., Simpson, R.J., Moritz, R.L., Cary, D., Richardson, R., Hausmann, G. et al. (1999). The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proc Natl Acad Sci U S A*, 96, 2071-2076.

Zhang, Y. and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*, 15, 2343-2360.

Zhu, Q., Maitra, U., Johnston, D., Lozano, M. and Dudley, J.P. (2004). The homeodomain protein CDP regulates mammary-specific gene transcription and tumorigenesis. *Mol Cell Biol*, 24, 4810-4823.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M. and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res*, 35, 5532-5544.

9. Appendix

9.1 STAT5A library sequences

STAT5A Library (non-redundant, non-repetitive DNA sequence dataset)
Clone ID Sequence

11-M13for_3
TGCCCCCTACCGGNTTATTATGAAACTGCAGGCCGCTCAACTCACCGCCCTCCTGGNTTGTAGGTGTCATCTGTGGTCATCCGGTCCTCACCA
TGTGGACCCCGCGAGTGCCCCCTCCACGCCCGGNCCTGTCGTTTTTCCTGNTGTTGCTCCTGGGGGTCA
13-M13revfor
NNTTCNNTTCCTATTACNNCACAGTTAGAGACTGTGGGTTTGGGGTTTCNAAAGCTCACCCGTGTGTGAAAACAGGGAGACTTGGTACTTAAG
AATGGTGCCCAGACAGAACTCAGCATACTGTTNTTACAGGAAGACCAGGGTACACCATCCCTAAGTCACGTGCTTTTCTGGTGACATAATCCTCATATAATTT
CAACATTTCTGAAACACTGCAATCAGAGGCTAAGATCCAGCGATTCTGTTTTCAGATATTCTCAGGAAGCCAAAAGCAAGGGCCCGGGTTTCTGTGNGTTCA
GTCTTCTCAAACCCATCTAAACGTCCACTGGCGGATGAGCGGGGACACGAGCTCAGGCACAGGCCATGCACGGAAGGCGTGGAGTTGACTCAGCCNNAG
GAATGAAGACGATGCTGGGACAGGCCACAGTATGGACCAGTCTTTGGGCTCATATGAGCTGCAGGGNNGTTGAACTTGGGAGCAGAATGCGTGCAGGGGT
GGAGGAGGAGGCAGTGTTAGTGTTTAGACTTCCGATGGGGAACCATCTGGAAGCTGTGGGATGAGTGGTGGTGGCAGCCAGGCAGCGNGGTACAGTGAG
AACTGCTGAGNAGAGCAGGGTCGGTGTTGTGTGTTCCATCTCGAGAGAACGGTAGAGGGGAACCAGAGGCAAGGTGCTGGACACCTGTCCCGTTCTGTCA
GTTCAGT
14-M13for_1
CCACAGGCAGCCAGAGAATGGGGTATGGGGGTGTTATGTGAGACCCGGGCATATTGTCCCTCTGAGAATCTATGCTCTAGATACTGGACTCA
AACAAGGTTCTGTACAGGGNgTCCTCAGATCCAGGCCAGGGCTCATGGAAAGGGTAACTAAGAGCTGGTGACTGCTGTCTGTATGATGGGGACCCAAATGC
TAGCATAGGTGGGCTTGTTGCAGGATTGGATAGCTCTGTGAGGGGACGGCTTGGTCTTGGGGCTGGGTCCCTGCTCTGTCTGATACCAATGCTCATCTCCC
TAATACTGACCTTATCCCATGCCTTGGCACTCTGTGAATGTGCTTTGGGGAATGATTAAAGCTTATGCGGTCTCT
151-M13rev_3
CTCACATCTTCCTTATTAGCAGTGAGACCACAATTTAACGTCTCTGAGCCTCAGACTTTTCATCTGTAAAGTAGGGATAACAGCAGTCAGCATC
CTAACGTATGAGTCAAGGATGTTGGGAGAAAACTGTGTAGTGCCTGCTAGGAACCCAGTAAATGCAAGTTCTACTCTTCTATATTGTCAGGAAACACCAAGG
TCAAAGCAACAGGCAACCTTGGCTTCTAAACCTGCTGTCACAGGAGGCTGCTCTCTCTGGGAGAGTGTAAGAGCATCCCCAGGGAATGTTTAGAAAGGAAT
CTAGGCTGGCAGCGTGGTGGCCTAAGCAGACTGTGGGGAGTTCCATTCCAGGCTGCATTGTATAGTGAAGCATTTGTCTCAACCACCACCAACCACCACC
153-M13rev_4
GGGAAGGATAAAGGAAGTCTGACTCACANNGGGAGTCGGAGGAGGATAGAGCTATTGAAAGAACAGGAACAGGTATTTTCCGTCATATTCATT
TTCTTAAAATTGAATTAGGTTTCTATGAAAAGCCGTTTTGTAGAATAACAGCCTCTTNATGGTTTTACTAACTGAGACATAAAAGACTACCTGAGGCATGTTAT
GATTGCTGAACACACNNNAAAAATCCCNTACTTGCTGGTTATGCTGAATGATGATCAGAANAAGGCTCTTGTGAATAACCTGACC
154-M13rev_1
CATGTCTCAGTCAAAACCAATAAAATAAATAAATACAAAATCGAAAGATTAAAATGATTTGAAGATCAGAAACATTTGAGACTAAAAATATTCCTT
CTCCACACAACTTGGCTCTGTTTGTTTGTTTGTTTGAGATTGGGGCTCACTATACATGCCTGGCTGGCCTGGAACTCTTACCATATAATTCAGGCTGCCCTCC
AACTCACAGACATCTGCTTGCCTCTGCCTCCAAGTGCTATGGTTAAAAGTATGGATCCCCGTGCCTGGTAGCTGATTTTGATAGAAGGAAAGAAAAGAGTAC
CTCCTCAGGTGAGAGGTAGGATGCCACAGCTGAAAGATGAGTTATATATAGTCAGTGAACTAACTGGCTTCTGAATATCTTC
155-M13rev_2
GGACATCTCAGAAGAGGTCTCAGACTGTCACACAGCAAATAAACAACAGGGCTGGAGGGCTGGCCAGGTCTGCTTGGCTGTGTTCCTCCCTC
CTCTAGCTGCCCCTCTGTGGCCCTGATTGTTCCCAGGAAAGTTGCAGGAGTGGGGTGGAGAAGAGGTGACTCCCACAGAGCCTGCCAGGTCTAAGATACA
AGGACAGGACTCTAAGTTCACTTTCCTGAAGCGTTTCAGCCTGCTTTGGCAAGTCCTAACCAAGGATCCTCAGGTTAAAGACACAAAGAAGTCTGTGAGAGC
TGTCACCTCAGGGCCTGGCTCTGGCAAGGACAGGCCCTGGCAGGCCTCCTCCTCTTCCTCCCTCTCTGCCTTGGCTTCTCTCCTGCCGTAACAGA
158-M13rev_3
CCCCAATGGATCATGTCTCTGGAGCCCACCCCAGACACTCAGTGTAGATTGCTTCCTTACAGCCTCTGAGGAGGACCACAGCAACTCGGCCT
GCTTCGCCTGCGTCCTGCTGAGCCACGGGGAAGAGGACCTGATTTACGGGAAAGATGGCGTGACACCCATAAAGGATCTGACAGCTCATTTTAGGGGAGA
CCGATGCAAAACCCTGTTAGAGAAACCCAAACTCTTCTTCATTCAGGTGCTTTTGAGGGCAGTGGGACTGCATGTGGGGAAAGCTAACAGGGCAGGAGCGG
TATTATCTGTACAAATCAACGTGCCATTTCCCTGGGGAATGCACGCTGGGGCCGGGGCTTCCTGCCCAGAAATGGTTGGTACGGGCACTTGCTGTATTT
160-M13rev_4
ATGCACATGGGCTCTTCTGTCATTCCTGTGGTACTTCATAAATAGTGATTTTAGTACGTCACACACTAATTTGTCTTTGAGCCACATCCAGGAAG
TCAACGTCCAAACTGCTACATCAGTTATAACCTTTGTTGTCCATCTGAGTCCTCCAAATGGTCATCCTCGCCTTGTAATTATACAGATACTCTACTCCTGATAG
AGAATCAGTCACTACTTGAGATGAAGCAGACTAATCCTTGATCTTCTAAAGAATTTCCACGCTCTGGCCAGTCTATTTCAGTAGCTGTGTCCCCTTTCAAGCT
GCATGTGCATCACTAAAAAGCCTTTTAAAATTATCCTCCTTTTCTCTTCTCCTTCTCTTCTCCTTCTCCTCTCCTCCTCTCCACTCCtCTCCTC
162-M13rev_2
AATGTAGAGTGCAATTTCTCTGTTGATGACAAACTCGCCCACGTGGGCTTCGTCATGGGTTTTCAGAATGAAAACAAAGTGCATTCATGGGGGT
CACTCCAGGTCATAGGGGTCACTCCAGATTAAAGTTACCTAGATAGTCGTCTATACTGCTGAGATTTAGTTTCACATGTGTGAGACAAGAGAGGCGAGAATA
TTCCCATCCCCGTAGATCCAGCAGGATTGGCTTAAGATTACGGCTGCAGAGGGCTTTGGAAAGGCTAAGGGGCCACAGGCATCCAGCTGTGAGTTATAAAT
GCCTGGCACGTGTCCACAGCTGCCGCGTGCAGGTACTATTTTTGAGCAGCCTTCCAAGCACATTTCTAGAACCTTATTCCTGCATATTTGCCAGGGGAAAGT
TGATTTT
163-M13rev_3
TTCTTACCGTCTACCATGTAGGCCCCGGGGATCAAACTCAGTCCTTGAGGCTTGGCAGCAAACACATTTACCCACTGAGCTACCTTGCCAGCC
CCTCATTTGCTTTAATTCTTAGAGAAATTTGAAAATTCCCTAGACACAGGCCCAGAATGGTCAAGAAACTTGCCAGGAGTCGGTCAGAGAGGAGTCGTGGTG
CCCAGCTAGGAAGACCAACTCTAGGTTTTTCATTGACAACTCCAGATTCTCGGGGCCTGTGTCAGGAAGATAGAGATGGCTTCTCTGTCTGGCAATGTGCCT
TAGGTTCCCGGCAGCCACCCCAGGTCTGCAGAGACCTCTCTGCAGCCACAGCCCCCAGGCCCCTCCACCTCCTGCCTCCTGCTTCGCCTGTGGCTTATAA
CCCCATAAACCTTTTG
169-M13rev_2
CTTCCATGGTGTGGTCAGTAATAATTTGCCCATGGACCATTAAGTAATCACCAAGTCATGTAAATAAAATTAAGTACAATCAGTAAAGGGAAAAT
ATTTTTAAAAATCAAAATAGGATGGGACTTAATTGAAAAAGAAAAGTTTTTGTATGATGAAAGACAGTAATAGGGGTGGAAATGACCAATATTCATCATATACA
TGCATACACTTGCTAGAGAATAGAAGAAAATGAAACAATGCTTTTTATGTTTGTGGTTATCAAATGCCATTATATAAAATAAAATTTGAGCCAGAAACAGTCTTC
CAAGATGCCATAAAAAGTTTTACAT
175-M13rev_2
GAACTGAAGCAGAGGTTCTCTACCCCACTGATGTGAGTACCGACTGCTGATTCCTAAGGCTCAGGCTACTCGGTCTTCTCTGACTTCTCTGTTT
GGGGATTTCCTTTTGGCTGAAGGGCGTTGTAGTGGCTATTCCTGGTTGTCAACTNGACAATATTTGGAANGAACTACAATCCGGAATTGGAAGGCTCACCAG
TGATCCTTATCTGGAGGCTTGGAGATCCTTATCTGGATCTTGGTTTGAAGATCTTGAGCCATAGTGGCTATGGATTCCAGAAGATTGAATCTCCGAGTTTAAA
GAACACACCTTTAATCTGGGCTACGCCTTTCATCTGGGATTAAAGGTGTGGTGGAACACACCTTTAATCTGNGCTACACCTTCTGCTGGAGA
177-M13rev_3
ATTCAGGGAATCCCATATGTCTCTTTGCCGGTCTCGGNATTTCCCAAGTGTCTTAATCAAAATAACAAAGACAGCAGCGTGTGACAGTTCCCTT
TCTGTTCTGGNTGTCCTTGGGCCACAATGCAAATTGTAGTCATGTTTCCCCATCNTCTTCTATCTGAAGAGTTCTTGGTTTTTCCTTGTTCGCCCAGAGGAGG
AGGACAGCAGAATCCCNAAGAGTAGCTGGTCCTCAGTCTATGTGTTTTGTCACTATCCTCCTCTCACGCCATCAAGAATTCNCNGNGNNACCACTCAGGATG
CCGTCAGTCACGTGGGTCTCTTGGTTTTTC
17-M13revfor
CNNTGTGGCTTTGTTGCTGCCNNCCCTCCAACTTCCAGCATTTANGCAGAGGCAGCCACCTCTGTTATTCCTACATTTCCACCCNTTTGTTTTT
AGATTTTTACAGCCTAGTGGGAGCTGGGGTGCCATTGCCTCAGCATTGTTGATCCCAGCTCTGAGGATTCCCTGTATACTGGACTGTGCCTGTCTTATGTTG
TTCTATCTTCCATGATCTCACCCATCATTGACTACCAGCCCTTAAAGAGACATCACTCCAAAGGAATGTCATGCAAGGGGGAGATAAGGAACAACCTATATG
GTGGCCTCACAGATCTTGGCTATCCAGACATAGATCATTTCCTTGAGGTGGGGGCCACAATCAGGATTTTTAAGAGCCATCAGCACCTGGAGGGCTACCTTA
CTTATGGAGCACTGCCTTAAAATATGTGGTGAGAGATACTTTAATAAAATTATTATTAGTAGGACAATTCTTCCCATCATAACATGAATGGCTATCCATGAGGG
GTCAAATACCCTCTTAANATTTTCCCACACTTTGCTCAAGAATTCAGTATCTAGGAAAAGCACAGATACAATGGTTGCATTG
180-M13rev_4
CTCCAACCCCAAAGTCATTCTTGACCCTAAAAAAATTGAAGGCCAATCCGGCTCTCAGAGATCCCGTCTGAGCAAAAGGGAGGAGGCAGTAGG

TTTCTGTGCTCTGGGAAACCGTGGGACTCCACGAGGCCTAAGAACAGGAAGTTAGAGGTGAAGGGCAGAAAGCAGCCTTTGGCAGCTGACCTGCTCTGGG
GGCTAGGGTTTCCTCTGCACGTGTCCCCAGTATCTGTAACATTTGAGCCTCGCTCCTGCTCACCACCCTGCCTGCCTCCAGACAACACTCTCTCACGGCAT
CGTGTTCATTTTAACATCCGGCCCCCTCTCAGAGACGTCAGGCAGGAAACTTGGAACCAGGTGGATTACCAACAAAAAGATCTCCAGAGGGACAGGAGGGA
CAGGGAGA

186-M13rev_1

TCATGGTCATGATCATGGTCGTGACCGTCCTCGTCGTGGTCATGGCCGTGTCCGCCGCAATGCAGGTGCGTGCCGTGCGGGTGGCAGGATT
CGTATTCATGGCTATGGCCCGGCAGGCCATAGGTGCTGTTCTGCTGGGTGTAGGCGACGCCCAGATAGCCGCGCGACCCGATCCAGGATGCGCCTAGTG
TGGCGGTGCGGCTGTCGTTGGCGGATCCGTCCAGACGCTTCTCGCTCCAGTCAGGAACCTTGTAGTCGTCGGTCTTGCGCCGCGCCGCCTCGATGCGGA
CGGCGAAGCCGTTTCCGCCTGCGGTCAGGCCGGCCACCGCCGCCTCTTCATTGTCCGCCGAGCCGATGCGGACTTCGGCGGCGCCCTCTCT

188-M13rev_2

TNNCTTGCACGTGAAAAAGGAACTCTGAATTCCCCTAGGCGCCGGTTGCGCGGAANTAGGAGACCAGGCTCCCGCCCCTTCCACACTCCCGC
CGGGTTGCGGAGGCCATNCAGTGGCCCGCGCAGCCTCCAGGGGGCGAAGCACCCCAACCGACAGCAGTCTGTCGGTCCGAGCTCCATCATGTTTATTCC
CCAGCANAGGCTAACCGGTGGCAGAGGTCCTTGCTCTGGGTGTNANCTCTACCTTCCCCACCTGCACCCCAAGNTGACTTGGTCAGGTTCCAGTGCTCTG
CCGTCGAGTTGGCCATGTTTNNNTAGGCCCCACTATACTTTGTGTTGACACCGTGTCCCTCAACGACAGAGGTTTGAAT

189-M13rev_3

ATGATAGACTGAACCTCTGAACCTGTAAGCCAGCCCCAATTAAATGTTGGGTTTTTTTTTTTTTTAATAAGATTTGCTTTGGTCATGGTGTCTGTT
CACAGCAGTAACTAACTAAGCAGTACCTTAAGACACCATCCAACCACACACACACACATCAATCCACCCAACCAAATATCATCCATCCATTCACCCACCCTCC
TGTCCTTACATGAAGACAGGCTGATTTTTTAAAATTATGTATCTATATATCTATATGTATGAGCTATCTTCAGACACGCATCAGATCCCATTACAGATGGTTGT
GAACCACCATGTAGTTGCTAGAAATTGAACTCAGGACCTCTGGAAGAGCGGTCAGTGCTCTTAACTGCTGAGCCATCTCTCCAGCTCGAAGA

190-M13rev_4

CTAACGGCCNNCAGTGTGCTGGAATTNCGCTTGCGGNGACANNGGAGATCTGACATTCCTAGTCACCCACTCTTGTCTTGTTTCCTACAAAAT
GAAACCGCTACCACAGTGCCTGATTCCATCACTTGTCTTGAAAGTTAATTGTTCCTCACATAGTCCAGGCTTCAGCTCAATACACAAGTAGCAGATACTGCAG
AAGAATTGCATGTTCTATCCGTCTTTAAGTAATTCTAGTGCACACTGAGAGTAAACGAACAATGATACCTGTAATGAGGAAACAATTGTCTTATGTAAACAAAC
TTCCCCCAAAAACATAGAATGANACATGTTGGTATTCTTTCNTTCTTTGTGGGTGCANAAAGACACAATCTCAATTTTCTAATTTTTATTCCTCANCTTGAACTN
CTTCATGGGNATGATCANATCTGACAGTGTCGCTTTCTG

194-M13rev_3

AGGAAGTAGCTGAGGAAGTCTGGGAACTGCTGCTGACCCATTGTTCTTGCCTCAGGCCAGGTGGCAGGGCAGCTTCTGAGGCCTGGACTTGC
CTGGAATTGCTCAGTTCTTGGAAACAGAGATTTAGGCCTAGTCTCCATAACTGCCGATCTGAAGCTTGTCATAGAATCAGCCTAACCTTTTCATTCCCCCCTT
TCAGATGGGCCATCTTGAATCCTTGAGATGTTTCTGAGGCACCTACACTAGCCATATTGTGCTTCAAACCACTCTCGTTACCAAGCATGGGTCTTCTCCAGCT
GAGACACTCAAGGCCATAGGGCCATCATATTCAAAGAACCTGCTCTTCTGGATAGCTGGGCTGTTCAAAAGAAGAAAGCTGGATTCATTTATTTAA

195-M13rev_4

GTTATACTCTCAGTACCTTACAGAACCTCATTATCATGCCTGGGCCAGGCAGTGTTTTTGCAAAAAACTTTACTGCATATGTACACATTGGTTGT
TTGTCCAATCTTATGCATGGTGGCCAGCAGTAGTCAGTGCCACTCAGCAACGGCACATGTGGCTTCCTACAATTGATTGCTTGGAGACACTGTCTTAGTTAG
GGTTTTACTGCTGTGAACAGACACCATGACCAATGTTAAGTCTTATAAAAACAACATTTAATTGGGGCTGGCTTACAGGTTCAGAAGTTTAGTTCATTATCATC
AAGGTGGGAGCATGGCAGTATCCAGGCAGGCATGGTGCAGGAGGAGCTGAGAGTTCTATGTCTTCATCCAAAGGCTGCTAGTGGAAGACTGACTTCCAGG
C

19-M13for_2

GAACCTCTGAGAATACGGTAGGTGCTTCACCTGCTTCTGACCAGGTCCCTGTCACATGACACACTCCCTACATAGGAACGTGACCAGTGGTTA
GGTAGATCAGAACTTAGTTCTCCATGCTAATGAGTTATCCAGAGGGTTTAAACAATAAGCTTTCCTCCTCAGGACATTCCTTCCTGCAAAAGGTATTTAATCTC
TGGTCCATCCCGAGAAGGTGGTATGTACCCACTTTCCATGATGAACAATCAATAACCAGTTTGGAACCCAGGACTGCCTCTCTGAAGAACTACTGTGGGGAG
CGTGGAGGAAGCCTTTGTCTACAGAGCTTCTACAGAGCTACACCATCTTAAGTGTCCTGAAAAAGGCCCCTCCACAC

19-M13rev_2

CCTGCACCACCATTGAGTACTGGGCTTGAGTGAAAAAACCACTGGGCTGAGGCTAAAGCCGGGGGTTCCCAAGCTCACGGGCATCCAGGTGCC
ACTGGAAACCACATGGAGTCAGGAATGAAGGGGATCCATGGGCCTGGGGGGGGGGGCCAGGGCGGGGGGgGGTACCAAATTCCTGA

1-M13for_1

CCCTGCTGATGCGCAAGGCAATCGCCGCCTGCCCGGATCTCAgAATCCTTATCCTCGATCCCCACAGCGAGTTTGCTGCTTCCCTGCCCGAG
CACAGCGTACGCATCGACAGTTCGACGCTCGACCTGCCCTTCTGGATGTTTCATCTGGAAGAACTCGCTGAAGTGCTGTTTCGCGGGCGCGAAGCCATAGC
TGAAGAAGTTGACCTTTTGCGGGATCTCGTTCCCGTCGCGCGCCAGATGTATCGCAATGGCGGATCATCGGCCGTCACGCGCCGTGGTTCGGATACTCTG
ACCGCCGACACGCCCGTGCCCTATCGCATCGCAgATCTGCTGAAACTGATCGATGAGCGGATCGGTCTGCTGGATGGCAAGGCGGAAAGGCCG

205-M13rev_1

AACCCCGACTAAGACAATTACCATTTATAGCCTTTTGTATTTAATTGTCCCAGAGAGCAATGCTTATATGCACAGATTATGATATCATGTACCCC
CTTCTACTGATACTGAATTTCTTAGTGACTTTGTTCAGCTGAGTGCACTATATGATGCAGCATTGGAATGTGAGAATCAATATTTAATTGATATTGTTTATAAT
ATCTATGAGAAGGGTGGAGTCTAACAAGACACACACAGAGATCTCTTAGAATGCCTCTAACACGTTATTCCCTCCAATTTTGATTCCAGTTACCACTTAGGAC
TGAGAAAAATTAGACTCTCCTAAACATGCTGTTTCATTCAAAGTTAAAAAAATAACCTTCATAATTGATTTAAATAGTA

207-M13rev_3

TCCTCTGCAAAAGCAGCCAGCACTCTTACCTGATGAGCCATCTGTTTATTCCATATTTGTGTATGTATGACTGGACACATGTAGAANTCAGAAAA
CAACATGTGGGAGTCCGTCTCTGTCCACCCTGAAGGTNTCAGGGATCGAACTCAGCTCCTCAGGATTGNCAGCATGCNCCCTTACTGTCTGAGCCATCTTT
NAGGNC

208-M13rev_4

CAGGGTTCGACCTTCGAAGACCNGCGTCgCCAGTCGTCGGAAaAGCTTCAGGANATCGNCTTCGGCGGCTACGCCATCGGCGGCCTGGCCG
TGGGCGAAGGCCATCAGGCCANGTGCGAGGTGCTGGATTACGCGCCCGAGTtCCTGCCCAAGGATCGCCCCCGCTATCTGATGGGCGTGGGCAAGCCGA
TCGATCTGGTCGAGGCGGTGTATCGCGGCGTGGACATGTTCGACTGCGTCCTGCCGACGCGCGCCGGGCGCCaCGGTCAGGCCTGGACCTGGGACGGC
CCGCNGAACCTGAAGAACGCCCGCTTCGCCGAGGATCAGGACCCGTCTGGACCCGACCATCGACGGCCCGTCGTCgCAATATTCAAAAGCCTATCTGCAC
CACCTGATCCGCGCCG

209-M13rev_1

ACAAGGCTAGTTGTTCCCCTTTTCTATATGTTTGGCTCTGCCTTGTGATGGATTATTAATGATGTGCTAAATCCACATTCTCTTCTTGTGCTATTT
GCTCTAACACGTGCAGTTTTTTCTAGACTAGAGAGATGCTGAATAGTCTTGTGGCCCGCAGATATGATTCACTGTTCTTTCATTGTAGATGGAGACAGAGGCA
GACAGACAGTAGAAATCAGAAGTGGAAGTAATAGCCATCTCAGCAGCTGCTTCATGGTTCTCTGGGAAATGGAGCAGCGAGTGGGGCTGAAGACAATCCAC
CTGACAGGCTTAGTCAGTACAAGATTTCAGTCTTACACTGTCAGGGCATATTAGTCACAGAGTGTTTCCCAGGTATGAGTTAT

20-M13rev_3

CACCCCTGAGTGCCACAACCAGTCAGGTACCCTCAGTGAAAGGACAGGTGTCCTGCCAGTAGATGGCCTATGTATGTGGTCCCATAAAGAGA
AAGGACATTGAGGTGTAAAAGTCCTTTTGGGATTCCTGTGATTCTATATCTACTTGTTCTGCCTCCTAGAGTGAAGAGGCCTCTCCTGTGGGGTCATGTGAG
GGTGCCAAAGCCTCTGtACGACATGTGGATTTGGGGCTAAATAGATAGCCCTCACCCCTTGGGCCTCTTGGGGACCACTAGGATTGTGGGAGACCATCCCC
CCTGGCAGTCAATGATTAGCTTTGGAGACACCAACTTGGGCTATCCCTGCAGGTCCTCAGTCTCCaAGAGCCATGAGTATCCA

219-M13rev_1

CTGAAGAATGACAGCAAGAGGGACTTGGCAAAgATGCTGTTAGATTGCTACCAGGTGGGGTCAACTGCAAGCCATGCACAACCAGACGAGAG
GCAGGTGANAGGGTCTGAGATGCTGTGTGATCCAGGTCATGCCCAGCAATACACAGGCTGTACAATAGTGAGTGCCTGCTCTTGCCCCACTGTAGACNgGC
ATTGTGACCCAGTCCTGACTGGCATAAAAAGCTTCCTTTGTTTAAGAGTGACTTGCCAGGTGTGTGGTAACCATGTGACTGGCTTACTTGATACGTATTCCCAA
TATCTAATACTTCCTGGTTATTAGCATCCTAGACTCTTGGCTAAACAAAGGCCAATTTTTCAATCTGTTCAGTCTCACCCCAACT

233-M13rev_4

GGCCTTTCTGTTGAGTGGAAATGCACCTTAACTTGGTGCTGCCAGCAGCCTGCAGGCCCCAGCCCTGTTGAGATGAACACAGTTGACAATAAA
GGAATGAGTATCATTGTGGAGTCGCTTGCCTGCTTCTCTCCCTTTGCCTCTATCCTGCATTCCCCTTTTCAAGCTCGCCTTTCTCTCCAGCTTTTTCAAAAAGT
GTTTGGAGGCTGCCCACAGCGATCAGAACATCGGAGGAGAAATAACTGCTCAGTGAAAGGATGTCTGTAGAGCTGCTGGATTCGCTTGGCTGTTAGGGGG
CTTGTGTAGCACGTTCAAGACACTGCTCCATTCCTGGCACGGGAAGAAAGAAAAGGATAATAGGCTAGGGCTGGTAACAGAGCACAGTGGGCAAAGTGCCT
GCTGCCAAGGCTGGCAGCCTGATTGGATCC

234-M13rev_1

TACAGGTGGTGTTTGAGCAAGCTCAGCAGCTGTCTCACCGAGTTTGTGCGAGAGGCAAGTATTTCAGAGCATGAACATGGAATGTAGCCAGAT
CCCAACGTCGAGACAGGAAAGATGGGCTCACCTTGGTTTTAATGAATCAGCGAATGAGGCCAAGAATTTCAGAAGAGTCAGAAAACATCCAGTGGTGGAGA
AAGCATTACAGAGGAGTTTCTCAACCTGTGGGTCGCAACCCCTGTCTCAAACCTCTCTCTCCCCAAAAATATTCACATTATGATTCATAGCAGCAGCAAAATT
GCAGTTCTAAAGTAGCAATGGAAATACTTTTAGGGTTGTGAGTCACCACAACATGAGGGACTGTATTGAGGGGTCTCAGCTTTGGGGAGATTGAGAACTGCT
GCATTAGAGGGTCTACTTTTGC

236-M13rev_2

GAAAAATTGCCATCTATTTGAGCTTTTTTTTTTCCTCTATAGGAGTAGTTCATGCCTCTAAAATAAGTAGAAAAAGAAAAATGACAGGAAATTTGA
AATTCAAACTTCCAGGCTGCTACCAATACAGTTTCTAATAGGCTTTCAGAGTTGCTTCTAAATTAAATCAAAACAAAACAAAAACCCATTGAGTCCAGAGTTTC
TCATATTTGGTAAACTATTTGGGAAATTTCACAAACTCTGCACGCCAAGGGGCTTTGTTTATTCTCAGACTTCAACCTTGCCATTTTGTTTAGCTTCAGCAATT
GTTATAAAACTAAGGAGACAGCTGGTATAAACCTAGGAGCTCTCTTTTCCAAGAATAGACTACACATACATCCCAGGCA

237-M13rev_3

TACATGTCAAAAGCCAAAACCAACAAAAATTTTATTCAGAAATTAAAAAATAAAAGAGGTCAAGGTGGTGGTGGCACATGCCTTTAATCCCAGCA
CTAGGGAGAAAAGTGAGTTCTAGTACGTGCTGGGCTACACAGAGAGAAATTCTATCTTGGAAAACAACCAAACCAACCAAACAAACAATAAAATCAAGAAGA
GGTTCTGGAGAGATGGCTCAGAGGCCAAGAGCACTGGCTGCTCTTTCAGAGAACCAGGGTTCAATTTCCAGCACATGGCAGCTCACTACTTCCTGTAATTC
CAGCTCCAGGAGTTCCAAGAGCTTCACACGAAATACACATGTGGGGCAAAACACCAATGCACATAAGATAAAAATAATTAACAAAACAAAACAAAAAGAAGAA
AGGTATTCCAGAGTAGGTTGTTTCCTGCACA

240-M13rev_4

ACATGACAAAGGCTCATTGCTCAGCCATGGGAAAGGCCTTGTAAATTTCCAGAGCAGACTTGGCAGGGGCATGTACAAGCAGGCCTGTCTGA

GAGATAGTTCAGGTTTGATTTTCTGAAACTGTTTCTTCTATGGAAGCATTGTTCCATAGAGAGAGGAACCAAGAGCAGGGTATCTATACCAGAACACCATGGT
AGACAGCAGGGATTTACCAAACCAAAGGGTTCAATCCCCAATCTGCAgGCCTGG

