ABSTRACT

Title of Dissertation: QUANTITATIVE TRAIT LOCI AND PROMOTER ANALYSIS OF THE BOVINE BUTYROPHILIN GENE

Abiy Zegeye, Doctor of Philosophy, 2003

Dissertation directed by: Professor Ian H. Mather
Department of Animal and Avian Sciences

The butyrophilin gene (*BTN*1A1) has been described in all mammalian species so far investigated. *BTN*1A1 is a likely QTL candidate that affects an economically important trait in dairy animals because it is specifically expressed in lactating mammary tissue and the gene product BTN1A1 may function in the secretion of milk lipid. Five PCR-RFLP intragenic markers were identified. Three of the five markers are bi-allelic, however, the two 5'-most markers may be multiallelic loci. The markers were further used to conduct a QTL analysis to examine any allelic substitution effects on economically important milk production traits, namely, total protein, percent protein,

total fat, percent fat, somatic cell score, herd life, and milk yield. One significant effect was detected for percent protein. Other effects were not significant, but this could possibly be due to the skewed allelic frequency distribution, or because the variable nucleotides were either intronic or coded for a conservative amino acid substitution.

The 5' flanking region of b*BTN* was cloned and sequenced. Computational and transient transfection assays were conducted to identify regions of b*BTN* that are important for its expression. A computational analysis was conducted to compare and analyze the 5' flanking region of cow, goat, human and mouse sequences. Several regions of homology were identified that code for shared binding motifs. However, neither the transcription start site (TSS) nor the other segments of similarity in the proximal promoter region were analogous to other 'milk protein genes' such as the caseins and α-lactalbumin, and other widely studied non-mammary-specific genes. Transient transfection assays in HC11 cells indicate the region from -1kb to -0.45 kb is sufficient to drive the expression of a reporter gene. b*BTN* appears to contain a novel set of elements that define the (TSS) and proximal promoter modules. A more incisive examination will circumscribe the major *cis*-acting elements.

QUANTITATIVE TRAIT LOCI AND PROMOTER ANALYSIS OF THE BOVINE BUTYROPHILIN GENE

by

Abiy Zegeye

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:

Professor Ian H. Mather, Chair
Professor Albert Ades
Professor Stephen M. Mount
Professor David A. O'Brochta
Professor Tom E. Porter

ACKNOWLEDGEMENTS

I am indebted to many family, friends and colleagues that have been supportive through the years. I thank Dr. Ian Mather for his guidance and mentorship and all the committee members who were always available to consult with and offered understanding and support. Dr. Larry Douglass was of great help in deciphering the statistical analysis. The many graduate and undergraduate student lab mates all deserve mention for making the work place pleasant and provided a respite from the routine. Of notable mention is Dr. Sherry L. Ogg for her patience and guidance throughout the years. Friends and family have been a constant source of encouragement, assistance and understanding. Many in the department have been a source of ideas, lab material, and friendly advice. I convey special thanks to Elizabeth Yoseph for being a true friend and counselor. Gratitude is due to my father, brothers, sisters, aunts, and in-laws for their constant belief and trust in me. Most of all, I would like to thank my wife, Fikir, who has been my unending source of energy, comfort, friendship and support. Tsinu, my son, has made all the effort worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	artificial insemination
AL-CAM	activated leukocyte cell adhesion molecule
ADRP	adipocyte-differentiation related protein (adipophilin)
BoLA	bovine leukocyte antigen
BRE	TFIIB recognition element
BTA	*Bos taurus* chromosome
bBTN	bovine butyrophilin protein
b*BTN*	bovine butyrohilin gene
BTN	butyrophilin gene
BTN	butyrophilin protein
CAT	chloramphenicol acetyl transferase
CDS	coding sequence
CLOX	cut-like homeodomain
CPRG	chlorophenol red β-D-galactopyranoside
DBDR	dairy bull DNA repository
DDEV	daughter deviations
DHIA	Dairy Herd Improvement Association
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DPE	downstream promoter element
DTT	Dithiothreitol
ER	endoplasmic reticulum
rER	rough endoplasmic reticulum
EVI1	ectopic virus integration site 1
Ermap	erythroid membrane-associated protein
FA	fatty acid
FABP	fatty acid binding protein
FBS	fetal bovine serum
FMF	familial Mediterranean fever
GAS	γ-interferon activation sequence
g*BTN*	goat butyrophilin gene (g*BTN*)
GDD	granddaughter design
GNIP	glycogenin interacting protein
GR	glucocorticoid receptor
H	hydrocortisone
h*BTN*1A1	human butyrophilin gene
I	insulin
IE	insulin and epidermal growth factor
Ig	immunoglobulin
IHP	insulin, hydrocortisone and prolactin
Inr	initiator
JAK	janus tyrosine protein kinase
LAMP	limbic-system-associated membrane protein
LAP	liver-enriched activating proteins

LB	Luria-Bertani
LIP	liver-enriched inhibitory proteins
MAC	mammary alveolar cells
MAS	marker assisted selection
mBtn	mouse butyrophilin protein
m*Btn*1a1	mouse butyrophilin gene
MAS	marker assisted selection
MEC	mammary epithelial cell
MFGM	milk-fat globule membrane
MHC	major histocompatibility complex
MTs	microtubules
MOG	myelin oligodendrocyte glycoprotein
N-CAM	neuronal cell adhesion molecule
NF-1	nuclear factor 1
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
PIC	preinitiation complex
PMN	polymorphonuclear
PP17	placental protein 17
PRL	prolactin
QTL	quantitative trait loci
RBC	red blood cell
RBCC	RING-BBox-coiled-coiled
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RPTK	receptor protein-tyrosine kinase
SDS	sodium dodecyl sulfate
SoS	segment of similarity
SPRY	sp1A/RYanodine domain
SSC	somatic cell score
TEB	terminal end bud
TIP-47	tail-interacting protein 47
TF	transcription factor
TGN	trans-Golgi network
TLC	thin-layer chromatography
TSS	transcription start site
WAP	whey acidic protein
XO	xanthine oxidase

INTRODUCTION

The purpose of this project was to examine the possible involvement of the mammary specific gene, butyrophilin (*BTN*), in lactation and its potential gene level regulation. *BTN* has to date been described in several mammalian species. In those species tested, *BTN* is expressed primarily during lactation, and also immediately preceding the onset and following the cessation of lactation. A natural upshot, therefore, is to ask what lactation-associated trait(s) it might affect, and what control element(s) confer it mammary/lactation specific expression. Answers to these questions will provide potential tools to customize the composition of milk by either selecting for a 'desirable' allele of *BTN* or introducing recombinant genes under the regulation of the *BTN* promoter.

Members of the class Mammalia are distinguished from other vertebrates by the possession of mammary glands. Natural selective pressures may have led to the evolution of apocrine sweat glands in the skin into a gland that secretes water, nutrients and antibodies solely for the sustenance of the neonate (Oftedal, 2002). Processes that allowed the newborn to suckle and digest milk also co-evolved. In a natural setting, the mammary gland's capacity for milk synthesis tapers and finally ceases, coincident with the infant's diversification of diet and thus reduced dependence on milk.

In contrast to other mammals, the customary use of milk in the human diet has two distinguishing peculiarities: the use of cross-species milk (cow, goat, camel, *etc*.) and the consumption of milk past infancy, and in some cases into adulthood. The nutritive value of this 'unnatural' dietary habit is not a matter of contention. In fact, milk and other dairy products are an integral part of the diets in many parts of the world. This demand

has led to the multibillion-dollar dairy industry of today.

In order to meet the growing demand for milk, improvements in dairy animals has mainly focused on selection for increased production, higher feed conversion efficiency and health. Traditionally, this was achieved by measuring quantifiable traits and preferentially breeding those animals that meet or surpass a set threshold.

Today, technological and scientific advances provide tools, not only to improve the quantity of milk production, but also to customize (alter) its composition. Molecular biology tools, such as genetic markers, have given researchers the means by which to select for desired traits. Further, the potential economic windfall from the expression and harvest of large quantities of recombinant 'pharma-proteins' is becoming increasingly apparent. Most recently, achievements in the cloning of ungulates (Campbell *et al.*, 1996; Wilmut *et al*., 1997; Cibelli *et al*., 1998; Kato *et al.*, 1998; Baguisi *et al.*, 1999; Kubota *et al.*, 2000) by somatic cell nuclear transfer makes the mass production of such custom designed dairy animals promising. Interestingly, the viability of cloned embryos derived from cattle fibroblast cells increases after several passages of the donor cells (Kubota *et al.*, 2000); and unlike sheep clones, cattle clones attain a phenotypically youthful state as compared to age-matched cells from whence the nuclei were derived (Lanza, *et al*., 2000).

In the present study, the mammary-specific bovine butyrophilin (b*BTN*) gene was used as a genetic marker to investigate allelic substitution effects on several quantitative milk production traits, including milk fat content. b*BTN* is a good candidate to serve as a genetic marker because of its exclusivity to the mammary gland and its purported role in milk fat secretion (for review see Mather, 2000). Of the major components of milk, milk

fat is economically the most expensive and also the most variable respondent to environmental factors (for review see Bauman and Griinari, 2003). Therefore, identifying a genetic basis by which to manipulate milk fat content has the potential of reducing production costs and meeting the increased demands for low-fat milk.

Not only is BTN expression limited to the mammary gland, but it is also dependent upon the physiological state of the mammary gland. BTN message is observed in the period prior to and during lactation (Banghart, *et al.*, 1998; Ogg *et al*., 1996). The tissue and temporal limitation manifest of *BTN* is indicative of a well-established control mechanism that directs its expression. Therefore, delineating and deciphering the *cis*-acting regions that confer specificity to *BTN* expression pattern can potentially be used to drive expression of exogenous genes. The potential commercial benefits of using such a lactation-specific promoter offer new possibilities to utilize the mammary gland as a bioreactor for customized milk.

The onset of lactation is a culmination of a series of developmental phases. The events that define the development of the mammary gland are analogous in most true placental mammals (Eutheria), although species-to-species differences do occur. For the developmental biologist, the mammary gland is a made-to-order model system since most of the developmental changes occur post-partum in the adult, during puberty, gestation and lactation. Another defining feature is its dependence on endocrine hormones for development and differentiation.

At birth, the anlage of the mammary gland comprises a rudimentary ductal structure. The surge of hormones that accompany pubescence further increases the branching and expansion of the ducts into the mammary fat pad. Cells at the terminal

buds of these ducts are populated with precursors to two types of cells: secretory epithelial and myoepithelial cells. Another round of proliferation is initiated during pregnancy and continues through parturition, which involves extensive arborization of the ductal system and formation of lobulo-alveolar structures. The alveoli are roughly spherical structures lined with terminally differentiated mammary epithelial cells (MECs) with their apical surfaces facing centrally located lumini. At parturition, the differentiated MECs synthesize and secrete milk into the lumini that subsequently leaves the gland destined to be consumed by the neonate. In most species, the sequence of growth and differentiation is asynchronous, so that at any given time the gland is populated with cells at different stages of development (Robinson *et al*., 1995; Andres *et al*., 1995).

Milk synthesis and secretion rapidly shuts down at the onset of involution. Upon cessation of lactation, the secretory MECs undergo apoptosis; this major remodeling causes the mammary gland to revert to a state morphologically comparable to that of an adult virgin. The cycle of proliferation and differentiation followed by synthesis and secretion of milk recurs with each subsequent gestation and parturition (for review see Dosogne *et al*., 2000).

Milk is a heterogeneous mixture of a colloidal suspension of proteins and milk fat, in an aqueous medium containing lactose, minerals and other minor components. The proportion of each constituent differs from one species to another, reflective of the needs of their respective offspring. However, except some minor differences (mentioned below), the qualitative make-up of milk is very similar among mammals.

Lactose is the major sugar in milk and a crucial osmole that draws water into the

milk that is being synthesized. This sugar is unique to the mammary gland and a source of readily available calories for the newborn. Another major caloric source is milk fat. Almost all milk-fat is in the form of globules of triglyceride enwrapped in a bilayer membrane called the milk-fat-globule membrane (MFGM) (Welsch *et al*., 1990).

There are two classes of milk protein: the caseins and the whey proteins. Caseins make up about 80% of total milk protein (McKenzie, 1970 & 1971) and are found as micelles in suspension. Two of the caseins, α-casein and β-casein are calcium-binding phosphoproteins that form the core of the micelle; while κ-casein forms the outer casing that stabilizes the micelle. The micellar structure is what makes caseins water-soluble. Lowering the pH of milk removes the calcium ions from the caseins thus making them water insoluble phosphoproteins. After casein precipitation, the remaining protein fraction is made up of what are known as whey proteins. The major whey proteins are α-lactalbumin, β-lactoglobulin (found in ruminants, pigs and some other species) and whey acidic protein (WAP) found in rodents. Minor components such as immunoglobulin, lactoferrin, transferrin and serum albumin are also present (Levieux and Ollier, 1999).

Another source of proteins is the MFGM. The MFGM has both integral as well as peripheral proteins associated with it. Because proteins associated with the MFGM do not comprise a large proportion of milk proteins, it is unlikely that they contribute much to the nutritional value of milk. However, some MFGM proteins may play a protective role against possible infections in the neonate (Peterson *et al.*, 1998; Peterson *et al.* 2001). Three MFGM associated proteins: namely, the integral membrane protein butyrophilin (BTN), the soluble protein xanthine oxidase (XO) and the lipid-embedded protein adipocyte-differentiation related protein (ADRP) (Heid *et al*., 1996), could play a

structural and/or functional role in lipid secretion. TIP47 (tail-interacting protein 47; also known as PP17-placental tissue protein 17) is a close homolog of perilipin and adipophilin, which is associated with intracellular lipid droplets as well as being one of the components of the MFGM (Cavaletto *et al*., 2002). Although its role in milk-fat biology is not defined, TIP47 retrieves mannose-6-phophate receptor from lysosomal compartments back to the trans-Golgi network (TGN) (Diaz and Pfeffer, 1998). Other proteins found in the MFGM are CD36 (previously called PASIV) (Greenwalt and Mather, 1985) and fatty acid binding protein (FABP) (Bohmer *et al.*, 1987). Co-immunoprecipitation data indicate possible interaction between these two proteins in the MFGM (Spitzberg *et al*., 1995). The development of CD36 null mice (Febbraio, *et al.*, 1999) is further evidence of the importance of this protein in the metabolism and regulation of the cytoplasmic fatty acid (FA) pool. Proteins such as the heavily glycoslated integral protein mucin 1 (MUC1), lactadherin [also known as breast antigen 46 (BA46), or periodic acid Schiff 6/7 (PAS6/7)], and trace quantities of other soluble and membrane proteins, conceivably arising during the secretion of milk-fat droplets complete the list [for a review see, Mather, 2000].

Despite the fact that the synthesis of milk constituents is accomplished by the MECs lining the alveoli, not all constituents are synthesized in the same subcellular location, nor do they all exit the cell by the same mechanism. Milk proteins (soluble and colloidal) and lactose originate from the rER and Golgi complex, respectively, and exit into the alveolar lumen via exocytosis from secretory vesicles (Sasaki et al., 1978). On the other hand, milk lipids exit the cells by an evagination process directly from the cytoplasm. Notwithstanding other cell types which also produce and accrue lipid

droplets, the secretory epithelium of the mammary gland does not utilize the conventional secretory pathway for fat droplet secretion (Mather and Keenan, 1998). Hence, efforts to understand the extrusion process of lipid droplets from MECs has not benefited from tissue comparative studies.

The synthesis of neutral lipids is imperative for lactation, as evidenced by the di-acylglycerol transferase knock-out (Dgat-/-) mouse in which lactogenesis was wholly arrested (Smith *et al*., 2000). There is now growing evidence that synthesis of fatty-acids and their concomitant esterification to glycerol takes place in the vicinity of the ER; quite possibly in the hydrophobic core of the ER membrane (Zaczek and Keenan, 1990; Keenan, *et al*., 1992). Once synthesized, the triacylglycerols nucleate, by an as yet unknown mechanism, to form droplets enclosed in a proteinaceous material and conceivably the cytoplasmic amphipathic monolayer of the ER membrane (Zaczek and Keenan, 1990; Wu, *et al*., 2000). Therefore, membrane spanning proteins with hydrophilic lumenal moieties, such as BTN, are presumed to be excluded from the monolayer surrounding the lipid microlipid droplets. Mammary lipid droplets are associated with the chaperone/translocon plug BiP (Ghosal, *et al*., 1994), ADRP (also known as adipophilin) (Heid, *et al*., 1996), the calcium binding chaperone, calreticulin (Ghosal, *et al*., 1994), and protein disulfide isomerase (Ghosal, *et al*., 1994). Lipid droplets in other non-MECs are also associated with other proteins including: the adipocyte specific protein, perilipin (Blanchette-Mackie *et al*., 1995), the ATPase, lipotransin (Syu and Saltiel, 1999), and the cytoskeletal protein, vimentin (Franke *et al*., 1987; Lieber and Evans, 1996). BiP (Prattes, *et al*., 2000) and ADRP (Ye and Serrero, 1998) are also associated with the lipid droplets in non-MECs.

The precise mode by which nascent microlipid bodies are able to dissociate from ER membrane into the cytoplasm is not known. However, Keenan *et al*. (1992), using an *in vitro* assay, showed that a small cytosolic factor stimulates their release from the ER. Henceforth, the triglycerides coalesce to form larger and larger droplets that vectorially move towards the apical plasma membrane. The factors that mediate the progressive coalescence of the lipid droplets while in transit to the plasma membrane remain to be clearly identified. Some evidence from *in vitro* assays suggests that calcium, cytosolic proteins or exogenous gangliosides might be involved (Valivullah *et al*., 1988).

Although there is lingering dissent among researchers studying milk lipid secretion (Wooding, 1971; Wooding, 1973; Kralj, and Pipan, 1992), the widely accepted model (Mather and Keenan, 1998) posits that milk-fat globules abut the apical plasma membrane possibly through interactions with membrane-associated proteins. The ensuing sequence of events that leads to the progressive and complete envelopment of the globule by apical membranes, and finally its pinching-off and expulsion into the lumen are less understood. Since the MFGM originates from apical membranes, apical-limited membrane associated proteins, such as BTN, ostensibly play an important role in the steering and/or extrusion process of the milk-fat globules.

The name for the mammary specific protein BTN was coined from the Greek terms '*butyros*' and '*philos*' to reflect its co-localization with milk-fat globules (Franke, *et al*., 1981). In the years since the first cow cDNA clone was isolated (Jack and Mather, 1990), *BTN* gene has been described in several mammalian species ranging from humans to dolphins (Hare *et al*., 2002) and shows a high degree of nucleic acid sequence similarity across many species. Genomic organization of *BTN* is such that the exons

closely correlate to functional protein domains. BTN is a type I transmembrane glycoprotein with two exoplasmic Ig-like domains, a centrally located transmembrane domain (Banghart *et al.*, 1998), and two predicted Ig-folds on the cytoplasmic side (Seto, *et al*., 1999). The cytoplasmic Ig-folds are encoded by the so-called B30.2 domain. This domain is present in many genes and will later be discussed in more detail.