241-M13rev_1

CATGGCTGCTGCTCCTAACCATCTGGTTTAGCGTGAGTGGCTTGAATGACAGGAGCATCATAAAAAAGAAAGTACTCAGATGGTAAAGTGTTTG
CTGTCCAGCATGAGAACATGCATTCAATCCTTGGCACCCATGCACATAAAATGCTAGGCACAGTGGCATCTGCTTGCAATCCCAGCACTGGGGAGATAGAG
AAAAGAAATTCTAGGGTTCACTGGCTGGCCAGCCTATCTGAGTTAGTGAGCCCCAAGTGCCTAGCTAGTAACAATAGCTAGTAGCAACAGCAAGAGGCAGC
CCTGTCAGGGCCAGTAAGGCTTCAGAGAGCAAAGCATCATTTTGAGACCTTATCTCAATAAATAAGGTGAAGCAGGCTGCAGATACCTTAGGAAGAAAACAT
GCTTGCTATGTAAGTACA

242-M13rev_2

AGNTCATCANCTAGCTTCATGAATTTGATCAGTAATTTGGATGGCCTAACATTATTTCTTCTTTCCATACCCAGAGGATAAAAATATCTTTTAGGA
ATTTTTCCCCTGGGCCCATGTCTTCTCGAGACTTTTCCCCACTTTCTCCTCTATAAATTTCAGTGTCTCTGGTTTTATGTGGAGATCTTTGATCCACTTAGACT
TGAGCTTTGTACAAGGAGATAACAATGGATCAATTCGCATTCTTCTATATGATAACCACCAGTTGTGCCATCACCATTTGTTGAAAGTGCTGTCATTTTCCACT
TGTTGGTTTTAGCTCCCTTGTCAAAGTTCAAGTGACCATAGG

243-M13rev_3

AGGAAGTCCCTCAGGTTTGGGAAGGCGTCCAGGCACTTGGGCTCAAACATACGGTACTGGTCAAGAATGTNaTAAGCAAGGAAATCCACATAG
GTGACCTGCAGGAAGCCAAGGGTGAGCGCTCTGGAATCTGACACCAGGGAAGGATGGCAACCCCTCCTCCCCACCCCTCCCTTTACCTTGTCCCCTGCAA
ACCATGGCCTCTTGCCCAGGAACTCAGAGTAGAGCTTCATTTTCTCAGGGATGGTCTTCAAGAACTCTGGCTTCTGCTTCTCCTGAGGCAGAGAACAGCTG
GCTGTCACCACCTTTAGACTCAGGCTCCCCAGCAAAGGGTTTGGCAATGGAACAGCCACCTTTCCAAGAAAGGGCCA

25-M13for_4

GGCTCACAAGCATGGACCGCACACACATGCAGGAACTCTCACTTGACAGCATCGCGGTCAGAAAGACAAAGGTTTGAAAAGACAAGGTCAGA
AAGACAAGAGGAAAGGTTTGGCAGAGTGGGGATTGATGAGTGTTTTCCAAGACCAGGGNnNACTTTTAGAAAGAGGAAGGTGAAAGATTCCACCCTGCAGG
GACGGGCCTTAAGTTCTACAAGTAAACAACTCAATGTCTTCAGGGAATCCTTGAAACTAACCAGATTCACCCAGGCTTGTCCCCTGAGCATATGTCAGAAGT
AAGGACCACTGAGAGATACTTTCAGACAAGCTAACCTGCCTTGAAGAGGCTCAGGCCAGCTGAGTGATCTACAGGTGCCTGAGTGCCCACC

260-M13rev_1

GGCCCCATTGAAATTCCCCTTTGGAGCCCTGAGTTCTGAAAACTCCCACGGAGAATGAACCAGCTTGGAGTCCTGCAGGCCTTGGGCCACCC
TGACACATTCCTGAAGCTTCATTGCTGTGTGGTCCCTTTCAAAACCTTTGGAAAGTTCCAGGAGATCCATTTCTTGATCAATGTTTTCATCTTTGCCTGGAGTG
GAGCCAGAAGAGAGCAGAGCAGATGTCAGATCCAAAGCCTTTCTGACTCCACCACTTGACCGCTTGGTGGTTAATTTGTGCGGATGTAAATGAAATGCCTG
GGAAATGTTTCTCATTCAAGGGAAGACATGCTGTGATAAAACTCATCACTTATGTGCTGACTTTAAAAAAAAAGAGAGAGAGAGGCAAGCAAGCTCCTGTAG
AAGCTGG

261-M13rev_2

CAATACTGGCCCAGAACCAGGAGTTAGTACATGTCCCTGGGTTTGTTGAATGAATGACTGAATGAATTGCTGATACTGTTAGTGTTTGGGGGG
GGGGTCTTGATCCTTTCTCTGACATTAAAAGAATTAAAAGGCAGGGCAAATCTGCCAACTGA

263-M13rev_3

TCTATAGTCaACACTTTAgGTTGTTNNATTTTTCAAGAATTGTAGCTTAAACGTTTTGTTTATTTGTCATGTGTTTATACATATATGCTTGTGTCATG
CTATATGTGTGGCACTCAGAAAGTANCTTGCNNTCTGAAAAACTGGGAAGTATTTATGGGGAAGGAACAATTGGATAACTCCTACATTTACCATTTGTATTGC
AGAGAATGAACTTATGCTATCAACCTTGGCAGCAGGTGTCCTCATAGGCTGAGTCATATTACTGCCCTAGGTTGtTTCTTAGtATGACCCTAATATCAAGAAAT
GTAATATGTAGTAG

264-M13rev_4

AGATCACCACAGGCTTTTGTACTAAGTGTCACCTGCTTATTCTTTAGTTGAGTTCNNAACCAANAAGCATGTAAgATGCCCCCCAACTCTGGGA
GACCCAAATCACA

265-M13rev_1

NCCCCTAAACCTGAAGTGGTGTTTGTTGCTCTTGTTGAATNCCAAACCGGCTCCCGACATGGGACTAGCGAATGAACGAATGAtGCTCACTGG
ATCCGGAGACCCCAAAGAGGAAGAGGAGGAGGAAGAATTAGTGGACCCCCTAACAACAGNGAGAGAGCACTGTGAACAGCTGGAGAAGTGTGTAAAGNCC
CGGGAACGACTAGAGTTNTGTGATAAGCGCGTGTCTTCCCGATCACAGTCGGAAGAGGGCTGTACAGAGGAGCGCTTTNACTTCCTGCANGcacGCGGACC
ACTGTGTGGCCCcACAAGCTCTTTAAAGNCTTGAGTAAATATGCAGATTCGCCTGCCTTG

267-M13rev_3

AAGAAcTTATAAATGGAGCTGGAGCCTCACCTAAGAGTCTGAGGAAGACTAATTAGAGATGCACAAAACCAGAANNCAGAGACCAGTGCCNTA
GGGATTTTCACCAAATTTATGACNnCAGGATCTGAGAGACATGGAGTATCATTGTAATGAAAGTTTTCTTCATTAGTGCTAATAACAACTACAGGGGATAAGG
GAAGTGAAGCTGACTGAAAGAGGTCATTAAGGTTTGGTATGGGGATGGAAAAAAATTCATAAACAGAAAAAAAAATATTCCCTCACATTGAGCCCATGAGAA
GCTAGATAGTACAGTTTTTTAAAAAAGGAGTTCTTAATCAAGTACACTAGCATTATGTTGCATGATTATTGCAGTTTTTAATAACTCAATTAAGCAATTTCCATTT
TAAAATGTAAGCT

268-M13rev_4

CAAAGACGGGACTCAGAGCAGCCTGGCCTTCAACCCCTGNNCTCAGCCTACACCACCTATTGGAGGTGGGCCCTGGCTGTGTCTGGGACAG
AACTACCCTTAACCTGCTTCTGAAATGCCTTTGCTTGTTCCTGGCTACCCAAACAGTGACCCTGACACAAACCTAGTGTTGGCTATAGAGGCTATATGATGAC
AGCCCATTCCTCCTTCATGACATTACTGTCCACAACAATGAGTATCATGTCCTTATCAGTATCAGCAGCCAGNTGATCGAATTNnAGCGAGAAATGCCCTTTAGA
ACTTGATCTGCCCCACTTTGTCCTTCCATTGTGACCTTAGGCAGACACCCAGAGAGTTAGTGCAGGGGGCTACAAGAGGATGAGAGTACTCAGTACAaAGC
AgGGCAGTCACTGTGACTTC

270-M13rev_1

AGTAGAAGGGAGGGGCATTCCATTTGAAGGTTTTTAGAGATATGAGAGAGGGGATTTGTTCTGGATGTCAGCTGAAGAAGAAAGTCAGCTGGG
TGCTTTCTCTGCCTTTCTGAGCTAGCAGGCTTTCACCCTAGCAGTTTGGCTCCTGAGTCTTCACTGGTAAAATCAATCGATTGAGATTTTGTTCAAAACAGCT
ATCTGCCAAAACTAAAACAAAGGCCTAACCCCATCAAAGTCCATAAAGTTCTGGGAAAGTCTGAAACTGTACTAACCTTGCTTTTTAGCTCCTATAGTTCTATC
TGGCTAACTGCTCTTCTTAACTGAAGAATAACAACCCAGAACATGGTTTTTGTACTTAAAAGTTGACCCTGAGAAAGGCTCAGGATCCCAAGCACCCAGTGTA
GTTGCCA

271-M13rev_2

AAGTTTTCTTCTATAATTTTATTGAAGATATTCCCTGGCATCTTAAGTTGGGAGTCTTCACTGTTGTCTATACCTATTATCTTTAGGTTTGGTCTTC
TCAAGGTGTCCTGGATTTTCTGGATTTTAGGTTAAGATCTTTTTGCATTTTGCATTTTCTTTGACTGTTGTGTCAATGTTTTCTATGGTATCTTCTGCCCCTAAG
ATTCTCTCTTTTATCTCTTGTATTCTGTTGGTGATGCTTGCATCCATGTCTTATGATCTCTTTCCTAGGTTTTCTCTCTCCAGAGTTGTCTCCCTTTGTGATTTC
TTTATTGTTTCTATTTTCATTTTTATATCANnGATGGTTTTGTTCTACACCTTCACCTTTTTA

272-M13rev_3

AGCTTTATaATGCTGTCTTCTCTTCAGCTTGAACTTACTAAATAACATCAGTTTGAGATTCCTCATTTAACATTTAAATTTAACATTAATTGAAATTC
CTTTCTATAGGAGTTGATGTAGTTGAGGTAGAGAACAAGGTGGGTTTTAGTTGCTGTTGGTATCTACTTTCTAATTTGTGTATGTTGAATGGGCCCGTAATAG
CATGTGTGAACATCAGAGCACAGCTTGCAGTAGTGGGTCCCCCTTTATCAAGTAGGTCGCAGGCATGGAATTCAAGTTGTCAGNTGTTGGCTGGCACTTTTC
CTACCAGGCCATCTCATTGGCTCACGGTGGAATAAGTGTTAATAAGCAGCAGTAGTTGGTGCTGTGAGCAGAGCTTTCACTGCTGCCCAGCCTTGTTGCCC
ACTGCAGTGTG

273-M13rev_4

CCTTCTAGCCGAACAGTTTCTTTCATAGCAAAGTAtTAATGTTAGAGAGtGGGAGGAGGAAGTCCTGGTCCCACCTCACCTGCTCTGGATCTTG
GCAGGCGGTGAGACAGCCCTGAGAACCTGGACGCGCTACTTTTGTTGCTGGAGTCTTCCTGTGTACATTGTTGAGTCAGGAGCTAGGTTAGATCAGCATTT
CCNNAACATTGTTGTAAGAAATTCCAGTACTCTTTGAGGAGTTTTGAAGTATCCATANGTAAAAAGTTTGCTTGATCGAATTNnAGCGAGAAATGCCCTTTAGA
GACTCCCAGTGCGCCTTGGCATATATAGGATATATGATAGTCCNAAATAAAGTTGTCTTTTCTACAGATTTTCGCCCAgTGAGTCTCAAATGCATT

27-M13for_1

AgGAGGTGGGGGAATCATTACACTATCAAATCTTTTTAAGGGAGAGGGTAAGACATCTACGAACTACAGCAGGGTGGTTTCCTTTGGGAGATAT
TTTGTAAGACCTAGATCTGTATATAGTAAAATAACATAACCTGACCAGTTGACATTGGTTGTCAGGAGGAGCAGAGGCAGATACTTTATTTGGTTCCTTTTTAT
CGGGCAGCCCCTTTTATTGAGTACGTCAATAGGAAAGCAAAGTGGTATGGATAGGGGGCCCAGTAAAGACAACACAGGTCTTACCAGAATCTTGAGTGGAG
GAAGCACCTGAGGGTCTTGTTCAGTACTGAGATGGGCAGTGAAGCTCTGCCGGGTACAAAGATCTCTCATTGTACTGAAAATATACTCACTTCTG

27-M13rev_2

CAAACTCACTTCATCACACTTCTTATTCTCCAGTTTCTTATTTTAAAATGAATATATTCCTGTCAAACTCTAACTCATTGTAAGTACAAATGACTGG
GAACAGGTTCCTCTTTAAAGGGCAcCACTATATTTCTTGTaATTCCAGCTTTGGACCACTCTAAAACACAGCCATATGTTTCCATCCCTGGnNCACATCCATGAA
GTCCTCCTGTCCTGAAATTCCAATGAGCCAGTCAAAGTTACTACCAGCGATTAAGCTAGAAGAGGAGATCCCTAAAATGACAGGCCACCAACAAGACAAGTC
ACCAACAGAGACAGAGCTGAAAGTCAGGGAGATCCTGGAGAAGCTGGATCGGCAGATACCTCCCAGACCCTTCACCCACCTCAACTACACCACAAGCGCC
ACACACAGCACGGCCACCATCCTCAACCCTCAAGATACATACTGTGTGGGGGACCTGCTGGACATCCTGGTAGAGGCTAGAGACCACCTAGGAAACAGGA
AGGGGTATGGTGGGGACTTCATGAGGGCCAGGATGTTCTCCCCAGCCCTGAAGGCAGGCGCTTCTGGAAAGGTGACAGACCTCAACAATGGCACCTACCT
CGTCAGCTTCATTCTGTTCTGGGAGGGCCCGGGTCTCGCTGTCTATTCTGCTCGTGCACCCCAGTGAAGGGGTGTCAGCTCTCTGGAGAGCAAGGAACGA
AGGCT

28-M13for_2

CAGCTAAAAGCTAAAAGAGTGTGTCTAGTTATCATGTTAAGTAAAAACAATATGAAGGATTCTATATAAGTATGGTACCATATGCCTTTCTTTATG
TTAGTGTCACAGAATACTAAAGCTAGCAGAACATAAAGACTACATAATCGAAGCCACCTTTTAAAAAAAACAACTGAAAATAAGGAACCAATGGGATTTGTTG
AACATGGATAGTGTTAAAATAGATAACTGAGGCCTTAATTCCAAACCCAACTAGTAAAATGTCTAAGGGAAACACAAAGAACAGTCTTCAAGCTATATGCACA
AGCAGGAAGTTCCCAAATGGGGCTCTGGTCACACAGGAACTAGAAATGACAAATGGACAAATGGTAGTGCATAAAA

28-M13rev_2

GAGACTACTCTAAAGTGATTTTCACAATGGCTACACTAGTTTGCAATCACACCAGCAGTGAANAAGGGTCTCCTTTTCCCCTCAGCCCTCTGGC
ATTTCTTATCTGAGTTCTTGATGACTGCTATTCTGAACAGGCTAAGGTGGTGTCTCAGGGAAGTTACAATTTCCATTTTGCCAGTGGCTGTAGATGCTGAATA
CTTTAAAAATGCTCATTGGTCATTTGTTTNnCTTAAGAAGTGTCTATTTATACATCAACTCACTTTTGAGCGAGAGCAACAGCGAGAGGTGCGGTGGTGAAGA
GAGAAAAGAGAGAAGAGAGAGGAGGGGAGAGGGGAGAGGAGAGAGAATGTATGCGTGTATGCATGTAATATTTTAGTTCTTTACAGTTCTGGAGAGTAGGC
CCCTCTGAAGTATAG

2-M13for_2

CCCCCTCACCCCCCCATCCCTGCCTACTTCCAAGGCCAGGGTAGCAATGTCTGGGCCAATCACAGATCAgATCACCGACACAAACTTGTGCCC
CTCCCCTGCAACTAACATTCACATCAACTCTGTTGATGTGATGGCAGGGACCACCTCAATACATAATGGGAAAGTTAAGGTATGGAGAGTGAAGTCACATCC
TGCCTTGTTCAGGTTCGGGCCACCAGCAAAGAAATGTTACAGTTCTGATTTTAGGTGTCTCTCNnTCAATCCCCTGACCAAACATTGGTCATTGACCAGATGG
CCCCTCTTGGATGAACCCTGCTCTCTGGCTCCATGGTTGCATTGTTTCTGCCTGAGAGTAAGTATAGTACCCTGCCTCTGGAAACCTGGGCTTGAAA

303-M13rev_1

ACACTTTATACTTATCCATAAGTACAAGGATAAACATTGAGAGTGTAGTTATGGATTATGCTGGTTTAATAAAGTGGTANTTTTAGATTCATCTAC
TCCATGATTAGGTTTTCAGTACCAGGAATGTTTTCTTTCATGTTGAGAGGGTCTTATGTCaAATTAGAGAGCTGTTGGTTACTGCTGATGTGTATACGCCACTA
CTGCACCCTTAGGGATATTGCCATGCTGGTCATTGTGATTTATTGGCCTAATAGCTGGGTAGGGCTGCTGGTTGCTTCCCTCCTTTGGAAGCTTGCATGGCA
ATTTCTGGTCCTTGAGGGAGGAAGCTTTCGGATTAGATTCAGCTTGGGTCCTCTGGGCTCNGAGTCTGAAGTGCCTGATGTCCTCATTGTTTCTTCCATTCTT
TAGCC

306-M13rev_3

GCCCCAAATCCAGCATGGAACTGGATCCACACAGTGGCCTCTAGCACAGCTGTGAAGAGGCTTTACCTCTTTAGATGATACACTAGGAGCATG
AGGACTGGAATTCCTCTCTCTAGAACCTACATAGAAGGCTGGGCTTAAGGGAGTACATCAGTAATCCTAGTTTTGGGAACTAGAGACAGGTGAATCCCCAGG
GTTTGCTGGCAAACCAGCTGGTGAGCTCCAAGTCCAGCTAGAGACCCTATATCAACCAGTAATGGGGAAAAGCAATTATGGAAGAAAACTAATGTCAATCTT
TGGCCTACTTGTCAATGGACATACATATGTATATGCACATACAGTGGACCCATACAAGATAATCACACACACACACACACACACACACACACACACACACA
CACACACACTGGATTTTG

307-M13rev_4

GCCCAGTCCCCATGATGCCTTGTATTTCTCCAAGGCTTTTACTTCCTGGGGAGTTNNTCAACACTAACAGACACAGGCCTCAGGTATTTTGGGA
TGGGATCACACCTCATCAGTTTAAGGCTGCACACAGGGACCCAGGTTCCATGAGCAAATGAAAGAAGGGAGGGTCCCACTGTGACACTCTAGGTGACTTTT
CTATTTCAAAATTGAATCCTGCAAACCAGAGAAACACAGTGGGCTCCTAGGCCTGGAgCTGGCACA

310-M13rev_1

GGGTGACACCCAAGGGGGGCTCACCTTCTCCTCAGAGAAGAGGAAGGGTAATGGGGGGGAGGGATTGGTAAGGGTGGAACTGGGAAGCGA
GGAGGGAGAGGGGCTGTGATGGGGATGTAAAGTGAATAAAAACAAAAACTAACAAACAAAAACCCCTATGTTTTCTTCAACTGAGTAAACTCCTTAGAAAAC
CACACCAACATTTTCTAATTTTATAGCTGTTTTATAACTATGTTAGAGAATGTCCTAATTTTTAGGAAACACATGATGAAGTACTATGAAGAGACATAATATCTG
GATTCTC

312-M13rev_2

TATCCGGCCTTCGAGATTCCCGATGCCGTGCTGGCCGAATGGCGCGAGCTCGGCGGCCGNGGCCGCNATGCGCGGCGCGCCNGGATCGAG
CGGTCGCGaGCCGCCTGCAAA

313-M13rev_3

ATGGCAAGAAGGAGAATCTATCTTAAAAGGGTTTGAAGAACCATTGGCTCAAGGTTGTAGCTAAATGGCAAAACAAAGTTGCATCCCTGTCTAC
AGGAAAGCACAGATAAGACCTGTTGTTTATAACACGCCCACCCATATTTCTGAAGTGTTAATTCATTCCAGAAATCTGGAACAAGCTGCTGCAAGCTGGATTT
GTCTAACATGGCAGTACTGACAAGGTATTCAACCATGCTGCAATTTACATACCGAATGCCTGGATGCACCCTTCTAACAAGTTAACTAATTATCTTTCCAAATC
AAGCCTGCAGACAGTATCCACGTGCTTCTGGGCTCATGGTAGCTCCCATGACAGTCTAGGTTTCCCTCCTCAATAGGCCACACACACTGCACACAAGAATG
GTCCTTCACCCTTCTAGGG

315-M13rev_4

AGCTGAACTCTAAATTCTGAATAACTATTCTCCAAATGCAAAGACATACACATTCATCAAGGAAACTTTTGTAAAGCTCAAAGCACACATTGCAC
CTCACACAATAATAGTGGGAGACTTCAACCNCACTCTCATAAATGGACAGATTCTAGAAACAGAAACTAAACAGACACAGGGTGAAACTAACAGAAGTTATGA
AAAAATGGATTTAACGGATATCTACAGAACATTNTATCCTAAAACAAAAGGATATACCTTCTTCTCAGAACCTCATGGCACCTTCTCCGAAATCAACCATATTA
TTGGTCACAAAACAGGTTTAAATAGGTATAAAAAAATTGAAATTGTCTCATGAATCCTATCAGATCACCACAGACTAAGGNCATCATGACCTTCAATAACAACA
TAAATAATAGAAAGCCACATT

318-M13rev_3

CCTGTTTCTTACATTTAACTCAGTCAGGACAACTGTATGGTGCCAATTTAAATAGTTTTATTATTCCATGTTACTTGTTTTGGGTCCCTGTGGAAC
AAAGCTACACCTTCTTCTTGGACTGACTTAATCCACCTCTTCAATGGTGGGGCCTGAAGAAGCACCACCAGATGGGGGAGCTCCTCCACCTGGGAAGCCAC
CAGGCATTCCCCCAGGCATGCCACCTGCACTCTGGTACAGCTTGGTAATGATAGGGTTGCAGACTTTCTCCAGTTCTTTCTGCTGATGCTCAAATTCTTCCTT
CTCTGCAGTCTGAGAAAAAAAGAAATCAAATTTATTAAAAGATCTCCAATTACTTTGGGGGGGGGGAGACTGACAACTAAGCCTAAAGAACCCTTAGCTATTG
AGTCACTCAGACATCCAAGGA

320-M13rev_1

CTGTTTGAAACAAGGAACAGACACTTAGTCAGGTTCTTAAGACACNNTTTTTGTTATAGTTATCCCAAACTATAGTGTTGGGAGCTATTAAGACA
ACTCTTGATCTCTGAATTGGGCCTCTCCCCCTGAGAAGAAAAGGGGGTCAAAAGCGGGCCACCGACACATGGATTCCGAGAACCACGGGAGGTTCCAGCC
CTAGGTCATCACCGATGTCCTGACCACACCCTGATACCACCAAGTTCCTGCTTCCCGGTGTAGCTGCCAAAAGAAAGAATTTCTATTGCCCCCCCTCCCATA
AATACTTCTCCTTTTGCTTGTATTGCCCCATCCCTCCCACTGATAAGTATCTGTACTTCCCCGtTTTCCTTGTGCAAnTTCTACTGCCCCATCCTTCCTGCTGAA
AGGCACTTCCCCCTTTTT

321-M13rev_2

AAAATGCCTCTTAAGATCTGGCCATAGGCAGGCATGGTAGCACATGCCTTNAATCTCAGCACTCANGAGGCATAGGCAGGNAGATTTCTGTGA
GTTTGAGGACAG

322-M13rev_3

AAAACTTCTTCAAGTTACTCTAGAAATTGCCAAGAATATCCCCGTGGTTCGTAGTTTGAGAAATACAAACAAGCTGCAAAAGAGAGGTGTGAAC
CTGGGGAGGCTTCCTCATGAATGTCATTCTCTTCATTCTTTTACAGCACTACAGTGTAAGAGAACAATGAATAAACCCTGAGTGAAGTAAGATGTAGGGATAA
GGCAGAGAGCTTGCCAAGCATACACAAAGCATAAATGGAGAGTGCTGGTGCACATCTAGAACTCCAGCACTCAGGAGGTGGAGGCACAAAGATCACAAGTT
CAAGGCCATCCTTGGCTACACAGTGAGTTCATGGCGAACCTGAGATTTAAAAATAAAAGTTAAGTGAAACACGTGCCACTGCTTACAGAGAATTC

323-M13rev_4

TGATTTAGTAAAGATGTGGAGATCAATATAACACGTTTGCCAACAGATGTTCAATAACATAACTGGGGTGTCATGAGCAGAATTCAACTTGTATA
AgCTGCACAGGAAAAAATGG

324-M13rev_1

TATTTCCTGCCGTCTACTCCTCTTGGGTGAATTTGCTTCCTTTTTTTCTAGAGCTTTTAGATGTGTTGTCAAGCTGCTAGTATGTGCTCTCTCCC
GTTTCTTCTTGGAGGCACTCAGAGCTATGAGTTTCCCTCTTAGAAATGCTTTCATTGTGTCCCATAGGTTTGGGTACGTTGTGGCTTCATTTTCATTAAACTCT
AAAAAGTCTTTAATTTCTTTCTTTATTCCTTCCTTGACCAAGGTATCATTGAGAAGAGTGTTGTTCAGTTTCCACGTGAATGTTGGCTTTCCGTTATTTATGTTG
TTATTGAAGATCAGTCTTAGGCCATGGTGGTCTGATAGGATACATGGGACAATTTCAATATTTTTGTATCTGTTGAGGCCTGTTTTGTGACCAATTATATGGTC
AATTTTGGAGAAGGTCCCGTGA

32-M13for_4

GGGCTAATGACTGGAGGGGGGCATGGCCAGATCTGGCTTTAGATTAGACGTAGTATAGACAGGAAGAGCCAAGTCCTAGCAGATAAGACAGA
GACTACAGAAAAGAAGGAAGAAATCATAGTGATTTGATAATGGCCTTGGGATTAACAAAAGTTTAAGGAGGTAGCCTCAGCAGGAGTTGGCTCTTAGTCGAT
AGAGCCATTATCTAAGACTGTGAAAATAAAACACTTGAGTGGGGAGAGAATTAATGAGAGTCCTCCATACGTTAAATCCGAGTTATCTTGAGATATCATGATG
GACACATCTTGTAGTTGAATAGGAAGGCTCTAGATCTAGGCTGCACTGCATATCTGGATCCAGGACTAGAGAGAGTATTCCGGACAGG

32-M13rev_4

CATGTCTTGAGTTTCTTAAGAAGTTTTTCCTGATACCCAGGAAGTTGAGACAGGAACCTCCGGAGTTCAAGGGCCTCCTAGGCTATGCAATGA
GACTCTGTCTCTAACAACAACAAAAGGATACAGGCATTTGGcACTTTATATCAGGCCTGCCCAATACAaCAGCTCCATCTCTTTATACGTGTCCATGGCTGCT
GGTTTGTCCCTGTCCGGAATACTCTCTCTNGTCCTGGATCCAaATA

349-M13rev_2

GTCAAATCTATTATCCTGTTCTGAACTACTCAAGCCATATGAATNNGAAAGAAGAGAAATGATTCCACAAAGTTGTCATCCTACCAAAACATCAT
GGACATACTAACAATAAATACAAATAAATATTTCTGCAAATCTCCATTACTTATGTGTTCCATAAAAGCCTAAATGCTGACTGGTAGGTGTAGTGGTTTGAATA
AGAATGGTCTTCATAAACTTTTATAAGTATTGCCTTGATAGAGGAAGTATATTACTGGGACCAGCTTTGAGATTTCAGAACTGCATTATAGGCCAAATTTCTCC
CTGTTGCCTGTAGATCATGATGTAAACTCTCAGTTACTTCTTCAGCACTATGTCTGCCTTTATGTTGTCATGTTACCCACCATAAGAATAATAAACTAAGCTTC
TGTCACTGTAAGAAATCCCCCAGTT

34-M13revfor

ACTTGCCTATTCCAAAGTGCAAAGGGCAGCCTGCTTCCTCCTCCAACCTGCTCGCCAGCCCGCTGCCATTCCCTGATACTTCCTAATTCACCTT
CCACTGCCATGACCACTGGCATGTTTTCATTCTTCATTCCTGGCTTCTCCTTCCTCATTCCTGGCTTTATCATCTGTTTAGTAAAACTGAGAGAGGCTTGGCA
GGATTCAAGGCTACGCTGGAGGACAAAAAACAGTGGTGAAATCCTAAGACATCAACCAGAGTAGGGTTGGGGGAGCCAGTACGCTGGTCTTGAGGTCTGC
ACGAGATAGGCCCTGCTAACACTCTGCTGAATTATGCTCTACAGGAGCTGGTACGCAGTCTTACCAGGCGCCAGGACTGTCTGGGATTAAGCAACCATAGG
TACGCTGGGTATTTATTCCATCAGATATACAAAGCAGAACGTTTCTCACACAGACATGGTGCCTTTTGCTTCAGTAAGGACACTACGGATTCACATGTGTCGC
CTCTCTATCTACCTGTCATCTCCCAAAGTTCAGCA

355-M13rev_1

CACCGCATATCACCTCCTCCAGGACCTCCACCAAACCCACTACCATCACCTCTGCCGCCACCTCCTCCAGGACCTCACTATCTCTCCTCTATC
ACCTTTACCAATAACCTTTGCCATATTTCCCACCACCACCTTATTCATCGGTTCTGCCAGGACTTCTACTATTATTCTTCACCATCACCTCTATTTTCCACAACC
ATATTTCATGCTATCCCCTCTACCATTCTGTACCACCACCTT

357-M13rev_3

TTAAAGTCATACAGATCACAGGGAAGGTGGCGCATGTATTTAACCCCCAGCATTGGAAAGATGGAGGCAGGAGGCTCTCTGTGGGTTTGAGG
TCAGCCTGATCTAGACAGAGTGCTCCGCAGATAGCCACACAGAGAGAGCCTGCCTTAGAGAAATAAATACCTGATGAAATAGAATTGAATTGAGAGTCCAGA
AATTAACCCACTCAGCTATGACCAACTGATTTCAGATAAAGGTCCAAGGTGTACTGAGTCAGCAAACCCTGCTGGGGCAATCTGACATCAGGTGCAAAGAGT
GCAGTCCACATGGTGACACCTGCCTGTCCCCTACTCGGGAGGCTGAGACAAGAAGATCAGTAGTAGTTGCAGTAAATCTCCACCCAAATATGCCCTGGCAA
TGAAAACACAACTCAATTAATATGAATACATGCTGTGCGCCTAGATTGGG

358-M13rev_4

TTCCTCACACTATATTAAATATGGCTGTAAGCTTAAGCAAAAGTACCAGTTTATTGGAATGCAAGATCTACTTGTGTATATATGGACTACAGAAA
GCAATTCCAACCACTTAGGCAGATGGGAGATTCCCACAGCAAAGATATTTCCAGAATTAGTTCATAAGTACAACAATGTTCTAAACATCAGATAATTCTTCAGT

GACAGGCTCCTAAGCAGAAAGTCTCAGAGGATGAGAAAATTAGCAAAACTCGAAGATGGAAAACATAAAAGCTGCAGTTGAGATGTCTGTTATTTTTAATAAA
TTCTAAATCATTCAATTTGGTCAAGGAAGTGTCAGGAGCCAGATGCTGGGTGGAAGCCTGGTAGCTCAGAGAAGCAGAAAAAACATCTGTTCACCTTTCTCT
TCATGTGAAATCCCAGAAGGAAATCTCTTTCTGCC

359-M13rev_1

GACAGGATAAAGATGCCCTGACTGTCTTGGGTGGTCCTGAGAAGAGCANTGTTCAGAGCTTGTTGCTGGCAGGGATTTATCCCTACACACTGT
GGTAGACGCCCTGTGATGTGCATGTAGGCTGTTTCTGGGATATCACACCACGGAGCCAAAGNAAGTGAGTGTAGCCATCTCACCAGGGCTAGCCTCATCAG
GGTCCTGGGAATGCACACACATGTTTGCAGAAGGTAGTGTGGGGAAACCTTGTGGCAATGTGCTGATCCATGCAATCCATCCCATGCNAGTCATTGCGGAT
CTCACTGTCACTTGGCCCATGAAAGGCACTGACAAACATATACATATTtTTATGTTTTAGAGTGGACTTTAGACATACTCATTGCTTCATGACTAGAATATTGGA
AAACGCC

35-M13revfor

GCTCCTCTTTCTGACTTTTAAACTGCCGACATAACAGACCCCGAGAAACTGTAATCTCCCCCAGGAACATCATCTGAGGCCACGTGGGAACCC
GTGTAAGAAAGCTGTGACTGTCCCGTGAGCTGTGGTTGGGGCAGAGTGCAGAATGTTCTTCCACAAGTGGCCTTCTAGAAATAATTGCAAACCCCATAAAAA
GTGGATGCTCTCAAAAGGGGCGACATGAAGGGTATGGCTTGGGATGTCCACGTTCTCTCTGTGGGAAGGGTGGGGTCCTGGGAGGAATATTCTTGGCTTTT
TCCCACAGGTGCAAGGATGATAAGACATAATACCAAGTGTTTTTTCACTTGCTGTTCTTGTATTTGCTTAAGCTCTCTGTAAACATCATCTCATGTTTTGCTTG
CTTTAATAATAAAAGAGTGAGAAGTTCTTTATTTGCTCTATTCTTGAGTTCAAACGCTTAGAGTCCCCAGGTTATCCCACAAAGATCTTTGGCTTTCAATCCAG
GGAGACTTTGTTAGCCTTGCTGGTCGCTGTCCGCTGCTGGGTTCGGATCTAAAGAGCCAGTAAATGTCAGAAGCCTGGAGCCACTGTGCCACCTTCAGAGG
AGCAGAGATTTCACACCGTCTGACAAACGCCACATCA

363-M13rev_4

CATATACACAATTAGAGTTAATCTTGGTTCAAGAAATAATAAATAGAAAAATTTGTTTGACTTGCTCCAAATAACAACTGCCATGATTTTCTCCAA
CAGAGGGTGAAAAACTATTTAGAATATGTTTTAGTATCTGTGCCAAGAATTTCATAAACCTGATTTTCTAGCTTGTTTATGTTGTTCTTACTAGCCAATATGA
CATATGTCCTTGAGATACCTGAGGACATGGTTGGAAGGACAGACAGACAGATACACACACACAGACACAACATTTCATTAATAACAATTCTTTAGTTATAAGA
CAAAGTTCTACTGGAAGTGTTGATTTCTAATCAAACAATGATGCTAAGTAGTTGTAAAATTTCAGGAGACATACTTGAATCTTAGGGACCTGAAGGTCATAAG
GTTTCCCACCTTATTATGATGATATACC

371-M13rev_3

GTCTCAAAAAACAAAACAAAACAAAACAAACAAAAAAACCAAAAACAGTAGCAAAATTATGGTTATGAAACAGCAATGAAAATTATTTTATGGCT
GAGGGTTACCACAACATAAGAAACTGTATTAAAAGGCTGCAGAATAAAGAAGGTTGAGAAACAGTACTATACATGATCCCAATGACACACTAATTAAATATTA
ACACATTAGTCCTCCAGTATATAAAAGGGTAAAGAACTGTTACCTCTTGTGGCCTTTTCATTTACTTTTATAGCATAATAATCCTAGCTTATATTAGAACACTCA
AAAAGAGTACTAAAAACAGCTTTATTGTCTTTTTGTTATGGGTAAAGAACCAATCTAAGTCTCTCCAGAATCAAAACTCTAGAAAACTTGTCGAGAACACAGAC
AAACAGTAGAACTCAGCCGGCCTGGGATGCATTACTTACTTTAT

372-M13rev_4

TCTTATTGGGTTCAATATAGTCTGTTATCACTGTTTTGTTCATTACTCATTTTAATCAGCTTGTCTTAGATGTATAGCTCATGGAAGAGTGATGCC
TGACCTACTTTAACAGGAACTGTGGCCTCCAGAATTGGCTACTTTCAGTGGTAAAACTTCAAGAGGCCTGGTAAATCCTGAGAAGCTAGTCAAGTCTTCAGG
CTGGCTCTCATGTCTTCGGGACAAGGTCTTTGTCATTCACTGAGTAAGATCTTTCTAGCTCAGTAAGATGTCCCAGGCTCATCCTATACTTTCTGTGACCCAG
ACCTGAAATCATCCATTGGTACTAGCATATTTGAAACCAGAGACATTTTGAGTATTAGAGTATGTATGCCATGGCACAGAAACTTGAGCCAGGTGGACAGGC
TCAGCCAGCCAGGAGACTGTGTTCTTGTCTGGCTCTGGGCCTTGGTCCTCACTGAA

376-M13rev_2

ATGCAACCAGCACCTACAATGTCGGCTTTATAAAAAAATCGTTCACAGGTGTTATCCATTTATTTATTTGTATCTGCGTGGGTGACCTCTGTAGT
CACCCGTGGAGGTCAGAAGACAATAGGAAACAGTTCTCCCTTTGCACCATTGAACTAAGGTCGTCAGTGCCTTGGTGGCAAGTGACTTTATCTACCGAGCC
GTCTTACACCCATCTACAAACTTTTAATACTTAGAGCAGGCTACATATAGCATGCCAAGTCCAATGACTGTCTTCAGACTTGAGTTGATTAGACACAGAGAAA
GACTGGATGCAGAAGTAACTATTGATCTGGACTCTAAGACAAGCTGCTTCCAAACAGCATATGATGTAGCTTGCTCACTGCCAACAGATAGTCACTCTCCAC
TGGATTTC

377-M13rev_1

CTAACATGATATGACATCTAGCCTTCCAACATAGGGAACATGAAGTGAGAGTGGCCAACAGACCCTCTCACACTCTGCTGTCTTTGTCCATGAA
CATGAAGAACTCCAGGTACCAACATTTGAACAAGCCACGGGGCCGTCCATGTTTATGTTGTGTTAGGCTTAAACAGATAGGACAGAGCAGTGGTCAGCTAG
GATCAAGTAGACTCCATCTGACACATAGGTGCTTGACAGATGTTGAGTGAATGATTGAAAGGGGGGGGGGGGGAAGTACCAAGAGACTTAGGAGGAGCTCA
AACAGGGAATCTCCCCTGTGAGGAGTCTCCACTTAAGCCTTCTCTTTTTACAAGGCACCCCTACCCTCTTGCACCCCAATCTCACACCCTCTTCAACACTCAC
AACTTCCTGCAAAGGTGTGTGTGTTGGGGGTGGATCTCTGTCTGATCTCTTAAGTTTTGCCCTTATGCCTGAACCAACTTTTCCA

37-M13for_1

TGTNCACAGCAGTAAAATCTTAACTAGGACAGAAGTTTTACCAGGGAACTCGGGTACTGCTGTTATAGGCCTGTGATAAAAAATTAGTAGTGAT
TACGAAgAGACCACCATCACTCTGGTGAAATAtTCTAAGAAGTGTTTTCTAAGATCAATAAGAAACTGTGTTCCAGCCGGGTAGTGGTAGCTCATGCCTTTAAT
CCCAGCACTTGGNAGGCAGAGGCAGGTGGATTTTTGAATTAGAGGTCAGCCTGGTCTACAGAGTGAGTTCTAGGACAGCCAGAGCTACACAGAAAAACCTT
GTCTTGAAAATCAAAAACAAAAACAAAGAAaCaGAGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

37-M13rev_1

CTATCTATGAAACACAATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTTTCTTC
CTTCCTTCCTTCCTTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTTCTCTCTCTCTCTCTCTCTCTCTCT

380-M13rev_3

TCCTTTCCACGGGTCTTGAACCCGAGGAAAAGTTTAGGTCAGGTAAGAATACCTGGGGAGAGATTAGAAGGAAGGAAAAGAAAAAAGAAAAGA
AAAAAGATCGATTAGCGGAGGTCTCTAGGAGATACTCGTCACCTAGATGAGCTCAGGAAGCCAGCTCTTAGTAGCTCTGAAGCAGATGAAGAATTCTCCTCTG
AGGAAACAGACTGGGAGGAAGAAGCAGCCCATTACCAGCCAGCTAATTGGTCAAGAAAAAAGCCAAAAGCGGCTGGCGAAAGCCAGCGTACTGTTCAACC
TCCGGGCAGTCGGTTTCAAGGTCCGCCCTATGCGGAGCCCCCGCCCTGCGTAGTGTGTCAGCAATGCGCAgAGAGACAATGCGCAgAGAGGCAATGCGCA
gAGAGGCAATGCGCATAGAGGCAGTGCGCAGAGAGGCAGTGCGCA

392-M13rev_4

CCACAATGACCAAGTTTGATCACTTCTAAATACCCCACACCTCTAAGCGCTACAGCTAGACTGAGTTTCTGCCCTTTTAATATCACTGGCATAAA
ACTTTGGTTATTAAACTTCTGTTAGTTTAGGAAGTGTGGTGGTTTGGATAATGGCCCCAATAGGTTCATATTCTTGAATGCTTAGAGAGTGGCACTATTTGA
AAGGATTAGGAGGTGTGGCCTTGTTGGAAGTGTGTCACTAGGTGTGGGCTTTGAGGTTTCAAAAGTCTAAGCCAGGCCCAGTGGCTTTCTCTCTCCCTGCC
GCCTGTAGATCTAGATGTGGAATTCTCACCCTCTTCTTTAGCACCATGTCTGCCTATGAGCTGCCTGCTTCCCACCCCAGTTAAATGCTTTCCCAATACAAGC
CCCAATCAAACACTTTATAAGAATTGTCATGGTCACGATGTCTCATCAC

395-M13rev_3

GGCTGTGAGCCCAAGATAGCCCCTTTCTCCCTTAACTTGTTTTTTGTCAGGGCATTTATTACATCGATAGGATAAGTAACGAATGCACTCTTTTTC
CAAATCATTTAAATGACTAAAATCGGGTTAGTAGGCATAGCTGTGAATCAAAAGCAGCTTTTCCTGCTTCTGGAGTATGCGTGCTTCATAAAATATGGCCTCT
CCCATGCCAAGACTCAGCAACAGTTCAGTTCTTTAGGAGCAGCTCTAGTGTCATGCTTACAGTAGATTGACTAAAGTATCCGTTTTTGCATAACTCCTGGAAG
AGCTGAAGTGGTATGATTCTTGATTTTGCAGATATTAATCACAGCGATCCTTCATGTTGTAAGGGGTCTAGGAAGACTTTGTAATCCTAACCAGCTCATTTAAT
GCTTTGCTTTTTATTAAAATATTCATACTTATGGTTTTTGGCTACTCAAATTATTTTATTTTGAATATGTTACTTTTGGGtGTGTGGGAGATGCTGTTTAAGTAAA

398-M13rev_2

ACGCCCGTGCCCGGCCCCGGCCCCCTCACACCTCACACCACAGCGGACTAGGACCAGGCGNGCGGCCGGGACTCCCGCCTCCCCTCCCCC
ACGCCCCCGCGGGTCCCCCAACGCCCATGCCGCCGGCCCGGCGCGGCCCGGCGCACTCACATCGGCTCCAGTAACTTGGCCAGCCAGGACTTGAGGGC
ATCCACATCTTCTATGAGCATGGTGCAGGGAAGGCGGGCTCGGCCGCCGCTCCTCCTAAGGCCCGCCGCGGGTCTACGTTCGCCGCCGCCTCACGTTCC
CTCACAGCTCCGAGTGCCCTGACCCACCGGGCATCTTATAGACTTCACCTACTCCCTGGGCCGGGGCCTAACAGGAACTAACCCAAGAGCCCCACTCAAC
CTAGCCCGGAGGCCCAGGTGGGC

3-M13revfor

CTTAATGATGCTTTAAATTCNTTGCAGAGAGGACACCCTTTGGCCTCNGTTCCAGCCAGACTGTGTTTTCTGTCTGCTTTTGGCACGTCACTGG
AAACACCAGACTCGTTCTAAACCTTTAGTCCTTGGATTCATTAAAAATTAATTGACTTTGATGGGCATAATATTTCTACACTATTTGGGAGGTACAGTACATTG
CTCTGACNCNTGCCTGCATTGGGCAATGATCAACTGAGANCAATTAAGCGGCTCCATCACCTGCAGCTTTGTCAGGCCTCCATGGTGAGGATATTCAAATTC
CTCTCCATTAGCTGTTTGGATGCCAGTCCCCGGATGTTAAAAGATATTTTAAAATTAATGATTTGCTTTCTCAGTGAAGAAATTTTCAGTGTTGCTCAACTCTG
ACTTGGCTCTTATCATCTTTAATCCAAACACAATAGGAGGGAGAAAAAAATATCNNTTCTTTGATCTGAATAAAGCATTGTCTGGAAGATTAAAATAAGTTCCA
CCAAAGGTCACCAGAAGGTGATCCGGCAGTAGCTTAGGCATGCTATCTAATTGAAAAAATAGAAATCCTTAAAAATTAAAACTAGGACAATACAGGAGTTTTC
TAATGAGGCAAATCTCAATAAAGGANNTAAGAGAGCTAATAGCTTGAGGCCTTGGATACCTGCTT

400-M13rev_4

GaAATTAAATAGCTCCCAGAAAAgTAGCTGCAGGTTTGTCAATTCCTGGCTCTGTCAAACTAATCATCAGAAATTCTTGACTCTTTGACATTGGC
AAAGTCCAAGGCTTCAGCTTGTAAGACCTACTCCTGGAAGGCCCAGCTCAAGCCTACCTCTACCACGAAGATCCCCAACTCTGCCCANGACTCNCTGTGCC
TTCTCCAGAATGACTTCCTTCTTCCTTTTTCATGANAAAGGATNCCCTCCTCCTTCTTTGT

406-M13rev_1

GGATCAACAAACTATAGCTCAAGGACTGAACTGTACATAACACCTGTTTCTGTCAAATTCTGTGGACGCAGACATGGTCACAAACATACAGATT
GTCTATGGCTGGTTTTGCACTGCAAGAGTTGAAGAGAATTAAAAGAGAATCCAAAATTTATAGTTAAAAGCATTTGGTCCTTTACAGAAGTTTCATAACCTTTG
ACCTATATCATTAACCGGGTTTGGGTTGTTATACCTTGAGTACATATACATGTGTGTGTATGCATGCGTGTACAGACAGACAGACAGACAGATACAGAGTTAT
ACTTTGCATTTGATATCCCTTCTGCTTCCAATAGATGGTCCCATATACTTGCACATTCTGGCAGCAATGAATAGACTCAgTGCATCTTTAAAATAAGAGCACAT
GAaATGAAgAGGTAAATGTGTTAgAAGTGTGTGTTAAAACAGAAACACTTTTGTCGCATAGG

43-M13rev_2

CAATATGTTCCAGACTCATTCCTCTGTGACAAGCAACCCCATTTGTTCCTTAGGACAGCCTCCTTCTAGNNTATGGTGTGCCCCAAAATCAGCT
CATAAATATATCTTTGTCTGGAATAAAAAGTCGAACTGTTACTTCTTGTAgGATCTGTCCTTGGCTTAACACATCGTTACGATAGCGAGGTGTGGAGGTGTCG
TTGTGTGCAAAGTGGGAGAAGCTGCCGAGGCTGAAGTGATGGGGGCTCTTGACGGGCTTGTTGACCTGCAGGCAAGTATGTAACTGGAGAGAGCTGAGCA
AGAGGCAGCTCAAACTCTTCCTCCAGGTCTCCAGTGCGCTCAACTGTGGTCAAGCCAGGAGGCTGCGGACCAACCTCTGGCTCACACAGCCTCTGCCAG
GGCCCTCACACTTCCAAGCGCTCTGTTTTCCTTCAACCA

45-M13rev_3

CACGGCAAGCGTTACGCCGTCATCGTCAACGAGTTCGGCGAGATCGGCATCGACAACGACCTGGTGGTCGGCGCCGACGAGGACGTGTTCG

AGATGAACAACGGCTGCGTCTGCTGCACGGTGCGCGGCGACCTCATCCGCGTGGTCGCGGGCCTGATGAANCGCCAGCGCCCCGGCAAGCCCGCCTTC
GACGCGATCATCGTCGAGACGACCGGCCTGGCCGATCCCGGCCCGGTGGCCCAGACCTTCTTCGTGGACGAGGACGTCAAGGCCAAGACCCAGTTGGAC
AGCGTCACCACCCTGGTCGACGCCAAGCACGTCATGGCCCGTCTGGACGACAGCAAGGAGGCCCGCGAACAGGTGGCCTTCGCTGACCGGATCATCTTG
AACAAGGTCGATCTGGCCACGGCGGACGAACTGGAGGCGGTGGAGGCCCGTCTGC
465-M13rev_2
GGAAAGAACAAGGACCTTGATAGATCCAGGAGGAACCCCAGTGTGCAAAAGGTGCCATGTGCCAAACTGGCTGTGTCCAGGGATCCTGGAAA
AGTGCCCTGCCAAACCCCTCTTTCTCCCTGAGCACCCATTTCAGATTTTGAAGTGCTTGTGGACAGCATAGCTAGAGAGATCCCTTCCATTCTTTCTGGGGG
AAGATGACCTCTCTCGTAGATCTGGGA
466-M13rev_3
CACACACACACACACCAAAGAGGTCTGTAGAGCTGAGATCAGATGGATGGACTCAGGATTCTACTCCATGAGGTTAGCATTGTCAACACATCA
ATCTATATGACACCCAGAGGTACACCCTTCATCCCTGAGTGTTCTAAATGGTTTTGAACATAGGCATTTCACTATCTTCTCTGGAAATGGGCACTCCTAGCAA
CTATAATTCTTCTATTGGGGTGTCTAAAGCTACTTAGCATGGTTTAACTAGTCCTGTGGGTTAAAAATCAGGATCGTAGTGGAGCATGTGTCCTTCTTACCAG
TTGGGGCATCTTCTGGATATATGCCCAGGAGAGGTATTGCTGGATCCTCCGGTAGTACTATATCCAGTTTTCTGAGGAACCACCAGACTGATCTCCAGAGTG
GTTGTACAAGCCTGCAATCCCACCAACAATGGAGGAGTGTTCCTCTTTCNCCACATC
467-M13rev_4
GTCGTTGCTCAAACCTCTGCATACTTGATCAGAAACGAGGTGGCCTTTGCTGTGTTTGTTTGTTTGTTTCTTTGTTTGTTTGTTTTTTGGGGGGGG
TTGTGATCAAAGTTTTAGGGTAAATTATTCTCACAATGGAGGTTTTAAATAAGCAATCTAGTATATGTTGAAGGCTTATCAACTGAGACAAGAGAGCAGGAAAC
ACATTTTAAGCAGATTTACTATGTTATAGGTGACTTCCATGCAGTGTGCAAATTCAGCAGTAGAAAAACTAGTATTTCCTTTCTGTCTCTTAGAAGAAGCATAT
TTGGCTATAATTTATAAGGTTGGTTGGGCCTTTAATCTACTTTCTTTCTAGAATACATCATTAATGTTAGTAAAATTTTCTCTCCAGAGAGATACGCAAGGATAG
GACACTTATAAGGGTATTTTTAAAACTTGCTTATTTTCTGGAGTTTTGATCATGTT
468-M13rev_1
CCGACGGGACTGTGCTGGGGGGACAGTTCACAGAGGAAAGGGTAGAAATGCTGAGACTGGAGTTGGGTTTCAAAGATAATCCTTGTGAGGTT
AGCATTCCAAAGCAGGGGTACTGGGGGGAACAAAGGCAGATGCCATCTCAGGGATCCAGGACACTCACCTTTACCTNCCTGTCTCTNGGTGGTGATGCACT
GGATCCCGTCCCAGAGGCGTGCCATCACCCTAAGTTATTTTTTTTCTCCTTTCTGAGGTTAATATATTTTTTAAAATCACTATCAGCTATTGACTTGGAGTNAA
ACTTCTGGGCTGGGGTGAGCCATATGGTCTCATCANCATCACCATCACCATCATCATCATCATCACCATCATCACCATCACCATCACCATCATCACCATCATC
ATCATCACCATCNTCNT
469-M13rev_2
AGGAANANTGGGGTGCGTTTGAAAGAGTAAGTTTTAGAGGAAAGTTTCTTATAGATTCGTGGATGGAAACTTAACAAATATGTCACTGGAGTTG
AGCCAGCAGAGAGGTAAGCCACATGACAGGAAGGGAGCTTCAGAGAAAGCAGGGAACACNGAAGCCCAGGGTACCAGAAGAGTATCTGGCCTCTGGGCA
GCACCCAAAAGAAAGAGACAGAGGAAAGGTGGCCCCACCTTCCTAAAGCATGCTCCCTTAACCAGGCCATTGACCTGCCAAACTCGAANGTTGGGTCTCTA
CCAGANATTTAATCTTCCTGTCTANAAGGAACTACTTTCCCCTTGGTGNGATTTTCCGGAACTGNGATAGCTTTCCCAGCAAGATGTTACTCTNNTAGGGCTC
TGCT
46-M13rev_4
CCCTCTCAGCCCCTCAGGGTGTGGAAGTTGCTTTGCAGCACTCACCCTGAACTCTTCCTCCAGCCCGATCTTGTCAAAATGGGTGGCTGTGCT
ATGGAGCGTCTGCAGGTGCCTCTCCAGGGAGGACACATCCAGCTCTGTGTCCCCATAGAGGTAGTATTCCAGTAAGGCTTGGTAGATGAATGTATACTGAA
CCTGTAGGGAGACAGGGGGTGCCACTCAAGACTGCCCCATCGCTGGTAAAGCAAGGCCGAGGAGCAAGAGTGATCATGTGGACTGGTCGTCTAGGGTAAA
ATAAGATGACAGTAATGACCACTACCTCCGAGGGCTCCAGGGGCTATGTCAGCTGCTTAGAATCACCACCATGGCATATGAACATGCTAAGGATCAGGGTC
ACCAGCCTTTGTGACTCCATGGGAGGTAATGACAGAGTGGGAGGTATACCAG
470-M13rev_3
GTGTGCCACGCTGGCGAGGCAAGACCCACAGCTAGATCTTGGAAAACGGAAGCAACTGGGGTGGGACAGGAGCTCAACAAAGAGGAGTGTG
GGTATTCAGAAAGCCAACTGCCATCTTCTCGGTGAAGGAAGTCGTCCTTCATTTAATGCTTCTGTCTGCTCTCAGAGACCTGGTTGTGAGAGTCTCTACCAG
TACTCTGTGGGGGCCTGGACCGTTCCCTGAGGAAGCATGGTTCTG
471-M13rev_4
ACTGCTTTGATAGGTGCCACCATTTTGACAGATGCTGTCGTTTTGACAGCTGTGGTGGTATCTTATGGTGGCCTCAGGCTGGAGCCGCTGCTT
TAACAGCTCTCCTGCTGCTCGGGGGTTACAGGGATCTCAAAATATATCGCTTTCTCTTTCCCCTGTTTTCTCCTCCCCATCTTTCCCTTTCCCTGCTCTGGTG
ACTGAACCCAGGGCTTTCAGCACACTAGGCAAACCCTCTGTCTCACTGTCACATGCCCTANTCAAAGGCAAGACTCCTAACCCAGCTAAAGGGCANGTTCA
473-M13rev_2
CTTGCCAATCACTCTGCATGTCACCATTCTCTGACCCTGTGGTTTTGGCAGGATGTAGGGGTCCTCACAAGTGCTGCTGTAGATACAGCTCAG
TACATTGCTGTGTGTGATAGAGAGGTCAGGCAGAGCTAGCCTGCTTCGCTCATCTGCTGCAGCCAATGGAGACATTTCTGAACTGCTGTGGGATTGACTGC
ATCCTCGGTTGTTGCTTCAGTAAGGAAGTGAAGAAAGTGCCTGTCCGTATGTTCTGGGTTAGGGTTTTCCAGGCCATGCCTGTCACCTGCTGTTCTGCCTCA
GTCACTTTGGGACGTAGCAGTCATGTGCTTTGCAGATGTCAGAAGTAATCTCCCGTCCTCGTGTTTTTCAGCTGCACTGGCGTCTGTTGAGAGCTCACAGAG
CTGAAGCTGTGGACCTCAGTGGTGTAANAAGATGTACAATATACCTTAACTCTATGAAATTCATAGTTCGGATGC
53-M13rev_1
TCTTCGGTCGCGTGACCGGCGTGCGCACCACCGAACGTACCGCCGAGGCGTCCATCATCCAGACCCGCCACCGCATCCCCGAGACGCCGCT
GCGCGAGGGCCAGACCATGGTGTTCCAGGTGCCCATTCCGGAGCCGCTGCGCTGGCTGGAGCCGAGCGAAAGCGAACCTCACTCAAGCACTCAAGCCCCGGGACA
CCGAATACGGCGCGATGCATGTCAAGCTGTATGAGGACATCGCGCGCCACGGTCATATCGCCACCACCTACGACTATCCCGTGATCGTCAACCAGCGCTA
CATGATGCGGCCGTCGCCGATTCCGAAGTTCGACAACCCCAAGCTCGACCGCAGCCCCGCGCTGATGCTGTTCGGCGCGGGCCGCGAGAAGCGCGTCTA
TGCGGTGCCGCCCTATACCCCGGTCAAGAGCCTGGATTTCGTCG
550-M13rev_3
CGTCCGCAAAACCGAAACTGCGATCGCACCAAACTTAGAAACAGAACACAAACATCAGACCAAAACAGAACAAAGAATCTTACCTCTAAGTGG
GTCTAGGACCTCTGGGTGGTCGCGCTGATTCCCGGCCAACGCACCAAAATGAAAGACCCCCGAAGGGAGACCCTCACTCAAGCACTCAAGCCCCGGGACA
GCCGCGGACCCAAGAAGACACTGAGACCGTACTCAAAATGTAGAAGGCAAGATTTAATAGAACAACAGCAAGAAAGCACATGGAACATAGACCAGAGCTCT
GGGATCGAAACTCATGCACCTAGTGGTGTAGAGGAGAATCAACGGCCAGCCAAATTTTTCACAGGCTTATATAGTAAAACTCAAGGGGGGAGAGCTGGGCA
GGGGAAAGTACAAGTTTACATCAC
551-M13rev_4
GTCCCCCAGACCTGGCTCAAAGGCCAGAGAGCTGGCATATGTGTCAAATGTCTTCCAGCAGCCTTGACAGGCAGAACAGGAAAATCAGGCCA
GGGCGGGTGTCTGCAAGCTATCCGGG
556-M13rev_3
GAATTCCCTATCCCATCCCCCCTCCCTCTGCTTCTATGAGGGTGTTGACCCACCCACCAACCCACTCCTGCCTCACTGCCCTAGCATTTCCCTA
CATTGGGGCAATGGGACTTCCTAGGACCAAGAACCTCTCCTCCCGTTGATGCCCAACATGACCATCCTCTGCCATGGATCCCTCCATGTGTAGTACTCTTTG
GTTGAAAATTTTATTTTTTTTCTATTAAAATCGATTCTTCCCTCATACAATAGAATCAAGCCCCACTTCCCCTTTCCTCTTCCTCTCTTTACACCCCCTGTCCCC
AAGATCCACTCCCAATCTGTTTCCTTTTCAGCAATAAGCAAACATCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGAGAGAGAGAGACAAATAAGACCA
558-M13rev_4
GCCATCTCACCAGCCCTTGTTCTCATTTTTTAAGTTTATTATTTATTTTTTTACTTGCATACACGTGTGTGTGTGTATATGTATGTTTGAGTTTGGC
ATGTGCATGTGAGCGCAATGTCCAAGGAAGCCAGAAGAGGGCGCTGAATCCTCCAGAGCTCAAGTTACAGGTGACTGGGAAGCCACCAGAGCAGGTGGTG
CTCTTAACTGCTGAGCCATCTGTCCAGCTCCTTAGGTTCAGTCTTTAGAAATGGCGTGTATTTATGGTACAAAGCACTGCTATCCATTAGGACATTTTTTTATA
TGTCTGTCTTTTACTTCCTTCACATTTCTGGCATTACCCTCTCTTGTGTCTGACCCCCCTTTTGATTTCCTTCTCTTTCACTCCTGCTTTCCCCCCAACTCCCTT
CTGGCCC
55-M13rev_2
ATCTTTTTCAATCCATTTATCATTTTATCATTTGTATAAAGCAGGGCTACTGATTTTTTTAAGTTAATTTTATATCCAGCCACATTGTTAGAGACATT
TATCACCTGTTGGAGTTCCCTGGCAGAAACTTTGGGCTCaCTCATCTGTACTACCATATTATCTATGAATAGCTATACTTTGACTTCCTCCTNTCCAGTTTGTA
TCCCCTTTATCTCAAGTTTTCTTATTACTCTGTCTAAATCTTCAAGNAnTATATTGAATAGATGTGGCATCAATGGGAAGGCTTGACTTGTCCTAtATTTTAGAG
GAATAGCCTTAAGTTTCTCCCCATTTCATTTGATTTGGCTATGGCCTTGCTGcATATTGCCTTTATTGtGTTTAGGTtTCTGCCtTG
560-M13rev_2
CTCCTGTGGTCTCAGCAGGCACCTGCTCTTGCATGCACACAGTCACACGCACACACAACTTACACATTCGAAAAGTAAACCTTAAAGAAAGAAC
ATAATTCCCAGTGACTTGAATGGCTTGGTATTAGGGCCTTTTTCTCCAAGTTCAGAAATACTTTTCTGAAATGTCGCGCCTCTGTGGCCTGCATGCAGGTCTT
GGCAAGAGTTTAACCTCACAGAAAAGTACCTGGCTGCTCTGGTCTATCCCTGTTGCTCAACTGTGCCAGGTTTTATATGTGAAGTTTCGATAGCATCCAGCC
ACACTCATGAGCTTGCTTGGTATCTATGGCTACAACTCCATAATCAAGTACTGCAATAGTGGCTACCAAGTATAAAACTATTCACTATCTGGACATTTCAGGG
AGTTTGT
568-M13rev_3
CTGCTCTTTGGCTGNCACGTTGACATCCTTACTTCCTGTTTGCTTGGTGTTTTTTGCCCTTCAGCTTTTTAATGGTCATGCATTACCAATACCCA
GGGCAGCCATGACAGAGGACCTTTAGGGAAAGGATTTAGAATATTGCTGAATGTTGGGAGATGGAAATAATAGATATACAGGGGAAATTCTTCCTGTTGAAA
ATTTGCAAGTGTCTTGAGCCCTGGGTATGTAAACTCTTCCTGTTGCCTGCAGGGGATATAGGGACTGACTGTTGTTACTCACTCCTGGGGATGCTGTCCAGA
TATGGGTATTGTTATCTTGGCTGGGATGCTGTCTGGTGTGGGGACTGACTCTTTTTTTATTAATCATTTTATTTGCTTACATTTCAAATGTCATTCCCCTTCCCN
AGTCTCCTCTCCATGA
569-M13rev_4
GTCTTGTGCTCCCACCCCCTGAAGACATGGCTTGGGTAACTCCCTGTAGCAGCTCAATTGTGCAAGAGGGAAAAGGAAGAGCCCAGGGGCAA
GTGGTTGAAAAGCAAGCTCATTTGTTTCAGTAGAACCTAGGCAGCCCCACGTTCAGGGTTGGCAGATCCCCCAGGTAGCATCCCCCAATTACCCTCTGGCT
CTTCTATTTGCCCGGGCACTCTTAACAGTGCACTTCCCCGGCTGGGCCCACTCACCGTCTCTGCGGACACACAGGCCATCAAGCCCAAGGTGAGCAGGAT
CCACGCGCGCTGCATATTCCCCAGGGGATGTGGCTCCTACCCGGCATAAAAGGAAGCCCATATCAAGCC
571-M13rev_2
GTACACTCATGTATCCGCCCTGGACATTTTGATTGAGATGAGGTAGTTTGCCAGCTGTGTTCTAACCATGTCTACTGTGAACATATGGTCTGTA
TAAAGCTTACCTGGGTAGCAGGCACTGTTACCATGGCACAGAGGAGACTCTCAGATGCAGGACTCCTGTCCACCACAGCAATGATGCCAAATATTATGCTG
ACTGTGGCCATCCGGGACCATTAGGATTGTTTACTAAATAATGGCAGCTGGGGTCCTACCGTCAGATCCTTTAAGTCAAAATCTCTACAACTGTTACTCTTAG