Labeling studies have shown that BTN preferentially sorts to the apical surface of secretory MECs (Franke *et al*., 1981; Johnson and Mather, 1985). Moreover, although there is no evidence that BTN is a major integral protein of the apical surface of MECs, it is by far the most abundant protein spanning the cow MFGM (Mather *et al*., 1980; Franke *et al*., 1981; Mondy and Keenan, 1993); indicating a preferential accumulation of this protein to exit sites of milk-fat globules.

Several isoforms of BTN have been detected that arise partly due from variable glycosylation reactions (for review see, Mather, 2000). A novel species of BTN, derived from an alternative splicing reaction that removes the IgC domain, is also present in human MFGM (Cavaletto *et al.*, 2002).

Completion of the first sequence maps of the human and mouse genomes provides the opportunity to hunt for paralogous genes to butyrophilin. In addition to the eponymous butyrophilin gene (recently designated as *BTN*1A1 in humans - h*BTN*1A1), there are several other strongly related and closely positioned genes. Sequence analysis of the human extended major histocompatibility complex I (MHCI) region revealed six other *BTN*-like genes telomeric to h*BTN*1A1 that are now collectively grouped into the *BTN* gene family (Rhodes *et al.*, 2001; Ruddy, *et al.*, 1997). The subfamily *BTN*1A only comprises h*BTN*1A1, while subfamilies *BTN*2A and *BTN*3A each comprise three

members. Based on the cDNA sequence of the *BTN*-like genes, polypeptides of the BTN family share about 40% primary sequence identity (Ruddy *et al*., 1997). Unlike the prototypical *BTN*1A1 gene, expression of the other BTN-like genes is not specific to the mammary gland (Ruddy *et al* 1997; Rhodes *et al*., 2001). Mouse *Btn*1a1 is not in the MHC region of chromosome 17 but rather is in synteny with other MHC-associated genes on chromosome 13 (Amadou *et al*., 1995).

Mammals are thought to have arisen sometime at the end of the Jurassic or the beginning of the Cretaceous period (Oftedal, 2002). If the arrival of *BTN* into the genomic repertoire is ancillary to the emergence of lactation, this in an evolutionary time-scale, is quite recent. Henry *et al*. (1997a) describe *BTN* as a mosaic which originated by exon shuffling of two distinct progenitor genes: one progenitor contributing the two exoplasmic Ig-domains, and possibly the transmembrane domain; and the second progenitor supplying the cytoplasmic end, comprising a pair of heptad repeats and the highly conserved B30.2 domain. The ancestral exons that combined to form the *BTN* open reading frame (ORF) also contributed to the configuration of other ORFs (Vernet *et al*., 1993; Henry *et al*., 1997a). In the following sections I discuss proteins that share sequence similarities with BTN: described here with the prospect that similarities at the level of the primary structure will convey some insight into the possible function(s) of BTN.

By virtue of its two exoplasmic immunoglobulin (Ig) folds, BTN shows strong to weak similarity with some members of the Ig-superfamily. Notable among these are the B7 family of co-stimulatory molecules (Linsley *et al*., 1994; Henry *et al*., 1999) that are expressed on the surface of antigen presenting cells and the activated leukocyte-cell

adhesion molecule (AL-CAM). Moreover, protein level sequence homology searches in the SWISSPROT database also reveal shared homology with several neuronal specific proteins such as myelin oligodendrocyte glycoprotein (MOG) (for review see Henry *et al*., 1999), neural cell adhesion molecule (N-CAM), and the limbic-system-associated membrane protein (LAMP). There is now experimental evidence that molecular mimicry of human MOG by bBTN could be the cause of a pathogenic immune response. Antibodies raised against bBTN from cows' milk cross react with hMOG possibly leading to the neurodegenerative autoimmune disease, multiple sclerosis (Stefferl *et al*., 2000; Kennel *et al*., 2003; Guggenmos *et al.*, 2004).

A class of MHC associated proteins known as B-G (chicken) (Vernet *et al*., 1993) and B-G-like (crane) (Jarvi *et al*., 1999) exhibit similarity to BTN within the exoplasmic domain. The homology between B-G and BTN covers the first Ig-loop, while that of B-G-like extends to the second Ig-loop as well. Both B-G and B-G-like genes code for a transmembrane domain. Taken together, these and the previously mentioned homologies to the exoplasmic domain of BTN indicate that BTN in the mammary gland has additional functions unrelated to milk-fat secretion.

The transmembrane domain of BTN is located near the center of the protein, dividing the polypeptide into two nearly equal halves. The cytoplasmic side possesses a pair of putative heptad repeats, the B30.2 domain and a short C-terminal tail.

Of particular interest is the highly conserved B30.2 domain which was first identified as a 166 amino acid region found on the carboxy terminus of four proteins; of which BTN was one (Jack and Mather, 1990). Three highly conserved motifs, named LDP, WEVE and LDYE, anchor the homology of the B30.2 region (Henry, *et al*.,

1997b). Computer based analysis programs for secondary structure and putative folds predict that the B30.2 domain comprises up to 15 β-strands that form two Ig-like loops (Seto, *et al*., 1999).

While the function of the B30.2 domain remains to be conclusively established, since its initial description, numerous other proteins with seemingly unrelated functions have been found to carry this domain. The domain is found in both membrane bound as well as soluble proteins, but always towards the C-terminal region of the proteins so far identified. However, the N-terminal ends of these same proteins are manifold. Besides the paralogous BTN-family members, the only other protein that possesses a B30.2 domain and an Ig-like domain on its amino-terminus is the transmembrane protein, erythroid membrane-associated protein (Ermap), which share an overall amino acid identity of 36% and an additional similarity of 22% (Ye *et al*., 2000).

Several proteins that possess the B30.2 domain in the carboxy-terminus apparently associate with microtubules (MTs). One such protein is MID1. Mutations in the C-terminal region of MID1 cause the genetic disorder Optiz syndrome (Quaderi *et al*., 1997) by abrogating its MT associating property (Schweiger, *et al*., 1999; Cainarca *et al*., 1999; Cox *et al*., 2000). Another related protein, MIR1, also has a B30.2 domain and associates with MTs in a cell-cycle dependent manner (Stein *et al*., 2002). In both MID1 and MIR1, MT association appears to be regulated by a phasic phosphorylation/ dephosphorylation event (Liu *et al*., 2001; Stein *et al*., 2002). The amino-terminus of MID1 has a zinc-binding domain known as RING-BBox-coiled-coiled (RBCC) domain. The RBCC domain (also known as TRIM – tripartite motif) is also a feature of some other proteins that possess a B30.2 domain (discussed later). MIR1 has only the coiled-

coil domain on its amino-terminus, and thus does not constitute all three motifs.

The recessive genetic disorder known as familial Mediterranean fever (FMF) is caused by a mutation(s) in the B30.2 domain of the pyrin/marenostrin gene (The International FMF Consortium, 1997; The French FMF Consortium, 1997). Interestingly, the MT inhibitor agent colchicine has been the most widely used drug treatment for FMF since the early 1970s (Zemer *et al*., 1974; Dinarello *et al*., 1974). However, contrary to this apparent link, work by Mansfield *et al*. (2001) indicates that the interaction between pyrin and MTs is not via the B30.2 domain.

The kinase sp1A and RYanodine receptor were found to have triplicates of a B30.2-like domain which was given the name SPRY (for sp1A/RYanodine) (Ponting *et al*., 1997). Sequence comparisons to the B30.2 domain reveal that the amino-terminus of SPRY is truncated and preliminary fold analysis predicts that SPRY lacks one of the two putative Ig-like folds (Seto *et al*., 1999). The SPRY domain is found in the proteins of a multitude of organisms and proteins of equally diverse functions. Although no functional properties were attributed specifically to the SPRY domain, recent *in vitro* interaction and *in vivo* functional assays show that the SPRY domain of RanBPM is implicated in interactions with MET, a receptor protein-tyrosine kinase (RPTK). Unlike the B30.2 domain, SPRY is not limited to the C-terminal segment of a protein.

Another protein, GLFND (not an abbreviation) was recently isolated and shown to bind to MTs partly through its C-terminal SPRY domain (Manabe, *et al*., 2002). Yeast two-hybrid screens using zyxin, (Schenker and Trueb, 2000) and 14-3-3ζ (Birkenfeld *et al*., 2003) as bait also revealed a novel protein named BSPRY (for B-Box and SPRY domain). The functional significance of these interactions has yet to be determined.

The putative coiled-coiled domain of BTN potentially plays a role in self-aggregation or cross-interaction with other proteins, as these domains typically serve as protein oligomerization motifs (for review see, Burkhard *et al*., 2001). This possibility is supported by the identification of high molecular weight aggregates of BTN from purified MFGM (Banghart, *et al*., 1998; Goldfarb, 1997; Cavaletto *et al*., 2002), and isolates of aggregates of GST fusion proteins of the cytoplasmic domain of mouse butyrophilin (mBtn1a1) (Rao and Mather, unpublished).

At this juncture, it is important to mention homologues of BTN whose function is attributed to their non-B30.2/SPRY domains. These comparisons may not lend direct deductive parallels as to BTN's possible function. The N-terminal non-B30.2/SPRY region of estrogen-responsive finger-protein (efp) (Inoue *et al*., 1993), xenopus nuclear factor 7 (xnf7) (Reddy *et al*., 1991), stimulated *trans*-acting factor 50 (Staf50) (Tissot and Mechti, 1995), ret finger protein (RFP) (Takahashi *et al*. 1988), pAw33 (Bellini *et al*., 1993) and Sjogren's syndrome antigen (SS-A/Ro) (Chan *et al*., 1991; Itoh *et al*., 1991) all constitute an RBCC domain that is a putative transcription regulator (Saurin *et al*., 1996).

Despite the longer N-terminal end of the B30.2 domain as compared to the SPRY domain, the level of similarity between the two domains is unmistakable (Seto *et al*., 1999). Even as we discuss B30.2 and SPRY domains as being separate entities, it is quite possible that one was derived from the other, or else, that both domain families belong to the same super-family.

The cytoplasmic B30.2 domain and/or the exoplasmic Ig-folds present in BTN are doubtless important for the function of BTN in lactation … most likely in milk-fat secretion. Milk-fat-droplets exiting secretory MECs are enveloped by the apical plasma

membrane that forms the MFGM (Mather, 2000). Franke *et al*. (1981) have observed that lipid droplets from the cytoplasm travel towards the apical cell surface and upon extrusion are enveloped by MFGM that is enriched with BTN and XO, implying that BTN and XO concentrate at designated milk-lipid exit areas; or that BTN and XO concentrate in areas of milk-fat globule secretion. BTN makes up the major integral protein of the MFGM in many species and is thought to be a major player in the milk-fat globule extrusion process (for review see Mather, 2000).

In order to gain an understanding of the functional role of BTN, Mather and Ogg (unpublished) have recently generated a mouse model in which the m*Btn*1a1 is ablated. Microscopic examination of mammary tissue of these mice reveals a dramatic alteration in both the histology of secretory MECs as well as the structure of the MFGM. The milk-fat secretion process was severely compromised, despite a seemingly unrelieved milk-lipid synthesis. This is believed to explain the large lipid droplets observed occupying nearly the entire cytoplasm. The phenotype exhibited by these m*Btn*1a1-ablated mice is the first compelling evidence that BTN is required for the proper secretion of milk fat. The exact sequence of events and all of the cohorts involved in this process are not fully understood. Still, XO was demonstrated to be one of the principal components of milk-fat secretion process (Vorbach *et al*., 2002) most likely by interacting with the cytoplasmic side of BTN (Ishii *et al*., 1995; Rao and Mather, unpublished). Preliminary morphological examination of mammary tissue derived from a mouse model where only one allele of XO has been ablated is a close phenocopy of the m*Btn*1a1 ablated mouse (Vorbach, *et al*., 2002).

As most present day genes probably arose from the rearrangement of a limited

number of primogenitor exons (Dorit and Gilbert, 1991), a linear comparison of the homologous domains of proteins/genes, can only offer limited insight towards deciphering the various functional elements of a protein. Nonetheless, the usefulness of syllogistical comparison of certain conserved modular domains cannot be discounted, particularly if taken in the context of other relevant factors such as tertiary structure, cellular localization, interacting partner(s) or the nature of the active site.

The potential role of BTN in milk-fat secretion does not parlay one into ascribing it as its sole function, nor that BTN's function is mediated via interaction with only XO. The evolution of BTN from two distinctly unrelated genes, whose present day descendants exhibit varied tissue distributions and functions, is suggestive of other tasks in milk-fat secretion or even secondary unrelated functions.

The expression profile of *BTN* is, nonetheless, a strong indicator that BTN's function is related to lactation. Genes important for mammary gland development, milk synthesis and secretion, and involution are subject to precise control that allows for the ordered episodic and cyclic physiological stages that constitute a functional secretory organ. Changes in gene expression profile, such as that of *BTN*, that occur commensurate with the developmental and/or functional stages of the mammary gland present a unique opportunity to study the temporal and spatial control of mammary-specific genes.

Despite exceptions in certain species and for some genes, the expression of most 'milk genes' is initiated during pregnancy and terminated shortly after the onset of involution. Therefore, not only are these genes tissue specific, but they also display temporal patterns that coincide with the physiological status of the mammary gland.

Multiple studies on the expression of milk protein genes, both *in vitro* and *in vivo*, reveal an elaborate axis of control that involves many hormones, growth factors and growth inhibitors (for example, Vonderhaar and Ziska, 1989; Stocklin *et al*., 1996) and the extracellular substratum (for example, Streuli, *et al*., 1991; Aggeler *et al*., 1988). The mammary gland is a complex system, thus ascribing the role of transcriptional effector to one or a few hormones is inherently flawed. Nonetheless, a trio of lactogenic hormones, *viz* insulin (I), prolactin (PRL), and hydrocortisone (H), have been identified as major *trans*-acting factors required for the transcriptional induction of 'milk genes' *in vitro*. These studies focused on the major 'milk genes', *i.e*. the caseins and whey proteins, primarily because their relatively high level of expression makes them ideal candidates for such investigations.

In addition to the major constituents of milk, such as caseins and lactose, which are exclusive to the mammary gland, less abundant proteins which are required for the proper synthesis and secretion of milk constituents, such as α-lactalbumin (Vilotte and Soulier, 1992) and BTN (Banghart, *et al.*, 1998; Ogg *et al*., 1996) are also restricted to secretory MECs.

In species so far investigated, and similar to the traditional mammary specific markers such as β-casein, the expression of BTN is confined to the mammary gland. This is in contrast to the other BTN-family members that have a wider tissue distribution (Ruddy *et al*.1997; Rhodes, *et al*., 2001). A corollary can thus be drawn that a regulatory element(s) dictates the expression of BTN distinct from the other BTN-like genes and possibly similar to other 'milk genes'.

BTN and casein are purported to be transcribed in the same subpopulation of

MECs (Molenaar *et al*., 1995). So far, characterization of *BTN* gene expression has largely relied on the detection of transcript from tissue (Aoki *et al*., 1997; Banghart *et al*., 1998). However, a more incisive investigation of the elements that direct *BTN* expression necessitates other approaches. Variation amongst animals and the lack of an effective and reproducible means of introducing reporter genes into the mammary gland has necessitated the reliance on established cell lines. With all the shortcomings of not being a facsimile of 'true' physiological conditions, established cell lines provide a means by which one can delineate important regulatory elements with a modicum of exertion and a fair degree of reproducibility.

There are no well established bovine MECs. The only widely available MEC line was obtained by transforming Holstein primary mammary alveolar cells with the simian virus-40 (SV40) large T-antigen (Huynh *et al*., 1991). Although this cell line is commonly used as a model to study mammary gland infections, efforts to characterize it have shown it to comprise a non-homogenous population of cells (Zavizion *et al.*, 1995).

However, there are better characterized mouse MECs. COMMA-1D cells were isolated from mid-pregnant Balb/c mice and characterized for epithelial and mammary specific properties (Danielson *et al*., 1984). HC11 cells were later subcloned by Ball *et al*. (1988) from the COMMA-1D cell line and exhibited the same mammary specific benchmarks as COMMA-1D cells, without the extracellular matrix requirements (Chammas, *et al*., 1994). HC11 cells display characteristics of terminally differentiated lactating mammary epithelial cells as measured by expression of the β-casein gene in response to lactogenic hormone (IHP - I, H, and PRL) treatment (Ball *et al.,* 1988). Moreover, they also make their own substratum providing a condition that many

investigators have shown is a requirement for the proper expression of milk genes (Chammas *et al*., 1994).

Work done by Aoki *et al*. (1997) using COMMA-1D and HC11 cells demonstrated that there is a marked difference between the level of expression of *BTN* and β-casein. Further, BTN mRNA is expressed in HC11 cells in the presence of basal media supplemented with the trio of lactogenic hormones; but the addition of lactogenic hormones elicits a less dramatic enhancement of expression as compared to that of β-casein (Aoki, *et al.*, 1997) pointing to a possible separate or parallel mode of regulation. Therefore, it is probable that the promoter of *BTN* may possess elements that are both shared with and divergent from that of β-casein.

The cloning of several genes integral in the signaling pathway required for lactogenesis has allowed the use of non-mammary cells in efforts to define important *cis*-elements (Jolivet *et al*., 1996). However, it is prudent to rely on well characterized cells that resemble mammary cells in some benchmarks of differentiated MECs. Therefore, all transfection assays in this study were done with the HC11 mouse mammary cell line.

The identification of *cis*-acting regulatory elements of b*BTN* has implication in its use for the expression of chimeric genes in the mammary gland, providing a means by which to alter the composition of milk. Further, genetic variation in the b*BTN* gene or promoter has potential use as a marker for quantitative trait loci (QTL) analysis. Over the past several decades, classical selection and breeding techniques to increase the quality and quantity of milk production in dairy animals has registered measurable and steady gains. In 1950 the yearly milk production per cow in the U.S. was 5,314 lbs in contrast to 18,204 lbs in the year 2000, while the population of dairy cows went down by more than

half during the same period (Blayney, 2002). The advent of molecular biology applications has brought with it new possibilities. For example, quantitative trait loci (QTL) analysis, marker assisted selection (MAS), cloning of whole organisms, *etc*. present new tools to improve the genetic make-up of dairy animals.

The theory of particulate inheritance elements (Mendel, 1865) does not hold true for many traits. In fact it is now an established fact that many traits display continuous variation. Phenotypic manifestation of continuous quantitative traits is due to effects and interplay of multiple loci with environmental factors. Each of these loci is a QTL. Ever since the pioneering work of Tanksley and Rick (1981) linking enzyme variation in tomato to a quantifiable trait, both the qualitative and the quantitative aspects of QTL analysis have evolved to suit other species and needs. A case in point is the circumstances peculiar to large dairy farms, where (1) commercial breeding programs are driven by economic benefits and are not purely research oriented, (2) most, if not all, economically important traits are sex limited, and (3) heavy reliance on artificial insemination (AI) technology from a few high genetic merit bulls, has necessitated a custom-tailored approach.

Even though there is a lag in the compilation of genomic sequences of farm animals, efforts are underway to identify and catalogue genetic markers. Since its inception, the Cattle Genome Mapping Project (Keele *et al.*, 1994; USDA, http://www.marc.usda.gov/ genome/genome.html) has grown to provide searchable databases for markers on all 29 autosomal chromosomes and the X chromosome.