CTAGGAGTGACCTTGTGTTCCTCCTTCTCCAGGAGACTGCAGCCCTGTCTGTGTGCATTTTGAGACGTCGCAGCTGGTAAGGCAGAGTACTGCTGGCATCC
AGTGGGAAGA

574-M13rev_3

AGCTCCTCACCTCTGAGTTCTCTTCTTTTACTCTTCGTGACTGCAAAACAAGGATATGTCCATTAGAATCCATGGAATTCCACAGAGCCACGATA
CAAGACAGCCAGTGGTGGCCGAGGGTGGTTACCTTCTCTTGGGAGGCGGCTGCCTTCCCCTCCTCGCTCTTCTCATCCATCTCGTCATCGCTCCACTCCCA
GTCGCCGTCTTCAGTCTTATGAAGGTGACCGCTCGACCCAGGAACTCGCCTTACCTGCACCAAAATAAGCCACACTTTCAAGGGTCGGATGGACAGAGCAC
TCACTGGTACTGCTACTCCAGGTTATCCAGCANACACAAGCCTCTGAAAGGTAACTCAAGAGGGTCTCAACCCACGGGATAACTGGAAGCTAACATACCAAA
AGANAAANNNCCTTAGACTTCCCC

579-M13rev_4

ACTACAACTCCCGGCAGCCGCGGCGGCTCGCGCAGCTTTGTGGGAGCAAGAAGCCGCTTGGTCCAATCTGCCGGCTTCTGCCACCGGAGAG
AGCACCTAGGGGAGCGCAGGGCTGTGAGTGCCGGACGTGCAGAGCTGGAGTATCTAGGGCTTGCTCGCGATAGGGCCTCGCGAGCGCCGCGACTACTTG
TCCCAGAAATCCTCGCGCGCGAGGTCTTCTTCCGGTTGGTCGTTCCTGACTTCCTGGCAGCCCGTAGCCGCACGGACCTGGACGGAAACTCCGAGGCAAC
TCCTCACTTGTCGGAGATCTTTGCAGTGCGAGCGAGCGATCTCCGCACATGCGCAGTGGGTGTCTGCGGTGCCCACTAgTCGTCTAAgGCTGCAgCGCTGC
CgTTCCgTGTGTGTTG

580-M13rev_1

CTCAGGGAGAGACCCTGCCTCAAGAGAATAGGACAGAGTGTAGATAGAACAAGGTACCCAACTTACTCTGGCCTCCTTTCTTACCCACATGCA
CACATTCTCTCTCTCTCTTTCTCTCTCTCTTTCTCTCTCTCTTTCATACACACACACACACACACACACACACACACAGACACCACAAATGTGGGGG
GCACACACATGTCCATACTTGTGGACACACAACAAAACAGAAAAATGAAAGGCAAAAAGGCTCTTTAAAAGCAAAGCTACTGAGGATAGATACTGTCAGATG
AAATCTCAGTAACGACGGCAAACTGGTACAGAGATGCACATCTTGAGGAAGCTGTGGATGATATTCTTGCAGGGTAAATGTTAATAATCTTGGAAAAAATTG

581-M13rev_2

CCTTGCGGAGAAAAAGAGATCTGAATTCATGTTGGAACAATGTANGTAAGGGATTTCCNCAAGCCGGATCCGTAACTTCGGGATAAGGATTGG
CTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGCCGCGGCTGGACNAGGCGCCGCCGCCCTCTCCCNCGTCCGGGGA
GACCCCCCGTCCTTTCCGCCCGGGCCCGCCCTCCCCTCTTCNCCGCGGGGCCCCNTCGTCCCCNGCGTCGTCTT

584-M13rev_4

TTATTTGTACATATGTATGTGTACACGTGCACATGTGTACAGTGCTCAAGCTTCCATTTTTGGGCTCCCCCATAATGATGGACTATATCCTANAA
TGGTTGGCTCAAATAAACCCATGCTCCTCTAAGTTAATTTTTAAGTTTAACCAACAACTCGGAACTAAAATGGAATTTTCCAAgTTTTACACAGGGGAGAAACT
GaGAGCACAAAACAAGCAACTTATCTGTGAAGTCTTAAGTCAGGATNNANACCCAGCTACTCTGCATCACAATCCAgaTTCTACCNTGTCACCACACA

585-M13rev_1

CTTCACTTCCCAGACACAGGAAAACAGCAAGAACGCCTCTGAATAAAGGGGACGTTTGAAGGAACTCTTTCACAAGCAGGAGAAAGCCagNAG
TCACTTGTCAGTAGAATGAAGATCACATCTCATACACTGagAGAGATCCCAGGTCATGTGAGCAATNAAAGATTTTTCTCAGGAAGAAACAAAATGGGAATAG
ATTTGAATGTAGAAGTAAACAGTGCTTTGGAGTGTTGTTGTGTTAGCAAAATTAGATCATTTAAAAATGAGACCCAGGAAAAGTTAGTCCTCCCCCTGTGGAC
AACCCTGCACATTAAGTGAACGGTGTACATAAAGTCACCATCAGTGTTGGCTTTTATGTTGGTGCCACTCCAGAAGAAAGGACGTCTTCACAGCCCACCACC
ACACTTA

591-M13rev_2

CTTCATCTTCTGCAGAAGTGCGTCTACTTCTGGAGCTGCTAGCATTTGATTCTCCACTGAGGGCTAGGGCTGCTTTGTGGTTACTCTTCTCTAA
CTTGGGGGAGCACAGATCATCACTAAGGACGTTAGTACTTTGACAGAAGCCCTTGATCTCTGATGGGCTAGCTCTCTGCTAGGCACTGGGGAGGGTGGAAT
AAGAAATAGCTCTGCTCAGAAGGATCTTATCAGATAGGAAGACAGTCTAGAGGGCACCCAGCAGAAGGGCAGCTTGAGCTTCTTCGAAGTGAGCTTCGCTG
GGACTGAAGCTGAGGGTAGATGGGgAAAGAACCCCCAGCcTCAgGGCACAGGCTGTCCAGAATACTTTGGAGAGATTGCTGTAACAtTTT

592-M13rev_3

TACTTAAAATGTAAAAAAGAAAGTTCTGAAGGTCACAGGCCTAAAGACANACTTAAGCCAAAGCTACAACTTCGTAAGAAAGTCAAATCCTATAA
AAGTCCTGTTCTATCAATACCGCTGGCATCCTTTTAGTGACGACTCCCAGCCTGCATCCTTGCTACAAATTCACCTGTCGCCCCGCCCTCCCCGATGCAGTA
GCGGGACCATGGGTTGTCCCGCCTTCTTCCATGTTANTAGACCCGCCCCTGATGCTGATTGTCTAGGGNGACAA

593-M13rev_4

ACTCGTCACCCACTTCCTCCTCCACCAGTCTTAGAGACTGGCATTTCAGCTTCATTTTAAATGGCCTCAGACAAAACCATTAAAAATTCCTAATG
CTCTATCCCCTTTCTTGCCACTGAGAAGATGGAGAAAATGGCCTTTTCCTTGAGTGGAGAGCAGTGAGGTGAGAAGGAGGTTAGATGGGAAATTGAACAGC
ACCCTCGCTTCCTCTGATGACTGAGCTATTTGGATTGACACCGTCTCCTCCCCGGCCAGTCAGTGTAGAGCTGAGCTCAGAATGCCATTAAAGACCACACCA
TTGGATATTCCCATCACGACTTGGAGGGTTCAGGATTTTAATACAATAAAACTTTATGTCATTTGCCATTTAAAAATACCCCCCCCCCCCAGCTACCCTGCTATA
TGCAGGTGAGCC

594-M13rev_1

GCATCGAAcCAGCGCCTGTCCGCCGCCCCGACCATCGGCCNGGACACCTTCGAGGGCCGTCGCCAGGCCCAGTCGAACGTCCTTCTGCAGA
CCATCCACGAGGCCCTGGCCGAAAACCGCGCCGACCTCTACCTGCAGCCGGTCGTTTCGCTGCCCCAGCGCCGGACCATCTTCTACGAAAGCTTCACCCG
TCTGCGCGACGCCTCTGACCGCGTCATGATGCCGGCGGAATACTTGTCTCTGGCCGAGGGCGAGGGTCTGGTTCCCGCCATCGACAATCTGCTGCTGTTC
CGCTGCGCCCAGATCGTGCGCCGTCTGGCCCGTCAGGACCGCAAGGTCgGCGTCTTCTGCAACGTGGCCnnnnng

597-M13rev_4

AGTCAGGATTTGAATTTATATTGTGTTGTCTGACTGAGCCTGCAAGGGATATTGTGAGAGAATCCAGGAGTCTTTTGTTTTTGTTTTTGTTTTTGT
TTTTTTCAAAACTAACCCCTTTTTATTGTCAAACACCGAATGCTTTTCANAACACTCTCCCATCTAAAATCTCCAATCTTCTCGACAGCCCTAGGANTTATATAA
AGCCAGGACTGAATGGCTCATTTTACACCACAAAAGCGCC

598-M13rev_1

TTCAAGCAGCCTGTGCTGGACCTTTGCTATTATAAGCCTGGGATGACTTCTAGATGATATAGCAACCACTCCCACCTGGGTCAGAGGGAGATT
ACCATAGACTTTGATGGCGAAATAGAAGTAATGCTATGATTGGTGTGGCTACTACCCTCTGTGAATGTGAAGTGTGCTGTGACCTAGCATCCCTCCTCTGGC
TAGGTGTAAGGGTCTCATCTTCAAAAACCCTCCTGTGCTTGTTGCTCCTGGTGTTGCAACTATAGGCTTnnnnnnnTTTCAAATTAgAATGGCCCTCACCTCTAA
GCTTCNTCCCCTGCACGCTAG

599-M13rev_2

AGACTCACAGAGTCTAGTGACTGGAGGCAGGGCAGACTGGCTAGTCCAGGATTGAGATCATCCAGCCCTGGTTTCCTCCATTCCTGTACTGAT
TCTTCCAACCTGCCCTCACAACTACACACACCCTGACCGGTTGGGCCTGTCCCTTGTAGGCAGTCTCCTGCAGTGCTCTTTTCTGTGCTCTCTGGAGGAAAG
CTGAAGGTTTATACTGACATGGAAATGAGTCACGGAAGGGACGGACTGGAGACATTTAGGAAAGAACCCCTCCTGTGTTCCTGGTCACTTACTAGATGGGA
GAAGGTAATGGGAGAGTCTAAAAAGATTAGTGGCCTGACGTATGAGACACTGGCCTCCTTAAATGg

59-M13rev_3

CTTTGAGAAACTGGAGAAGTGTTGCAAAAGTGCAAATATTTTGAATTTCCTGAGATTGAAAGGTTGTAGTTATTTTCTGTGGAGATTTTTTTTTTC
CAAAATGCATAGTGTTCTACCTGAAAGCCATTAAAAATTCATATTTCCTTCTCTGCGTGGAGTTTTCTATGTGTTGCTGAAAATGCTATCCCAGAACACAATGA
AAAAGGGAGAAATGGGCAATCCATCTCAGAGTTCTCATTTAGCCTGCTGTTAATTTGTCTCAGTTTTATATAGAGAGGTAGTAGAATAAACTCTTAGAATATTA
TGAATATAAAGAAAAGCCTTACCGTTTTCCATAAGTGTTCTCCTAATTTTGAAAAAAAAAATTATTTACATCTTTCAGCTCCCAAGCAAGTTGGCATCTCTCAAA
GTCTCTTTTTAAAGAACAGTAGGTGGGTTATGCCTTGAAAAAT

600-M13rev_3

GCCCTTGAGCACATCTGGTCTATAGGACTTGAAGAACCAGCTAACTGAGACTTGAGCANAAATTGCTTTCCTTGCAGTATGTTCTTGTGTAGAT
ACTGCTGCTTCTAAGGTTTTTATCAACTAAAATCAATTTGTCTGACAAGCCTGCCTTTGTAGGACAGAGAGGATTTTCTTTGTGTGGGTTGGTAAAACTTGTAA
TTTTTTTTTTTGGAGTTACCAGAATATTCCAGTTAGTGCAATCCCTNANCAAACGAGGNGCTCCTGATAACCATTGCATTTTCTCTGAATTGAGTTATNAANAT
GCATATTTCCTTGCGG

602-M13rev_1

CTGTCCGGTTCTGTGGCGCACCCTCTCACCTGTTCAGACTAATTTCCTAAGTTCGGCAGAATCCTGGAACCAAGTCCCGGGAAACCTCCCATC
AAATAGGAAAACTAAATAAAAGTATGGACTGCTCACTGAAATGATAAGAATAACTATCAGTATATGAATAGGCCTAGTGTCATAGGATATGTGCTAAAGTAATA
TCACCCTTTCGAAAAAGACTTAAATTCTATTTGTAATTTTCTGTAATTTCTTCTAACCAAGTAAACACCATGCAACTTTTCATAAGTGGCTTGGCACTATATGTC
AACTTTTAAAATTGTCCACCTTAATTCAATAATCTGCTATCCAAGCCGATTCTCTACAAATTATTGGTTTGGTGACCTAGTTTGGTTTCTGTTtCTTGTGATGAAC
ATC

604-M13rev_3

TCCAACGCAAGGAAGGAAACTATGCCCTACAAAAGCAAGAAAGTAATCTTTCCACAAACCTACAAGAAGATAGCCACATGAACAGAATTCAAAC
TCTAACAACAAAATAACAGGAAGCAACAATTAATTTTCTTTAATTTCTCTTAACATCAATGAACTCAATNCCCCAATAAAAAGATATAGAGCAATAGACTGGCTA
CACAAACAGGAACCAACATTTTGCTACTTACAGGAAACCCACCTCAGGGATAAAGACAGACACTACCTCAGAGTAAAAGGCTAGAAAACAATTTTCCAAGCA
AATGGTCCAAAGAAACAAGCTGGAGTAGCCATTCTAATATCAAATAAAATTGACTTCCAACCCAAAGGTTATCcAAAAAAGANnnnnnnnnGGGGTACTTCATAC
TCATCAAGGGTAAAATCTACCAAGATGAA

605-M13rev_4

CGGGAGGAAAGTGGATTTTGAAACTCTCCCAGGGTACAGAGAGAAAATCGGGGAATCCCCTCTTGTGTCTCGATGTCTTCGTTCTAAGTGTCT
GTCTTATGTTTCCGGAAACGAACATGTGCTCGTTGTGTTGGTTTGTTTTCCCCGTCGCCCTGGGGAGTGGGGAGAGCCAGGGATGCCTGGTGGTCTCCAGA
CATCGGTCAGAGGGACAAAGTTTTGTTTATGAAATGGAAGCTTACCCTTCTAATTCTTTTGCCTGCTCGTGGAGGACGAGTAAGGGTTGCACGTGTCTGGAG
CTGTTGGTTTCTGCTTTGTGTATTTTGTGTATGTGTTTAGAGATATGTGAAAATTTAGAAAAGTGCTGGTGTAGCTGCTTCATTTGGTTTTTTTACTGATAACGT
GCTCTGCGTGTCTTCCCA

606-M13rev_1

AATTGGAGCATGAGGGTTAAGGTTCCCTAGGTCTAGAGTGGCGCCTGCGGCGTAGACGAAGCTTAAGAGGATTGTCAGTGGCTGCCACTGTC
CAATTGTCGTCGTGGTTTTCTTCATCTTGACTGGCGGCCCTCTTGACGTGGGAGGCATGGATCCAGGAGGCGATTCCGTCAACTTTTACCACAGTAGGGGT
TGTCAGTAACACCAAATAGGGTCCCTTCCACCGTGGCTCGAGGCTGCCGGTTCGATGCCGTCTGACCAAGACTTTGTCTCCAACTCGGAACCCGTGGGGTA
CTGCGGTGGTCCCTGGGGTATAGGCTGCTTTCAGTTGGTCCCAAATCTGGGTCCTAATCACTTCTAGGGCCTTTAAATGAATAAGCAAAGAGGAGGGTAAG
ACAGGGTCAGAGTCGGAAACTAATGTTCCACCAGCTT

607-M13rev_2

GGCTTCAGGGCCGGAAGTTGTGGTAGCTGGAGATGCCAGAGGCCGCCGAGCTTGCAGATCCTGAGTGGCGAGGCGGGGTGCGTGGCTCGG
GAATGTCGTTTCCTTCGCCCCAGGTTGTGACAGCAGCTCTTCTCGAACCCGAGTACTTCCTGGGTACCAGGTCAGGTCTTGCTACTCGGACCCCGCGCCCG

CATCGCGTCTGGTTGCTATGGAAAACAGCTGTCGCCTTATGAGAGGCGCTGAAAATTGCCGCCCTGGGTTGTAGATGCTCATGGGAATCCCGCGGNTTAAC
CAAACCGAAATAGTCCACCCTGGGAAAAGCGGCTGNTGGAAAAGACTCCTTTGCTTCNCATTTTCTCCCTGCATGAAAANANAANCCTTGGAACC

60-M13rev_4

ACCTGCtGTAACAATCCCTGGATGGAGCAgAGTGTGTCTTCCTCTTGCCTGACTGGTTACATGTGACCTTTTGTCCATATATATATAGCTGAAAG
TAAATGTGTGTGTTTTATGACCATGTAAAGCAATAGCCTCTCCAGACTGTCAGACTTGCTCATTAATCTAGCTGTCCTGGATAATTTTTACATGGATTATTTTAA
TGtTGAAAGCaATCTTGAAAAGCCACAGGGTaCTTCTATCATGCACAtAtCAGNTTTGTCCcATGCAAAACACgTATACaACACACGCGCGCACNCACACACACA

611-M13rev_3

GTGCCGCAGACAGGACCCGGTCCCNACATCTTTGGAGAGGCGCACTGGCTATTGGCAGGCATCTCTTTCCAACCTGACGTCAGCTGAGTGCA
TTCAGCTTCNAAGTAACTGCATGCTTTCTTCCCGCTGACATCATTTCCAGGAATAGACACACCAATTA

612-M13rev_4

CCATCTTTGACTTTTATACAGATTATGCAGGGTGAGGGCTTTCCAACTAGACATTTACACGATTAGTTGACACTTCAGAAATCACAGCTTGCACA
AAATTTTTAGCTAGTGACAAATTTAACTTACTATATATAGTAGTGGTGAACTTGAGTCATTTGCCTTTGACAATGGTCATTTCCCTATACCCAGGGTGCAGTGG
CTGTTGCTTGCTGGTGTCTTAGTTCCTGGAACAAGTTGTGCTTTTTCCAGTTACTAAGTTCACACTAAGGACATGGGAGACATAAGATGGAATCTGAACACAA
AATGGAATTAGTTAGGCGTGGTCCTTCAGTCTGCACATCCAGGAGCATTTGTTCCATCTCCGTCCCTCCTAAGACATGCAACCAATAAAGAGAGTAGAAACT
TAAAGGGTCCTTTGGGAAGTCNAAGTGTGTTGGCTGGC

614-M13rev_1

AAGGGCTGGAGAGGATGAGAAGGCAGGCTCTCTGGCACTTGGGAGGATGCCTGAACAACTGCACATCCCCTAAGCCAGGAGCAGCTGCACC
CAGGCTCTGTGCTTCTGTTTCAGAGCCAAGCTGGCGCCCTGGGCAGTATGCTGTGGAATGACATTATTCTCTGTGACAGAGACACTGAGGAATACTGAATG
TTGTCCACATGGATGGCTATGTGAGTGGAAAAGACCCAGAGCAGATGTGAGCATTATGGAAAGACACCAGTAAGACAACAGACTCTGACCTCAGAACGGTG
ATGCCTG

617-M13rev_3

GCTGGTGCCTTGAGAATGACTTCAGCTTATGGGTCTTAATACTTGAAGTGGACACAGGGCATCAGCAGGATGGATTTCGGGTCCTTGGAGACC
GTGGTGGCCAACTCTGCCTTCATCGCTGCCCGGGGCAGTTTTGACGGAAGCAGCACACCGTCTTCCCGAGACAAGAAATATCTGGCCAAGCTCCGGTTGC
CCCCACTGTCCAAGTGTGAGGGTCTCCGTGACAACCTCAGCTTGGAGTTTGAGAGCCTGTGCTCAGAGCAACCCATCGGCAAGAGGCTCTTCCAACAGTTC
TTGAAGACAGACGAGAGACATGTGCCAGCTCTGGAGCTCTGGAAGGACATTGAGGACTATGACACTGCTGATGATGACCTGAGGCCACAGAA

618-M13rev_4

CATAGAGCCGGTCTCCTTGCTCAGCCAGTAAGGTCACCTGGGCTTCAGTCTTCTGCCGAAGACACACCACCTCTTCTTGGCTGCTTGACAGAG
CTGCTTCTGTGGCCTGAAACCTCCTCTGTTCAGAGCAGAGAGACAGAGGGTGGGATGCAAACAGTTAGAGCAGGTCTTGTGCCAAGATGGGGCACATGGA
CTTCNCAGTCTGGGAAAGAAAGAGGAAGAAAATAGGGGGCGGGGTGGGAGTGTGTGCCCCACGGACTTCAGAGATCTCACAGTGATCTTACACTTGAT
GTAAAAGCAGAAAGGACCTAAGTTAAAATTCCTAGCNNCCCNANGATTTTTA

619-M13rev_1

ACACTTCCAAATTCATTTAGTGCTCAATCCCTTAGAGCTGTAGTGCTGATATCCTCAGGAGCTTCTGTTTTCCAGCTAAAACACACACCGCACA
GCCCCGCCCTCTGACTGGCTCCAGCCTAAACCNGCCACATCTTTCTGATGTCCATAGGAAAAGTTCCTGTAGGTTGAAAATTCGATGCTTAGACTGGACCCN
NNGAGAATATACAGAGCAATCTTTCTATTTTAAGGTCCCCACCAGCAGTGTATTTGCAAAGTCCCCATGCCATGTAAACGCANATTGAGATATCTTGGTGATT
A

620-M13rev_2

AAAATGACCGCCTGTTCTTCGGGCCGTCGCACGCATCCCGGCCCGTGCCGCTGCACGGCTTGGCCAGCGTCTCCCAAATCAACCGCCTGAC
AACAGGCAGCCTTCTGACGTGCGCGCCCGCTTGCGTGGTCGTCATAGGGGCTCAGGTA

621-M13rev_3

AGCCCTGTTCCATTTCATGAAGTATGGTTATTTCATGGGAATGGTGGTACAGGGGCAGATGTACTTATCTGTGAGAAGTATGTGAAGATTTAAA
TTTGGTACTTTAAGCATCTGGTGCTATAGGGTGGTGGGCTGCTCGCATGACAGACCACCTTGTTCCTGCCCAGCCTTTGTGAACAACTGAGCAATTACGTAG
AGAGAGAATGGGATGGCATAGGGAGGGGTTAGTGTGCCACTTTCCTTTAATGGGCTTTAGAAGTGTTTAACAATAAGACTTAGAAATCTTCAAGGACTGAGC
TTTTCTATAAATGAATGATTTATTGTAAGTGTGCCTTCCTAGGCTTTTTAAAGGCCAGAATGTATTATCTCCCCTTCTTTTGTGGTACCTTAGTTTGCCATGCAT
GTTATCTGGG

622-M13rev_4

ACAGGACCAAGGGCCTCTCCTCCCATTGATGACCGACTAGACCATCCTCTGCTATATATGGAGCTGGAACCATGAGCCCCACCATGTATACTC
TTTGGTTTGTGGTTTAGTCCCTGGGAGCTCTGTGGATACTGGTTAGTTCATATTGTTGTTCCTCCTATGGGCTGCAAACCCCTTCAGCTCCTTGGGTCCTTTC
TCTATGTCTTTCATTGGAGACCCTGTGCTCACAGTCCAATGGATGGCTGTGAGCATCCACTTCTGTATTTGTCAGGCACTGGCAGAGACTCTAAGGAGACAG
CTATATCAGGCTCCTGTCAGCAAGCACTTATTGACATCCACAGTAGTGTCTAGGTTTGGTGATTGTATGTGGGATGGATCCCCAGGTGGGGCAGTCTCTGGA
TGGTCATTC

623-M13rev_1

TGTGAGACCCCACTCATGAGTCTCTGGGCTCCTGGGGCCTGGGCAGTGGCACTTCTCTAGCTCTGCCATCCACACTGCTTGTAGCTTGTCCC
CTAGGTTCAGGCTGGCCCCACTCCCCAGCAACTACTGTCCTTGGCATTCATCCAGTTAATACTGGCATCTCCAAAGTATTTGGGTCTCCAGGGTAACTGGGC
TACATATTGTGGGCACTTTTCAGGGACTCTAACTCTGCTGCTTGATGTCAAGCTTCAACTTCTCTCCATGTCCTCTTCAAATTCCTGGGGCTTCTACTTCAAG
CACATTCATCACTGACCTTCCTGCCTCTCCCAGTACAAAGCCTCAGCTACTGCAACACACCTTCAACATTCACACCACCTGGGAACCTCACATTGCCGAGTT
TGGCT

624-M13rev_2

TCGATCCTGGTCAGCCAGGGCTATTCGCCGATCGCCGCAGGCTCGCTGCACGGGGTGCTGCAACTGGCCACCGCCCTGCCGGGGCTGCTG
CTGCCGCCGTTGACCGGGCGGATGAAGGATCAGCGCGTCATCGCCGCCGGTCTGGGCGTGATGGGGCTGGCCGCCATACTGGGCCTGTATCTGGCGCC
GACCCTGGCGGTGGTGTGGATCGCCCTGTTCGGCGCAAGCATAGGCGGCGCCTTCATCCTGGCCCCGACCTTCATCGCGCTCCGGGGCGCGCAACGCCC
ACCAGACGGCCCGTCTGTCCGGCATGGTTCAAGCGGGTCGGTTATCTGAATGTCGGC

628-M13rev_1

GTAGTGTGCACAGTACCGTTTATGTGTAGTGTGGAGAGCCAAGAGAGCCAAGACAAGAGGCTGGGGCTTTAGGGTCCCAATAAATGAATGCA
AAATGATTAGAATGTTTGGAGAGTTAGTAGAAGTCAAGGGTCATTGTCTCCAGGTTTATTTGTGCAGACTTGTCTCGGAACCGTGCCTGCTGACACCTCTTCC
CTGAGTGCAGGATGGAAGGGAAATCTGTGGGCTTCATTAGGAAGACCAGGAGGGTGGAGGCTGCTTTCGCTTCAATGTCTTCAGCTCAACACATTCTAATG
TGACACTGGCCTACTTGGAATCTGACGTAGCCTGACTCCCCTCATAAGCATGAGGAAGGGAGGGAGCACCTCAGACTTGAGAGC

62-M13rev_1

CAGGATCCAGCTTTGAGTTTATGCCCTGACTTCCCTTTATGATAAACTGTGAGTAGTAGGTGAAGTAAACTCCTTCTCGGTCAGTTGGTCATTTT
ATTCCACAGTGTTTTTCACAGCAGTGGGAAGCAAATTAGTTCATCAAGTTTTTCATTTGATTTTCATTTCACTTCGGTGGCTTAAAAATTATTAGAATGTTGGTC
TGGTGAGATGGCTCAGCAGGCTAAGGTGCTTGCCACGCAAACCGGGTGACCTGAGCTAGAGCTCAGTAATTCATGTGCAGGGGGAAGGGGAGAATTGTCA
GATGCCAGAGTTGTCTCTGACCTCTGCATGCATGCCGTGACGAGCGTGGATCATAGACACAAAAGTAATACATAGAACTTCAAAGGCTTTGGCATGTAAGAA
CCTTAACATTTACTTCTGCAGCCAGAAAACACACACA

630-M13rev_2

CACCCAGCGGCCGTGCCCTTTGACTTGACCTCTGATCCTGCCAGCGGGTAGGCTGCAGGACCAAGGGGCGTCCTGAACTCGCGCACACCCG
CCAGGCACTTCTGTTCGGTTGGGCTCTCCTGCCCCCATCTGGGAAGGGTCCATAAAAGCCTTTCTCTTCCTTCCCTGCTCAAAACCACTAATTCATAGGACT
CCAGGGGTTCTGGGTTGAATACTGAATCAGTATTTCCTCACATCCCCAAGGATCAAACAACGCCAGCCGCGGAAAAAAAGGAATCTCCTTGAACCACACTTT
GGGTTAGTTCACCTTCCTCCTCTCTTTCCTTTTTTCCCTTTTAAACCAAGACCGAGTCGAGTGGTCAAGAAGAAAGCTGGGGACAACTTAGTTGGCTAACCTG
AGAAAAACCGGCGGCTGTCACTAATGAAAATGCTGTGTAATAGAAGACAGAGACAGTCACCACAATCCTGCGTC

633-M13rev_4

GGTTTGTTtATTTGCTCGTCAGCTTCTGTCCTGgAGCCTCANCCTGTCCCATGGNTCATTTTCCTCAACANGTGCCAGCACCTGCCCCAGAGTC
AGCCTCANAACAAAGAGGCTTTTAAGCACGTCTCATATGTGCTATGTAGGCAAACATTACATCACTCCAGGTTACTTTGCTTGCATCCCAGTAGCAAAATGAG
CTCATTGTTGCTTTGCTGTCTTGGTTTCCATGACNCTGCATGACCGCATAGGTCCGCTCCTACATAGCCATCCTGCAGGCTTTGCCCCANGTCATAAAATGG
AGTGCTCCTTGCACATCAACCATCAACACAGAAAGTGTGAGTTGAANAAGTTTATCTCACATGNGCTATAAATGNTGGCTCTACCNACAAGTATTGCCAAATA
CTTNTATGAANGTAC

634-M13rev_1

CNTGNAATTAATTATTTCTATCACcACCAACAGACACACACAAGCCAAGGAAACAAACGACTTCCCTATGAAGCAAGGGCTGAGGGCTGATAAT
AAAAATGGCAACTGAATGACAGCAAACTGTTTCTCCCCCCTGAGCTTGCCTGATAATTCCTTTTTTTTTCCTCCCTGAATAACTAAAGGAGAGGCAGCTGCCAC
AGCAATCAGGATTCAAAACCCCATTAATATTGGTGACTCATCAGCCGGATTCTTTCTCTGTGTCTGGAAATGTCAGGATATTAAAAAAAGCTTTGTGGCATCT
CTANGAGTCAAATCCCCAGTCAGAGCTGTAACCACTTCAAAATATTCTTGGCCTCGTTTGGAAAATGGCTGTCACCTGGGACTTGGCCTGGTTTGTTTCTGT
CAGTGAAGGGCGAAN

637-M13rev_3

GCTGCACTGGCACGAGTCTTCAGGAGGGATTGGGATCAGCAGGGATGCTAGAAGTTCTCCTGTGTTCCTCTCTCGACATGCTTGTCATTAGTA
AAGAGCCCTTCATCACACGCCCTTTCATGTGCTTGTTTTAGCAGAACATCCTTCCACGTGTCTGCTTCGGCAAAGCTTTCCTTCACTAGTCTGCCTTAGCGAA
ATGTCCTTTCACCTGTGTGTGCTTTCCTTCATGTGTCTGCTTAGCAAAACATCCTTTCAGCTGTGTCTGCTTCAGGAAAACATTCCTTCACATGTTTGCCCCAG
CAAAACACCATCCGACACAACTGACTTTCCAAAGAACCCTTAACTTTCCATGTCAGAAGTCCATCTATGTAATGATAGGAAAGGCATTCTCCGTGCTGGGCAA
AGATATTTCAGTGTGGTTGCTTTGGAGTCACTCATGCTATGGTCAGTAGCCCTAAGCCATTCTTTTAAGTAAGCCCA

638-M13rev_4

GTGAGCACACACATAGCACAGCATCCATATGGAAACCAGAGGACAAGTCTTGGGAGTCGTTTACTCTATCCCCTGTGAGATCTTGGGCCATCA
GACTTACATGGCAAGCTCTCTAACCCACTGAACCATGTCATGTCCCCTTTATACTTTCATGTCTGTATGTCTATTTGGTATGTACACATACAAAACATTCGCTA
GCTGTGGATAGTGATGGAGTTTTTAGAAGGACAGGCACATGTTTATGTGGCGGATCCAGGATTGTTGGGGCTGGCCTGTGACTCTCATGTCCGGGTTCGAG
CCTGGGAGGCATCTTGGAGCTGAAAGAGAAGAGGGAATCTAGGGGGGTAGAGAAATAATAGAACCAAGACAAGCTAGTCAGCTCTTGCACCCCACGAGGC
TAGTCTCATTGCACGGGATAGAGTGAGTTGCTCAGCAGCCCGAGAGAGTTGCACGGCTAAGCACTGCAGTAGAAGGGCTCTGCGGCAT

639-M13rev_1

CGTCAGCGGCCGGGGCGCAGCGTCATGGCCGAGGCGGGGGCCAGTTGCAGGCCGCGCTGCTCAAGCCCGGCGGTCCAGCGGGCGGCGG
CCTCGACCGTCACCGGATAGGAGAAGCCGGTCCCCAGGGCCGAGCCCCGAACCTTGGCCTGGGCCTCCAGACTGTTCAGGGCGGCGATGATTGCGGCG

GGCGACTGGGTCTGGTCGATGATGCGGTCGGCGCTGGCGCGCGCCCAGGCGCCGGGGCGGCGCTGGAAGGTCCCGTCGTCGACGAAGGCGACGCCGC
GCTGGCGCAGATAGGACAGATAGGCGCTGGACGCCGTGTCGGAGGCCACGAACTGGTCGCCCAGATAGTTGGTCACGCCGAAATAGCCGGTGGCGCGG
CCCAGCAGCCAGTCCATCTTGGCCTGGACATCGTCGGGCGTGCTGTTGGCCAAGAGGGTGTAGGGGCCGGGG

64-M13rev_2

CTCCTTGCCTGCCTTAGCTCCTGGACTTTGCTTTCAATTAACGGACCTTATTCAGCGCCGCTTCAGGTTTAAAGGAGGAGTGAACAGTGAGTCC
TCTCGCCCCTCCCCACCCATCTCCGGTTCACATTTGGTGCATGTGTGTGTTTGCCAGTTGCTTAACACATTCATGTTTCCCCTTGGAACTTACCATTCACTCC
AGGCTGGTCTGTCACTGTGGGACAGACACACATGACACACGACACACACCACCATCTCCTGCCCACCCATGCTCCTGTGCAGTACACAACTTTGGATATAA
ACAAAGTAAGGGGAGCTGTTAGGTAGGCACAGTGAGTGGACAAATCAGAGAACCACAATCTGCCTCTGGGTGACAGTGACCGTTGTGCAAGCGTTGGGTG
TGGTAATGGAACCTCAGGCCAGGCTCCGTGTCCCAGAACTCGATCC

70-M13rev_3

CCTTGGCGTCGACCATGGCCTTCTGCCACTGCTCGGCGGCATCAGGCTGGCTGACCATGACCGTCTCGATCGTCATGGCGCGACGTCCGAC
GTTGAAGGTGTACTGGATCGAAGGCTCCACCGCATTGCCGCACATCATCAGGCAGTCGGCGGTCAGGACGCCGTCCTGAAAATTGTCGAAGATCTCGGCG
AAATGGATGATGTTGGCCGTGCCGATGGCGTCGCTGCCGACGACGGCGTCCACATTGAAGGTGCAGCGGGGCGAGGTTTCCATCGCCAAGGCCTCGACC
TTGCCGCCCGTCACCAGGGCCTGGACGCTGCATCCTTCGACCTCATAATAGCGCACTGAGGCGGCGTTCTGGCTCGCCCGCCCTGTCACCCGCTCCAGCA
CCGATGCATCCGCGCCCAGCATGGCGGAGGTGAAGACACGGTCGATCTTGGACGCCGCCC

711-M13rev_4

GCTTAACCGTTTCTTTACACTCAAGGAGGCGCTGCaCAGCCTTGGTGCCATGTAAAGCCAGTGAGATCGGGCAGACAGCCTGACTGGGCTAC
TGGTCCTTGCTTTGATTTTCCTCTTGTTAATTTATTTGTTAGCTTTGGTTTTGTTTCCCAGCATGGTCTATACATTTTCCATTGCCTTACACTTCTATGTAAGTTC
TAGAGCTCATCACACCCAACAGTTGATAATGCTGCAGTTCAGTAAGTGTTAAACTCCTGAGTCCTCCGGTGGCTTCTTAATTCTCTAACCTTTCCAAATTCAAT
TGTTTTCAGGAGCTCCTAAGACAACATGATGAAGACATCCTGCAAATACAGGAGCTGAGAAAATCCAAAGAGGTAAGGGTTCCCATTCAGCTCTATGAAAGT
AGTGTCTT

717-M13rev_4

GAAACTCCAGCTTATACTGCTGACANANTCTAGCAGAGGAAAGCACTTGCCGGACTCTGTGATTGGCTCCCTGATCCGTGCTATACACNAACC
CCTCTGTGGCAGNCAGTCTTCCTCAATTGCNTGGCCGAATTACATACAGATTTAGGTCCTGAGACTGACCTGTGGCTTCATGTGACCTCCCCATGACACACC
TGGACAGAGCAGCNAGGGGACNGGNACAAGTGGCTTTTATCAGACAGGAGATGGGAATAGACCAACGGGAGGCCGAACAGGCGCTCA

718-M13rev_1

CCCAGTCCCCACTATTTTATTAAAGTGCTTTTGTATGTCTTCTCTGACTTTCTCTACTGTATTCTACTCACAGAATGCAGACTGAGGTATGGTTTG
ATGGAAGATGCCCATTCTCTCTTTAACAAATTTCCGGTTTTGCCAGCTTGTTTTGTTTGAAGCATGCAACCTTGAACAAACCGTTTACATTCCAAAATATTTTAT
CACTGCATTTCAGATATATGCCTTGCCTACTTTTTGTGATTATTGAAAGCAGATATGCAATCAAGTGCTTGAGAAAGTAAGGTGTTTATCAAATTGCTTTGTTTT
ATGTTTATGCAGCCTTTCAAAGCAGTCTGTAAACTCTTGAGGATAAGTCTATCCCGTGTGTTTATACACTTCATCACGGGCCATGGCGGTAAGAGCCATTGG
CGTGCGT

719-M13rev_2

AAAACAATGACCGCATATATGACATCTGCTGGAGAAACACAAAGTATGCCTCTATTCACACACAGATTGGCACCATTCATATTGTCTCATCCAGA
ATCTCCTGTCAGCTGTTTAAGCAACACTACCTTCAAGAAGAATATAATCCATACTCCAGAAGAGGTGGAAATCCTCACAATAGCCAGTAAACCCCTATCACAT
CAgAAGGTAGTTTGGGTTTTGGCTTGTCATTCCAACCATTCCTAGACAAGCATTAAGCCTTTTTAAGAGTCTGGTGTAGTTGTCCCTGGAAGGAGAGTTTGGG
CCAAAGTTCAGTAACCTGTGTAACCTTTCAGAGAAAGCCTGCTGTTTCGAGAATTTTGTGCAAAGCCCAGTATACCCTACTTTGTCTTGAGGCAACAGGTGGg
CATAaATTCAATGTTAGTATT

720-M13rev_3

TAGGTTTGTGAAACCCCTCCTATTTAAACACGACAGAAATAAACACCCTCTTCTGTAGACCTGTGGGTCACCAGCCCTTTGGCAAACCTCTATC
TCTAAAAATGTTTACATTACAATGAATTCACAACCGTAGCAAAATTATAGGCATGAAATAGCAACGAAAATAATTTTATTGTTGGGGGTATTTAACTGTATTAAA
GGGTGGTACCATTAGGCAGGTTGAGAGCCATGGATCTAGGGGAATCCTCTGGCTTCCTGCAATAATGACCTTGGCCTGCTGGGAGGCCAGTTTATAGTCCC
TGCTCCCACCCCCACCCCGACCTCTGCCCTCCCAAGGCTGCTTCCCTTAGCTCTGACTATGTGAACTCAAGGCTCACTTTCCTAGCAAGCTCCAATCCATGT
GACAAGACTCAGGGAGTC

721-M13rev_4

GAGTACACATGAGAACAGCGGCCTCGACCTAACACGCCCCCTGCCTGACGCGCAGCCTAGGAACCTAAACCTAAGCCGGCTACACACAAGCT
TAGGCGCTGATCTGGAGAACAGCCTGGGTGCGGACAACCCTCCCGCGCAGACTGGTCCATGGCTTTTTCTGCCAGCGCTTACCTGAACCACTTTAGTGAAC
TCCTCATACACCCGCATTTTTCAGGTGCGCCGCCATGGCTGCACCGGACCTTCTCGGCtTCTGTACTTGCTGAGAACGTGGAGAGAAAC

722-M13rev_1

CCCTACCAGACAATGTTCCTGCAGAACCTGTTACCTTACTTGGCAGGGATTTTTGTCAAGGGATCTACTACTGTGCCTAGTATGGTAACTACCA
CAGTGATATATCCAGCAACAAAAACCCATTATGCCCAGCAGACTCTTCAGCTCAAAACAGTGAGTCAAGAGGAGCTTGCACCTAAGGGGCTATTTCTGAATC
TCGACAACTGAGAATCATTAGTTAATAGGATCCAACACAGAGCTTTATGACTGCCCACAGACAGGTGAAAGCACACTAAAGACAGATCCCAATAGTCTCTCAT
CCAGAATCCTAGTACCCCATTATAAAAGCACTCAGCTCTCCACATGGACCAGTGCCTCCTTTCAATAAGAAGGGCCCAGAGATCTTTTTCTCAGAGAATCTTG
GGAAACACTCCC

728-M13rev_4

ACCACGGaaACATGAACATATGGAGGCTGGTTATTAAGTGCTGGCTTCTCAAAAGATAAGGAGTCAGGGCTGCCTTCGCCACTTACCCAGATT
CCTAACCTATCTGTTTACTGTGATGCTGGTTGTAAGAATTACAACCCCATGTGTTTCATATGTGATGGCCTGGCCAGCTCAATAATAATTGTCAACACTAATG
GGATGTCAAACCAAACAATGTGGTTTTTCCTAAGAAACCCGGAAATGGCAAATACAGtTTTCAGTTTAGCAAGGGGGGAAAAAAGGTTCAATTTGCAGTTTCA
AACATGAAgGGTTGGGATTCCTTAAAGCCTTGGGCTGACCTGGCATTGCTTGAGTTTGTTCGCCCAGACTCTATACTGACcCAGGGAAGGACCTCTCACAGT
A

729-M13rev_1

CCCAGTCAAGGAGAAGGTACTTTGAAATCTCTTTACTCAGCCTGCCAATGACACTAAACATGTCTGAATCTAGCTGTGTTTCCAGATCCAGTGC
ATGGCCCACAGCAACATGGTGAGATTCCACACCTCTCCCTCACTGGTCTCATGGCTATGGAAATCATTATGAATACTTAATGTTCCAAGTAATGAAAGTAAAC
TTCTATAGAGAACCATGGTTCCTGAGTAACATTCGGAGCTGAAAAGAAGTGATCCTACCCAGCCCAGGGGGAACGACCTACTATTAACCCTAGCTTCTGGCT
CCAGCTGGaAAGCTGGCATGCATGGGCAGGAAGCTATTCTATGTAAACTAGGACCATTGCTGAGAAGTCAGCTCTGCCCACCCTCAACATTAGCACCCAGT
CCAATCCTTTCTAGGAC

731-M13rev_3

AAGATNTGAATTCGTGTGCGCCACCATCGACCAGTGCCCCANTGCNCATCGAGGAGACCTCGCCCATCGTCCTGCGCTACAAGACNCCTTAC
TTCAAAGCCTCGGCCCGTGTGGTCATGCCCCCCATCCCCCGCCACGAGACCTGGGTGGTGGGCTGGATCCAGGNATGCAACCAGATGGAGTTTTTCAACA
CCTACANCGACC

732-M13rev_4

CGGGATCTGAATTCTCGCTGACCCGTTGCGTCCCGGCTTCCCTGGGGGGGACCCGGCGTCTGTGGGCTGTGCGNCCCGGGGGTtGcGTGTG
ANNAANATCCTCCACCCCCGCCGCCCTCCCCTCCCGCCGGCCTCTCGGGGACCCCCTGAGA

739-M13rev_2

GAACTAGGACACTGAACCTTCATGCAATCAGACGTGCCTTCCACAGGATGCCAACAGCAAGCTATCCTCTGtGCgGGGCTTAGCCTCGCACCC
CACGGCATAACGCANCATAATTTCTTTTAGAGCGGCcAACCGAATCTGAGGAGATTGTCCCGCCTCCACTGCAGCATANTTCTANGCTACCATGGCGAAAGT
GTTNTCCCCCACACTCTGCACCTAACACGNGTCAAACACANAAGANACAC

73-M13rev_4

ANnTGAATTCCCTCACTGGATACATCCTCTTATAATGTGTCCAGGAGAGTAGGGCAGGAGGCTCAACTGGATCTTTGGCTGGGATTCCCAGCC
CTGAGTAGTGTGATTGTCTAAGAGAATTTTGTGGGGTGGGGAAGAACTACTATAACATGGAAATCCTTTTACATTTGGACTCATGTCTGATGTTCCTCTATTTC
AGAAAATGTTCCTCTATTTCAGAAAATGAGAGACACTTCTAAATTACCTTATCAATGGTTCCTTTACCAATTATGTAAATTCCCCTTCTAAGCTACCCCAAAATT
TCTTGGACAAGTTATGGGAATCTAAACTCTCTTTTCATCCTTCAACATCCNCTCAGTGTTATACTCAGTGTTGCAATACTTAGGTTAACAATCCATTTGATTTCA
GTAGGAGATATCTGTGAGTTCAGAGAACTTATGAGTCTTTTAAACACTCAACCACCAAAA

740-M13rev_3

TGGGGTGGGGGGTGGGGCTCCTGGTACACAGTCAGTAAAAAGGCTGACTGGTCAGGAACTTGCAACTGGCACCTAAAAGGGAAGTTGTATAA
GGCTGAGCTCCTTAACTGGTCAGATCTGACAGTAACTCAGAACCTAGTGTCACAAGTGAATTACATTGTTGGACACCCAACTTGTGTTCAAGATTTGTTAATA
ATAGAGAAAATTCCCCAGAGACATTCATAACTTCACCTTGGGAAAGACTCAGTCACTGGGCCCCTACTATAGATAGCTCCAAAATTTCCGAAGCAGtAGCATG
CCTCATTTTCCAGGATCCAGTTCATTTTTTTAAATCCCTTTGTCAATGGGAGACAAAGCTAAATACACTCTCCTAAAAGCTATTACATCAT

742-M13rev_1

NAAATGAAGATATTATGAATAGACCTCAGTCTAACTATTAGACTATTGAAACAATGAATCTCAGAAATGAAACTTTCTCCAAGAATCTAATGCTGG
GAGGAGGAAAGGCCACCAGCCTGAGGGGGACTTAGCTCCTGGAAAGACACTTTACCCGGGGTTAACAATGAAACCTTATGGGATTCTACGGATGATGGAAT
GGGCCAACGGCAGCAGGACCACACCGTGGCTGAGACTGCTCAGAGTGGATACCTCTAGACTTCTATTTAAACGTTCATCTTGGACTGGAGGGATGGCTCAG
TAGTTAAGAGCACTTGTTCTTGCAGAGGACCCTGGCTCAAGTCCCAGCACACATGTGGAGGCTCATAACCATTGATTCTTCTGATCTCTGTGGGCCCCAGAC
A

76-M13rev_3

TCTTGGtGGTATACGATCCCTAAGAAAGTGGCTGAATTTGGTATTTAAATTAAACCATGAATTGTCTCCCTTGTGCTCCTTGGTCATCCCTTTCCC
CAGCAATTAACATAAATTTCTGGCAAACAAAAGAGCTNATGGAAGTGGGATAAGAAAAATGCACATTTGCTTGCTGGCCTTTCTGCAGTGGCATGGGGCTGT
GGAGTGAAAGGAATGTGGCTTTGCCTTCCCTCCACTTAGGAATCCTGTGCTGTCTGTGTGTCTGGAGACTTTAATAGCTGACATATTCTCGGAAGGAAAATA
GCATGCTTCTCATGGCGATGATGCTGTGAATCTCAGCCTCTCCACTCTGTTTTCATTAGAGAGTTGCCATAGATATTTCATTAAAAACCAAAGAAAGCTTTCCA
CATTAGTTTGAAACTATAGAAAATCC

77-M13rev_4

ATGAGGCATTAGTATGCTTAGTTCAATATATAAAGACTCTATCCAAAGAAGTGTCCAAAACATGGTAGAGTTTGTTCATGTGGATGTCAGAAAGA
TACCCTTTCTCCGACTTCCTAATGAATGGAATGAAGGGAAGAATCTACAAGAAAGGTAGAGGCAAGTATGAAGCAGACCTTGTACTTGCACACTTGCAAACC
TTTGCTAGATGCTGTAGGTTGCACAAACTGCCTTGCTAACATCTCATGAATACTCACAAGATCAACTAGCATCCATGAGGTGTTGAAATTGTCTATTGGAACC
AGGGAAGTGTGGGTTGAAAACACAAATTAGGTGCCGTACTCTGGACAGTGATTCTGATTAGGAATACTTCCAACCCTTAGAGAATCATTTGGACATGACACT
ATGGCATGAAGCCAATGGTCCCTAAGCACCATA

78-M13rev_1

GCTGCTATGGATTGAAAATATTTACAGAAATGAAGGCAAATAATGTCTGTACCAAACATTTCCAAGCAACATGTTTCCTTTGTTATTACTCCCTAA
ATAATANTACCATGTAGCATATTTACATAGCACTTGCACCTAGAGCCATATTAAAGNTTGCAGGAATATGCACGAAGGTTACAGCAGATATTCNCCCTTTTGTT
GGGTTGTTTTGGGTTTTCAGGACAAGGTCTTTGGCATTCCCCACCAGCCGGTAACTTGAATGGCCTTGAATCCCACAGGCACTCTTGCCTCTGCTGGCTAG
GCCCTTTTCATGAGGGACCTCAGGCTCTTTAACTTTTTGGTGGCTCCTGGTAAAGGAATGGGCCAATTTTCCCCCTGCCCAATATCCGTTAAATCCcGAAGA
GCAACTGTTGATAA

7-M13revfor

TCTGAATTCCCGATGAAGCAGTGGGACCCGGGCACTTCTTAGGGCTAAGAGACTGACAGCAGTCCATAGCAGAGGCGAATATATTTTATAGTA
CAGAAAATTGGGAGATTCCAGCAGTCTGAAGTTTTATAAGCTTAGGTAGGTAACTGTGTCAGCAGAATCATGGCTGCTACATAAGCTCTTAGTGCTGCTTGCT
TGGCTCTGACACTGTCACGGGAGGGACCAGGCCTCAGAGGGACCGTCTGGGAAGAAAGTGTAGAATTGTACAAACAGAGTGGAATTGTACATCCAAGGTG
AATTGGTGGTCAGTGGAGGGTTTCTGTTCTTGTTTTGTTTTGTTTGCAGTGCTAGGTGTCAAACCTAGGGCTGTGAACGGTCAAGGTGAGTGTTGTACTTCT
GATCTCGGCCTTACCCCACCCCGAGGAAAGGTTGACGCTGAGGAGGTGGAGGCTCTGATGGTTTTCCCTCCGAGCTCAGGGAGCGAGTGGTGACGATGG
CTTCGTGGTGCCCGTCGCTCCTCTCCACGCACATTGTAGTTGCTGTGTAGTGGACAGNTGCTGCTAGGCCGTCTTTGGTTTTAAAATTTGCACA

809-M13rev_2

CTGCATATGCTGAAGTCTTAATGCTTTTTTGTCCCTTCAAAGTTGCTGTGCTAAAATACAAACCCCCATGGTTATAGGTTAAGAGGTGGAGCCTT
GGGGCAGTGATCAGCCCACAGGAACACAGGCCTCATGGATGAGATCAACAAGAGGGAGATTTCAGACCTAGGTCTACTGATTCTGCGCTGTCTGCCACGTA
GTCATCATAGAAGGATACTATATAAGACTACCTATAAACTAGGACGCAAGTCCTCTCCAGACAGTAAGCCTGCTAAGGTCTTGATCTTGGGCTACCCAGCCT
CCAGAACAGCGGAAAAACACATTTCTATTGTTANAAGTTGCCTGCCTCTGGTGGTTTATTATAGCAGCCTGAATGGACTGAGACAGTGGCCTGTCTGTGTTA
GTTGTAAGTCCTGTTCAATAGAGAAGCCTCCATTACAGCACAATTGGGCTTTTTCAAGTGTGGTGAG

80-M13rev_2

CGGAGAGTACTCTTTACTGCTTTATTCCCTATGGCTTGCTCAGCCTACATGGACTGGGCCTTCACATATCAATCACTAATTAAGAAAACAGCATA
CGGCTTGATCTTATAGGGGCATTTTCTCAATTGACATTCCCTCCTTTCAGAGAACTCCAGTTTGTATGTCAAGTTGACACAAGATTAGCCAGGACAGTAGTAT
GTTTTATGATACATGTGTTCAAGACTTTCTTTAGGGGCTTAGAGAGATGAGTAAGTAGAGAGGATCTGAGTTCCTATTCTCAGGACCCACATAAAAAGTCAA
GTGTAGTTTCGTGTGTGCCTGTAACTCTCTAGTTTCGTTCAGGGAAGAAGTCTGTGAATAGAGGATCTTTGTGGTTTGTTACTACAAACCTAGCTCCAGGTTT
ACTGAGAAACCCTGTCTT

814-M13rev_3

GAACCTACTTCCAGAGCCCCAGTCACATCTAGGACCACCCTACATGGACTTGACCTGCAGCACCAGTGtCTGCTATCACCCAGTGACACCACA
TTCCTGAAGTTCTCCAGCATCACGTCACAGTACAGCTTCCTCCCAGCAGCGTCCAGCAACCCCAGTTCTTCCTCATTGAAGACCTATGGTCATGTCCTTGAN
nGTCGCCATCTCCTACAACACCAAACACACAACTTCAACCTTGGTCTTCTGCAAGAAGGGTGAAGAAATCCAGGAAGGATGAGCAGGAAGCAACTGAGCTAA
GAACCCAGGGTCCATGTTTCTGTATGAGTCACCCACTTTTTGCTACTGCTCTGTTTTCGTTTTTGAGACAGGGTCTTGCTATCTAGCCTTGGTTGGTCAGAAA
CTTGCTATGTACACCAGGCTAAGCCTCAACTTGTACTAATCTTCCCACCTCTGCCCCTGCC

826-M13rev_4

AGGGGNNGGAATTCNNCCTTGCGCNANAAAAAGGAGATCNGACTTCCACNGTAGGTGCTTTTACCCTATCANCCNNTTCACTGGCTCCTGAGT
ATTTTAATAACACCGTTTAATATTGAACTCTGCTTTCTTTGACGTGTTGATTCTCATGTTAGCAGGCAGTTTAATTACTGCCTGCATATTTTAAAGCCGAGTAGA
TATAATGGTATACTTTGCTAATGGAGAGTCTATGCTGTTTCTAAAGCCTCCCCCCCTACACACACACACACACACACACACACAAACTGGTAGGATTTA
ACTTAGTTTCATGTTTCCCTCTGGGTTTGCCNGGTTTGTTTACAGGCTTTGCTGAGATGAGTCTACTGTAGCCCACACTTTGTTTGGACATGGTTTCCATCGC
TGAAGCCAGCTTTTTTGGCTTACCAGCCAAATGCCAGTAGTAGTAACTACAAGGTCTCCATAGATGGGCTTGGCTAGCCTAGCTCTCTTGTAATCCTTG

831-M13rev_3

TTAGAGATGATCCTGTCAGCAGCAGGTCTTGGCTTCTGCAGTTAGTCACTGGCTTTAGCTGCCCAGAGGAGTTTACAAAGATCACAAGATAACT
GCACCGTTATCTTGGTGGTTGTGTTACTCTGACGGTATCGCAGCCACGCCAGTTCCATGAACCTCCCGAGTAAACCAGTGCTACGTCTGGGCTCACCAGGA
AGTTGATGCTTTTTTTATTTGTTTGTGCTCACAGATTCATTGCAACATTTCGATGTTGAGATTAAGGCCGGCTTCCAGGCTTCCAGGTATAGAATTGCATTACC
CAGCAAAATAAGCACTAGCTGCGAGTGATTATTTCAATTTTAACTAATTCTTTTAAATAACATTAAACATTTGGTTCCCACTTTTGCATTGCATGTGGTTAGTGT
ATTGAATAGCTCAGGTATGGAATGTTTCATCATTGGAGAGATGTCTGCTAGACAAAGAC

832-M13rev_4

CTGGCACCTAGCATGAAGCAAAAGGTTTGGGATGCTGTCAGATCACACTTGATTAAGTGCATGGAATTCATAAGAAAAATTGAGAGTTAAGGAT
TCAGTCTAAAATAACCACACGGGGCTCACCCTAAGACTAGATTTAAATAGGTAACACTTGAATGGGTAGCAATTAGGGGAGAGACTAAAGAGTAAGGCATGT
GGTATCAGCCACAAATAATTTTGTACAGAACGTCATTGATTCTGTTCCTTCCCCTATGTGAGCTGCTCTGATTCTATGATTTATAANCAATAATATGCTCACTG
CTATAGAAGACACAGCTCCTGACAGCTGCTTACTAGTGCATTGGGTTTTCTGTCTGCTTCTGCTTTTTGTTCTTTTTGGCCAAAGAAAATAGTTGTTATAGTCT
CTGATTCCTGTGACTGTTATTGTTACCATTCATCTCTCCAATGAGTTTTACA

836-M13rev_2

CGAAGCCTGAAAGTGACCTAGGCCCTCTTCCTGACAGCAGTAGAGGGAATCCTCTGGAGAGAGGCCATTGGGTTGGGTGGCTACTGGATGG
GGAGGTGGATGGGCAGGCACAGGGGATGGGTATGCCCTGGGCTGAAAATGACTGGCTGATCGCTGAGGAGTGTATGGAGACCCTGTCTGCAGCATCACA
CTGTTGGGGTGGGAAAGCAGATGAACCAAAAGTCCTTCCTGAACTCATAGATGGTGGTGGANTTCCATAAANATAGTCACTGGTTGACAGTGAGTTCTGCCN
CTTGGTANCTGCATTATGGTTGTCGAGATCTACAAATTCTTTGGGACTCTCTGGCCNCTCCACNGATTCATCCATCATTTTCTTCTTCTNGNACCAGAATTGT
GTNCNATNAGGAAGAAACAACTTTATGAAATTNAGATCTCCC

838-M13rev_3

TTTGGATTTCTTCAGGATCCGTTGTTATGTCTCCCTTCTCATCTCTGATTTTGTTAATTAGGATGCTTTCCCTGTGCCCTCTAGTGAGTCTGGCC
AAGAGTTTATCTATCTTGTTGATTTTCTCAAAGAACAAGCTCCTTAATGATCTGAATGGTCCTATATCCCCTAAAGAAATAGAAGCAATCATTAATAGTCTCCCA
ACCAAAAAAAAGCCCAGGACCAGATGGGTTTAGTGCAGAGTTCTATAAGACCTTCAAAGAAGATCTAATTCCAGTTCTTCACAAACTATTCCACAAAATAGAAA
CAGAAGGTACTCTACCCAACTCATTCTATGAAGCCACAATTACTCTGATACCTAAACCACAAAAAGACCCAACAAAGATAGAGAACTTCAGACCAATTTCCCT
TATGAATATCGATGCAAAAATCCTCAATAAAGTTCTTGCTAACCGAATCCAGGAACA

839-M13rev_4

TGTGTGTGTGTATGTGTGTGTGTGCATGTGCGTGTGTGCGTGTGTGTGTGTGCGCGCGCGTGTGTGCATGTGTGTGCACGTGTGCGCGTGAG
CGCGCGTGTGTGTGTATGTGTGTGTGCGTGTGCGTGTGTGCGTGCGTGTGTGTGTGTGCGCGCATGTGTGCGTGTGTGCGTGTGTGTGTG
TGTGTGTGTATGTGTGTATGTGCGTGTGTGTGTGTGCGCGCGTGTGTGTGGGTGTGTGTGTGTCTGTGTGCACTTGGTGTACATAGAGCTCAAAAGGAC
AACTTGGGAGGGTTGATTTACTCCTGCCATGTGGGTCTTAAGAATTGAACTCTGATCCTGAGGCTTGGTGGCAATCTCCATTACCCACTGGGCCCTCTCATT
GGCCCCANTAATCTTGTTCCTTCCTCTCCTGCTTGCGTGTGCGTNTTTCCNAATCAAGA