Although some QTL analyses in dairy animals have relied on the use of variants of already known genes as markers (for example, Bovenhuis and Weller, 1994; Lagziel,

et al., 1999), this approach limits the discovery of putative QTLs to regions of the chromosome that are closely linked to the known genes. Traditional genetic markers such as microsatellites, restriction fragment length polymorphisms (RFLPs), tandem repeats, *etc*. are the preferred choices. In particular, the use of RFLP markers for QTL mapping has been extensive since the early 80s (Botstein *et al*., 1980).

Previous searches to identify genetic markers in *BTN* have yielded a single-stranded conformational polymorphism (SSCP) in the 3' UTR (Karall *et al*., 1997) and an RFLP in the coding region of the 3'-most exon (Taylor *et al*., 1996). Lee *et al*. (2002) also reported five RFLPs in Korean Holstein-Friesian and identified two allelic forms of b*BTN*.

In order to identify the locus that influences a certain quantitative trait in livestock, current QTL analysis approaches utilize multiple genetic markers to narrow down the likelihood of a certain locus being statistically associated with a defined quantitative trait, and subsequent comparative analysis with publicly available human and rodent genome databases.

In livestock, genome-wide scans to detect regions of chromosomes that influence a trait of interest are limited in their effectiveness depending on the rate of recombination between the putative QTL and the marker. This approach, although of great value, is prone to 'noise' effects that could arise from the linkage disequilibrium of a true causal locus to a gene of interest.

Given a choice of available markers, the most commonly used method to improve the power of detecting QTLs, in a scenario where the putative QTL and the marker do not map to the same locus, is interval mapping (Lander and Botstein, 1989). When a marker

which lies at the same locus as a candidate gene is obtainable, it practically ensures linkage of the two and thus accords several advantages. If marker and candidate gene lie on a single locus, a QTL mapping study (1) does not underestimate the phenotypic effect of the QTL, (2) does not require as many progeny to be scored; and, (3) is independent of tight linkage/small effect or loose linkage/large effect conundrum (Lander and Botstein, 1989).

The design used to apply QTL analysis in dairy animals has to take into account the specific conditions that are peculiar to the commercial dairy industry (see above for list). Weller *et al.* (1990) developed a method of analyzing the inheritance pattern of genetic markers/genes in multi-generational dairy half sib families. Dubbed the granddaughter design (GDD), this method was particularly pertinent to dairy animals since it employed the scoring of genotypic/marker information from sons of heterozygous grandsires, and the collection of phenotypic data from granddaughters. The GDD is by far the most widely used model for QTL analysis in dairy cattle world wide.

The GDD affords a comparable power of detecting QTLs for less number of marker scorings as opposed to a daughter design, where both the genotyping and phenotyping are done on the same animal (Weller *et al*., 1990).

BTN is one a few mammary/lactation specific genes so far identified. This study tried to address two questions: what confers mammary specificity to BTN expression? and how does allelic variation in BTN affect milk production traits? To these ends, two specific aims were formulated:

I QTL analysis of b*BTN*:

- to identify intragenic polymorphic markers of b*BTN*, and

- to utilize these markers to determine if b*BTN* is a QTL affecting economically important milk production traits.

II Analysis of the *cis*-regulatory elements of b*BTN*:

- to clone a sizable region of the 5' flanking sequence of the b*BTN*,

- to perform computational modeling of the putative promoter region of b*BTN*

- to utilize transient transfection assays to delimit important *cis*-acting regions that potentially control b*BTN* expression.

METHODS

1. Materials

Dextran sulfate, sodium hydroxide, polymyxin B agarose, acetyl-CoA, porcine insulin, epidermal growth factor, hydrocortisone, Dulbeco's phosphate buffered saline (D-PBS, Ca^{+2}, Mg^{+2} free), chloroform, isoamyl alcohol, 2-propanol, diethyl pyrocarbonate (DEPC), ethidium bromide, casein enzymatic hydrolysate (NZ amine), $MgSO_4$-$7H_2O$, glycerol, boric acid and polyethylene glycol were all obtained from Sigma (St. Louis, MO). Phenol was obtained either from Sigma or Amersham Biosciences Corp. (Piscataway, NJ). Prolactin was a gift from the USDA Hormone Program (Beltsville, MD). RPMI-1640 media, PCR Supermix ®, trypsin, gentamicin, sodium bicarbonate (cell culture grade), TRIzol, FAST Track ® mRNA isolation kit, and agarose were acquired from Invitrogen Life Technologies (Carlsbad, CA). The radiocompounds [^{14}C]-chloramphenicol and α[^{32}P]-dCTP were obtained from American Radiolabeled Chemicals (St. Louis, MO), and Dupont NEN (Boston, MA) or Amersham, respectively. DNA polymerase I Klenow fragment, DNA ligase, S1 nuclease, DNA polymerase and DNA size marker were from Promega (Madison, WI). While all endonucleases used were either from Promega or New England Biolabs (Beverly, MA). RNAse, proteinase K, and chlorophenol red -β- D- galactopyranoside (CPRG) were obtained from Boehringer Mannheim (Indianapolis, IN). ULTRAhyb hybridization buffer, formaldehyde loading dye, and RNA size marker were procured from Ambion (Austin, TX). Characterized fetal bovine serum (FBS) was obtained from Hyclone (Logan, UT) or Atlanta Biological (Norcross, GA). The Prime IT ® II kit and NucTrap

® push column were obtained from Stratagene (La Jolla, CA). Yeast extract, tyrptone, and bacto-agar were acquired from Beckton Dicknson (Sparks, MD). Gene Porter cell transfection kit was from Gene Therapy Systems (San Diego, CA). Magna Lift nylon membranes were obtained from Micro Separations Inc. (Westboro, MA). Ethyl acetate was obtained from JT Baker Chemical Company (Phillipsburg, NJ), while thin layer chromatography (TLC) plates were acquired from EM Science (Gibbstown, NJ). ORIGEN freeze media was procured from IGEN International Inc. (Gaithersburg, MD). BSA ® protein assay kit and sodium dodecyl sulphate were obtained from Pierce Biotechnology (Rockford, IL) and Biorad Inc. (Hercules, CA), respectively. All other reagents were obtained from Fisher Scientific (Newark, DE). Micron-100 and 0.22μm filters were obtained from Millipore (Bedford, MA).

2 Isolation and characterization of the 5' flanking region of b*BTN*

A bovine genomic library was obtained from ClonTech Laboratories Inc. (Palo Alto, CA). The library consisted of bovine genomic DNA fragments generated by partial digestion with *Mbo*I and cloned into the *Bam*H1 cloning site of the EMBL3 SP6/T7 phage vector.

XL1-Blue host cells were used to amplify and screen the library. The screening process was essentially as per the company's protocol with slight modifications. NZYM agarose plates and LB (Luria-Bertani) soft top agar were used on 150mm plates; this was found to improve the ease of manipulation of phage DNA in down stream processes. Plaques were replica-lifted with Magna Lift ® (Micro Separation Inc., Westboro, MA) nylon membrane. The membranes were then placed for two 2 min steps each on

Whatman ™ paper soaked in denaturation (1.5m NaCl, 0.5N NaOH) and neutralization (1M Tris-HCl, pH 7.4, 1.5M NaCl) solutions. Excess denaturation and neutralizing solutions on the nylon membranes were blotted on fresh Whatman ® paper (membrane not allowed to dry) in between steps. The membrane was placed on a 10X SSC soaked Whatman ® paper in a Stratalinker ® (Stratagene, La Jolla, CA) to cross link the DNA to the nylon membrane.

The library was screened using a 0.9kb b*BTN* cDNA probe representing the 5' end of b*BTN* extending from exon 1B to exon 4. The probe was radioactively labeled by random priming using the Prime IT®II kit (Stratagene, CA) and α[32P]-dCTP (Dupont NEN, MA). Unincorporated nucleotides were removed by NucTrap ® (Stratagene, CA).

3. Construction of chloramphenicol acetyl transferase (*CAT*) reporter plasmids

The p*CAT*® -Basic vector (Promega, Madison, WI; GenBank accession № X65322) lacks eukaryotic promoter and enhancer elements (Figure 1) and thus was selected to assay for sequences that could potentially have *cis*-acting regulatory elements (example, Feltus *et al*., 1999; Grønning *et al*., 1999). The p*CAT*-Basic® was the initial vector from which all *CAT* reporter constructs were ultimately derived.

3.1 pb*BTN*1.7/*CAT*

A 1.7 kb piece that encompasses 1042 bp of the 5' putative promoter, 750 bp of the 5' UTR and intron A was excised from its pBluescript® II KS phagemid vector by digestion with *Xho*I and *Bam*HI. The *Xho*I site resides in exon IB of b*BTN*, while the *Bam*HI site is part of the multiple cloning site of the pBluescript II KS vector. The 5'

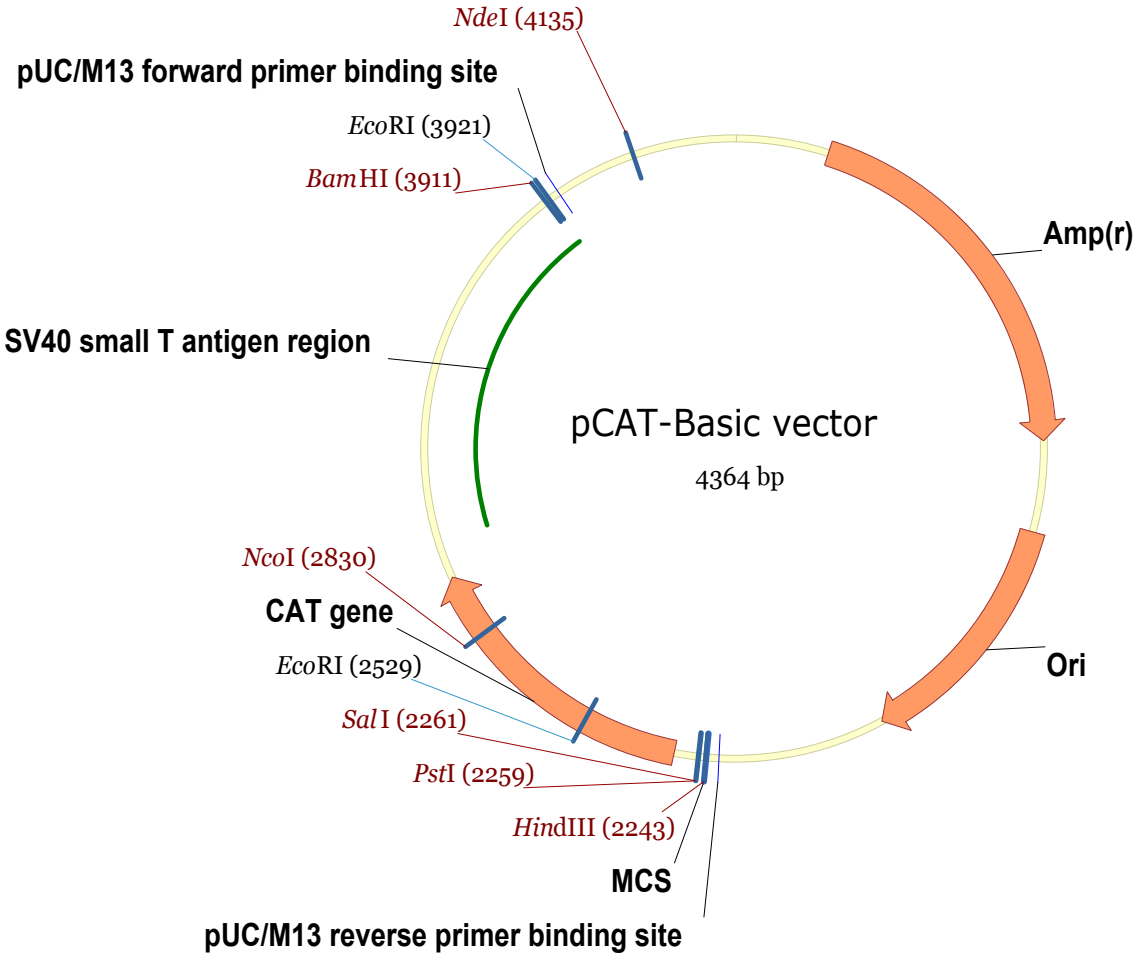


Figure 1: pCAT-Basic vector plasmid features. This plasmid was the starting material for constructing pCAT-Basic/+1.7 and pCAT-Basic/-1.7 by inserting 3' sequence of b*BTN* at the *Nde*I site in correct and reverse orientations, respectively.

overhang ends were blunt-ended with DNA polymerase (Klenow) and inserted into the p*CAT*-Basic® vector that had been linearized with *Sal*I and blunt-ended.

This reporter test construct was named pb*BTN*1.7/*CAT* (Figure 2, Table 1). The 1.7 kb insert comprised sequence that extends up to 49 bp upstream relative to the b*BTN* native translation start site and includes ~1 kb of 5' flanking region.

3.2 pb*BTN*1.2/*CAT*

This test construct has a putative promoter that is approx. 0.45 kb situated 5' of the *CAT* gene. Since the pb*BTN*1.7/*CAT* was used as a starting material to generate pb*BTN*1.2/*CAT*, all other features were the same. To this end, approx. 0.6 kb of 5' sequence pb*BTN*1.7/*CAT*'s insert was removed by digestion with *Apa*I and *Pst*I, located in the insert and the 5' multiple cloning site respectively The ends were blunt-ended with S1 nuclease and religated creating the construct named pb*BTN*1.2/*CAT* (Table 1).

3.3 p*CAT*-Basic/+1.7 and p*CAT*-Basic/-1.7

These two constructs were generated from p*CAT*-Basic®. A 1.7 kb *Nde*I fragment (courtesy of H. Davey, AgResearch, Ruakura Research Centre, Hamilton, New Zealand) that encompasses 273 bp of 3' UTR, the polyA element and 1394 bp of 3' flanking region of b*BTN* was inserted at the sole *Nde*I site 1.2 kb 3' of the *CAT* gene (Figure 2). Two clones were isolated in which the inserted sequence was either in the correct or reverse orientation and named p*CAT*-Basic/+1.7 and p*CAT*-Basic/-1.7, respectively (Table 1).

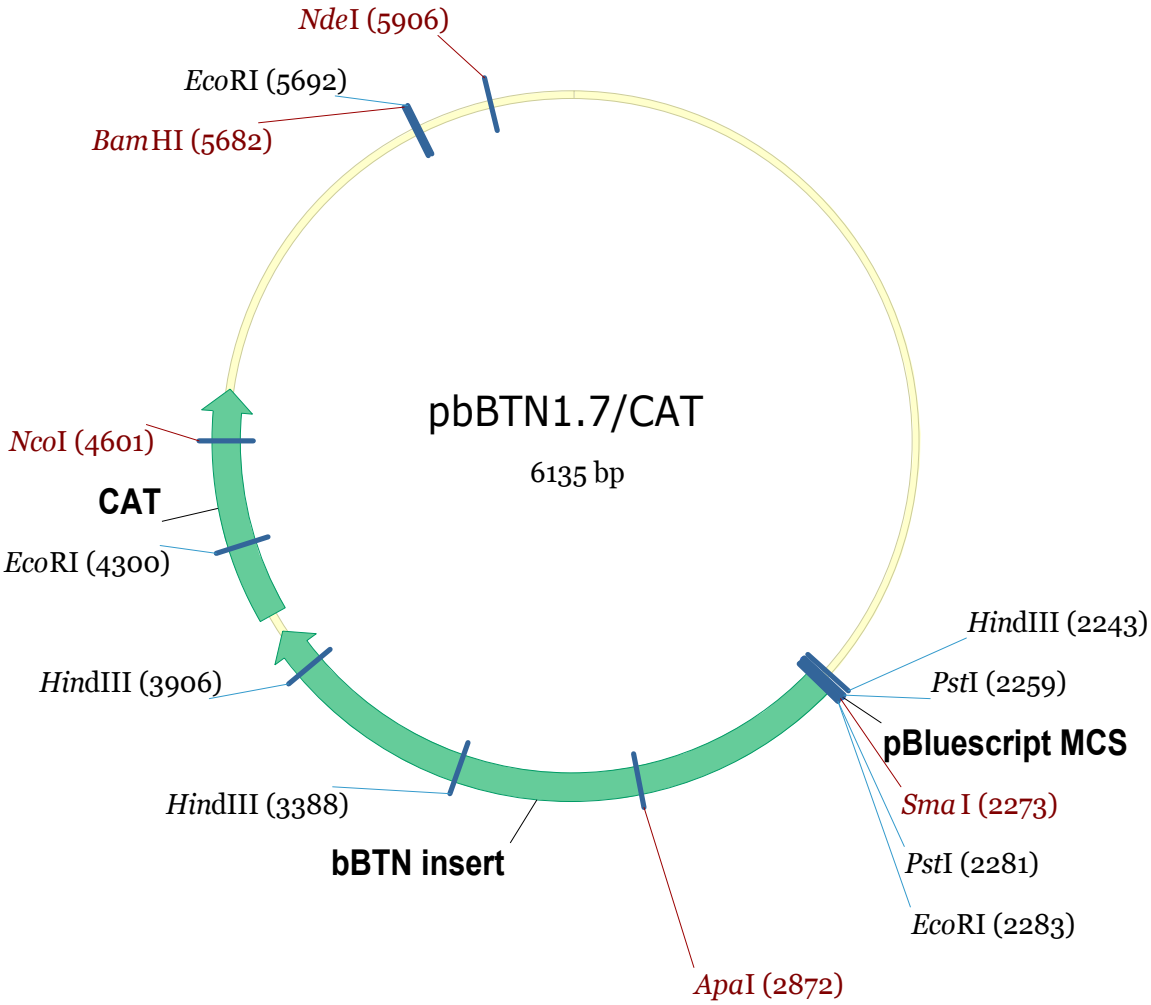


Figure 2: pb*BTN*1.7/*CAT*. A p*CAT*-Basic vector with the 1.7kb b*BTN* insert. Notice the pBluescript ® II KS MCS is at the 3' end of the putative promoter. Although not depicted, all other vector plasmid features are the same as in Figure 1 above. pb*BTN*1.7/*CAT* was the starting material for constructing pb*BTN*1.2/*CAT* by the excision of 0.6 kb of the insert's 3' end between *Pst*I and *Apa*I.