843-M13rev_2

GGCTCTGCTGTCAGCAAAACATGTAATTTAACTCTGTTTCTGCCACTGCTACAGCAGCAACTATGCATCTCCCTGGCTGAGCCTAGAAAATGCT
TCCTTAGGCAATTGTTCTCAAGTATGGAAAGACTAGGGCAGCCTTCGATGGATGGGTCTTGCTATGAAGATGCTTTTCCTATTGTCCAATGGCAAAGTGGGC
AGCATCTTGTTAATGGGTCCGATCTTGAACAACTCCAGCTGTCGGCACCCAGCCTGTCTGCCACTTTAGCCACGCTCATGCTTCTCTCCTCACCAAAATACA
TGTTTTTCACATTTCCCTGACTTCAATCCTCACTCTAAAGCCTTGTCAGCCCAGTGAGCATTAGCCATGCCACCGCCTGAAATAC

854-M13rev_1

CTGTCAGAATCTAAAGCTCACTGCCCAAACACGATGCCAGCCTGGCAGCTGAGAGAGGGAGCCCAGCTGTGGAGCTCCGGAGGGTTTTCCA
GGCCTCCCTCCGGGCATCCTTGTGCCATTTATGCAGGATCTCCCAATAAATACCAAAAGTCTGAGTAGCGGGAGCTCCTGCCAGCGGCTCACAGTTGGAAC
CACGCTTTAGAGCAGCAACGTTTTCCAGATGTACAGACTCTGAGCGGGGAGATCTGCTTAGAACAGGATAGATGTTTATCCTGCTCACTGAAGAGGGTGTGT
GGTAGACGCCCTTCTGGATACAAATGAAGGAATTCTCTTGAATATATTTAAGGGAGTGATAACCTGGTGAAGAAAGTACTAGAGGATTTTAGAAAG

85-M13rev_3

CGTCCTACGGCCATGCGCGGCGAGGGCAGGTTATAGCGACGCGCATAGTCGGCCAGGCAGTACAGGGCCGCCGAATAGGGCGTCGGGTTT
GCCGTGACGGGCGCNTTGCCGATAGGCGTGGCGTAGAGACCCGACGGCCCAGCGGTCGGGCTGACACAGCCTGCCAGCAGCAGNACGGATGCGGNGG
CACCGGCGACGCTGCGAATAAGGGTGCGGCGGATCATGGCAGGTTGGTGTTTCCATTCAGGCTGGTGCCTGCGTTGACGGCGCCGTTGTTGGTCTGGGT
GGAGTTCACGATGACCACGTTGTTTCGGCCCTGGACCACGACGTTCAACTGGTTGCCGATGGCNGTTGAGCCGCCGATGGTCGTGCCGTTGCGCGATCCG
CNGCCTGCACCGGCGTAgGCCGAGGCGGCTCCGCCCGA

860-M13rev_2

TTTCACATATGAGAGAGAGAGTTTAGACCTAAACCTGTCCTAAGTGACCACCCAGATCCCAGCAGGGGACAGGCTCTGATCCAAGTGAAGCAG
GCTCCCCAGGCAGAGCCCTGAACACTGAAGGAGCCTGCAATGACCTGCATTCAAAAGAAATAAACACAGGTTGAAGCCCAATGCCAAGGCCAACACCCACA
TCTCTTTCCCTCACCTTTGCCCACTTCGATAGCCAATCCCAAGGCAGCCTGGGGTTGAGATAGCAACCAAAATGGGCCAAACAAACTCATCTGGGGTTGGG
GCTGGGTTGTTTATCACTAAGGAAGAACCAATGCCAAAGGCAGCTACCATCGTGAACTTGAGCTCCGACTTCCTAAACAAACAGCTTCTGGTCGTTCTGGAT
TCAAAGGGCACTT

865-M13rev_4

GGTTGGGGGATCAGGTTACAGAGGTCCTGGGAGATCTAGTTGAGAGTGTGTTTCTGTTCAATGTTAAAAAATATATGTGGTATGTGTATGTGAA
TGTGCTCATGGTGTGTGGGTACAAGTGTGTGGTGGCATGTACAAGCACATGTNNNAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTAAGTG
CACATATGTGCATATGCATGTGAAGGCCAGAAGTCAATCTTGGATGTCATTTTTCTGAAGCTACCTGCCTTGCTTGTNGAATACAGGGTCTCTCGCtGA

872-M13rev_1

CCAGAAGTGTCAATGGCTACTCAGTTTTGGGACACTTCCCCACAAGAGCCCCAGGTGGAGCAGGGCCATTGGCATTGGCATTTGTGATAGGA
GGGCCTGCTGGGCTCCAGGGTCTGCCTTGGCAGGAGGCAGCTTTCCTGGGTCTCTCCATACAGACCTCCACTGCAGTCTCCAAACTGTGGTGCACTGCCC
ACTAAgGGTCTTTCCCTTCCAGATCCNnCCAGAGTTCTACTGCAGAATAGAGACTTCCTCTCTGTCTGAACCGGTGTTGGATCCCGTTCCTGTGTAACACACT
GCAGCTCTGTGGAAAGTCGTGATACTGTTGACA

873-M13rev_2

CGGCGACATGATCGCGATAGACCGCTTCATGCCAGACAAAGGCGGTGGGGTCGTATTTGCCGTGAATGGTCAGGGTCGGGGCCACGATCAT
GCCCGACAGGTCGGCGTCATAACCCAGCCGCGCCAGCGCCATCGGATCAGCCGAGAACCTCTGAACGCCCGCATTCAGCGCCGCATCATCATGCGATCC
GGCATAGACGGTGTCCGCGTTCGAGAAGGGATTGAGGCCGCCCAGCCGACGCTGAACCATGTCACGGAACAGGAAGGTCCCCATGCCAGGTGGCCGAC
CAGTTGATCCTCCTCAACGCCCGTAACCGCCAGAATATCGGCCAGATTGCGTTTCTGGATCGCCGAACGCTGGGCCTGAGGCAGGCCGACGCCGGTGCAG
GTTTTAACACGCTCCGCCAATTGGGCGCGCGTCAGGCGTGAACCTTCCGGCAGACCCTGCCACAGCGGA

874-M13rev_3

CAGGCTTGCCAGGCTTAAAGCCCACAGACTGATTTGTGAACTAAGCAAACCATGAATTAGGGGGAATCACCCAAAAAGCTCCTCACACTAACC
CTGCTGTAGCCTTTAAAGGAAAGCTCCCCTGAGGAGACCCATGGTTACTGCAAAAGAAAACCGCTTCTGTCCTCACTTCACATCCTAGAGGGCAATTCTTGA
ATCACGGATGGTCTTTTTAAAGCCAGTGAACTGGGAGTCAACAATTCCTCATTTGCACGAAATCCAGAATGTGTTTAGCCCTTCACCTGGGAACAGCAGGAT
GAAAGCTGCCTGAACACAGAGGAGCTGGATGCCACCAGCCACCAGCCACCAAGCTGAGAGAGAGCCAATGAAGAACTCAGCACTTCACTTCCAAAATACC
CAAGTCGTGCCCTGAATTTTTAGGGCATAAAAAACATGTTCTACGTTTTTTTCTAATCTA

875-M13rev_4

CAGCCAGGGCTACACAGAGAAACCCTGTCTCGAAAAAAGAAAGTGCTGGTCCATCCATCCTTCGTCCCTTCCTCCCTCCCTTTCTTCAGTCAG
TCATGTCTTCACCACACGAATCCCAGGGAATGATTTGTGGCAAGCAGAGCCACTTGCTAGCCCTGTTTTTATTTCTCTTGGCATTTCTTCTGCTAGACAATCC
AGAGACTAACCTGGGTATCTATTTTTGTAGCTAAGCTGATCAACATGTTTATTAAATCTGTTTTTAAATTCAGAAAAGATTAGCATGAACTGTTNTTACCCAGAG
TGTTACATGGTCTCTTGATTGTNNACCCTTCAAATCTAATTATCTGTCATCAGAATATAGTTCTGATACTGGGAGTCTCACCACAAGGCTCTGCTATGTATGTA
GC

876-M13rev_1

TTGGCTTCCCTCAGTGTATATTATGACCTGGGATATGAGCCAAATAACCCCTTTTGTCCTCAAGTTGCTTTTGGTCATGCTTTTTATCACAGCAG
TAGGCACCCAAAGACACATATTAAAGAAATCATACACCATAATCATGTAGGATTTACCCTCGTAATGCAAGGTTAGTTTAACATGCAAGGAAGCGTAAACAGA
TCTCCATGAGAAACGCATGCAGGAGCGTTAACGGCAACATCATTCTCCAGAGCCAGAGCTATGTACGGCATCTCCACGCAGGGAAACACTGCTCTGAAATC
ACTCCACGCGGGGAAACACTGCTCTGAGATCACTCCACTCCACACAGGG

878-M13rev_3

ACCCACTTTCCACGCCGCCTGCTCCGATGATAGGAAAATGGCCCTGTTATCTCAGAGCAGGGGCCCAGGCTCTCACTGCAACAGNTCACCCA
CCATGCACTTTCCACAGNCTGGAGACTGCTGGGTCTGAGAGGGGATGGGAGCCAAAATCACCCAGCAGCAGTGACCCAGACAGACCACCTGAGCAGCGA
GTGTCACCACACACACACACACACACACACACACACACACACACGCCTCCTAATCCTTTTTTTAAGCTAAGAAGCTTTCTGAGAGCCCTGGGTTACAGGTCTAT
CTTTCAAGCCCACTGTAGTATTGAAGAGATTCTTGCTTGGCTGGCAAGATTGCTCAGTGGGTAAAGCACTCACCACTCAACTCTGATGCCCCAAGTTTGATT
CCCAGAACCCAAACAAAGAAGAGAATCAACTCCACCAAG

879-M13rev_4

TCAATGTTTCATCCTTGCAACAAAAAGGCAACTAATACACATGGGTTATCAGAAGTAGGGAAGGTGATACAGGCTTTAAGTGTGTATACACAGG
CATACAACCTTGCCAAAAGGAATGGTAAAGAAAGAAAAGAAAAAAAGGACACAGATAAGAAAGTTGTATGCTCATAATGGCATTTCCAGAACTAAGTTCCCTT
TTCTATAATTGTAAATAATTTTTTAAACTGTCCTTAATAAACACACAAGTGAATGACTGTTGCTTATTGATGTGAGCCAAAGTATAGTTCTGATGTTTTATAATTA
GTCATGGGTTATGATACTATAAAACTACAATGAGCTAATTATAGTAATTTTGGTGGTGGTGGTGCAGGGTGCTTTGTGAACACCCTGTGCTTATAAGCTCACA
ACAGCAACAGATAAGAATGCTTATCTTCTCTTAAA

87-M13rev_4

TAAGGNATCAACATCCTTGGGAAGCTGGGACAGAACGTTACCGCTTGGCCTGGGGCAAAGGCCCACCACCATATTAATTCCAAAGGGCACCT
ACTCCTGCCCTTTGGAATTACTCCCCTCAGTTCTTCTTCCCTGACCCNGATGCCAGNACTCTCTGCAAGAGCATGTCTACCAAGCAACTGCATTTAAAACTAT
ATATCAGGGCTAGCAGGAAAGGCTGGTCTCCTGGGAAATCANATACTACATTTAAGAGGTATATAGATTTTGCCACCCCCTACTTCNTCGGAACTTGAGTTT
GAATGACTCCATGTGACTTCTGGGCATACTGGNATGCTTGTGCTNGGCCCCCTGCAAGGANAGAANAGAAAGATCNNC

880-M13rev_1

TGCCCCTCTGTTCTCCTCTCCAGAAGTTTGTAGGAGTGTCTGCATGTCTTTGTCTCCATAAGATCTTGATGCTTTGACATACTGACCTTAGTCTG
ACTCCCCNnCAAACATTATGTTTACTGCTCCTGAGTTCCCTCAGACTCAGCAACCTTGCACATTCTCTCTTGCTCATGCTCTCTCAATCCCCCCTCTCGCCCT
TGCTCTCTCCTGCACTCTCTTGCTCTCTCTCCCTCACTCATGCTCTCTCATGCACTCTCTCACTCTCCCATGCACTCTCTCGCTCTCTCGCTCTCTTCTCTCAT
CTCCCTGCATCTCCTCTACATCACTCCTTCCGTATTGCAGGCAACCCTCCCCACTTCCCACC

881-M13rev_2

CCACGAGTATGCCCGGAAGCTCGAGGACAAGGTGAGGCGTTCTGTTCCCCGCTTTATTCCTTTGACACAGGTCTCTCCCTGAACCTGAGCCCT
ACAAGCAGTCAGCAAGCCCAGGGAACCCCCTGTCCTCTCTCCTCCACCCAGTGCTAGGAGTACAGGATTATATTCCTGCAGAGATCCGAACTGAGNTCCCC
ACACTCGCACATCAGACACTTTACTTAATGCGACATCTCCCAAGCCAACCATTTTTTCCTTGATTTTGTGGAGCTTTGAATGAGAAAGGCTTCTATGAGCTCAT
ATTTTGGAAGGACTGGGGGGAGTGTCCTCATGAGGAAGCATAGCCTTACTAGAGGAGGCGTGT

882-M13rev_3

AGGGAAGAAAGAAACTAGAGGTTGAGGGGGTGGCTAGAGCAAGCAGAGAAAACCCTAGAAAACAAGAGCACTGGCTGACCCAGCCTGCCTC
ACTTGCCCACATCAAACACCCACGTCTACCCTGGCTCGTGCCCAGGATGCCTCCTGACTTCCTCCATCCCCCACTCAGCCCCAGGTCGTCTGCTGCTTTCC
TGTCTCCTGGCACCACCAAGCCTTGGCCAATGCCTTGCANGTTCCTCTACTTCTACCATGCAATGAGGCAGCTCCCCTGGTCACGCCATTTCCCCACCTAAG
CCAAGGGTATGCTTCTCAAC

883-M13rev_4

GCTGTTGTTCCAAAGCGGGAACCTTGTCTTTCAACACCNAATTATCCAGTAAGGTGTAAAAATCCAGGTGTTCTTTCGGCTGACGGTACGCCTT
GCCTTGAGTCGTCACCGAGCCATACGGCTCCACCGCGATGGGCCCTAACTCCTGCGCACCGGGATACCAATCATCAATGNTACGAATGGCGTTGCCAATTT
TTTGTGAGTGCATCGCCGCNnCATGATTGACTTGATAGANCGTCCGGCTTTTATCGCCTTTGCCCCGTTCCAGG

884-M13rev_1

GAGTGAGAGAGAGATGCAGGATTCTGACTTCTCAGTCACACGTTCGCCCGGGCCCTACCTCACTTCCCTTTCACCCCCCCCCCCCAACACTTCC
AGAAGCATCAGAGGATGCAAGCATGCGAGAATTGAAGGGTTCAGCGAGCTCATCCACTCGCTGCTTCTCAGCTCTCTTCGCTGTCAGCTTAGGGTGTTGTT
CTCCAAAACTTTGGGATCTTTGCCTCTATTTACTCTTGGGACCCAACCTACTTGTTTTTGAGAAAACAAATCCATTTAAGTAGAGCTTCAAGTACACACAAACG
GAAGTGCAAATGGAACCAATCGCTAAGATAATATTCTTCATGCATTGAGCAACTCTCAAGAAGAGGTAAAGAAAAGTTAGGACACCTCGTCTTTCACCCTCCT
CGTTGCCTGGCCATTCCCACGTTTGCATGGCTTCCCCAGTTACGTTGACTCTGCAGCATCTCCCGTTGGCTCATGTTTTTGACCACTCATTGGTCTGGCATG
TGGCAATACTCTGTGACTGTAAAACTTCAGGAGGTAGGGTCTAACTGGCAAAACTAGTTCACTAAAGACAAGTCTTTGAAGTTTATACCTGGACCTGGGGCC
CTGTC

888-M13rev_1

ATGGATGGGCTTTGTAGTTCTGATATAAGCTCTCTCACCTACTGCCTAAGCATCATGCCTGACTGCTTGCTGCAGTACAGCTTGCCATAATATT
CATGGAAAACAACACTCTGAAATCACAGGCAAGCACCTTTAAATGTTTCTTTTTTATGTTCCCTTAGTCGTAATGTCTCTTCACAGCTATAGAATAATAACTAA
GACAGGAGCTGGTGCCAGGAACTATAATACAGTATTATTGTTTGCAAAAAATATGAAGTATTTGGGAATGTTGGATTAGAAAAGCAGTTAAACATTGTAAGGG
TAGTTTAACAGGGCATCCTAGTAAGAGCATGGAAGCTGGTGTCATTGAGAGCAAGGCATACTGTGGAGACCC

889-M13rev_2

ACATTTACTCTCAAAGCCTCACATAATGCCTGGAGCCACATATCCTGCTCTCCTCTTGTTGAGAGTAGCTGTCTGTTTTGTTTTTAGGTAACAGC
GGTCACCTGTCATGATCGATCACTCTTGAAGAGTGAGTGTTGCCCA

890-M13rev_4

CCNAATCGCCAAGCTTGGCTCCGNAGCTCGGATCCCNTACCTAACGGCCNCCAGATGTGCTGGAATTCNGCTTGCGGTGANCNNNGAGATCT
GAATTCCCCCTCTAGCTGGTAGAACGGCTTGCCAGCGGCATACCTGCGAGTGTGACAAGAGAGCTGCTCTTTGCTTTCGCCACAACCTGAACACTTACAAC
CGCAAGTATGCCCACTACCCCAACAAGCTGTGTACTGGGCCCACCCCACCCTGCTGAGGCCCTGCTCGGCTCCATAGCCNCCCCAGGCTGCTGCAGTCTC
AGGCCCAGAGAAGCTCGGAACCCAGATTCCTCTCCCAGCAGACTCATCCCGCCCCCCCCCCCAGAGATCATGANCCCTGGTCTCTGGAATTCGGATCTCCC
GGNTCACCGCACGCC

892-M13rev_3

TCAATATGTAGCCAAGGATGGCCCTGAACTTCCCTTCCCTTCTCTACTTCTATAGCACTGGGATTACAGGCCTGTGTCACCATGCATAGCTAGT
TCATGTTAACATTCCTATTTCATATCTGGTAGAGTTATTTAGGGAGACTGTAAACCTCTAGGAGGTGGGAGCAAGGAGGAGGAAGTAGGTTGGCCACCGGG
AGCATCCCTTGGAGGTGTATTACCCTTGGCTCCCTACAGTCACACCCTCCCTGCTTCCTCCTCAGCCACAATAGGAAACACTCTGACACATTCTTACTGCTAT
CATGGACTGATGACTGCGTCCCTTGAAACCTTTGGCCCAGATAGATTCCCCTCTCCCCAAGTTGTTTCTGTCAGGTGTTCTGTCTCAGGAACACACAGCTAA
CACTTTGGTGTTATAGATCTGTTTGGCTATTGGTCATATGTACTCCAGGGTGAAATAATTAG

893-M13rev_4

TTATTTAAGTATCTTTCCTCGCCAGAAAGCTGTAGTCACAAACTCATATAAGCATGCCTGTTTTCTTTTAAACTAAGTTCATAGGGGTCAGGCTT
GGGGGCAGGGTACCTGCCTGGGGAGCGACGGAAACTACATTTCCCGGGATGCAATACGCGAGACGATGGGGGTGGAGTTGAGCCCGTGGAAATTCGGCA
TTTATTTGTATTATTATTTTTTGGGGTGTGTGTTTGTGGACTAGTGCGCGGTTCGTGTCTGGAGCTAGATTTCCAGCGTGGTCTTTTTGGTGGTCGTTTAGGC
GTTTCTGAGGTTTGTTTACCTAACTGGAAGCTCGGCGAGGTGGGCAGTTCTTATGGCAGAGCGCAAAGAGCTGCTCTGTGTTACAGAGGCCCTGCTGAGGC
CCAGCGACTCCGCCCACATGCCCCGAGCCAGGCCTAGGGAAGAGACCTCGGCGACCTGGTCCTTG

895-M13rev_2

CAAATATGTAGGCAGGGCTGGGCAGCCCCAGCGTGGCTCTCTGGGGGCCTCCAGGAGGATGCAGGGGTATCAGGAGTGTCATTTTTCCAAA
CATGGACAATGCCAGACCTGTTGGGTCCACGTTGCCCTGTCAAGCCTGCTCCTGGGCCTCTCACAACTGAAGAAAGGAGTTGTCCCACAAAGCTCTGGGG
GTATGGGGACAATATGAACTGCCCTGAAGGAGACACTATACTCTTGTTCCAATGGAGGTAGGCTGTGAACGGACTGGGGCACAGGGCTACCAAACTGGCG
TTACATGGTGTTCCTACCTGGTGTTACATGGTATCTGGTGCTGGCTGGTTTGTTGAGGTGATAAGGTCCTGGGTGACGTGTCTTTTTGTTTGTTTAGTTGGTT
GGTTTGGTTTGGTTTGGTTTGGTTTTTTTTGAGACAGGGTTTCTCTGCATAGCCCTGGCTGTC

898-M13rev_1

TGCCTAGGATCTTGGCCTGGGAGAAGAAGGGAGGAGAAAGAGATACAACACACATACGCAGATGAGCACGCGCACACACACACACACACACA
TACACATGCAGATTACCACACACACAGAGAGCAAGAGAGAGCCACTAGAAGGTTCTGCTCTCTCCACAGCATCTCTGCAGCGTCATCTGTACATACAGGAC
CTGCTAGGTACTACACTAACTCCTAAGAGGTGCTTCAGGTTTATCTTCCCAACATAGAAACTCCATGATTACTCCTGACTACAGGTAAGCACTATGTCACACC
GTGCCACACAGTGAAGGACCGGAGACCATGTGACACTGCCCCAGTGCCTGGCTGTTCTAAGCACTTCCCCCATATTAACTTGTTTAACAGGACAGGTACTTT
TATCTGTCCCAGGAAGCAAGAGCAGAAGGACACATGTTGTCCCAAAGAGGCCCCACT

89-M13rev_2

CCATAGTCATGCATCTCTCTCCCTCTCTCGCTCTCTCACCCNGTTACTCTTTGGATCTTGATAGTTTCTATTCAACATGTAACTTTCTATTATGTG
GTAGATTCAAGGGATGTGGAAACAAAACAGAAAAACCTGACCTTTCCCATGTAAAGTTTACACTTGATTGAGATGGTAATTATATAAATGTAACCTGTGTGCTT
CACCAAAGCATCACTCGTGGGCAGACAGTGNCACAGTCCTTCNGGAAATGTCCTCACCGCTTCTTTCTTAATGTGCCTTCTATCCCTACAGTGCACCCCACA
GCCACAGGACGCAACATAGAGCCCACCGTTCAAGATGACTCTGCGACTGCGCAGACGGGAATTGAATGAGAAATGGCGGTGATTTCTTTCTTCCTTTCTTTT
TTAATCACTAAAGCATTTTTCATCT

900-M13rev_3

TGGATATGCTCCCAACGCACCCGGTCGGCGCTCATCGGATTTTCCGGCGGCTCGGGCGCCTCGCCCGTGGCCAGCAGGGCCTTGACCACGT
CATCAGCGTCGGCGGGCTTGGACAGATAGTCGACGGCCCCGGCCTTGACCGCAGCGACGGCGGTGGCGATGGCGCCGTATCCGGTCAGCATGACGATA
CGGGCGTCGGCGCGGCGGTCACGCACGGCCTCGACCACCTTCAGGCCGTTGCCGTCCTCCAGCCGCATGTCCAGCACGGCGAAGGCGGGGGCCTTGTC
GCGCAGTGCGGCGGTGGCTTCGGCAACGGATTCCGCCGTCGTCACGTCGAAACCGCGCGCCTCGAGGGCGCGGCCCATGCGGGTGCGCAGGGCCTGAT
CGTCGTCCAGCAGCAGCAGGGACTTGTCGGACAAGGCGGCGATACGGGTCTCTAGGTCGGACATTATTCTCTCCTGCC

901-M13rev_4

AGAGACATAAGCAAGTGGAAATATTACCCTGCCAGAAATGAGGGGGGAAAAGAAAAGTGCACAGTGCTCCCTGGTGGCATAAAGGCTGCTGG
ATAAACCACCCTGAGACAGGGAACAGGAGAATCTGAAGGGTGGATTCTGCTCCATTCTGGGGCCAGATCTCAAGCACCTGCCCTTCCTACAAGCAGAGCAG
CCCATAAGTAACGCCTGCCTGTAGAGGGGCAACCTCCACTGGTGCCCGTGACCCACAAAGAGCAAACGTGCCCGTTAGAATATATCAGGTATGTATTTCCT
CCTTTCCGCACACAGGATCTGGCGTTCAGATCTCATGTTCACGTCCCTCAGTAGCCAGAAGCGGATGAATGAGAATCCTACTGAGCGGTCACAAAGATGTA
CACCACTTCCCTGAAGAACCCTTTCCTAAGGTTTCCGGCGCTCTGTTCACCCCGTGGTCCACTCTTAGC

902-M13rev_1

CTGAATTTTCAATACATGCTCAGAACTACAGATGTAAGAAATCCAGCCCCATTTTAATATGAGTGCTAGGGATCCAAACTCATGTCCCTTTGCTT
CTGGGGCTGGCAGTCCACTCACTGAACCATCTCTCCACCCTCTATATCTGCTTACCTTAGTATTCATTTAAAGTCTACAAGATAGATAGAAGTTTTATGACACA
GACACATGATGTGACTTTAACAGAGGGAGGGATAGATGAAGCATGGAATACTTTTAGGGTGTGAAACTACTCTGTATGAGGTAATGATGGTAATGGTATAAA
TGTCATTATGCCAGTAGCCAAATGTGTATAACACACAACACCTAGAATGAACCCATTGATGGCTAGTGTTTTGTTAACTTGACCAAGCAAGAGTCATGTAAGA
AAAAGGAAACTTAACAGAGAAAATATTCCCACCAGTTTGTCCTGTGGGGC

906-M13rev_1

TGGTAAAATTTTGGGGTCACTTATATATGCTATCATATCATCTGCAAATAGTGATACTTTATTTCTTATTTGTCAATTTGTATCCCCTTGATCTCTT
TTTGTTGTCTTATCATTCTAGTTAACACTTTGAGTACTATATTGAATAGATATGGGGAGAGTATATATCCTTGTCTTGTTCCTGATTTCAGTGGGATTGCTTCAA
GTATTTCTCCATTTAATTGGATATTCACTGTTTGTTTGCAGTAAATTGCTTTTATTATGTTTTGGTACAGCCCTTGAACTACTGATCTCTCCAATACTTTTAGCAT
GAAGGGATGTTGTATTTTGTCAAATGCTTTTTCTGCATCTAAGGAGATGATAATGCGATTTGTTTTCAGTTTGCTTGTATAGTGGATTATGTTAATGGATTTTCA
TATATTAAACCAACCTTGCATTCCTGGAATGAAGCCTACT

908-M13rev_3

CTCCAAATTCCCTCTTCCTCAGAACTACCAAGATCACTGCAAACAGACTTTACCATGAGTTGTAGTTTGAATATGGCATATCTGTTCCAACCCAT
AAATTTCATATGATAAAGTCTTAGTTTTTTACCTAATGGCCTTTGGGAAGTGATCCACTACAACGGTTCTAACTTCACCAATGGGTTAATCTTTTGATATATACA
AAATCGGATGACCATATTAGGAAGTGACAGAAACTAGGAGATGCACATCTAATCAGGAGACAACAATTACTGGAAGATATGCTGTAGAAAGGGATACCTTGT
TTCCAGACTTTTCTGTTCTTTCTCTGGTTACTGGGTACCATGTGGAGAAAATCTGGCCTCTGTCACATGCTCCTTGCCATCGTGTTCCATCTTTCCCAAGGCT
TACATCGGTGGAGCCAAGAGTCTCTGAACTAAATCAGTCCTCCTTTGACTTGTGCTA

909-M13rev_4

TCCACTCGTATGGATCAATAGAAGCAACATATTGTCTTCACAGAGCTTTAGATCCTTGCCCAGATTGTTTTTCCCCCTAAACGTGGGATGCCCT
GAAGCTTCAGCTGCTGCGAAGAATAGAAAAGCAGCCACAGTGTTTCTCCACCAGAGTCGGTCCTGCCTCAATAATTTGTGTTCGTAAGTGACTCCCAAAGAC
CTCCTCCTAGAACCTGAAAGCAGAAGAATGGAGCAACACGGGTGCTCAAACACAACCCTAACTTCCAACGTGCTCTTTTGTAACAAGGCAGCAGAGCTAAA
GTTAGGAGCACACAGCTGCCCAGGCAGGGTCCGAGCATGGTCCCAACAGGAAGGAATCAACATGCTTGGGCTGGCAGCTGGGTAAACATTTCTGCTAAAG
CCTCACAGTTGTGATGGAGCCTCGTGTGCGGAGGGTACAGTTGGATGTGGAGAGAGCTGTTTAAA

90-M13rev_3

CAGAGCCACCCTNNTCCTCGGCTTAACCACACCAGTATCACCCCAGGAGAGGAAACGCACCTGTGTGCTAACCACTGAGACATTAAGTGGCTT
TGATCCTGTAGCATAAGTTGGTCTGTCTAAATGGCTCCGTGTGTTNCTTACGATACAATACAACTCTTTTAGTTGGTTTCCTTTTTCTCTTTCCTTCTTATTTCT
TTTTGCTTTCTTTCTTTTTTGGTNGGGACAGGGCTAGAGGAGAGCTCAAACTCATTCAACCCNCCCCCCCACTATCACaAATTATGCAGTCGAGTTCCCTGAA
TTTAAAGAAATCCCAGGGTTCGACATATCTGGCATGCAAAGAATGATCATCACCTGACAAAACCACATTCGTGATCATGGCACGTACCCTNCC

910-M13rev_1

CATCTATTAAAACTGAGGACATCATGAAATTTGCAGGCAAATGGAGAGAACTAGAAAATGTCATCCTGAGTGAGGAACCCAGGCTCGAAAGGA
CATACATGTTATGTCCTCATTGATAAGTAGTTATTTCGCAAAAAGATCAGAATACCATTATACACAACTCACAGATCACATAAAATTTAATTAGAAGGAATGCCA
AATTGTGGATACTTCAACCCACATAGTAGGAGAAACAAAATAATCCCAGGAGGCAGAAGAAGGGAGGGATGTGGATAAAAGAGGGAGAATCAGGGACAATC
GGTGGCAGGATCAGGTATATGGGGAGACAGGAGAGAAGGCCAGAGGGCCAGGAGTATTAATAGAAACATACAACCGTGTGGGAATGAAGAGCAGGGCAA
ATATCTAGAAAGTCCCAGACAAGATGGATGTGGAAGTCTCCCAAACCTCATCG

911-M13rev_2

TACATACGTGTTCTACGAGAAGACAGGACAGGAAACTGAATAAATAAAAGACACAAGGAGCTATCAGGTAGATACCAAGAAGGACCCAGCCAA
ACCCATGAAGAGCTCAGAAGCTGAGGTGGTCTTTGAGGTGTGCCCTGAGATGGAGGTACAAGGCTGGGTCTTTATACTGCCTTGTCAACCATTGTATCTAG
GCCACTACGAGAGTGGAGAATGACTTTGGCAAACAGAAATCATNTTCAGCTGAGGGAAATCCAACAGCCCTTCAACATTGGGTACCTGATGACTCTCCAGTT
CAGGTTCGCCCAAGGTAGCCATAGAGGACCTTGGTGGCTAGCAGCCATCATGCCAGAACACCTCCTCCTGGGCAATTGGCCTCTCCTTGGGCAACTAGGT
CCTGGTCTGCAAACCAGCATCACTGCATGTAAGGGACAAGAAAGGANATATGAATTCAA

915-M13rev_2

CCTCATTTCCATGTGCTATAAGGACACAAAGTGTCATAGAGGAGAGAGTGTCAGAGGGAAGACAGCATTTTCTTAGAGGCATGAGCCATGCTG
GGCTGATCTCAAACACATCCCTCACTACAGAGAAGCTGTATTCCCATCCTTACTCACTGTTTACACTCCATAGTCTAGTAAATTCCACAAAGGAAGGGAAACT
CAAACCCTATTTTATCCATTTATACATTCCTGCAACCTAGTTCATACAACCACTGTGTTAATACTTGCTGAACGCATGCATGAGCTGTGAGGGTCTGATGGTG
GCTCTTCCCAGTGCACACCCNAACAAGAGAAGGATCTTTCTCCTCAGGCTTCTCAGGCTAGGTCTCGGACCTTAGAGCTCANAGCCATAGGCTTAGAGCCA
ACGCCATAGNCCCTGGTACATTC

943-M13rev_2

ACACACACACACACACACAAAATCTATATATTTACACTCTCTGGGAATTCTAANTACAGAATTTCCTCACTGACATAAATCCNAGTGAGTCACAA
CTAAGAATATTAAAAATCTTTGTGTTTTCTGTTTTCCTGATAAATTGTTCATTCTGACCTTATACATAATTAATGTGAATTCCAATTCTAGTGTCAGACTTACCTC
CCTGTTGTTATGTAACTGATGNGTGTGNTACAACTAAGGATGATATCTTGGGCAGGTATCNTGGGTGGNTTACCTGGNTNAGCATACATCACCTGGTTTTTCT
TCCTCCCAGGCTTTGCTGTTCCAAAAACCCA

945-M13rev_3

TTCTAAGACAGCACTCCTGAAAACAAGATGGAAGGGCTTAAATTTTAGGCTCTGGCTTCAGAAGTCACAGAGTGCTGTTGCCATGGTGGTCTTC
TGCTGGTCGAGGCAGAAACAATGCCAGCCAGGTTGAAACAGTGGGACATAGAGTCTCTGCTTTCTCAGTATAATGGGAGATACGATGATTGATGTATTTGG
AAACATCTGGAGATCTGAAGGATTGTAACTTAGGAGTGAGTGAAGGCATGGCCCAAAGTGGCCAGTGGTTGAATAAAGGAAGTATATAATACGCATTTGCA
TGCAGATATTTGCATATACACCCATGTGAATATATACTGTAAATTCATGCACATGTGCAATATCTTTGAAATGTTCTCCAATCATGAAATTTAATTGCTGTTCAT
CTTATTTGATATGGAGATGGGAGGTAACTACTGATGAAGGAGCTACTATAGGGTTTCG

946-M13rev_4

TTGCTATACCAGTACCATGCAGTTTTTATCACTATTGATCTGTAGTACAGCTTTAGGTCAGGCATGGTGATTCTGCCAGAGGTTCTTTTATCCTT
GAGGAGAGTTTTTGGTACCCTAAGTTTTTTGTTATTCCAGATGAATTTGCAGATTGCCCTTTCTAATTCGTAGAAGAATTGAGTTGGAATTTTGATGGGGATTG
CATTGAATCTGTAGATTGCTTTTGCAAGATAGCCATTTTTAACTATATTGATCCTCCCCATCCATTATCATGGGAGATCTTTCCATTTTCTGAGATCTTCTTTAA
TTTCTTTCTCCAAAGAATTGACATTCTCATCATATAAATCTGTCACTTCCTTAGTTAGAGTCACACAAAGGTATTTTATAATATTTGTGACTATTTTGAAGGTTGT
TGTTTCCCTAATTTCTTTCTCAGCCTGTTTATCCTTTGTGTACAGAAAGG

96-M13rev_1

AGCCATGCTGGCGCATTTCTCCACCGAGGGCGGTTATGTGGTGGAGGGAGAACTGGAGCTTACCGTCGGCGATGATGTGCGGGTTCTGAAG
GCGGGCGAATCCTATCTCTTCGACAGCCGTCTGCCGCACCGGTTTCGCAATGCCTCGGAGGGCGTCACCATCGTCATCTCGGCCTGTACACCCCCCTACC
TTTGACTAATGGCGCTCAATATATTGAGCGGTATGTTGCGCAAAATTTTGAACGGTGTTAGGAATTCCGCAGGGGGCAAGGTCGCAGATATGGCGGCATCT
GGGGCGCGGCAGGAATGGAGGCCTGCATCGCGGCATATCCGGAGCCGGGTCCCGCAAGGCAGGCTGAGGAGACGACTGCGTTGATTGCAAAGCATGGA
AAAATGCCGTCGCAAAC

97-M13rev_2

TGCTGTTCCTTGCCTCTAANTTCAAGGACACAGTTTATGTGAGTAAGACAAGCTGAACAGCATGTGCCACTTAGCCACAGAGAGACCATTTGgA
TGCTGCTAAGCAAACAGGCACTCAGATTAGGGTTTTAAGGTGTTTCATGCCTGTGTGGTTAGAACATCAAATGTCTACATGAATTGTATGAAGAAGTAAGGTA
AGATGAGAACAAATCCACAGTTGCGTCTGGTGTGTGAATTGGGTCTGTGCAGCTGCACTGGTTGCCATCACTGTGGACTGGCACTTTGAAAAATTGTCTATG
GAAACAATTGGCAAGTTGCTCAGGGAAATTCATGTGTTTTAGAGTTCTTTCTCTTTACGTGTTTTGTTATGGTTTTGGGGATTAGTGAGAAGACAAACAACAAA
TAGTACTGGACAATTCTGCTATAAAG

98-M13rev_3

TGGGTCCATGTGCTGGGATGTGTGTGTATTCAGTGACAGGGCGCTAAGGGTNTCTTTCAGTGATTCAACCAGCTTCCAGACAACATGGTGAGT
CTGTACAGTATGGAACCTGGAAATAATTGGCAAAGGTGGCTATTAATTAAATAACCATACAAACAAAAGTGTCATTCAATTCAAAGCTAAACTAATACCATGAT
TAGAGATGCACCTTAATGGTAGAGTGTTTGCATGTGCAAGGCCCGGAAAAACATGCATGTGCTCTCACATGAGTGAGTACTGACNCTCTCTCTCTCTCTC
TCNCTCTCTCACACACACACACACACACTCtATTCAAAGAAGGAATTTGGCACCCATGAGCTAATGCAATANAATCAGCCCTGCTTTTACAGCaaaGTGT
TCCTGGGACACTGAGGAGGAGGGACT

The Contig sequences belong to STAT5A Library (non-repetitive DNA sequence) dataset. Contigs were built for overlapping clone sequences by Dr. Ramialison Heidelberg, Original clone IDs comprising the individual Contigs are listed in parenthesis.

Contig1 (828-M13rev, 914-M13rev)

TCNTTTTCACNAGGNNNGNNANGGGNTGGAATTCCNCCTTGCGGAGAAAAAAGAGATCTGAATTCTTTAGTGNNTAAATGTGCATTTGTTGATT
TTTCTTAAAGTTTATTTCATTCACTAACAACATAGGATCATAGATTTTTTTCTTTGCCATTCAGAAGCTCATGGCATAGTGAAAATGGAGAATATTTTCACGCTT
GATAGATACATACTTGCTGTAAATTCAACCATAAGTGTACAAACACACATTTTATAAAAGACGCTGAAAAGATTTTTTTGAGAGAGGGTTTCTCTGTGTAGCCC
TGGCTGTCCTGGAACTCACTTTGTAGACCAGGCTGGCCTCAAACTCAGAAATCTGCCTGCCTCTGCCTTCCCAAGTGCTGGGATTAAAGGCAAGTGCCACC
ACTGCCCAGTCCTGAAAAGATTTTGTTTAGGATTCCTAATAAGTTCTGCACCAAGAAAAACATAATTACAAATGAATCTAGTCTATATTTGAAGTTGGAATTGG
AGA

Contig2 (583-M13rev, 316-M13rev)

GNNCCGCCCTGCGCCTCCACCTACTGGACTCTGCCGGAGAAGCTCCAGGTTGATTTTGACAAGGTTAGTGACAGATGTCTTTAATACGGCTGC
TCCAGACCTCAGTTTCCAAGAAGAGCTTTCATTGGAAGGGCAGAAATGTCACGCCTTCGTGGGGGGGGGGGTGCAGTACAGCTGTCGTATGCTGTTTCCAT
TTCTGTGTCCATCAGTCCTCCCCACCCCCCAACACACACACACACACACAAAGTGATCGGGTATGCCCTCAGGGTTGGAGAAAGCAGAAATGGCTCCAGCG
ATGTAGAAGAGAAAGGGGGTCCTTAAAGCGCAGGGGAGGCGGGGTCACAACTGGTATATCAGAGCCTACTCATTGGTCATTCGCTGCTGAGCTGCAGCGA
TCCTGGGCTTGCGTTTGCCTCTCGATA

Contig3 (364-M13rev, 863-M13rev, 390-M13rev, 897-M13rev, 712-M13rev, 897-M13rev, 206-M13rev, 897-M13rev, 472-M13rev, 259-M13rev, 871-M13rev)

AACTCTACATAAACCAGAGATACTGAAACTTATAGAGGAGAAAGTGAGGAAAAGCCTCGAAGATATGGGCACAGAAGAAAATTTCCTGAAGAG
AACAGCAATGGCTTGTGCTGTAAGATCGAGAATCGACAAATGGGACCTGATAAAATTGCAAAGCTTCTGTAAGGCAAAAGACACCATCAATAAGACAAAAAG
GCCACCAACAGATTGGGAAAGNATCTTTACCAATCCTAAATCAGATAGGGGACCAATATCCAATATATATATAAAAAACTCAACAAGGTGGACTCCAGAAAAA
TCAAATAACCCCATTAAAAAAATGAGGCTCAGAACTGAACAAAGAATTCTCACCTGAGGAATACCGAATGGCAGAGAAGCACCTGAAAAAATGTTCAACATCC
TTAATCATCAGGGAAATGCAAATCAAAACAACCCTGAGATTCCACCTCACACCAGTCAGAATGGCTAAGATCAAAATTTCAGGTCACAGCAGATGCTGGCGT
AGATGTGGAGAAAGAGGAACACTCCTCCATTGTTGGTGGGATTGCAAGCTTGTACAACCACTCTGGAAATCAGTCTGGCGGTTCCTCAGAAAATTGGACACA
GTACTACCGGAGGATCCAGCAATACCTCTCCTGGGCATATATCCAGAAGATGTCCCAACCAGTAAGAAAGAAACATGCTCCACTATGTTCATAGCAGCCTTA
TTTATAATAGCCAGAAGCTGGAAAGAACCCAGATGCCCCTCAACAGAGGAATGGATACAGAAAATGTGGTACATTTACACAATGGAGTACTACTCAGCTATTA
AAAAGAATGAATTTATGAAATTCCTAGGCAAATGGATTGACCTGGAGGGCATCATCCTGAGTGAGGTAACCCAATCACAAAAGAACTCACACAATATGTACTC
ACTGATAAGTGGATATTAGCCCAGAAACTTAGAATACCCAAGATATAAGATACAATTTGCAAAACACATGAAACTCAAGAAGAACGAAGACCAAAGTGTGGAC
ACTTTGCCCCTTCTTAGAATTGGGAACAAAACACCCATGGAAGGAGTTACAGAGACAAAGTTTGGAGCTGAGACGAAAGGATGGACCATCTAGAGACTGCC
ATATCCAGGGATCCATCCCATAATCAGCCTCCAAACGCTGACACCATTGCATACACTAGCAAGATTGTGCTGAAAGGACCCAGATATAGCTCTCTCTTGTGA
GACTATGCCGGGAAGAAAACACAGAAGTGGATGCTCACAGTCAGCTATTGGATGGGTCACAGGTCCCCCAATGG

Contig4 (616-M13rev,319-M13rev)

CATCCTTTCCTATGCACCCACACAGCTTCCCAAATCCATCTTTTACCGGAAGCTTTGCTGACTTCTTAGCTTTGAGCTTCACTGGCTGCTATGAC
CTCTTACCTGATGAATATGTAACTAAGGTACAGCGCACCCCCTACATCCCTTTAATCCCTTTACTGACTAACTGCTATTCTTCTGCCCAGTAGATCTCAGAGT
CTTTGGGCTGCCTGGGACACTCATTAGTAAAGCCATACTCTGTCTAAATCTCTGTGGAAGGATCCTTAGGTCTTGTTTTGTTGTTGTTGTTTTGATTTTTTGTT
TTGTTTCTTTCTTCCTCTTTTTGGATGGTGAGGTACTGTGTAGC

Contig5 (896-M13rev, 218-M13rev)

CAGCAGAGAGAATTGAAACATGCAACCTGGAGATATAGGTGGTTAGGAAGACCATCCAGAATGTAACAGAGACCTGGGAGGTAAAATATTCTT
AGGACTCAAAGAGAGGGACCTTAAATGAAATGTCCTATGTTGGAGAGAGGAATATTGTAGAGCCTACCTCCAGCAGAAAGAAAGAACATCAAGGAAGGGAT
GGGGTTGCCATCCCACAGTCAAACCTCTAACACATAATTGTTTCTGACTGAAAGAACTACAGGGATGGAAATGGAGAGGAGCCTGAGGAAAAGAAGGCCCA
GTGACAGGCCCAAAGAAGGATCCAGCTCAACGGGAGACCCCAAGACCTGACACTATTACTAAGGCTATGGAGTGCTCACAAAAGGGGTCTATTGTGACTGC
CCTCCNAAAGACCCAAAACACAGCTGAAAGAGTTGGATGCAGATATTTGTACCCAACCNATGGTCGGAAGCACTGATCGCTAATTG

Contig6 (317-M13rev, 635-M13rev)

ACAGTCATCAACTCCAGGAACAGACCGTAAATCCCAGGAAGTAAGCCATAAGTCACTGGCTCTGGAACAGATCATAAATTCCAGGAACAAGGT
CATAAAACCTCTCTCCCAAAGAACAGTTTGGACATAACCACAAGAACGGGAAATTATCTTATCTCAGGCTGAGCCATCCATGCTGGGCAGCCCGAATTCAGA
TCTCCCGGGAA

Contig7 (360-M13rev, 835-M13rev)

TTTCTTGCTTTTTCTAGGGCATGATTTCTGTCCTTGTATTGTTTTTTTTTCTGTTATTATCCTTTGAAGGGCTGGATTCGTGGAGAGATAATGTGT
GAATTTGGTTTTGCTGCGGAATACTTTGGTTTCTCCATCTATGGTAATTGAGAGTTTGGCCAGGTATAGTAGCCTGAGCTGGAATTTGTGTTCTCTTAGTGTC
TCTATAACATCTGTCCAGGCTCTTCTGGATTTCATAGTCTCTGGTGAAAAATCTGGTGTAATTCTGATAGGCCTGCCTTTATATGTAACTTGACCTTTCTCCCT
TACTGCTTTTAATATTCTATCTTTATTTAGTGCATTTGATGTTCTGATTATTATGTGTCGGGAGGAATTTCTTTTCTGCTCCAGTCTATTTGGAGTTCTGTACGC
TTCTTGTATGTTCATGGGCATCTCTTTCTTTAGATTTGGGAAGTTTTTTTCAATAATTTTGTTGAAGATGTTTGCTGGTCCTTTGAGTTGAAAATCTTCATTCTCA
TCAATTCCTATTATCCGTAGGTTTGGTCTTCTCATTGTGTCCTCGATTTCCTGG

Contig8 (304-M13rev, 356-M13rev)

TGCAAGGTCTGTAGCTGAAAATACTTCTTGCCAACTGCCTGGAGGGGCTGATAAGGGAAATGACTACATTCTGTAAATTAACATATTCTAAGCA
GGTCAGAAGTCTCCAGTTTATAAAAAAGACTGTGATAAAGATAGGTAAAATAATGCCCATTTTAGTTGCCATGCTTAAGTAATGAAGCTGTAAATTTCTTGACA
AGGATCACTGCCTTAGTGTTGGTAGTGGGTGGCAACTGAGTATAACCTTTAGGAATAACTACTGCTGAAAATTAAGGAAGGTATATTCTCAGACTCATTTCTT
AATTGCTTCTTCATTGTACTCTCTCATACCTGAACCATAAATCTGACTCTGGGTTTCTTCTTTTTCCTTAGCCAGCCAGATACTTTTGTCCCTGCACTGACACC
TGGGTCATTTCTCAAACACTAGTTTTTGTTTTTGTTTTT

Contig9 (941-M13rev, 361-M13rev)

CCCTGTTATAGCTGTCTCATATGAGGCTATGCCAGTTCCTGGAAAATACAGAAGTGGATGCTCACAGTCATCTATAAGATGGAACACAGGGCC
CCCAATGGAGAAGCTAGAGAAAGCACCCAAGGAGATAAAGGGGTCTGCAACCCTATAGGTGGAACAACAATATGAACTAACCAGTACCCCCCCCCCCAGAG
CTCGTATCTCTAGCTGCATATGAAGCAGAAGATGGCCTAGGCGGCCATCACTGAAAAGAGAGGCCTATTGGTCATGCAAACATTATATGCCTCGTACAGGG
GAATGTCAGGGCCACGAAGTGTGAGTGGGTGGGTAGGGGAGTGGGGGGTAGGGTATGGGGAACTTTGGGATAGCATTGGAAATGCAAATGAGGAAAATA
CCTAATTAAAAAAAANAGCTGCCTGTCACACAGGATAAATTTGGTATTAAAGAAAGTTNATGTGAAAGAAACC

Contig10 (33-M13rev, 886-M13rev)

TCTCCACTCCCTAGAGACCTCATCCGCACCTCGCACTGAATATCACCTGCTACTGGAAACCACCCCCGATCTCTAACTGAGCATCTCTGCAGA
GGTCTGTATACTTTTAACTAAATATGACCAACACTGAGCTGATTTCTGCCATCACCTTCCTCGTATTGTCCTATCTCAAACCAAACACCCTAGGATCACTTAAG
CCCACCTTTTTTGTTTGTTTGTTTTGCTTTAATACCCATGTCCATTACCTCAGCAAAACAAAGAAAACTTGTATTATTCCTACCTTAAAAATATAACTACAGGGC
TGGAGAAATAGGTCAGTGGTTAAGAGCACTGACTGCTCTTCCTGAGTTCAATTCCCAGCATCCACATGGTGGCTCGCAACCATCTGTAATGGGATCCGATGC
TCTCTTCTGATGTGTCTGAAGATAGCAACAGTGTACTCACATATATAAAATAAATGGAAGTGTAAGCTGAATAAACCCTTTCCTCCCCAACTTGCTTCTTGGTC
ATGATGTTTGTGCAGGAATAGAAACCCTGACTAAGACAGGATGCTACAAATTTATTCTTCACTGGCCTTGCAGAGAAGGAAAACCCAAATCTGTAACAGTAT
AGGACGTACACTGAGGACAGAACTATGTCCACAGGGAAGAGGACAGGGTGCTAGGGCAGGTCATAGAAAGGAAACACCCAATGATATGCTTGCCCCCCTT
CAACAAATGGCACCCAGCTGTGTGCATGATATTTCATATATGCAGTCTGTTTCAGTGGTTCAGAGCTATTTCAAATGAGTCAGAACAACTGGGGAGAAAATCA
G

Contig11 (4-M13revfor,476-M13rev, 41-M13rev, 476-M13rev, 631-M13rev, 07-M13rev, 631-M13rev, 404-M13rev, 575-M13rev, 404-M13rev, 824-M13rev, 575-M13rev, 899-M13rev, 383-M13rev, 168-M13rev, 354-M13rev, 548-M13rev, 603-M13rev, 548-M13rev, 552-M13rev, 36-M13revfor, 595-M13rev, 36-M13revfor, 904-M13rev, 15-M13revfor)

CTAAATTGTTCAACTTTTTCAGTTTTCCACGCCATATTTCACGTCCTAAAGTGTGTATTTCTCATTTTCCGTGATTTTCAGTTTTCTCGCCATATTC
CAGGTCCTTCAGTGTGCATTTCTCATTTTTCACGTTTTTTAGTGATTTCGTCATTTTTCAAGTCGTCAAGTGGATGTTTCTCATTTTCCATGATTTTCAGTTTTCT
TGCCATATTCCACGTCCTACAGTGGACATTTCTAAATTTTCCACCTTTTTCAGTTTTCCTCGCCATATTTCACGTCCTAAAGTGTGTATTTCTCATTTTCCGTGA
TTTTCAGTTTTCTCGCCATATTCCAGGTCCTTCAGTGTGCATTTCTCATTTTTCACGTTTTTTAGTGATTTCGTCATTTTTCAAGTCGTCAAGTGGATGTTTCTC
ATTTTCCATGATTTTCAGTTTTCTTGCCATATTCCACGTCCTACAGTGGACATTTCTAAATTTTCCACCTTTTTCAGTTTTCCTCGCCATATTTCACGTCCTAAA
GTGTGTATTTCTCATTTTCCGTGATTTTCAGTTTTNCTCGCCATATTACAGGTCCTTTCAGGGTGCATTTCTCATTTTTCACGTTTTTTAGTGATTTCATCATTN
NTCAATTCGTCAAGTGGATGTTTCTCATTATCCATGATTTTCAGTTTTCTTGCCATATTCCAGGANN

STAT5A Library (+/- 1000 bp extended DNA sequences of STAT5 confirmed clones. The BLAST was performed by Dr. Bulashevska, DKFZ, Heidelberg)

Clone ID Sequence

20-m13rev_3.seq|chr11|1|97880023|97882439

AAGAAGATCCTGAGTACTGCACAGTCTGTGGGCTGCAATGTGGATGGCCGCCACCCTCATGACATTATAGATGACATCAACAGTGGTTTGGTGGAGTGCCC
AGCTAGTTAAGAAGCAATAAGAAAATATTTCAATAAAAGACTATCTGATAACCAGTGTGTGTGTGGGGGGGATTTCCTATAATCCTCACTGCCCTGAGA
AGAAGGAGGAGGTCTTGTCCCACTTGGCAAATGAGGAACTATGTTCAGAAGTATGGAACAGATGAGAGTCTGGGTAGCCTTGCCTTGTGAGGTCTCGTATT
CTTCCCTCCCAGGGCTTCTCAGGTCTGATGATTGGTGAAGAGCTGCTTCTCTGTGCCCAGGTCAGGGTGCATGTCAGAAGACTGAGGTAGGGGTTGCAACC
AGGCAACTCTACCTGTCTCAAGCTGAGTTTTCCCACCTTCAACAAAGGCTGAGGCTGGGGTAGGCAGCTGGGACGGGAAAGATCAGAGATGCTGAAAGGG
CTGCGCTGAGGGGAGGGGGACCCGGAGCCCTGAGCCGTGCAGGGGCACAACCACTGTAATCCTGCCACCATGTGGTTGCTGAGAATTGAACTCAGGACC
TCTAGAAGAGCAGTCAGTGCTCTTAACTGCTGAGCCATCTCTCCAGCCCCAGCTGCTCCACTAAGGATCCAGGTACCCAGAGATGGAGACAACCAGGCCAG
GGAACATGATCCCATTTCTGCTTCCTGTGATAAGACATAGTGACTGGGGTATGTAGGACAGGCTACGTGATCAGAGGTGAATAGAGAGGTGGAAGGTGGGC
GAGGCGGGAGGAGGCCGGAGGAGGCTGTCTGTATAGAAGGACAACGGGGGTGGGGTGGGGGTGGGGTGGGGGGTGGGGGGCTAATGTGAGGAAGGA
AACACTTCCCCAGGCAGGTTGTCTGTACAAACCCTCAGTGGATCCATGCCCTAAGGCAGGGCAGGACAGAAGCAAAGAGACCACCCTGAGGAAGTCTCCA
CCCCTGAGTGCCACAACCAGTCAGGTACCCTCAGTGAAAGGACAGGTGTCCTGCCAGTAGATGGCCTATGTATGTGGTCCCATAAAGAGAAAGGACATTGA
GGTGTAAAAGTCCTTTTGGGATTCCTGTGATTCTATATCTACTTGTTCTGCCTCCTAGAGTGAAGAGGCCTCTCCTGTGGGGTCATGTGAGGGTGCCAAAGC
CTCTGTACGACATGTGGATTTGGGGCTAAATAGATAGCCCTCACCCCTTGGGCCTCTTGGGGACCACTAGGATTGTGGGAGACCATCCCCCCTGGCAGTCA
ATGATTAGCTTTGGAGACACCAACTTGGGCTATCCCTGCAGGTCCTCAGTCTCCAAGAGCCATGAGTATCCAGCCAACAAGACGCCATGCCTCACCATCTTG
TGGGTTTGCTAGTGTATGAAGGCTAGAAGGCAGTTCACTGATCCAGCCCAACCCTGTGCTGTGCTCTCTTTTAAACATGTCGTTTTGGTAGAAAACCTGTAAT
CCAGTCTGGGGGGAACTGTGTGTGGCAGCCCTCTGCAGCTTGTTCAAGGAAAGCTGGGTCTTCAAGAACTTGTTTGAAAGTGGAGAGAACAGACTTCCTGCT
CTTGGACTTTTTTGGAAACCAGCCCTTAGTTCTTAGTTCTTAGAATTTTTATTAGGCTCTGGGACATCTCAAAACACCCTGACCTCCCACTCTCCTGGGTTATT
GTTGGGCATGGCTGGATGTTTCAAAAGCCCTAGGAGTTACAAAGCCGGGGAATAAAGGCGGCTGTTGAGTGAGGCAGCCACATGTTAACAGGGAGGTTGG
AGGCCAGGGAACTAGGGGAGGGTTCTCTGTGCCTGACGGCTCGTAGGGTGTGCTACATGAAGCTGCAGGTTCTTTGGGGAAGAAGGGAGAAAGGAGTGA
GGAGAGAGGAGAGGAGAGGGAGAGGCGAAGGAGGGAGAGAGGGGGGAGGGGAGAGAGAGACCCTTAGGGGATGGTGGGAAGGTCCTGTTTGA
AGCCCCTGCCCTGTTTGAAACAGGAGCTGTCCAGAGATCCCCCCCCCCCCCCCCCAAGAATTGGCAAAGGCTAAGTATGACTGCTCCTCTGGGCCATAAG
CCAATGTCACCACTGTGATGTCATTGTGAAGTAGAACTCTCTGGAGTCCATTATTTCATGTGGCCCTTGCTGCTGCTCAGAAATGGGTGGGCGGGTGGGAA

CTGTTAAGGAAAGGAGCCCAGAAGCGAGGATTTTCTCAAGGACACAGAGTCGGTCGGTCGGTAGCTGAGCTGAGCTTTGGAATGGCTGATGGTGACGCGG
AGTTATTCTCAAGCATAAACTTTGCTGTCAGGTGACCTAGTTAGCGGCCATTCTGTGTCACGCCCATTGTTCTGTGACTGAGGGACTGGCTGTACACAGAAA

34-m13rev_2.seq|Chr14|1|26176070|26178495
CTGCACCATGAGAAGTGGCCAAGCAAACAGCATTGTGTTCTGCGTGAGCGTGCACACCCATCCTAGGGAATAGCATTTGGAACTATTTATAAACTTACTTGA
GGGGTGTGGGATGACAGAGGACCATAGAAATCTATTGCACGTCACCACCCTCACACTTAACTGAGCAGCCTACTCTCTACCTTATTTATTCACATACCATTG
CTCATTGGGTGGCATGGTGTCACCCCAACATGAAGGTTTAAATTAATACACACACATACACACACACACACACACACACACACACACACACACAAACACATA
CACAAACACACACACACACACACACACAAACACAAACACATACACAAACACACACACACACACACACACACACACACACACACACACACGTTTCCAGTTC
ACTGATTTTTCTCATGTCATTGTTTGGTTGTTTTTTCCACGTAGAATAACCAGATTTGGGTTTATCCTGTTCTCTCTTGGGAGCTGAATGTTCTAAAATGTTCTG
TATTCCCTGGTTGGTAAAGTGTGGGGACTTGCTGCTGAGGGCACACGTGGGCCAGGCATCTGAAGGGATAAAACGTGCCACTGGGTACAGCTGGCACATG
GCAGAAGCATCACACGTATTTTATGTTTCTTTCTGGAAGTCACGTCCCTGAAAAGCTTCCTTGCTCAAAAATCATTGTCAAAGCTAAGAAGTAACCAGAAGGC
AAAGGAGTGTATCTGCTCTCTCCTGCAGCCTGGGTCTCCCGCCCCATGACAATGCAGGTTTTATCTAGCAGCCAAATGTGATCCCTGCCTTATCTTCAGCCT
TATCTTCAGCCAGACTGCAGACCTGTCTGTCAGCTCAGCTCCGAGCGAGATGCATGTCATCATCCCAGATTCAGCTAGAACACTCTGTCCTACCCTTTGGGA
CACCAAGATGGAGATCTGCGCACGTGGCTCTTTCCATCTCTGTACTTCTATTTCATTTTCTGGGAAGCTAAGATAATGAATGCTGAACTTTGGGAGATGACAG
GTAGATAGAGAGGCGACACATGTGAATCCGTAGTGTCCTTACTGAAGCAAAAGGCACCATGTCTGTGTGAGAAACGTTCTGCTTTGTATATCTGATGGAATA
AATACCCAGCGTACCTATGGTTGCTTAATCCCAGACAGTCCTGGCGCCTGGTAAGACTGCGTACCAGCTCCTGTAGAGCATAATTCAGCAGAGTGTTAGCA
GGGCCTATCTCGTGCAGACCTCAAGACCAGCGTACTGGCTCCCCCAACCCTACTCTGGTTGATGTCTTAGGATTTCACCACTGTTTTTTGTCCTCCAGCGTA
GCCTTGAATCCTGCCAAGCCTCTCTCAGTTTTACTAAACAGATGATAAAGCCAGGAATGAGGAAGGAGAAGCCAGGAATGAAGAATGAAAACATGCCAGTG
GTCATGGCAGTGGAAGGTGAATTAGGAAGTATCAGGGAATGGCAGCGGGCTGGCGAGCAGGTTGGAGGAGGAAGCAGGCTGCCCTTTGCACTTTGGAATA
GGCAAGTGCTCTGGTCTCATCTGGGGGCTCCCACTGAGGTGGGCCTATGACACACACATTCCACAGAGTCTCCCATGCTCCCCCACACAACTCTAGGAGG
GTCATTGTACTTCACAGCTAGCAAACCAAGCCACAGGAGGTTACATTATTGGCTCGAGGTTCTAAAATAATAGTCTCTGGCTAGGAAGATGGGCTTATATTTA
TCCCAGTATACTTTGTAAAGAATTGCTACATTCTTTGATCAAACTAACAGGAACGTTGTTACCTTCTGCAGAGGCTTTTGTGAGTCTGTTCCTAAGCTCCCCG
GGAGCCTACAAAATGTCTTAAAAGTAGTCTTGAGATTTGTACACGCCAGTTTCCTCAACTCTAAAGCAGGGTTGGAAGGCCCAAGGCTTGCTTTTCTCCTAC
CCAGCCCAGAGGGGTTGGGGTGGAAGCCAGGATATCCCTTCCACAGTGATAAACAAGGACAGAACTTAGTCCCACCCACAAGTTCCAAGAAGTCAGTAGG
AAGTAGCTGAGGAACTCCAAAGCCACAAAGGCCTTCCTTAGGGCCCAGTGGCCACGATCAGGCTATTAGCCATCTCTTAGGTGTCCCATGGGTTTCATCTA
GCCGGTACTGTCCCCAGCTGTACCTCATTTTCCAGACCTGGAAATGGGACTAGCATGTGGGGCATATTACTTAACATGTAACATTCCTGCCCAAAGTGCCCT
TCCCGAGATTCAGAGCCTGGGGGCACAGTCCTGCCGGCAGGCCTCCTATCTATCTTGTCAGAAGGTCTGACTCCTGAAGTGACACACAGCTCTAAGATTCT
GGCCCTGTCTGCTTCAGTAGAGTAAAAAAAGGTGCTGCTTGAAGTTGGGCAGGCAGATGTACAGCCAGGCCACAGAGAGACCTAA