Table 1: List of promoter constructs and their components

Name of construct	Sequence element(s) inserted to pCAT-Basic Vector			
	Size in bps	**Location of insertion**	**Orientation of insertion**	**Description of inserted element**
pCAT-Basic	None	N/A	N/A	N/A
pCAT-Basic/+1.7	1667[1]	3' of CAT gene	Forward	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region
pCAT-Basic/-1.7	1667	3' of CAT gene	Reverse	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region
pb*BTN*1.2/CAT	1160[2]	5'of CAT gene	Forward	452bp 5'putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
pb*BTN*1.7/CAT	1750[3]	5' of CAT gene	Forward	1042bp 5'putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
pb*BTN*1.2/CAT/+1.7	1160	5' of CAT gene	Forward	452bp 5'putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
	1667	3' of CAT gene	Forward	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region
pb*BTN*1.7/CAT/+1.7	1750	5' of CAT gene	Forward	1042bp 5' putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
	1667	3' of CAT gene	Forward	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region
pb*BTN*1.2/CAT/-1.7	1160	5' of CAT gene	Forward	452bp 5'putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
	1667	3' of CAT gene	Reverse	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region
pb*BTN*1.7/CAT/-1.7	1750	5' of CAT gene	Forward	1042bp 5' putative promoter region of b*BTN*; plus 708bp of 5'UTR of b*BTN*
	1667	3' of CAT gene	Reverse	273bp 3' UTR and polyA element; plus 1394bp 3' flanking region

[1] GenBank accession number AF005497, nucleotide positions 7768 - 9434
[2] GenBank accession number AF005497, nucleotide positions 591 - 1751
[3] GenBank accession number AF005497, nucleotide positions 1 - 1751

3.4 pb*BTN*1.7/*CAT*/+17 and pb*BTN*1.7/*CAT*/-1.7

These two constructs possess both 5' and 3' sequence of b*BTN*. The 5' sequence is as described for pb*BTN*1.7/*CAT*, while the 3' insert encompasses 273 bp of 3' UTR, the polyA element and 1394 bp of 3' flanking region of b*BTN*. The two constructs were derived from pb*BTN*1.7/*CAT* reporter vector. The 1.7 kb *Nde*I fragment (described above) was inserted at the sole *Nde*I site 1.2 kb 3' of the *CAT* gene. Two test constructs were prepared in which the inserted sequence was either in the correct or reverse orientation and named pb*BTN*1.7/*CAT*/+17 and pb*BTN*1.7/*CAT*/-1.7, respectively (Table 1).

3.5 pb*BTN*1.2/*CAT*/+1.7 and pb*BTN*1.2/*CAT*/-1.7

These two constructs were generated from pb*BTN*1.2/*CAT*. They possess 5' b*BTN* sequence as described in pb*BTN*1.2/*CAT*. Further, the 1.7 kb 3' *Nde*I fragment of b*BTN*, described above, was inserted at the *Nde*I site of pb*BTN*1.2/*CAT* thereby generating two plasmid constructs in which the inserted 3' sequence was either in the correct or reverse orientation and named pb*BTN*1.2/*CAT*/+1.7 and pb*BTN*1.2/*CAT*/-1.7, respectively (Table 1).

4. Isolation of nucleic acid

4.1 Ribonucleic acid (RNA)

Total RNA from HC11 cells and lactating mouse mammary tissue were isolated using the method of Chomczynski and Sacchi (1987) or with TRIzol® reagent (Invitrogen Life Technologies, Carlsbad, CA) as per the manufacturer's instructions.

mRNA was isolated using the Fast Track ® mRNA isolation kit (Invitrogen Life Technologies) directly from HC11 cells strictly as specified by the company.

4.2 Deoxyribonucleic acid (DNA)

4.2.1 Mammalian DNA isolation

Bovine genomic DNA from sperm and polymorphonuclear (PMN) leukocytes was isolated for genotyping. Blood was collected into vacutainers from the tail vein of cows and kept on ice until processing. The samples were centrifuged at 1300g for 5 min to separate the plasma, buffy coat, red blood cells (RBCs) and polymorphonuclear (PMN) pellet. All of the plasma and buffy coat, and most of the RBC layer was removed by aspiration. The pelleted PMN cells were resuspended in the remaining RBC layer. The RBCs were lysed by adding 2X the volume of ddH_20 and mixing for exactly 45 seconds, and quickly reconstituted by adding an equal volume of 2.7% saline solution. The sample was then centrifugated at 1900g for 5 min to sediment the PMN cells and resuspended in 10ml of PBS. PMN cells were washed twice in 30ml PBS and lysed in a mixture of 10mM Tris-HCl pH 8.0, 10mM EDTA and 1% sodium dodecyl sulfate (SDS). An equal volume of buffer containing 10mM NaCl, 10mM Tris-HCl, pH 8.0, and 10mM EDTA was added, followed by RNase and proteinase K treatment as described (Sambrook *et al*., 1989). The DNA was recovered by treatment with phenol:chloroform and precipitated with ethanol.

Sperm samples were washed thrice with TEN(1) buffer (0.1M Tris pH 8, 0.01M EDTA, 1.0M NaCl) and the cells recovered centrifugation at 1500g for 3 min between washes. Sperm cells were then resuspended in TEN(1) buffer containing 1% SDS,

40mM dithiothreitol (DTT) and proteinase K (1μg/μL) and digested at 37°C overnight or until the sperm heads had been disrupted. DNA was recovered from the lysate by extraction with phenol:chloroform mixture and precipitation with ethanol.

4.2.2 Bacterial plasmid isolation (Maxi-preps)

Bacterial plasmid for routine molecular use, except for transfection, was prepared as follows. Bacterial cells grown in Luria-Bertani (LB) medium were harvested by centrifugation and resuspended in GTE-1 (25 mM Tris-HCl, pH 7.5, 50 mM glucose, and 1 mM EDTA). Cells were lysed by the addition of lysis buffer (0.2N NaOH, 1% SDS). Three molar NaOAc pH 4.8 was added to the lysate and the sample centrifuged for 10 min at 2000g. The supernatant was removed and a 1.25X volume of ice cold isopropanol added followed by centrifugation for 10 min at 2000g to sediment the precipitated nucleic acids. The pellet was resuspended in TE (100mM Tris-HCl pH 8, 10mM EDTA) and 1.25X volume of 5M LiCl and the mixture incubated on ice for 5 min. RNA was pelleted by centrifugation at 2000g for 10 min and 2.5 volumes of ice cold ethanol added to the supernatant. A mostly plasmid DNA was sedimented by centrifugation for 10 min at 2000g, and dried. The pellet was resuspended in TE and incubated with RNase A for 30 min. An equal volume of a cold mixture of 1.6M NaCl and 13% PEG (6000) was added to the digested sample, mixed well and incubated on ice for 5 min. Sample was centrifuged at 20000g for 5 min and the supernatant removed (repeated until much of the supernatant was cleared). The pellet was resuspended in TE and treated with two rounds of chloroform extraction, followed by phenol:chloroform extraction until a clear interface was obtained. Plasmid DNA was precipitated by the addition of 3M NaOAc pH 5.5 and

ethanol.

Plasmid DNA for transfection analysis was processed to remove endotoxins by a method modified from Ausubel *et al*. (1988). Bacteria were grown in LB medium and the cells treated with GTE-2 (50 mM glucose, 25 mM Tris-CL, pH 8.0, and 10 mM EDTA). To the resuspended cells, a 2X volume mixture of 0.2 N NaOH, 1% SDS was added followed by a 1.5 volume mixture of 3M potassium acetate and 1.18 M formic acid. The lysate was then centrifuged to remove high molecular weight genomic DNA, proteins, and other cellular debris; and the aggregates filtered out by passing the lysate through several layers of cheesecloth. Nucleic acid was precipitated by the addition of an 0.6 volume of isopropanol, and the precipitate washed with 70% ethanol and air dried. The nucleic acid pellet was then resuspended in a small volume of GTE and treated with RNase at 37°C. Plasmid DNA was recovered after treatment with mixtures of phenol, chloroform and isoamyl alcohol, and precipitated by the addition of an 0.25 volume of 10 M ammonium acetate and two volumes of ethanol at -80°C for 10 min. Plasmid DNA was recovered by centrifugation at 20000g, washed with 80% ethanol and air dried. The DNA was resuspended in 2 ml TE and 0.8 ml 30% PEG 8000/1.6M NaCl added. After incubation overnight at 4°C, the precipitated DNA was recovered by centrifugation at 20000g for 20 min, resuspended in TE, purified by treatment with phenol/chloroform and precipitated with one quarter volume of 10 M ammonium acetate and two volumes of ethanol as above.

Endotoxins were removed by adsorption on polymyxin B agarose (Sigma, St. Louis, MO). The polymyxin B column (0.75 ml packed volume in 3ml syringe) was washed with 25 ml of polymyxin buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.6, 1 mM

EDTA). The concentration of the NaCl, EDTA and Tris-HCl in the plasmid DNA solution was adjusted to match that of the polymyxin buffer. The eluant was collected and the plasmid precipitated in 10M ammonium acetate and ethanol, as described above.

4.2.3 Bacterial plasmid DNA (mini-preps)

Bacterial plasmid was prepared on the mini-scale by alkaline lysis using protocol developed by Birnboim and Doly (1979) and modified by Sambrook *et al*. (1989).

5. Detection of RNA by northern blot analysis

Total RNA (20µg) or mRNA (2 µg) was denatured by heating to 65°C for 15 min and separated by size by electrophoresis in agarose gels containing formaldehyde as previously described (Lehrach *et al*., 1977; Rosen *et al*., 1990). The gels were then soaked in 2X volumes of 10X SSC twice, for 20 min each, and transferred to nylon membrane by capillary action in the presence of 10X SSC. RNA was cross-linked to the membrane by exposure to UV light using a Stratalinker® (Stratagene). Membranes were then incubated with either ULTRHyb ® for 1 h or 2X Denhardt's solution/0.1% SDS/50% and 0.2µg/ml heat-denatured sheared salmon sperm DNA for 4 h.

DNA probes were labeled with [^{32}P]-dCTP random primers using a PRIME IT ® kit and m*Btn*1a1 cDNA fragments as template. DNA probes were hybridized to the membrane overnight at 42°C in either ULTRAHyb® (Ambion) or 2X Denhardt's solution/0.1% SDS/50% formamide/5X SSPE/10% dextran sulfate and 0.2 µg/mL heat denatured sheared salmon sperm DNA. After hybridization, the membrane was first rinsed with 2X SSC prewarmed to 42°C and then washed at 50°C three times with 2X

SSC/ 0.1% SDS for 20 min each time. Membrane was then exposed either to radiographic film at -80°C or analyzed with a PhosphoImager ™ (Molecular Dynamics).

6. Detection of DNA by Southern blot analysis

DNA fragments generated by complete digestion with restriction enzymes were resolved by electrophoresis in agarose gels as described (Southern, 1975). The gel was then soaked in a 2X gel volume of denaturation solution (0.5N NaOH/1.5M NaCl) for 40 min with one change of denaturation solution after the first 20 min. The gel was then soaked in a 2X volume of neutralization solution (0.5M Tris-HCl pH 8.0/3 M NaCl) for 60 min, with one change of solution after the first 30 min. DNA was transferred via capillary action in 10X SSC buffer overnight to a nylon membrane and crosslinked by exposure to UV light as described for RNA above.

Membranes were prehybridized by incubating at 65°C for 4 - 6 h in 0.76M NaCl/46mM sodium citrate/8.52% dextran sulfate and 0.25µg/ml heat denatured sheared salmon sperm DNA. The radioactive probe was prepared as described for northern blots above using b*BTN* cDNA as template. The membranes were then incubated overnight in a fresh hybridization solution (same as pre-hybridization) preheated to 42°C. Prior to the wash step, the membranes were first rinsed in 2X SSC preheated to 50°C. The wash involved two 5 min washes with 6X SSC, 0.1% SDS at RT°C followed by a 3 min wash with 0.1X SSC, 0.1% SDS preheated to 65°C. Membranes were exposed to x-ray film at -80°C.

7. Polymerase chain reaction (PCR)

A touch-down approach was employed for all PCR reactions that started the reaction with the annealing cycle set 3°C above the optimal temperature. This was followed by a drop of 1°C after two cycles until the optimal annealing temperature was reached; and final 30 cycles at the optimal annealing temperature.

For the initial screening for RFLPs, the PCR mixtures contained in a final volume of 25µl: PCR Supermix® (Invitrogen Life Technologies) (22.5µl), primers (125-130 ng each) and genomic DNA (100ng). For the larger scale genotype scoring, a 96 well system was used employing a PCR mix with 35 - 50 ng of template genomic DNA. Nonetheless, the PCR reactions still employed the touch-down approach described above.

8. Maintenance and microscopy of HC11 cells

HC11 cells were isolated as a subclone of COMMA-1D cells (Danielson *et al.*, 1984) by Ball *et al*. (1988). Early passages were obtained from C. Shmenko (Univeristy of Calgary, Canada) and D. J. Jerry (University of Massachusetts, MA) and cells of unknown passage were obtained from I. Vijay (University of Maryland at College Park, MD).

8.1 Media and incubation conditions

Cells were maintained on plastic tissue culture plates in an atmosphere of 95% air/5% CO_2 at 37°C. The base medium consisted of RPMI-1640 (Invitrogen Life Technologies, CA) plus 10% heat inactivated characterized fetal bovine serum (FBS) and 0.25mg/ml of gentamicin (Invitrogen Life Technologies, CA). In order to constitute an

HC11 growth medium, porcine insulin and epidermal growth factor (IE) (both from Sigma) were added to the base medium to a final concentration of 5 μg/ml and 0.01μg/ml, respectively. Induction medium was constituted from the base medium by addition of I (5μg/ml), H (Sigma) (1μg/ml) and PRL (USDA Hormone Program, Beltsville, MD; or Sigma) (5μg/ml).

8.2 Making HC11 subclones

HC11 cells were plated in growth medium until they attained confluency. Cells were then detached from the plastic substrata by digesting with trypsin. Briefly, at confluency, cells were washed twice with 5ml prewarmed Dulbeco's phosphate buffered saline (D-PBS: $CaCl_2$ and $MgCl_2$ free, Sigma) and treated with 3mls of trypsin (Invitrogen, Carlsbad, CA) diluted to 0.25% in D-PBS was added to 100 mm plates (1ml in 60 mm plates) and the cells were then incubated for 15-20 min to detach them from the plastic surface. Trypsin action was stopped by adding an equal volume of pre-warmed base media. The cells were then serially diluted to about 10^5 per ml and 2μl of this dilution plated on 60mm plates in the presence of growth medium (fresh medium being changed every two days). When the colony foci grew to be visible to the naked eye, the cells were bounded by cloning cylinders and allowed to reach confluency within the confines of the cylinders. Cells were then harvested by use of trypsin digestion within each cloning cylinder in a scaled-down version of the procedure described above, and each colony plated on a fresh 60mm plate with growth medium.

8.3 Expansion and storage of HC11 and subcloned cells

Cells incubated in trypsin (described in section 8.2) were harvested by gentle pipeting and the suspension centrifuged at low speed for 2 min. Cells harvested from a single 100 mm plate were plated onto four similarly sized plates and incubated in growth medium. Generally, cells incubated under such conditions attain confluency within 24 h.

Cells were expanded in this manner over several passages and harvested by trypsin digestion as described above. Cell suspensions from two 100 mm plates were combined and pelleted with low speed centrifugation. The pellets were then resuspended in Origen® (IGEN International Inc., Gaithersburg, MD) cell-freezing media and stored in liquid nitrogen.

A fresh sample from a frozen stock was rapidly thawed at 37°C and used for each subsequent transient transfection experiment. To the thawed stock, an equal volume of prewarmed base medium was added to remove DMSO present in the freezing medium. Cells were then centrifuged at 200g for 2 min.

8.4 Staining and microscopy of HC11

Cells were cultured on glass cover slips placed in 60 mm plates and grown until they attained confluency. Cover slips were then removed and cells fixed by treatment with glutaraldehyde. Fixed cells were either directly examined by phase contrast; or else stained with Mayer's haematoxylin stain and then examined by phase contrast.

9. Transient transfection and harvest of HC11 cell lysates

9.1 Transfection of HC11 cells

Transient transfection of HC11 cells with test constructs and the pcDNA3.1/HisB/*lac*Z (β-galactosidase) internal control were done using the serum-free transfection system Gene Porter ® (Gene Therapy Systems, Inc., San Diego, CA). Briefly, confluent HC11 cells from one 100 mm tissue culture plate were trypsinzed (as discussed above) and replated onto twelve 60mm plates in the presence of growth medium. Cells were allowed to grow to 40 – 50% of confluency. The growth medium was then removed, and cells washed twice with prewarmed RPMI-1640 medium that does **not** contain FBS and gentamicin. As much washing medium as possible was removed by pipeting. Each plate was then transfected with 1.5μg of test plasmids plus 0.5μg of β-galactosidase endotoxin-free plasmids using 1ml lipid-based transfection reagent as per the manufacturer's instructions (Gene Therapy Systems). After four hours of incubation at 37°C in the transfection reagent, an equal volume of prewarmed RPMI-1640 that contains 20 % FBS and twice the concentration of IE as found in the standard growth media (described in section 8.1), was added to the plates bringing the final volume to 2ml. Cells were then allowed to grow for 12 h and a fresh supply of 5 ml growth medium provided until cells attained confluency. Once confluent, the growth medium was replaced with an induction medium or a test medium where one or more of the lactogenic hormones had been excluded. After 24 h of incubation in induction/test media, cells were harvested as per the procedure described below.

9.2 Harvest of HC11 cell lysates

Cells were washed three times with 3ml of prewarmed D-PBS. After the final wash, 1ml of the TEN (40mM Tris-HCl pH7.5, 10mM EDTA, 150mM NaCl) was added and plates allowed to stand for 5 min. Loosened cells were harvested by gentle scraping using a rubber policeman and transferred to 1.5 ml centrifuge tubes. The harvested cells were then centrifuged at 12,000g for 10 sec to pellet the cells. As much of the TEN supernatant as possible was removed and the cells resuspended in 250mM Tris-HCl pH 8.0.

Cells were lysed by three rounds of freeze-thawing at –80°C and 37°C. The lysate was then centrifuged at 19000 g for 5 min and at 4°C to pellet cellular debris. The lysate supernatant was collected and transferred to fresh microcentrifuge tubes and the volume of the samples adjusted to 180µl by adding cold 250mM Tris-Cl pH 8.0.

10. CAT and β-galactosidase assays

10.1 Estimation of protein content

In order to standardize the amount of protein used in each assay, the protein content of HC11 lysates was estimated by the BCA® protein assay procedure as described by the manufacturer (Pierce Biotechnology Inc., Rockford, IL) (Smith *et al*., 1985).

10.2 β-Galactosidase assay

The measured activity of the co-transfected pcDNA3.1/His/*lacZ* served as an

internal transfection control. Based on the BCA protein estimation, lysate with an estimated 40 μg of total protein content was used for a single β-galactosidase assay.

The assay conditions described by Sambrook *et al*. (1989) were modified, particularly in regards to the substrate for β-galactosidase. A reaction mix was prepared to a final concentration of 1X magnesium buffer (1mM MgCl, 50 mM 2-mercaptoethanol), and 1.5 mM chlorophenol red β-D-galactopyranoside (CPRG), and 0.1M sodium phosphate buffer added to a final volume of 1.2 ml, minus the required volume for the lysate. The mix was added directly into a cuvette and warmed to 37°C. The lysate was added into the prewarmed reaction mix, mixed and the absorbance measured at 570nm for 12 min at 37°C.

10.3 CAT assay

A combination of two methods (Sambrook *et al*., 1989; Berger & Kimmel, 1987) was used to measure the CAT activity. In brief, an estimated 100μg of total protein from the cell extract was mixed with 250 mM Tris-HCl pH 8.0 to bring the volume to 159μl, to which 1 μl of [^{14}C]-chloramphenicol (American Radiolabeled Chemicals, Inc., St Louis, MO) was added. Twenty microliters of acetyl-CoA (Sigma) was added last to a final concentration of 0.4gm/ml to start the reaction and to bring the final reaction volume to 180μl. The reaction mixture was incubated at 37°C for 4 h, after which, 1ml ethyl acetate was added to end the reaction and to simultaneously extract the acetylated [^{14}C]-chloramphenicol from the aqueous phase. After a 5 min centrifugation at 19000 g, 900 μl of the upper organic phase was collected, transferred to a fresh microfuge tube and air dried under vacuum. The reaction product was redisolved in 25 μl of ethyl acetate and

immediately spotted on a thin-layer chromatography (TLC) plate. The plate was placed in a developing chamber with a running buffer of chloroform:methanol (95:5 ratio). The TLC plate was then exposed to a phosphor-imaging plate.

11. Computational analyses

11.1 QTL analysis

The QTL analysis was based on the design developed by Weller *et al*. (1990) in a mixed model analysis (SAS, NC). The design assumes QTL marker effect to be fixed and followed the model:

$$Y_{ijkl} = G_i + S_{ijk} + M_{ij} + e_{ijkl}$$

The Y_{ijkl} is DDev trait value of granddaughter l (daughter of sire k, granddaughter of grandsire i) that received marker j. S_{ijk} is the effect of sire k (son of grandsire i) that carried the marker j. M_{ij} is the effect of marker j of grandsire i. The design represents a hierarchical nested analysis.

Basically, sons of heterozygous grandsires were genotyped, and quantitative trait measurements of daughters from the homozygous sons analyzed. Each allele (marker designated by a notation) is scored in half-sib sons and traced to the daughter subgroups, and an analyzed utilizing a mixed model analysis in SAS. The least square means of the daughter deviations for health and milk production traits were estimated and an F-test conducted to evaluate the significance of differences between the least square means. A least square difference test was conducted to detect significant differences between the trait values of animals receiving alternative alleles of b*BTN*.

11.2 Quantitation of northern blots

Images from northern blots were captured with a phosphor-imager (Molecular Dynamics) and quantified using ImageQuant software (Amersham Biosciences Corp., Piscataway, NJ). The density of each band was measured by estimating the number of pixel values in each delimited area. The pixel values for each band of interest were normalized against similarly acquired pixel values for control bands.

11.3 Quantitation and statistical analysis of promoter activity

Bands representing single and double [^{14}C]-acetylated chloramphenicol were delimited and their pixel intensity values minus the background pixel intensity values measured. The net pixel values were then normalized to their respective average β-galactosidase activities.

Corrected pixel values were subjected to a one-way analysis-of-variance using the SAS statistical software. Basically, a procedure that handles mixed variables was used to estimate the least square means and to conduct pairwise comparisons.

11.4 DNA sequence comparison and analysis

BTN of four type genera, namely: cow, goat, human and mouse were the basis for cross-species primary structural analysis. Sequences of the four species were obtained from either cloned sequence or from the GenBank database. Transcription start sites (TSSs) were established by means of several approaches: (1) cow and human: based on previous published experimental evidence, and as annotated in GenBank, (2) goat: by comparative exon mapping against the cow sequence; and (3) mouse: exon mapping

against cDNA cloned sequence (Open Biosystems). Except for the 5' flanking region of the goat *BTN* (g*BTN*) which is circumscribed by the length of available sequence (704bp), both boundaries of the other sequences and the 3' end of the g*BTN* sequence were all based on these inherent properties relative to their axial TSS. Although, the sequences were delimited by these attributes, due to inherent qualitative differences the length of sequence extending relative to the TSS was non-axisymmetrical and non-congruent.

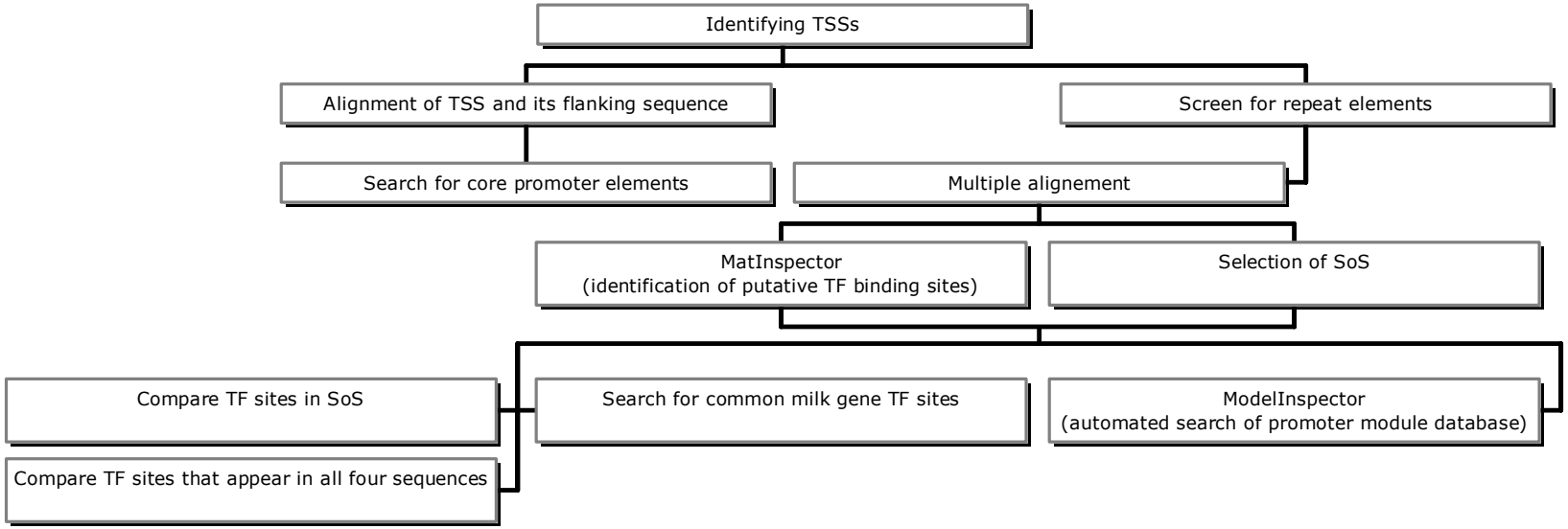
11.4.1 Identifying sequence features

The fragments selected for sequence analysis were scanned for CpG islands using the EMBOSS program of the European Bioinformatics Institute launched via Vector NTI © (InforMax). Search criteria were set at an observed/expected ratio of >0.60, length >100, and C+G > 50%.

Repeat elements were identified by submitting sequences to the National Research Council - Institute for Advanced Biomedical Technologies' WebGene program launched from the Vector NTI© platform. Searches for repeating elements in the human and mouse sequences were analyzed against a species-specific database, while the cow and goat genes were compared against a database for other vertebrates. A segment of g*BTN* that was not picked by the program but was highly homologous to a repeat region in the cow sequence was assumed to be a repeat region.

Repeat elements were precluded from subsequent analysis to obviate the detection of spurious putative transcription factor (TF) binding sites. Figure 3 is a flow diagram representing the sequential steps in the analysis of the primary sequence of the *BTN* promoter and adjoining 5'UTR.

Figure : Schematic of strategy of putative TF binding sites of butyrophilin gene



11.4.2 Local multiple and pairwise alignments

Local multiple alignments were conducted via the web-based DiAlign package of the GEMS Launcher hosted by the Genomatix © (Quandt *et al*., 1995; www.genomatix.de). Alignments were scored based on similarity against the shortest sequence, and assigned a percentage similarity value and a similarity index. Similarity values (relative to the maximum similarity) for each pairwise alignment were generated.

11.4.3 Selection of segments of similarity (SoSs)

Selection of segments of similarity for phylogenetic footprinting was heuristic and based on the degree of relative local alignment value. The selection was restricted to regions for which sequences from all four species were available.

11.4.4 Detection of putative TF binding sites and modules

The sequences were submitted to the GEMS Launcher program hosted by Genomatix. This program made use of MatInspector (professional version 6.2.2) to search the vertebrate TF matrix family library (version 3.3). In order to minimize false positive matches, the optimized matrix threshold was established as a criterion for searching and selecting putative TF binding sites. Generally, the output provides footprint and anchor positions in the sequence and the random expectation values as computed against the random occurrence of the matrix in a random 1kb DNA fragment. The aligned sequences were also submitted to Genomatix © for an automated screening of promoter modules that were previously described and held in a databank library.

11.4.5 Alignment of TSS region and search for core promoter elements

The sequence spanning -60 to +40 relative to the TSS in all four species was subjected to multiple alignment. A search for core contextually positioned promoter elements TATA-box, downstream promoter element (DPE), initiator (Inr) and TFIIB recognition element was conducted (See review by Smale and Kadonaga, 2003).

11.4.6 Collation and comparison of TF sites prevalent in all four species

Putative transcription sites that were present in all four species, including those present in segments that showed only weak similarity, were singled out and compiled for further analysis. The compilation was searched for modular patterns of occurrence.

11.4.7 Search for commonly occurring 'milk gene' TF sites

The output of all the putative TF binding sites was searched for sites that match those described for other 'milk genes' (See review by Rosen *et al*., 1999).

RESULTS

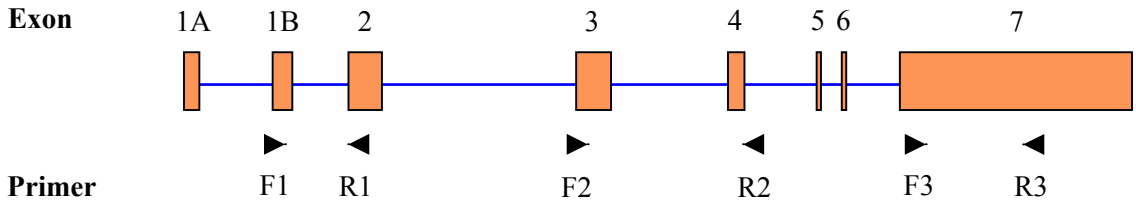
Screening for RFLPs

Genomic DNA samples from thirteen Holstein dairy animals were collected, *viz*, three bulls and ten cows, to screen for RFLPs using a PCR approach. Three primer pairs that encompass the intragenic region of b*BTN* were used to generate PCR fragments that cover the 5' end, central region and the 3' end of the gene (Figure 4). Polymorphisms were identified by digesting the amplified DNA fragments (amplicons) with a panel of six- and four-base cutting endonucleases.

Several RFLPs were identified in all three amplicons utilizing the endonucleases *Hpa*II, *Taq*I, *Hae*III, *Mbo*I, and *Tru*9I (Table 2). Subsequent sequencing of the polymorphic sites revealed a single polymorphism in the F2/R2 amplicon using *Mbo*I and *Tru*9I (Figure 5), while a two nucleotide substitution resulted in the polymorphism that was detected by *Taq*I (Figure 6) in the F1/R1 amplicon. Another RFLP due to a single base pair substitution was also detected by *Hpa*II the in F1/R1 amplicon (Figure 6). The only RFLP detected in F3/R3 amplicon was a *Hae*III polymorphic site (Figure 7). All the RFLPs were detected with four base cutters.

All of the RFLPs, except the one identified by *Hae*III, are intronic. The single nucleotide substitution in the F3/R3 amplicon predicts an amino acid substitution of lysine to arginine, and is in the B30.2 domain. However, the polymorphic position is not a conserved residue in the B30.2 domain. The significance of this substitution has not been biochemically evaluated.

A)



B)

Primer pairs			Region amplified
Name	**Sequence (5' to 3')**	**Annealing T^0C**	
1F	CTGAAGTTCCCGACAAACTCG	57	Exon1B, Intron 1B
1R	CTCTGCATCTTCACCCACCAC		
2F	CTTCTTCCCAAGGCTGAC	57	Exon3, Intron C, Exon4
2R	CTTACTGAGCTCTTCCAGG		
3F	TCCCGAGAATGGGTTCTG	52	Exon 7
3R	CACTGCCTGAGTTCACCTCAT		

Figure 4: A) Location of primers used for amplifying regions of the b*BTN* gene. Boxes represent exons (corresponding numbers shown above), and lines represent introns. Primers are indicated by arrow heads, with primer pairs sharing the same number. B) Sequence of primer used for PCR and their corresponding attributes.

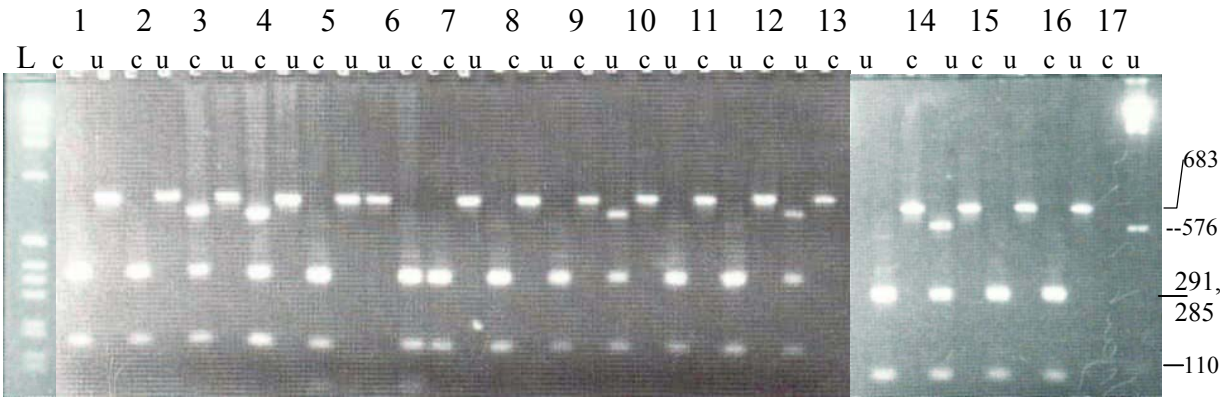
Table 2: Endonucleases and the respective polymorphisms

Primer pair	Amplicon size (in bps)	PCR-RFLP		
		Endonuclease	**Nucleotide substitution**[1]	**Ref. location** [2]
F1/R1	576	*Hpa*II	CAGG → C↓CGG	1938-1941
		*Taq*I	ACAA → T↓CGA	2074-2077
F2/R2	683	*Mbo*I	↓GATC → AATC	5495-5498
		*Tru*9I	TTGA → T↓TAA	5493-5496
F3/R3	893	*Hae*III	AGCC → GG↓CC	6804-6807

[1] Four-base-pair endonuclease recognition sequence with cut sites denoted by downward arrows and nucleotide substitution in bold.

[2] Locus on bBTN based on Genbank sequence accession number Z93323.

A)



B)

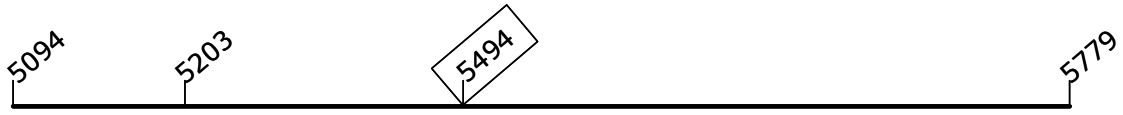
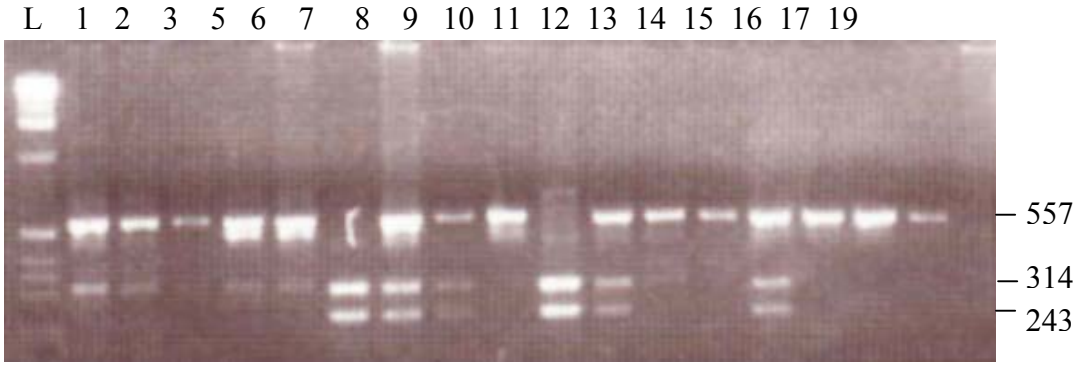
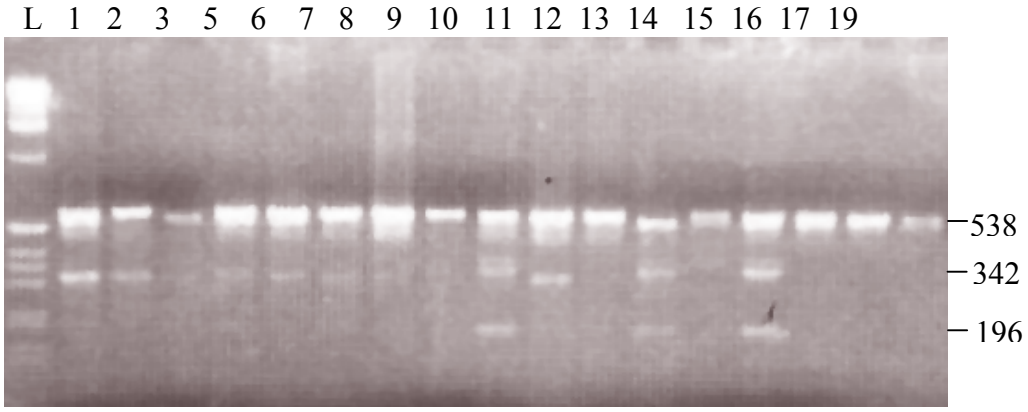


Figure 5: Amplicon F2/R2 (683bps): A) examples of animals homozygous (lane 1) and heterozygous (lane 3) for the presence of the *Mbo*I recognition sequence. PCR products denoted as 'u' are pre digestion; and 'c' are post digestion. B) Restriction enzyme sites of *Mbo*I. Polymorphic site is boxed. Locus numbers are as in Figure 2.