372-m13rev_4.seq|Chr7|-1|101210453|101212912
ACTTAGGGCTATGCTTGGGAGTTCTGGGTAGCCTCACACAGTCATAGTAGTCCATCCTTATATCAAGATTTGGAGACAGTTCAGGGGACACTTGGTCCCTCT
GGGACTCACTTGTAACATGGAAGCTGTCACATATATTCACACACGCTATATCCACACACACTATATTCACACACACTATATTCACACATGCTATTTCCAGACTT
TGAAAGCTTTACAGCCCAGCAGGGATCCTAGCACTCTCTGCCCCTGAACTTGGCAAACATCCTCCTCCTAGACTGAGCTTCTCTAGGCAGTCTGCTCCATAC
CAGCCCCTGAGGAGCCAGGCTGCTCTGAGCCCTCCCAGGCATGTTAGGTCACCTTCCTTGTCTGGAAGTTACAGTAAGCCCAGGCGTCACATTCACCTCTG
CCTCCTTAGACTCAGCACAGAGCTGGAACTGGGGGCTGCGAGTTAGGTGGTGGGATCCAGTCACACACTTTCATGGACTAGAAAAGCAAACGAGGCAGAG
GGAAATGACTTCCTCAAAGCCATGTTTTATAGGTATGGGTTGGAGTTCATTCTTGAATTCTTTTGATGCTGGTAGAACCCACAGCTTTGGCCACGCCAAAGCT
CTGTCCTCCTCTTGCCTGAACTTTAAAACTCTCTTTAAAATGAAATGTATTAATGCTGCAATGTGTTTACTTGTGCATGCATGTGTGTTTGTGTGCAGGTGCAT
GTGTGTGCATGTATGTGTATGTTCCAATCATCTTATTCAGAGGCCGAAGGTGGGGAAGGGGGTGAAGATGTGATCCGGTGTCCTCAGCCAGGCAGATCCAT
CTGGCCGCCTCCTCACTGACAGCTGGGACAGGCCTCTGCTAGGAAGGTGAGAATCCGGCTACTATTGGGAAGGGGTGGTGTGGGGCTCAGAGCAGCCTG
GCTCCTCAGGGGCCAGTGTGGAGCAGGTTTCCTGGAGACACTCAGTCTAGGAGGGGGATGTTTGCCAAGTTCAGGGGCAGAGAGTTCAGTGAGGACCAA
GGCCCAGAGCCAGACAAGAACACAGTCTCCTGGCTGGCTGAGCCTGTCCACCTGGCTCAAGTTTCTGTGCCATGGCATACATACTCTAATACTCAAAATGT
CTCTGGTTTCAAATATGCTAGTACCAATGGATGATTTCAGGTCTGGGTCACAGAAAGTATAGGATGAGCCTGGGACATCTTACTGAGCTAGAAAGATCTTACT
CAGTGAATGACAAAGACCTTGTCCCGAAGACATGAGAGCCAGCCTGAAGACTTGACTAGCTTCTCAGGATTTACCAGGCCTCTTGAAGTTTTACCACTGAAA
GTAGCCAATTCTGGAGGCCACAGTTCCTGGTAAAGTAGGTCAGGCATCACTCTTCCATGAGCTATACATCTAAGACAAGCTGATTAAAATGAGTAATGAACA
AAACAGTAATAACAGACTATATTGAACCCAATAAGAATCCAAAGTAGCCTTTCCAGCTTGGGGCCAGCAGAATGGTTCCCGCAAAAAAGGATGGCAAGAAGA
AGGGCCATCAACGATGTGGTGACTAGAGAATACATCACCAACGTTCACAACCGCATCCATGGAGTGGGCTTCATGAAGCGTGCTCCTTGGGCATTCAAAGA
AACCCGGAAATCCTTCATGAAGGAGATGAGGACTCCAGATGTAAGCATTGACACCAGGCTTAATAAGGCCGTCTGGGCCAAGAGAATAAGGAATGTTCCGT
ATCACATCTGAGTGTGTTTGTTCAGAAAATGTAATGAGGGTGAGGATTCACCAAACAAGCTCTACACACATTCAAAACTCTACAGATGGTCAATCTGGATGAG
AACTAACCTGCTGAATGTCAGATAAGGTTGCAGAACAGAAAAAAAAGGACAGACAGACAGACAGAAAGAGAAAGAACTAAATTGAAAAGAAACCAAAGTATA
TACAGTTCTGGCAAAATAGATGCAATGTATGATCAGGAATGAATTATGTATAGTTTCAAGGCACCGCCTATCCAATACTTATTCATTACAAGGGAAAATAAGAG
CCTGAGAAATGGCTTGTAAGTAAAGCACTTGGAATGCAAGTGTGAAGGCCTAGGTATGGTCCCCATCACCTATGCAAATGCTGGACAGGCCTGGCAGCCCT
CTGCCATCTACGAGACAAAGCTGGGGGATTCCTGGGATAAGTTGAACTGGTGAGCTCTTGGTTCACAATGAATCCTGCCTCATTAAACAAAGTGGAGAGTTA
TCAAGTAGGACATTCCATATCATCAACTCCTGGCCTCTTACACACATGTGCACACGCATGTGCATCCAAGTACAGGGGCAGAGCCTGCAGTCAGGAACAAG
GCCCAAAGCCAGACAACAGCATGGCGTACTGAGTTGCTGAGCGCGTTTCTGGCATGGTGTACATAGCCTAATATTCAAAAGGCCCCCGGTTCCAAACACAT
GAAGTGGGAGGAATAACTT

398-m13rev_2.seq|Chr18|-1|42400232|42402716
CACCCAAGTGCTAGGATTATAGAGGTGGGCTACCATGCTGCCCCAAATTTTAACACCAGTGCCAAGCCACTGATTTCAGAACACCAGTCAGAATTTCTCATC
TAATAAAAACTCAGTTACCTATTTTTGGCATATTCCAGGGTTTTTAAATTTTTATTTATTTATTTATTTATTTATTTTGCCATTATTCATTCTCATTCTCTCAGTCTC
CCCCCTCCCCCATGGTTTCCCAATGCCCCATTTTCGCCATTAATTTCCTTTTAAATGATGTTTGCTGATGAAAATTCCTCCTGAAAAAAAAAAACAAATAAAAA
TATTGACATTTAGAGATGCAATGATGAGATATTAGTCTGGTGCCTTTAAATCCAGCACTCGGGAGGCAGAGGTAGGTATCTGAGTTCGAGGCCACAGTTCCA
GGACAGCCAGGATTAGATGGAGAAATCCTGTCTCAAGAAAAACAAAACAAGCAAACACACACACACACACACACACACACACACACACACACACACACAGA
GAGAGAGAGAGAGAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGATTTTGTAATTGAAGCTAAGCAGCAAAAAATCAGAAGTATGAACGCCGATCATC
AAAATAAGTGTGTAAGGAAGGAAAAAAAAATAGCAAGATCTTGCTTCTTGGCATTATGTCCAACATAAACGCACCGCACGTTGTGTTGCAGATTTACTACAGC
TGAGTCATTTCGGGACATCGGAGCGAGTTTCCAGGAAAAAAACCAGCGGAAGTAACCAGGAAGAGGGGCTCTCCGGGTTCCCCGAGGCACTCCCAGCAGC
AGACACCGGTACTGCGGGCGGCGGTGGCGGCATGGTCAAAAGAGGAGTCGTGGACCCTTTAAGAGACACAGCACCTTCTCTAAGATACCATCTTGCACCG
GGAGTTCACTGAACTCTCCACCTACCTTGCTAGGTCCCTCCGCCCTCACTGGGTACGGCTTGCGGGCGGGGCGAACGAGTCTTACGATTGGCCTTCTGCG
CGCGGCCCTCTACCGCCTATTGGTTAGTCGGCCCGTCACTCACCCTACCTCCCGCCCACCTGGGCCTCCGGGCTAGGTTGAGTGGGGCTCTTGGGTTAGT
TCCTGTTAGGCCCCGGCCCAGGGAGTAGGTGAAGTCTATAAGATGCCCGGTGGGTCAGGGCACTCGGAGCTGTGAGGGAACGTGAGGCGGCGGCGAAC
GTAGACCCGCGGCGGGCCTTAGGAGGAGCGGCGGCCGAGCCCGCCTTCCCTGCACCATGCTCATAGAAGATGTGGATGCCCTCAAGTCCTGGCTGGCCA
AGTTACTGGAGCCGATGTGAGTGCGCCGGGCCGCGCCGGGCCGGCGGCATGGGCGTTGGGGGACCCGCGGGGGCGTGGGGGAGGGGAGGCGGGAGT
CCCGGCCGCCCGCCTGGTCCTAGTCCGCTGTGGTGTGAGGTGTGAGGGGGCCGGGGCCGGGCACGGGCGTCGGGGCTTCCCGAGTCCACACACACACC
CCGGGAGCCGGCCCGGGCTGCAGGCCGGCGGGAGACTGAAGGACGGTGGCGACCCGGCATCTGAGGCGAGTGGCGCGACTGTCGAGGGATTTTTAAAT
GTGAAAGCTGAGGTGTGGGACGGCAGAGCGGGCTGGAGAGGACTTTGCAAACAGTGTTCAGGAGCCCCGGTGTCGAGGGCCTTCTCAGCACACTGGGGC
AGCGAGGGGCAGAGAGGGCGAGCGGGGGTCCTGAAGCTGAGTCGTAATTAGGGGGGAGCTCCGACGGATGAGGATATTTAAGGATTTGAAAGACCTCGG
AGTGCTGAAGTGGATATGCGGGCGTGGAAGGGCTTTGGTAAACCATTAAGGAGCTCCTGGGTCCGAGAGCAAGGAAACTGAGGGATTTTAAGGTATTCTTG
AGGAGTTGGTGGTCGAATTGGTGATTTTAAATTCCGTAGGATGTTGGGGGACGGAGTCGAGATTTTTATGCTAATTGAGGTTGAAAGGACTAACTAGTGTGT
ACTATGGATGATGGTGAGGTTGAAGCGGGTGTACTTGCAGTATCTTTTTTTTTTTTTTTTTTTTAAGATGCAGGTATTAGAGCAAAATAGGCACAAGAATAAAT
AAGGACCTGGCAAACATTGTTGAGAGATGACAGGGCGACTATAATTGGGGTGTAGGGTGACTGTTGAGGGGCAAGGAGAATGAGAGACCGGTTCTCTTAA
GATATTTTGGGTAGTGTCAGTTGTACAGATATGGTGAGGCTTGACGAGGACAGCAGAGGGGGTAACTTGGTTTATAACTTTTAAGAAAGTAAAGGCCGTGTG
GAGTGTTTGAGGATTCTTGGGTATAGCTAGCCTCTAGCACAGTCTTGTGAGAAAGAAAGGATAGAAGGCAGGAGAGACCAGTCTTTAGCCGGTTCTTTGAA
GAGAGCATGGGAAGCCGAATAGGAGGGAAAGAAGCCCAGCTTGCCCCCTTTTAATTTCTGTTTCTTC

73-m13rev_4.seq|Chr8|-1|85008850|85011307
TTCCTGTTTGAATCCAATCCAAAACAACTAGTATTAGTATACCAGGAAGTTTGTGGGTGATGCTTCCTCTGAACAGAAAAGTTACAAGACTCCTTCCTCCACC
TTCTGTCATGAGATTTCCAGGAACACTTTCCTAATCTCACTTAGGCTCACACACTCTTCTCTTTCTCTGATACTTTATCACACGGTGCTTTGAGCCAGATGTTA
TCATTTTCCTTTGACCTACTCAAGTGGAAACTTCATCAAAGCAGAAATTTCATTATCCACTGTTTTATTCCTTTTCTGAAGCAGTGAGTGGCCCATGACAATAA
TTTAATGAGTGAATAAAATGTTCTCTGAGATTTACCAACAATGTAAGTAATTCAAAGCCAAAAATAAATAATACAGCACAAGATGCCTTATCACAGACACATTC
AAAGGTATTGATACTTTAAAACTGACTGTACTTTGAGGCTGGAGAGATTGTGACCTGGATAAGTTGCATCATTTTTTAGTGACAGGGTTCTGAGAGAGAGGAC
CAGGAAGTCAGGAGATAAGAAGATAGGACACAAAGAAGATACAGCATAGGGTCAGGAGGTCTTGCTCAAGTTCTGTCTGATGCCAACCAAACTTATTAAGAA
ATACAACATTTTACATACTAGTTCCGGGGAGAGGGGTGGGGGGGGGACAGAGTTCAAAGAATCTCTCATACAATGGGACAAAGTCAGCAGGAGGCTAATCC
AGCAAAGTTGCACAAAATGAGCACAGGCATTACAAACTGAGTGCACAATCACCTTGCGTCTGATAACCCCAGTCTCTTTAAAGAGAAAAGCCAAGCTCCCAG
GACAAACCTTAACTGCTGTGAAGCTCTCTACCTGTAGCTCTAGAGTGGCCTTGCCCAGTCTCCTGAGCAAGGAAAAAGGATTGGTTCTTTTTGTCCTCTCAT
CAGGGCTTCTGAGAAAGCCTTGGATTGGTCTAGGTCTCAATAACTCCCCTTGTTTACTTCTTGAGGTCAGATGAATTTTGGTGGTTGAGTGTTTAAAAGACTC
ATAAGTTCTCTGAACTCACAGATATCTCCTACTGAAATCAAATGGATTGTTAACCTAAGTATAGCAACACTGAGTATAACACTGAGTGGATGTTGAAGGATGA
AAAGAGAGTTTAGATTCCCATAACTTGTCCAAGAAATTTTGGGGTAGCTTAGAAGGGGAATTTACATAATTGGTAAAGGAACCATTGATAAGGTAATTTAGAA
GTGTCTCTCATTTTCTGAAATAGAGGAACATTTTCTGAAATAGAGGAACATCAGACATGAGTCCAAATGTAAAAGGATTTCCATGTTATAGTAGTTCTTCCCCA
CCCCACAAAACTCTCTTAGACAATCACACTACTCAGGGCTGGGAATCCCAGCCAAAGATCCAGTTGAGCCTCCTGCCCTACTCTCCTGGACACATTATAAGA
GGATGTATCCAGTGAGGAGATCTTAAGAGGTTATCTTCATTCAGACCTGGTGGGTGCCTCCAGGACTAGTCAACCAGCACTTCAGTTCACCTTCTTAGAGGC
ACAGAGTTCTTTTGAATCATGAAGCATGTCCCCAGGGGGGACATCATCCTATAACCAGGGACCCAGATTCTCACTGCATTCTGTGAAACATCAATCGAATGG
GGCAACCTGGATCATGAGAAACTATAAAACGCAAGTGATTAAATAAATTGACCATGTCTGAGAAACTGAACTTTAGACGTCTCCTGGATTCCAAAGCCATAGG
CTATACTTGTCAAACAGACAGAAGGTTGTACCTGTACTGACCCACCAATGACAAACAAGCAGGAATGGTTGATCATACCCCAGTCTGGCTAAAGAGCAGTGA
AGAACTTCACATGAGAAAAACCCGTAAGGTCATACACCAGTCTGATATATTTAACTATTGGGTAATATCTTAAGGCATTTTTATTAGATATTTTCTTTATTTACA
TTTCAAATGTTATCCCCAAAGCCCCCTATACTCTCCCCACATCCTGCTTTCCAACCCACCCATTACCACTTCCCGGCCTTGGCATTCCTCTGTACTGGGGCAT
ATGAACTTTGTAATACCAAGGCCTTTCCTCCCATTGATGGCTGACTAGGCCATCCTCTGCTACATATGTAGCTAGAAACACAGCTCTGGGGGTACTGGTTAG

TTCATATTGTTGTTCCTCCTATAGGGATGCAGACCCCTTGTGTAATATTTTAATAGCCAGTTATTTTAAAATAAACCTTGATTTTTACATGCATTATAGCCAGGT
TTTTGTAACTGTTTTGTTGTTGCTTTTTTTTCTTGTTTATTTTTTGTTTTTCGAGACAGGGTTTCTCTGTAGCCCTGGCTGTCCTGGAGCTCACTTTGTAGACCAG
GCTGGCCCTGAACTCAGAAATCTGCCTGCCTCTGCCTCCTGAGTGCTGGGATTAAAGGCATGTGCCACCATGCCCGACTTTGTAATTGT

721-m13rev_4.seq|Chr7|-1|97355236|97357517
TAAATAATAACAGGAACTCTATTCTACTCATTGTTCATGTTCTAGACTTATTTCCAGGGTCTTGATATTTGTACTTGCATTTTTTAGGCAAAATGAGATTTATCTT
TCTATTATGTAGACTGTTTTCCAAAGAAAGACTACAAAGACCTTCTGAGAAGCAGACTTAACTTGAAAAACAAGGTTTTCTGAAACCTACCAACTACAATGAAG
TTAAAAACCTGTAATTGAGTTTATGCTAAATGCTAATCATAATAAAGTTCCCTTTTTATTAAGCTTTGCTTGCTTGCCAAAGTCATAAAGAACCTTTGTAAGACT
TTGAGGACACCTCATCCTTCTACTGTGGGTGTGTGTGTGTGGGTGTGTCTGTAATGGTAACATTTTTGCCTTTCTATATCTCAGTGTTGGCTCCTGGTTTTTG
ACACCACAACACGTGGCTTCATCCTCAGTCTCAAGATTCTTTTTCACTGTTTCCCAGCTGTCAGAGGAAGCATTAGTCTCTTCACCAGAGATTCGTGGAACAC
ACTGACAGTCCTGATAGAGACTTGGCTCCGGGTGGACCACCAGCATTCTTAAATCTTCAGACTTGCAGAGTAGTACCATTCAGAAAAATGATCTTAAACCTTT
CATTGATTCAACAGGAAGATTCTTGGCTTTCTTTTCTTTATCTTTCCGTTTTTCTTCTTCTTCTTTCTTATTGATGGATAACTTCTAAATTACTGTAACAAAGTAA
AACGAACTTCGGGGAGGAGCTGGAACTATGGCAACACTCACCCTAAGATTCTAGCAGGGCACCTCCAGGATACAGCTCCGGGTTTTGCCTCCAGCTCCCAC
CAGCCCCGCCTTTTACTCCACCCATCTTCCAGGATGCACTCTAATTTCCGTCTCCCAGCAGTAGCGGCAGCCAATCGACTCACTGGGCAGTGGAGCTTCAG
GAAGAGGCCTCTCTATCCAGCAACCCCACTGCTTTGTGGGATTGCAGCGAAATCGGTGCTGTAGAAAGGTTTCTCTCCACGTTCTCAGCAAGTACAGAAGC
CGAGAAGGTCCGGTGCAGCCATGGCGGCGCACCTGAAAAAGCGGGTGTATGAGGAGTTCACTAAAGTGGTTCAGGTAAGCGCTGGCAGAAAAAGCCATG
GACCAGTCTGCGCGGGAGGGTTGTCCGCACCCAGGCTGTTCTCCAGATCAGCGCCTAAGCTTGTGTGTAGCCGGCTTAGGTTTAGGTTCCTAGGCTGCGC
GTCAGGCAGGGGGCGTGTTAGGTCGAGGCCGCTGTTCTCATGTGTACTCTGACTTTGGGGTTTGTTGTTGTTGTATTTAAAGTTCTGGGACTAGAATTC
GGGGTCTTAAGCATGCTAACTTATCACTCTATCCCAAAGTCACACTCCCAGCTCACTTGATAAAGACAAAACAAAATCATTGGGACAACGTGCTGTATAGTTT
TAAAAACTGTCAGGTATCTTATTTTAATGTCTCACAGCTCGGTTGTGCCAGATCGGTTATATTTGCATTCTCCTGTCCTCAGTTGGTGGCAGAACTCCTCTTAA
CTCGTGTACCTCTAACTTTCAGAGACAAGTTGAGCTGTATGCAAAACCCCAAGCTGACTAAAGGTCTGTAGAATTCAAGGCTCAGACCAAGCAGGCAAGAGT
CACAGGACGGCAAACCTATCATCCTCTCTCTTTTAAACATAGGTGGACCAAGAGAGCCCCTGTCAGAACTGTCAGTGTGCTCCACAGTGTCTCTGTAGAAGC
GATGTATTCTGCTTCTTAAATTTCCATATGTACACAGATACATACAAAATATACGCGTCTTGAGGAATACTTAAATTTTTTATCTTTTTTGTAGTAGGGGAGATT
GAGTGTAGGGCCTTATGCATGCTGAGGCATGTGGTCTACCACCGAGCTGTTACAGATCTCGCTCTTTTAGGGAGTTTGTTATGATGCAGATTATGTTCAATA
GGTTGGTGGTGTGAGGCAAGATCTGCATTTTGACCAGGCCCGGGGGTTTATGCCAGTGGATCCCCAGCAGTCTTAGATATTTAGGCAGAGTGGTCTCATTA
AGGAAAAAAAAAAAATTTTTTTTTTTCTTTGAGACAGGGTCTCACTATGTAGAGCAGGCTGATCATGAACAGGAAGGAATCAGCCTGCCTTTTCTTCCCAAGTGC
AGGGATTAAAGGCTTGCATCACCATGCCTGATGGAAAATGACTCTTAAACCTTCCATTCATTCATTAATTCAACAGGAAAAAATTAAAGTATTTTCAAATACTT
GAGTGATTTCTGTGTACCATGT

882-m13rev_3.seq|Chr3|-1|101535707|101538021
TTAATTCTTGCAATGATGGGTCTATGGTTCAATGCGTCCAAAGGATGGGCATTATAGTGTTTGAAGAGGAGCATTCTTGATGTGTGAGCCGTTGCCTTAAGTC
GTTTCTGGAACACTGTGGGTGGGAAATAAAAAAGTCTTTTTTTCTCTGTGTGTGTGGGGAGTTGGAAAGGGTGATGTTCTGGAAAGTGGCTGAGACAGCCC
CAGGTGCAGGCGCCCGGGAGGTGGCCAGGCTGTGACAGAGGGAGAAGGATGGCCAGGGCTGGCTACCAGGTGTTAGGAATGAGTAGAGGTAGGGCCCA
TTCTTAAGTGAGGTCAGGGAGGGGAGAACTTAACTGACTCCCGGGGGGCTAGAAAGAGAGCAAGGGGAAGCCACAAGGAAGGACTGTCTGGGGTGCAGG
CAGAGCAAGCTGAGGTAGAGAATTTATCTAGTGTCTAGAACTGTGTCTCTAGGTCTTGAGTCCAGTAAGACAGAATGAGTGATGGAGCATCCCCTGCTTAGA
GGCAGAAGGAAGACCTTCTTAGTTACCTTTTGTTTTGTTTTGTTTTGTTTTCAACACAGGGTTTCTCTGCGTAGCCCTGGCTGTCCTGGAACTCACTCTGTAG
ACCAGGCTGACCTCAGAAATCCACCTGCCTCTGCCTCCCAAGTGCTGGGATTAAGGCGTGCGCCACCACTGCCTTAAGCCTTAGTGGCTTTCATGAGCTGG
TTCTGGCTCAGGTTGCAGACAGATCTGGGGACCCTGGGAACAGGAGTCCGGGGCTGGCTCAGGCTTCTTCAGAGTATGGTGGACTTAGAGCCAGAATATG
CATGGCAGCTCCGGGCTCCAGATGAGCATCCTAGAAATTGCCTGCAGAAATCGTATAGCTGTATATCTAGACTTGAAGTCCCAGCGTGTGACTTCTGCTAGG
ATCCTAGCTCGTCAGACTGCACGGGGCCCACCTCTCTGGCTGGCCAGGAATCAGTGTCTGTCACCAGTCAGCTGTGTGAGGAAGCGGGCTCGTTGAGAAG
CATACCCTTGGCTTAGGTGGGGAAATGGCGTGACCAGGGGAGCTGCCTCATTGCATGGTAGAAGTAGAGGAACCTGCAAGGCATTGGCCAAGGCTTGGTG
GTGCCAGGAGACAGGAAAGCAGCAGACGACCTGGGGCTGAGTGGGGGATGGAGGAAGTCAGGAGGCATCCTGGGCACGAGCCAGGGTAGACGTGGGTG
TTTGATGTGGGCAAGTGAGGCGGGCTGGGTCAGCCAGTGCTCTTGTTTTCTAGGGTTTTCTCTGCTTGCTCTAGCTACCCCCTCAACCTCTAGTTTCTTTCTT
CCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTTTGTTTTTTTTTTTTGTTACCCAAAGTTTTTCTTTTTTAAGCAGTATTTTATTA
GAAAAGGTTTAAAATTATTTTTGCCCATTTTTACCCTCTCCCTGGTTTACTTTAGCTGGAATTTTTATTCATGTATTTTTGAGTAGGCAATGCATGCACGCTCGT
GTTCTTACCTAAAAGGAGTGTAAGAAAGCTGGGTGGGGGAGACACAAGCCTTTTACCCCATCACGTGGGAGGCAGAGGCAGGTGGATCTCTCTGAGATTAA
GGCCAGCCAGGTTCACAGTGTGTTCCAAGGAGCAGCATAGTGGGACCCTGCCTCAAAAAAACAAAAACAAAACAAAACCAAAAACCAAAAACCAAAAACAAA
ACCAAAAATGTAATCAGGTCTGGATTAAGTCCGAACGCTCCCCACATCTGGGTTCTGTGTATGCCCTCCACCAGTGTTAACTCTTTCGTCAACTGCCCACAGC
TGAGCTGAGCCCATCTTGAGCCTTTGGCTTCAGCCCCCCTAGCATGGAAGCCTTGCCAGGGCCATAGGCATCCCCTCCAGACCAGGCTACCCATCTTCAGA
GAAGGCAGAGGCTGCAGGCTCAGGGATATCTCAGCTCTGGATTCTGCCCAGCCGTTCCCAAACCTGCAGTTCTAAGTGTTGCGGCTTCTAACTCCAGTCAG
AGCCTTCCACGCTGTTCTTCATCTGGCTCTCTGTGGTTTCTCAGAGTTGTTTTGATCCCGAGACCATGAGTCTGCACGTCCCCCTCCCATCACGTAAGGGAG
AGTCTTTTTTCTGGCAGAGAGGAGTGTGTAAGTTAGGAAACACAGGTGCTTTGAATTGCACAGTTCTTGGAGCTCAGATGAGATTTTT
AAGGATTATCTATGTTCCTTTTTTTTTTTTTTTTTTAAAAAAAAAATAGGTATACAGAGAATTCATCCCTAGAAAGTA

9.2 Curriculum Vitae

Stefanie Eva Regina Kliem
Königsberger Str.2
61440 Oberursel

Personal information:

Date of birth:	07.01.1971
Birth place:	Bad Homburg v.d. Höhe
Nationality:	German
Marital status:	Single

Education:

1977-1990 Frankfurt International School, Oberursel, Germany
High school and International Baccalaureate Diploma

1990-1994 Bachelor of Arts, Connecticut College, New London, CT, USA,
Major: Biology and Environmental Studies

1994-1995 Graduate Special Students Tufts University, Department of Biology, MA, USA

1995-1998 Masters of Science. Boston College, Department of Biology, MA, USA. Masters Thesis: Investigation of *fosB* promoter *cis*-acting elements involved in mediating mIgM stimulated FosB expression in mature B lymphocytes. Under supervision of Dr. Thomas C. Chiles Boston College, Departmant of Biology, MA, USA.

12/2001-2008 PhD student at the Georg-Speyer-Haus, Chemotherapeutisches Forschungsinstitut Frankfurt am Main in the laboratory of Dr. Edith Pfitzner under the supervision of Prof. Dr. Starzinski-Powitz, Johann Wolfgang Goethe-University

Work experience:

1998-2000 Research Associate 1, Genaissance Pharmaceuticals, DNA Sequencing Facility, New Haven, CT USA.

2000-2001 Scientific Analyst III, Genaissance Pharmaceuticals, Bioinformatics Department, New Haven, CT USA.

Publications:

Litterst, C.M., Kliem, S., Marilley, D. and Pfitzner, E. (2003). NCoA-1/SRC-1 is an essential coactivator of STAT5 that binds to the FDL motif in the alpha-helical region of the STAT5 transactivation domain. *J Biol Chem*, 278, 45340-45351.

Pfitzner, E., Kliem, S., Baus, D. and Litterst, C.M. (2004). The role of STATs in inflammation and inflammatory diseases. *Curr Pharm Des*, 10, 2839-2850.

Litterst, C.M., Kliem, S., Lodrini, M. and Pfitzner, E. (2005). Coactivators in gene regulation by STAT5. *Vitam Horm*, 70, 359-386.

Scharf, S., Zech, J., Bursen, A., Schraets, D., Oliver, P.L., Kliem, S., Pfitzner, E., Gillert, E., Dingermann, T. and Marschalek, R. (2007). Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. *Oncogene*, 26, 1361-1371.

Kliem, S. and Pfitzner, E. (2008). Regulated expression of milk protein genes during mammary gland differentiation correlates with distinct histone modifications. *In preparation*.

Abstracts:

Kliem S. and Pfitzner E, (2005). Transcription factor recruitment and histone modification during mouse mammary gland differentiation. 2nd Alan Wolffe EMBO Conference on Chromatin and Epigenetics. EMBL Heidelberg, Germany 2005. Abstracts book p. 124.

9.3 Acknowledgments

First I would like to thank my advisor, Dr. Edith Pfitzner for allowing me to do a PhD Thesis in her lab and for her patience, guidance and support during the project.

I am grateful to Prof. Dr. Starzinski-Powitz for the supervision of this thesis and for accepting to be a referee.

My warmest thanks to the past and present members of the Pfitzner lab for providing a friendly work atmosphere and the many stimulating discussions. More particularly, I would like to thank Dr. Daniela Baus and Dr. Marco Lodrini for providing continuous advice, encouragement and friendship.

I am grateful for having had the opportunity to finish this work in the George-Speyer-Haus. I want to thank the members of the George-Speyer-Haus for their support, especially, Dr. Klaus Lehmen for help with diverse computer problems, Margot Landrez for the sequencing of so many plasmid constructs, Sabrina Lehmen for her assistance in breeding of the mice and Dr. Herbert Kühnel for advice concerning laboratory safety.

Finally and most importantly, I would like to thank my parents for their faith in my abilities, their unconditional love and support that gave me the strength to achieve this goal.