A)



B)



C)

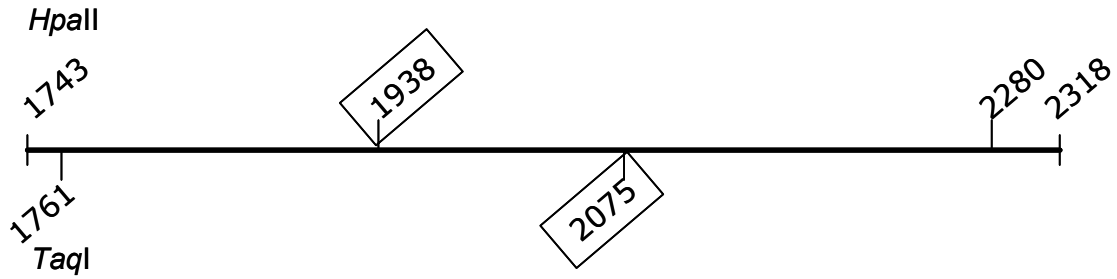
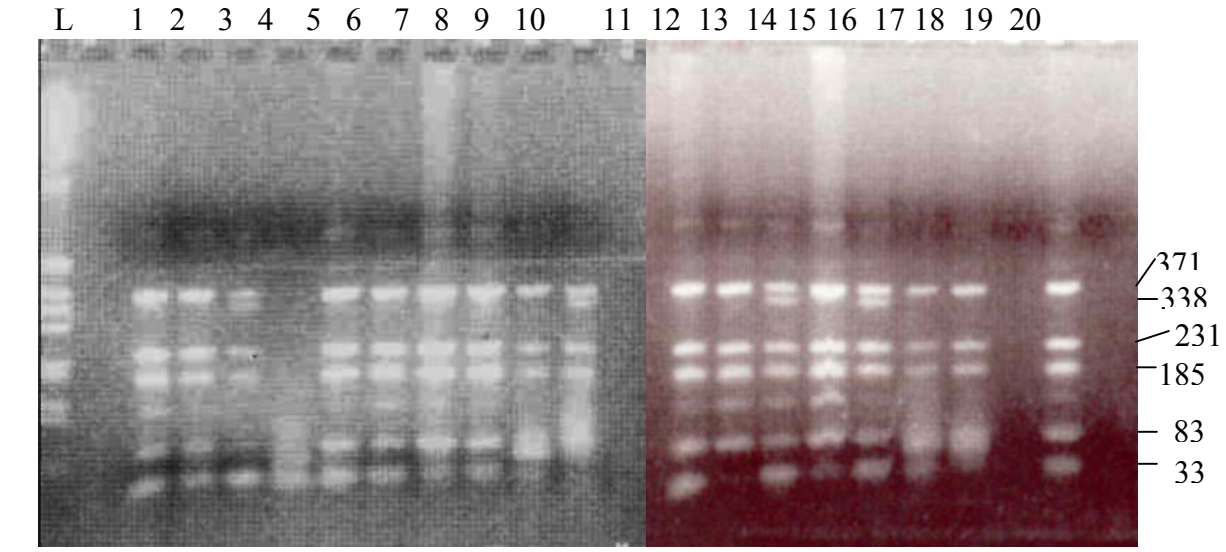


Figure 6: Amplicon F1/R1 (576bp): restriction digest profiles of grandsires. A) *Taq*I endonuclease; examples of animals that are homozygous for the absence of the endonuclease recognition sequence (lane 3), and for animals that are either heterozygous (lane 8) and homozygous (lane 7) for presence of the endonuclease recognition sequence. B) *Hpa*I endonuclease; examples of animals homozygous (lane 9), and heterozygous (lane 10) for presence of endonuclease recognition sequence. Notice also other banding patterns in samples 1 and 11. C) Restriction enzyme sites of *Hpa*II and *Taq*I. Polymorphic sites are boxed. Locus numbers are as in Figure 2.



B)

Figure 7: Amplicon F3/R3 (893bps): A) restriction digest profiles of grandsires. *Hae*III endonuclease. Examples of homozygous (lane 2) and heterozygous (lane 3) for endonuclease recognition sequence. B) Restriction enzyme sites of *Mbo*I. Polymorphic site is boxed. Locus numbers are as in Figure 2.

Genotype of grandsires

The presence of a large number of half-sib families in the commercial dairy population of the United States and the routine measurement and collection of quantitative and qualitative traits in these herds allows the utilization of the GDD for the detection of QTLs. To this end, genomic DNA from twenty proven grandsires was screened for the identified PCR-RFLPs (Figures 5, 6 and 7) described above.

Results of the entire genotypic profile of the grandsires are presented in Table 3. Grandsires 10, 13, and 15 possessed heterozygous marker genotypes and their families were thus selected for QTL analysis in a GDD. Both grandsires 8 and 9 are heterozygous at the *Taq*I marker locus; however, they both exhibited homozygosity at all other marker loci tested. The grandsire 9 family was thus included in the QTL analysis specifically for the one marker that exhibited heterozygosity.

Genotype of sons

Genomic DNA of sons of the four selected grandsires was genotyped for the identified PCR-RFLP markers. The number of informative sons per grandsire for each marker was greater than 50, while the total number of informative sons per marker ranged from 68 to 259. None of the grandsires had less than 45% informative sons (Table 4). Sons with homozygous marker genotypes were selected for QTL mapping.

Allelic distribution

The markers in amplicons 2 and 3 suggest that b*BTN* has two alleles; however, haplotypes generated from the polymorphic markers of amplicon 1 indicate more than

Table 3: RFLP marker genotypes of grandsires

Grandsire	Intron 2	Exon 5 to Intron 6		Exon 8
Id No.	***Taq*I**	***Mbo*I**[§]	***Tru*9I**	***Hae*III**
1	-	AA	AA	AA
2	-	AA	AA	AA
3	-	AB	AB	AB
4	-	AB	AB	-
5	-	AA	AA	AA
6	-	AA	AA	AA
7	BB	AA	AA	AA
8	AB	AA	AA	AA
9*	AB	AA	AA	AA
10*	AA	AB	AB	AB
11	BB	AA	AA	AA
12	AB	AA	AA	AA
13*	AA	AB	AB	AB
14	AA	AA	AA	AA
15*	AB	AB	AB	AB
16	AA	AA	AA	AA
17	AA	AA	AA	AA
18	AA	AA	-	-
19	AA	AA	-	-
20	AA	AA	-	-

* Families of these grandsires were selected for further QTL analysis.
† The selection of the allele letter denotations do not reflect dominance or recessiveness.
§ Only 5 of 20 identified *Mbo*I genotypes are heterozygous, *i.e.* an estimated heterozygosity of 0.25.

Table 4: Number and frequency of informative sons in each grandsire family

RFLP Marker	Grandsire ID	No. informative sons/grandsire	% informative sons/grandsire
*Tru*9I	10	68	57.35
	All	68	57.35
HpaII	10	72	52.78
	13	124	44.35
	15	58	37.93
	All	**254**	**45.27**
TaqI	9	52	55.77
	15	61	55.74
	All	**113**	**55.75**
MboI	10	73	54.79
	13	123	44.72
	15	63	39.68
	All	**259**	**46.33**
HaeIII	10	68	55.88
	13	90	47.78
	15	63	38.10
	All	**221**	**47.51**

two alleles have been detected. Particularly in the intron 2 region, the restriction enzyme *Hpa*II did not provide a clear pattern that corresponded with the sequence available nor with the allelic identification obtained from the other RFLP sites: nor did the *Taq*I marker loci of the same amplicon (Figure 6).

Gene diversity analysis using the *Mbo*I marker in the sons of grandsire families 10, 13, and 15 provides a frequency distribution of each allele. Although there are nearly twice as many **A** alleles in the population of dairy bulls examined, the observed heterozygosity for the b*BTN* loci remains close to 0.5 with the observed AA genotype at four times that of the BB genotype (Table 5).

QTL analysis

Records of economically important quantitative traits were obtained from the National Dairy Herd Improvement Association (DHIA). Data on seven yield traits, namely, milk, protein, percent protein, fat, percent fat, herd life and somatic cell score (SCS) were utilized as weighted daughter deviations to test allelic substitution effects.

Analysis of allelic substitution at the various RFLP loci within the b*BTN* gene was conducted. More than one marker loci was analyzed for most grandsire families for each quantitative trait tested.

Most allelic substitution effects tested for all the available economically important traits showed no statistically significant difference. Only one statistically significant effect was detected in grandsire family 15 for percent protein ($P < 0.01$). All other results indicate that no effects on the quantitative traits investigated are due to the allelic inheritance patterns of b*BTN*. In some instances, the standard error estimates were larger

Table 5: Observed genotype and gene frequency of b*BTN Mbo*I marker loci.

Genotypes	Count	Genotypic frequency	Allele Frequency	
AA	96	0.37	A	B
AB	139	0.54	0.64	0.36
BB	24	0.09		

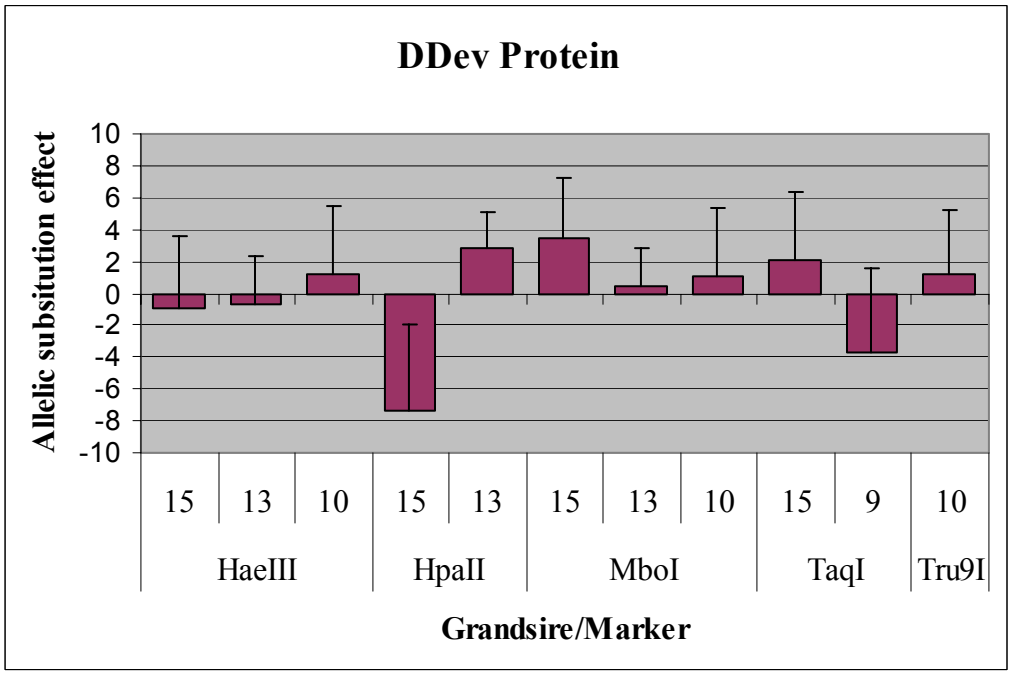
than the effect due to allelic substitution, indicating that variability in the measured quantitative traits is mostly explained by factors other than allelic differences (Figure 8).

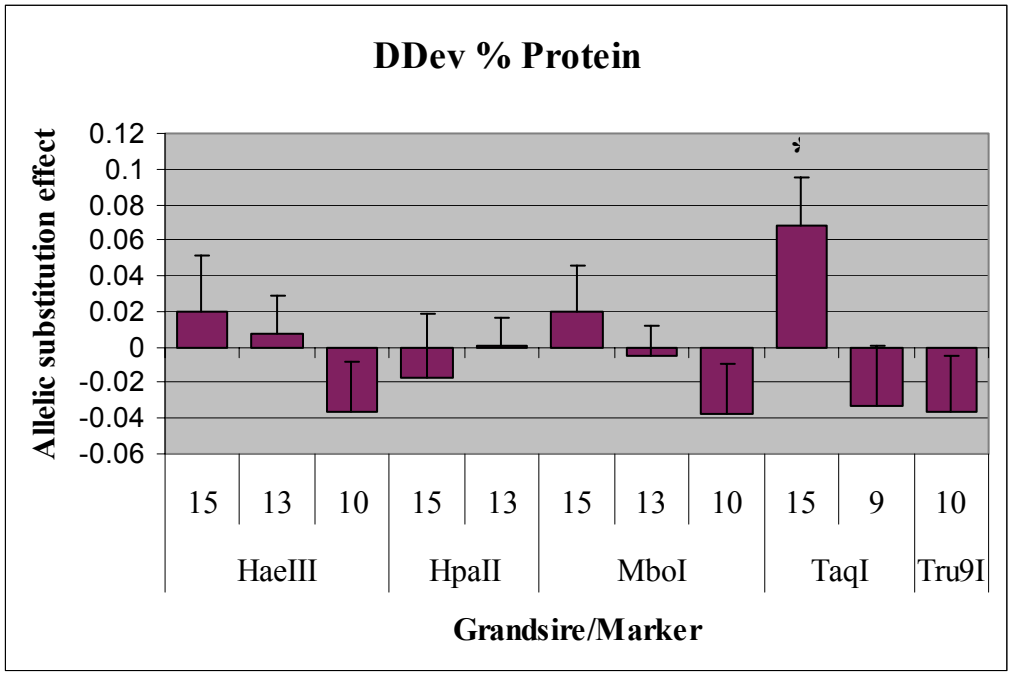
The single significant difference found was limited to one grandsire family and to only one PCR-RFLP in intron 2. In addition, grandsire family effects were assessed simultaneously with the QTL analysis, which revealed that tested significantly for sire effect on percent protein and other quantitative traits on the most part due to the grandsire 15 component. Therefore, the marginal significance detected for percent protein may be incidental to a grandsire family effect.

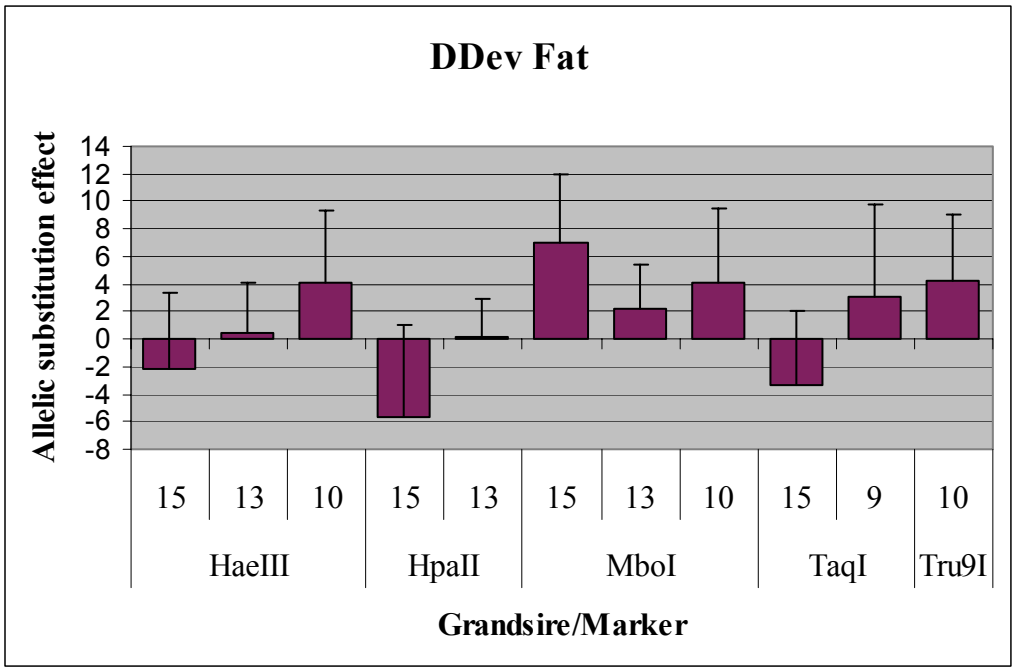
Since the *Mbo*I (intronic) marker strictly segregated with the *Hae*III (exonic) marker, any QTL analysis performed on the former is reflective of the latter. None of the economical traits tested showed any significant allelic substitution effects for these two markers.

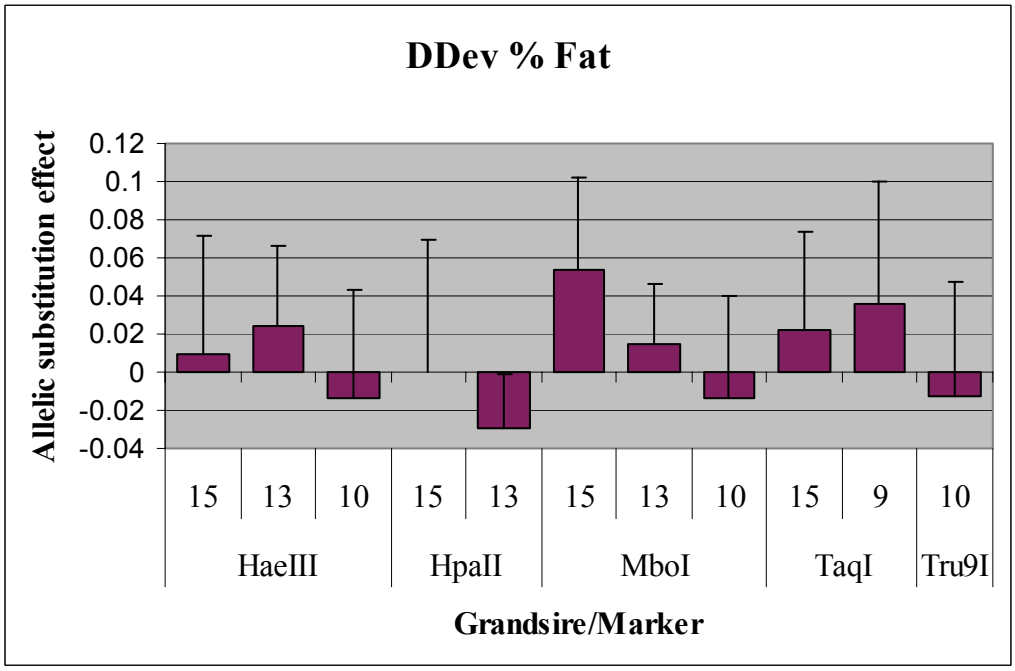
Cloning and subcloning of b*BTN* 5' flanking region

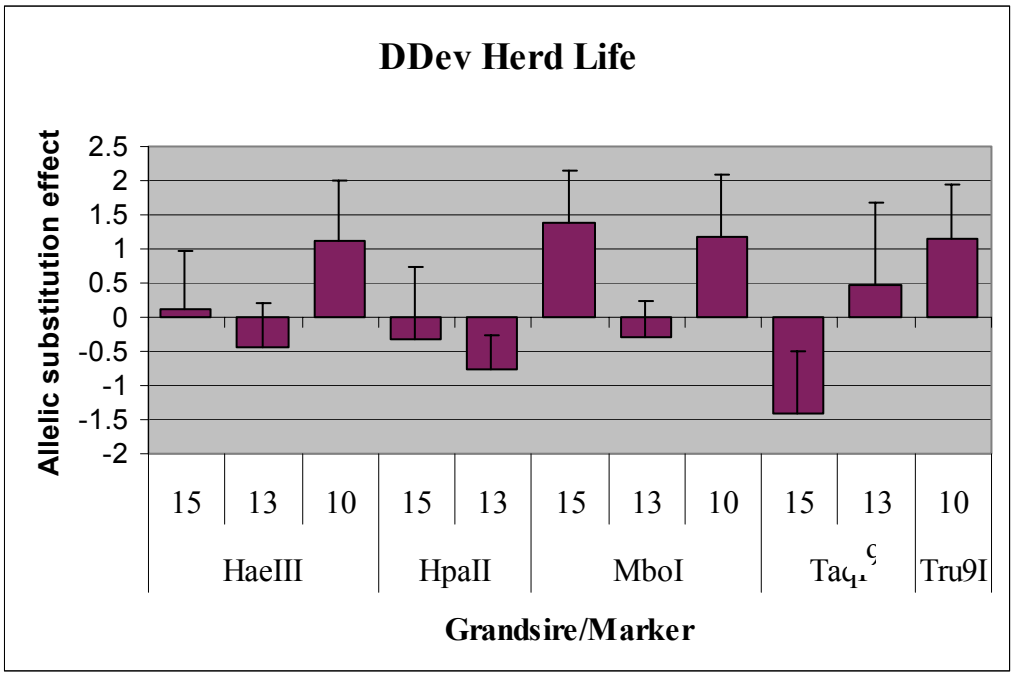
A 0.9kb b*BTN* cDNA fragment, which extends from the start of exon IB to exon III, was used to screen a cow genomic library (ClonTech®). A positive clone >15kb designated *5'bBTN* was picked and plaque purified for further use. For ease of manipulation, smaller fragments of 5'b*BTN* were generated by digestion with restriction enzymes and inserted into the multiple cloning site of the pBluescript II KS ® phagemid vector (Stratagene, La Jolla, CA). The cloned DNA was sequenced either directly from the phage vector, or from the smaller fragments that were subcloned into the pBluescript II KS ® vector (Figure 9).

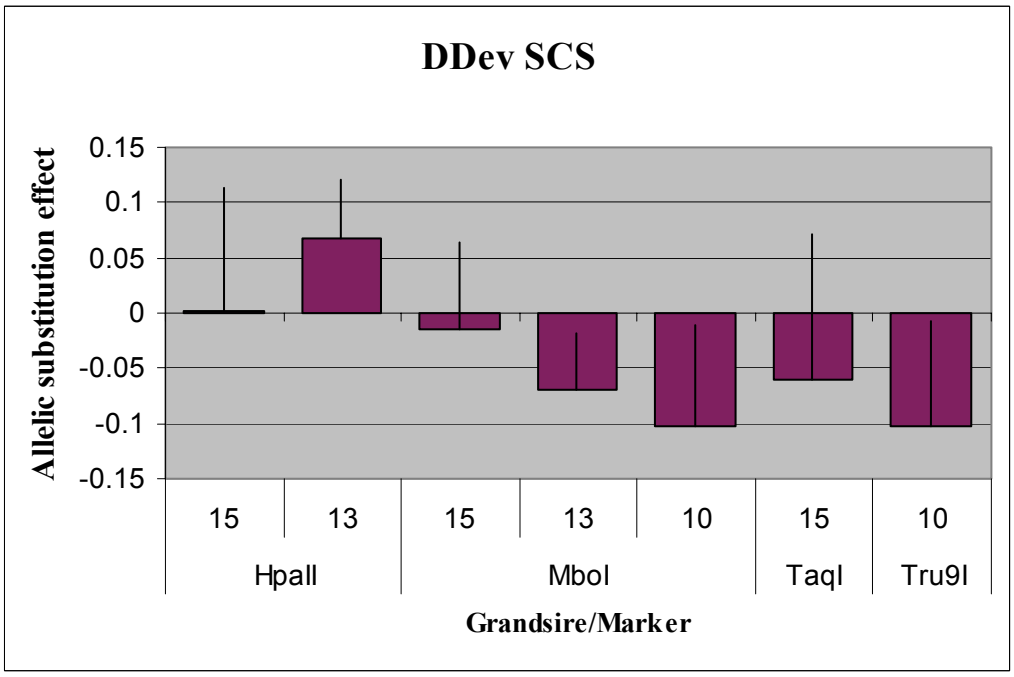

DDev Protein
Allelic subsitution effect
10
8
6
4
2
0
-2
-4
-6
-8
-10
15
13
10
15
13
15
13
10
15
9
10
HaeIII
HpaII
MboI
TaqI
Tru9I
Grandsire/Marker


DDev % Protein
Allelic substitution effect
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
15
13
10
15
13
15
13
10
15
9
10
HaeIII
HpaII
MboI
TaqI
Tru9I
Grandsire/Marker


DDev Fat
Allelic substitution effect
14
12
10
8
6
4
2
0
-2
-4
-6
-8
15
13
10
15
13
15
13
10
15
9
10
HaeIII
HpaII
MboI
TaqI
Tru9I
Grandsire/Marker


DDev % Fat
Allelic substitution effect
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
15
13
10
15
13
15
13
10
15
9
10
HaeIII
HpaII
MboI
TaqI
Tru9I
Grandsire/Marker


DDev Herd Life
Allelic substitution effect
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
15
13
10
15
13
15
13
10
15
13
10
HaeIII
HpaII
MboI
Tru9I
Grandsire/Marker


DDev SCS
Allelic substitution effect
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
15
13
15
13
10
15
10
HpaII
MboI
TaqI
Tru9I
Grandsire/Marker

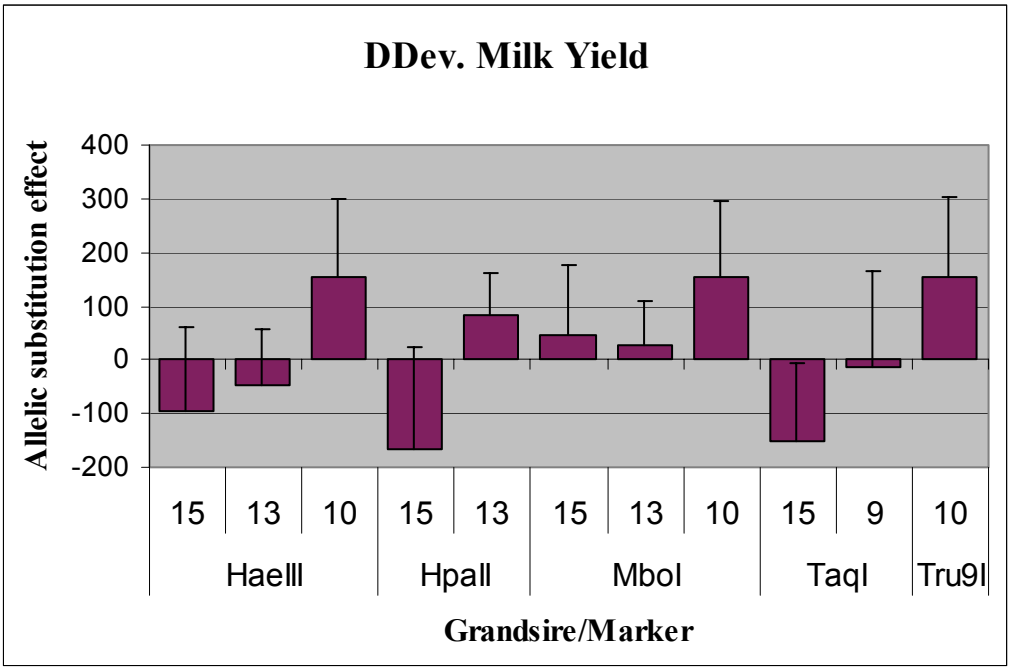


Figure 8: Allelic substitution effects on seven economically important traits by RFLP marker for each grandsire family. Traits are presented as weighted daughter deviations (DDev) adjusted for dam predicted transmitting ability (PTA). Statistical significance $LSD_{0.01}$ denote by asterisk.

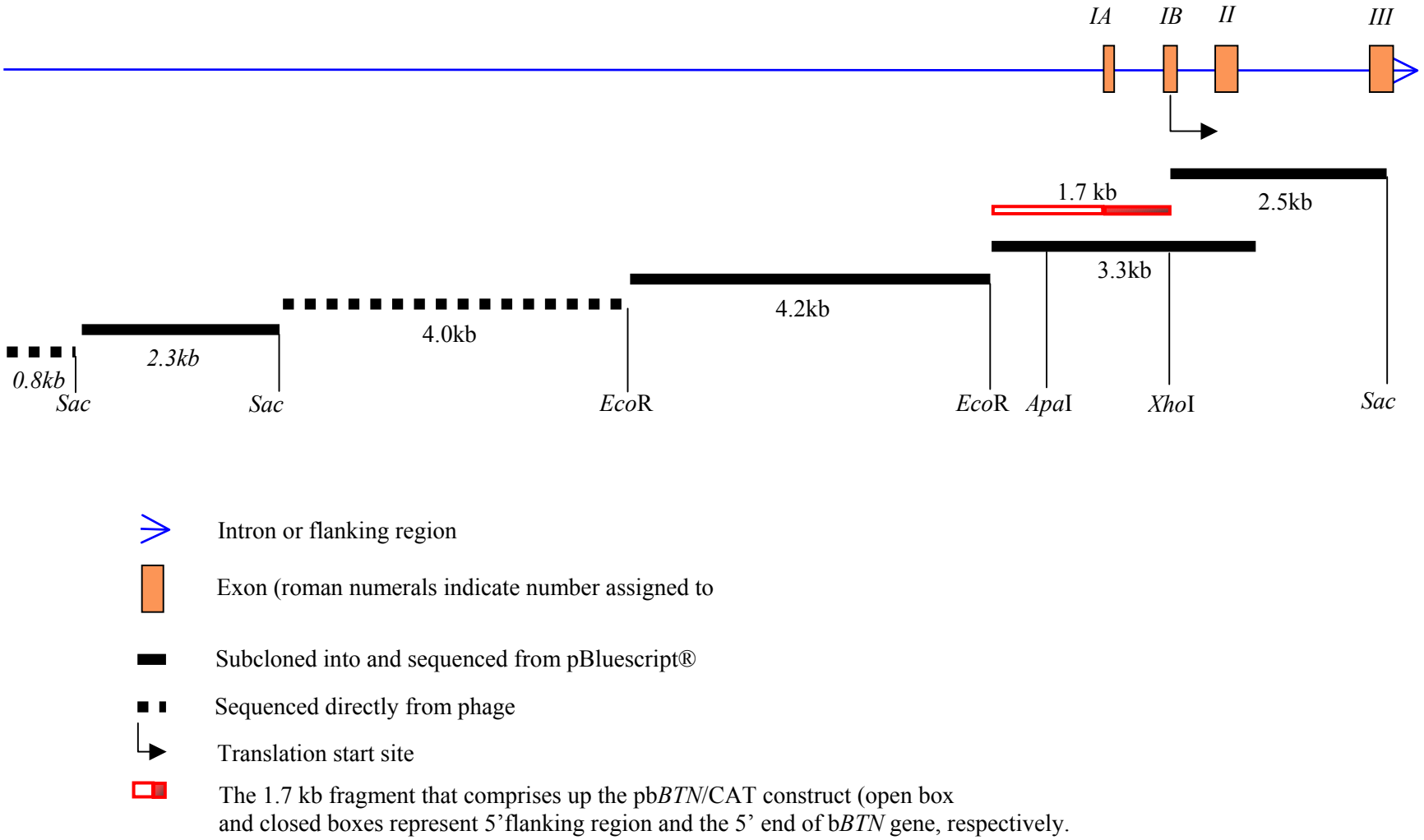


Figure 9: Map of the bovine genomic clone of b*BTN* and its flanking sequence, depicting the restriction sites of the various subclones, and sequencing strategy.

The 15 kb cloned DNA includes 12kb of 5' flanking region of b*BTN*. A single subclone of 2.5 kb was generated by digestion with *Sac*I/*Xho*I. Two subclones with sizes of 4.2 kb and 3.3 kb were generated by digestion with *Eco*RI and represent contiguous segments (verified by mapping against GenBank accession № AF037402, Husaini *et al*., 2001); while a 2.3 kb fragment was generated by digestion with *Sac*I. The 2.5 kb *Sac*I/*Xho*I subclone overlaps with the 3.3 kb clone. The 3.3 kb fragment contains all of the 5' UTR and 1 kb of proximal promoter region and was thus the starting material for preparing promoter reporter constructs (Figure 2).

The cloned cow DNA was sequenced to verify congruency and compared with the 6.8 kb sequence in the database (Accession № AF037402, Husaini *et al*., 2001) that encompasses 5.6 kb of the 3' end of the cloned 5'b*BTN*. Other regions of the cloned DNA were sequenced directly from the phage DNA after purifying the phage DNA using a microcon-100 filter (Millipore ®, Bedford, MA).

Cloning of CAT reporter vectors

The basic reporter plasmid, pCAT-Basic Vector ® (Promega: Madison, WI) (Figure 1), was used to make test reporter constructs to assay for promoter/enhancer activity of the b*BTN* 5' and 3' flanking regions. This basic vector lacks eukaryotic promoter and enhancer sequences. Eight reporter vectors were constructed that encode either, or both, the 5' or 3' regions of b*BTN*. The 5' fragment covers 0.7 kb of the 5' UTR and either 0.45 kb or 1.04 kb of flanking sequence and were inserted 5' of the CAT reporter gene. The 3' fragment encompassed 273 bps of 3' UTR and 1.4 kb of flanking sequence and was inserted 3' of the CAT reporter gene in the correct or reverse

orientation (Table 1). A positive assay control for CAT activity was provided by pcDNA3/CAT (Invitrogen ® Carlsbad, CA) which contains a CAT gene driven by the cytomegalovirus promoter that allows constitutive expression in mammary cells and the basic reporter plasmid served as a negative control.

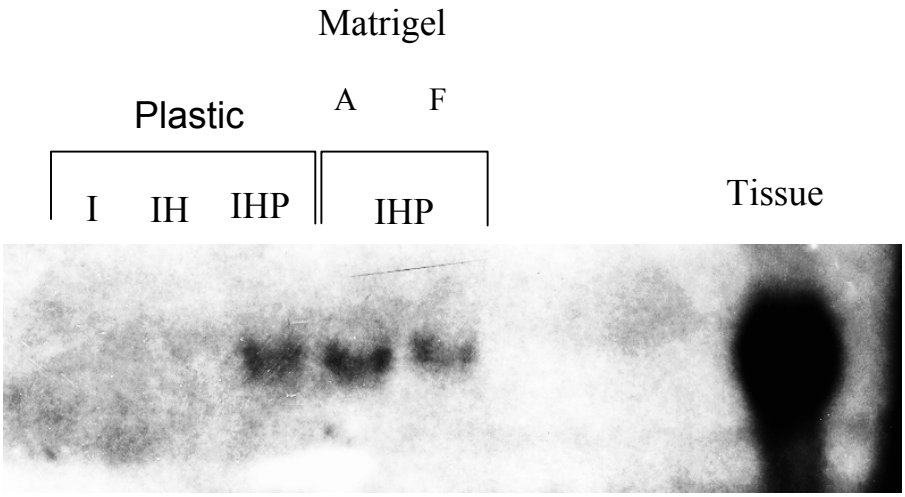
Endogenous m*Btn*1a1 expression in HC11 cells (initial cell line)

Currently, there is no well established and characterized cow mammary cell line that is homogenous and displays the hallmarks of mammary differentiation. Therefore, all transient transfections for ensuing CAT expression assays were performed in the HC11 mammary cell line, derived from mid pregnant BalbC mice (Ball *et al.*, 1988). In order to establish a lactogenic hormone combination for endogenous m*Btn*1a1 expression in HC11 cells, m*Btn*1a1 was detected by Northern blot of total RNA or mRNA from cells incubated in medium containing either insulin (I), insulin + hydrocortisone (IH), or insulin+ hydrocortisone + prolactin (IHP). An G3PDH probe was used as an internal loading control for the mRNA Northern assay. The HC11 cells express endogenous m*Btn*1a1 message on either plastic or matrigel substrata, but require the combination of IHP in the induction media for maximum expression levels (Figure 10).

Promoter assay for CAT expression

To identify the *cis*-acting elements present in the putative promoter and enhancer regions of b*BTN*, the reporter constructs were co-transfected along with a control plasmid encoding β-galactosidase, a reporter to correct for transfection efficiency, at a 3:1 ratio. At confluency, the transfected cells were incubated in the presence of the lactogenic

A) Total RNA



B) mRNA

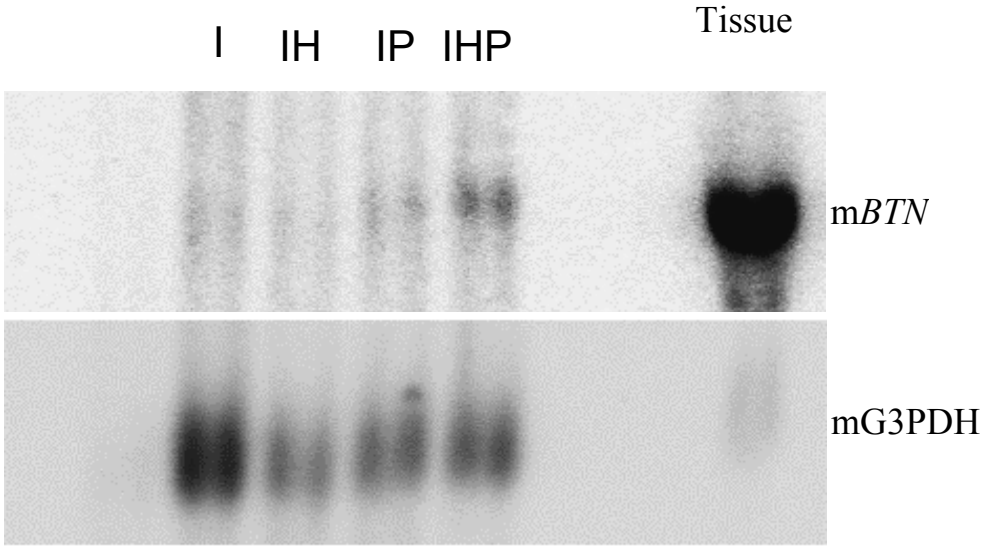


Figure 10: A) Northern blot analysis of total RNA (20µg per sample) from HC11 cells grown on either plastic, attached matrigel (A-matrigel), or floating matrigel (F-matrigel) for 24 h in the presence or absence of the lactogenic hormones insulin (I), hydrocortisone (H), and prolactin (P). Tissue from lactating mouse mammary gland as control is shown to the right of the figure. B) Northern blot analysis of mRNA (2 µg) from HC11 cells grown on plastic in the presence or absence of lactogenic hormones I, H and P.

hormone combination IHP. Cell lysates were assayed for CAT activity by following the acetylation of [C14]-labeled chloramphenicol. A PhosphoImager (Molecular Dynamics) was used to obtain quantifiable estimates of ^{14}C-chloramphenicol counts. Promoter activity was measured as relative pixel counts (surrogate measurements for acetylation) from a volume integration analysis (minus local background) using ImageQuant software version 5.2.08 (Molecular Dynamics, Inc.) and was normalized to the β-galactosidase activity measured from the same lysate.

Statistical analysis and results

Results of the CAT expression analyses are presented in Figure 11. Using a mixed model procedure, normalized CAT activity levels were subjected to analysis-of-variance followed by least square differences test to detect any statistically significant differences between the least square means (SAS Institute, NC). The statistical analysis conducted excluded the pCMV/CAT and no transfection controls, since these are experimental controls that would not provide a relevant comparison to the performance of the test constructs. Data from normalized relative pixel values were evaluated for test of normality prior to statistical analysis. Furthermore, since the variances amongst the samples were not homogenous, samples were grouped into variance groupings to the best fit model (using Bayesian information criteria).

The pb*BTN*1.2/CAT construct, containing the 0.45 kb 5' putative promoter fragment, had a small statistically non-significant difference from the pCAT-Basic control. However, the inclusion of the additional 0.5 kb piece in the pb*BTN*1.7/CAT construct yielded the highest level of CAT expression, indicative of an important *cis*-

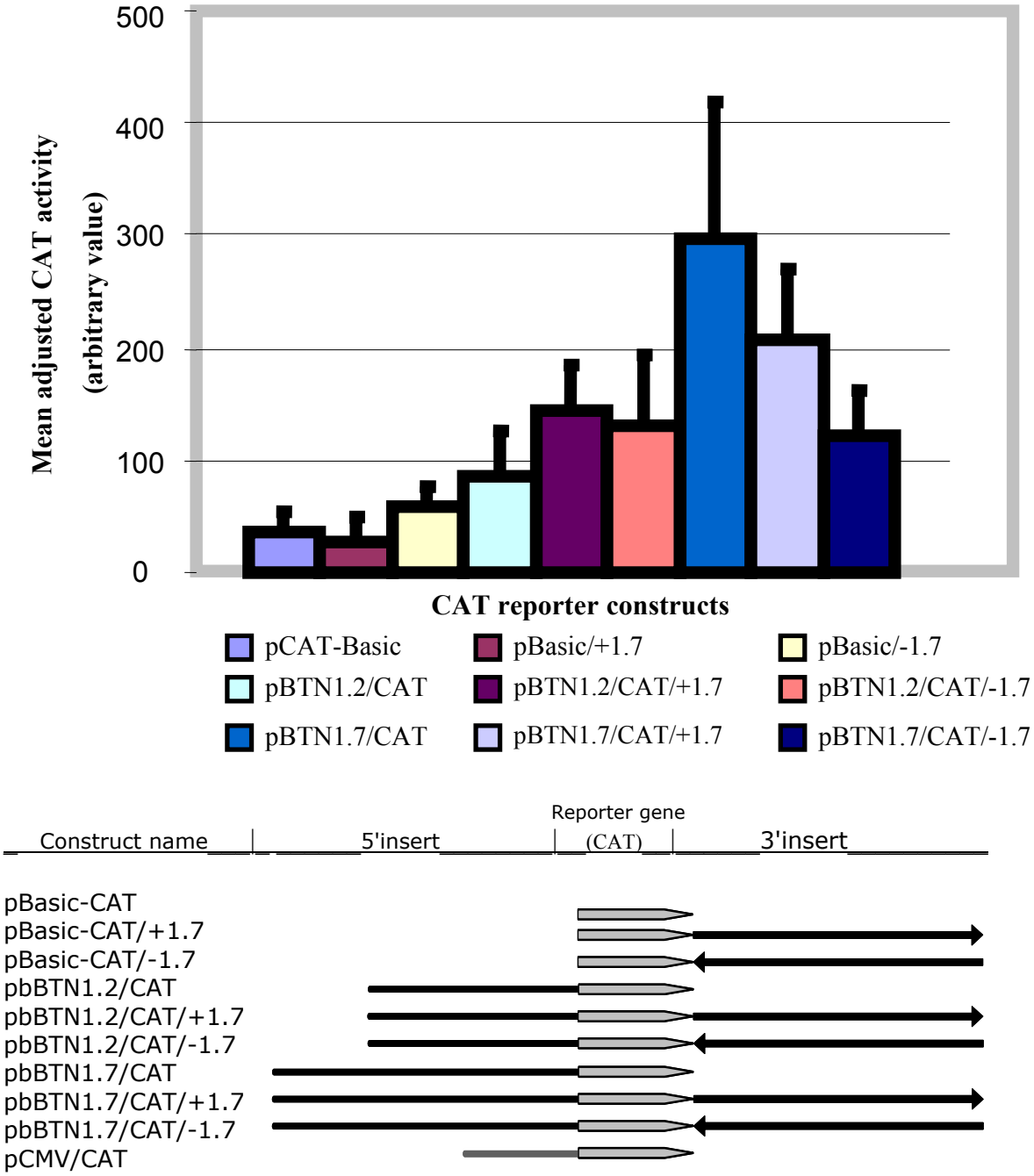


Figure 11: Comparison of potential promoter activity: CAT activity was measured indirectly from pixel values obtained from densitometric readings. The pixel values were normalized to the co-transfected β-gal levels. The two experimental controls, no transfection and pCMV/CAT, are not displayed. (Different letters signify LSD at $p < 0.1$). All assays were replicated five times, except pBasic/CAT/-1.7 which had four replicates.

element in this region. The CAT expression detected with the 1.2kb construct which included the 1.7 kb 3' fragment (in either orientation, *i.e.* pb*BTN*1.2/CAT/+1.7 or pb*BTN*1.2/CAT/-1.7) was statistically not different from the highest expressing construct, but distinctly different from the pCAT-Basic control.

The results obtained from the introduction of the 3' fragment to either the pb*BTN*1.2/CAT or pb*BTN*1.7/CAT vector seems bifurcated. While the statistical differences were not clearly evident, the 3' fragment appears to depress the enhanced expression activity that was attained by the 0.5kb piece in pb*BTN*1.7/CAT, while it increased the expression level of pb*BTN*1.2/CAT. This is suggestive of an interplay between the 3' region and the 5' region in coordinating expression of the CAT reporter gene. The clearest example of the depressive activity of the 1.7kb 3' fragment was seen with the pb*BTN*1.7/CAT/-1.7 construct, with which the ability to drive expression of the reporter gene was statistically non distinguishable from the least active pCAT-Basic negative control.

Clone 9: a subclone isolated from HC11 cell line

Subsequent tests on HC11 cells revealed that they had lost the property of expressing endogenous m*Btn*1a1 in response to IHP induction. Therefore, it was necessary to identify and isolate a new subpopulation of cells that have a higher expression level of endogenous m*Btn*1a1. A subclonal cell line, clone 9, was isolated from the previous HC11 cells. Low power microscopic examination reveals that clone 9 cells are more uniform in gross morphology as compared to low passage HC11 cells (Figure 12). A Northern blot analysis for the expression of endogenous m*Btn*1a1 reveals

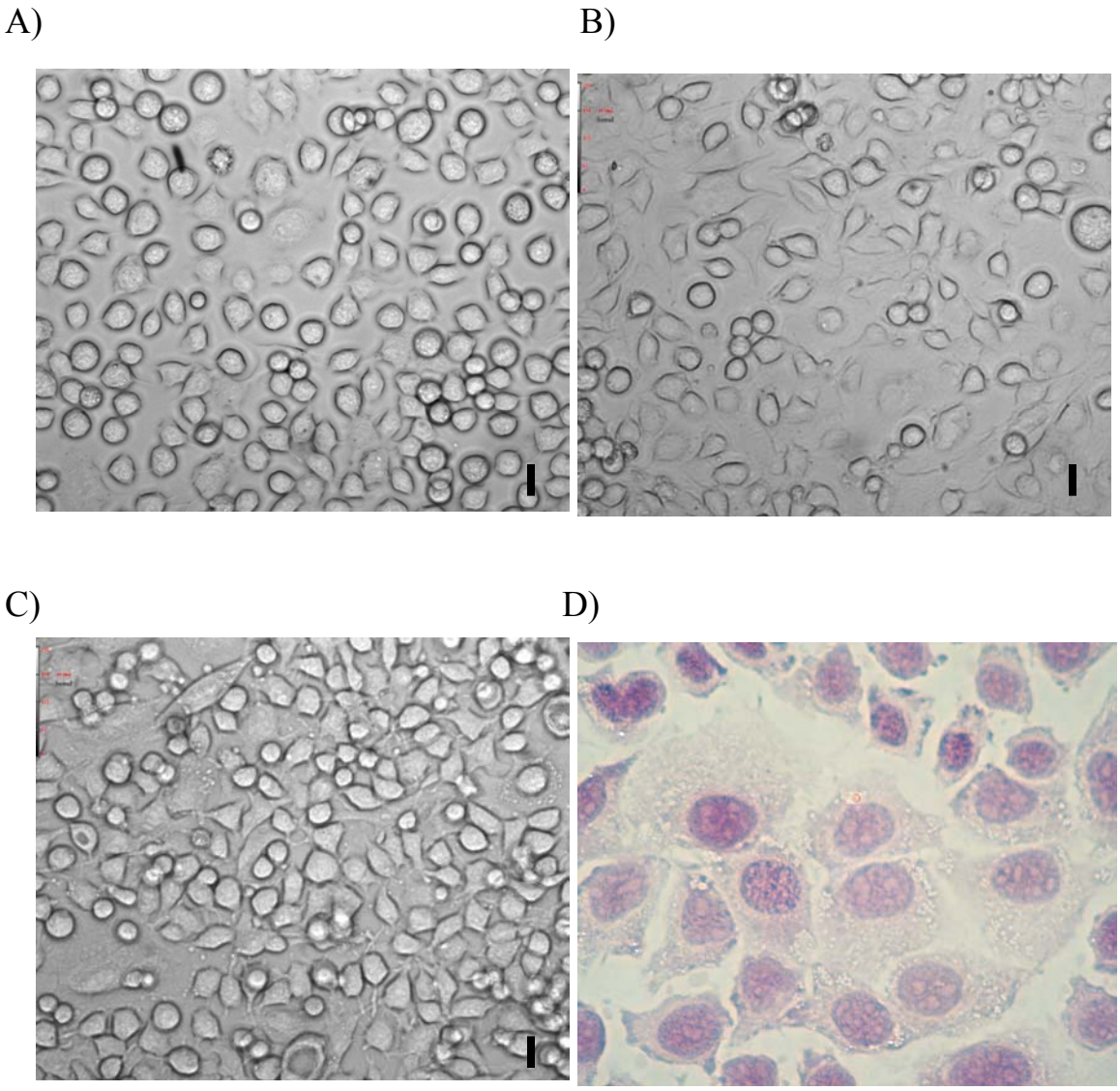


Figure 12: Light microscopy of a confluent culture of HC11 cells. Phase contrast (A-C) and Mayer's hematoxylin stained (D) of subclone 9 cells (A,D) as compared to low passage HC11 clones 1 (B) and 2 (C). Bars represent 20µm.

that, in clone 9 cells, expression is less dependent on the presence of all the lactogenic hormones in the induction media (Figure 13). Additionally, a subclone 16, isolated along with clone 9, also expresses m*Btn*1a1 in the presence of IHP, however, at a comparatively lower level when inspected in context of G3PDH internal control (Figure 13).

CAT assay in clone 9 cells

CAT assay using a limited set of the CAT test constructs and the basic vector control in clone 9 cells reveals statistically significant differences ($p < 0.05$) (Figure 14). Pairwise mean comparisons show that the pb*BTN*1.7/cat and pb*BTN*1.7/CAT/+1.7 constructs provided significantly higher CAT expression levels as compared to the basic control and pbBTN1.2/CAT constructs. However, the pb*BTN*1.7/cat and pb*BTN*1.7/CAT/+1.7 constructs were statistically the same as were the pb*BTN*1.2/CAT and the basic control. In congruity with results obtained in the original HC11 cells, the ~0.5kb 5' end of the putative promoter region provides a significant effect towards expression of the reporter gene.

Features of BTN sequences analyzed

The CAT assay results were complemented with a computational approach to identify potentially important *cis*-acting elements. Features and length that delineate the boundaries of the sequences set for comparison are presented in Table 6. The 5' flanking regions of the sequences ranged in size from 1042bp in cow to 704bp in goat; and 1000bp each in human and mouse. The 3' boundary of the sequences extends to the last

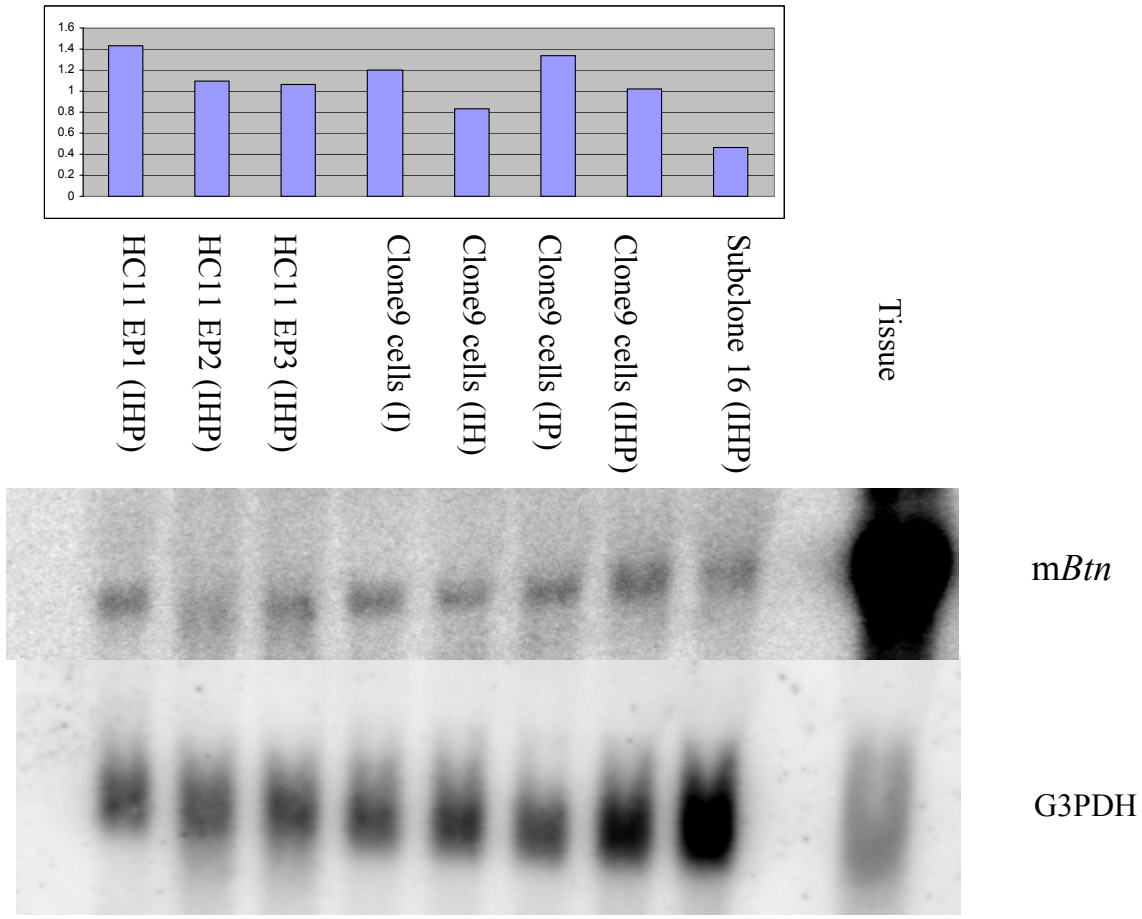


Figure 13: Northern blot analysis of 20 µg of total RNA derived from early passage clone 1, 2 and 3 (all cultured in the presence of IHP), clone 9 cells (cultured under I, IH, IP or IHP) and subclone 16 (cultured in the presence of IHP). Bar graph represents average m*Btn*1a1 normalized to G3PDH from two Northern blots; except clone 16 which was not replicated.

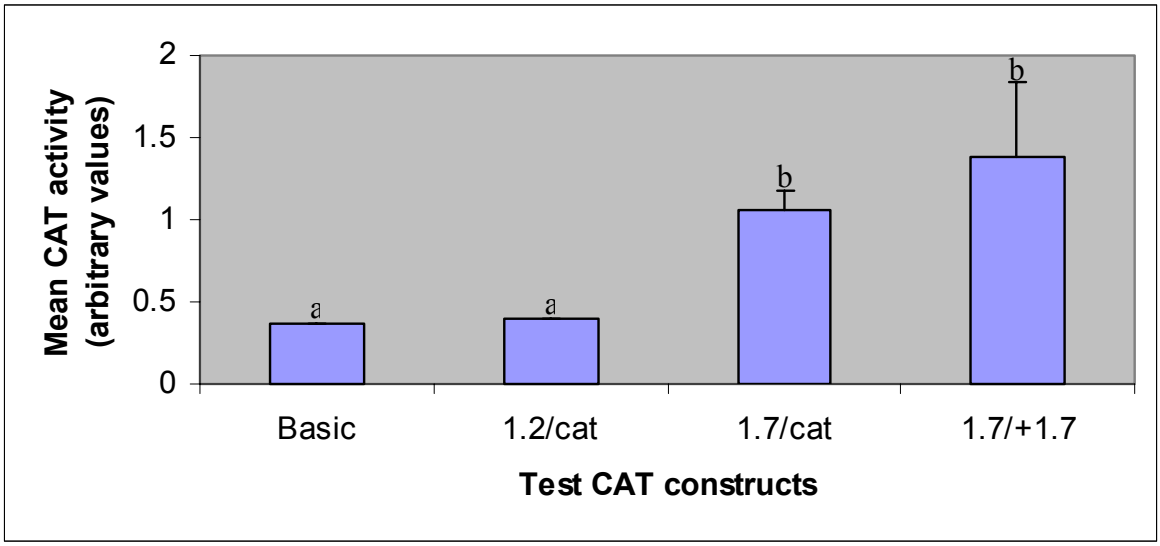


Figure 14: Comparison of potential promoter activity in clone 9 cells. CAT activity as measured indirectly from pixel values obtained from densitometric readings for some of the CAT test constructs and the basic vector. No transfection control and pCMV/CAT are not displayed. (Different letters signify LSD at $p < 0.05$). All samples were replicated four times.

Table 6: Sequence delineation of b*BTN*, g*BTN*, h*BTN*1A1 and m*Btn*1a1 used in comparative sequence analysis

Species	5' flanking (bp)	5'UTR to CDS	Total size (bp)
Cow	1043	707	1750
Goat*	704	733	1437
Human	1000	944	1944
Mouse	1000	695	1695

* Transcription start site of g*BTN* assigned based on high degree of homology to b*BTN* sequence and thus does not agree with the annotation with sequence deposited in Genbank

nucleotide before the start of their CDS, except in b*BTN* that stops 6 nts short of its CDS. TSSs in the sequences of cow (cloned fragment described previously), mice (accession № U67065), and human are delineated by exon mapping. However, in light of no published experimental evidence to substantiate the TSS of g*BTN* (accession № AY036083), TSS was determined by superposing it to the b*BTN* exon structure (therefore, it is not in congruency with the TSS annotated in the GenBank database).

Preparatory to the comparative analyses of the orthologous *BTN* genes, repeat regions in b*BTN* and its orthologues in goat, human and mouse were identified using species-specific repeat element databases for human and rodent, while that of cow and goat was based on a database compiled for 'other' mammals. The identified repeat regions are presented in Table 7 and figure 15, and are masked for all subsequent sequence alignments and analysis.

Several repeat elements were identified in the 1042 bp 5' flanking region of b*BTN*. Four of these repeat elements, ART2A, BCS, BOVTA and BTALUL1 overlap and together extend for 207 bp, while the fifth repeat element, MIR, spans 114 bp and is situated closer to the TSS at -237 bp (Table 7, Figure 15).

A similar search to identify repeat elements in the available 704 bp sequence from the 5'-flanking region of g*BTN* revealed a BCS/BTSC element covering 278bp. An additional 92 bp region contiguous to the 3' end of BCS/BTSC was found to be highly homologous to a repeat region identified in the cow sequence and was thus also excluded from subsequent phylogenetic analysis (Table 7, Figure 15).

Table 7: Repeat elements in the b*BTN*, g*BTN*, h*Btn*1a1 and m*BTN*1A1 5' flanking regions

Species	Type of repeat element	length (bp)	Remark
Goat	BCS	278	
	Unnamed	92	Homologous to repeat in cow
Cow	ART2A/BTSC/BTALUL1	207	Overlaps with a BCS element
	MIR	114	
Human	Monomer $(T)_n$	10	
	Monomer $(aaat)_n$	20	
	Alu	292	
Mouse		98	

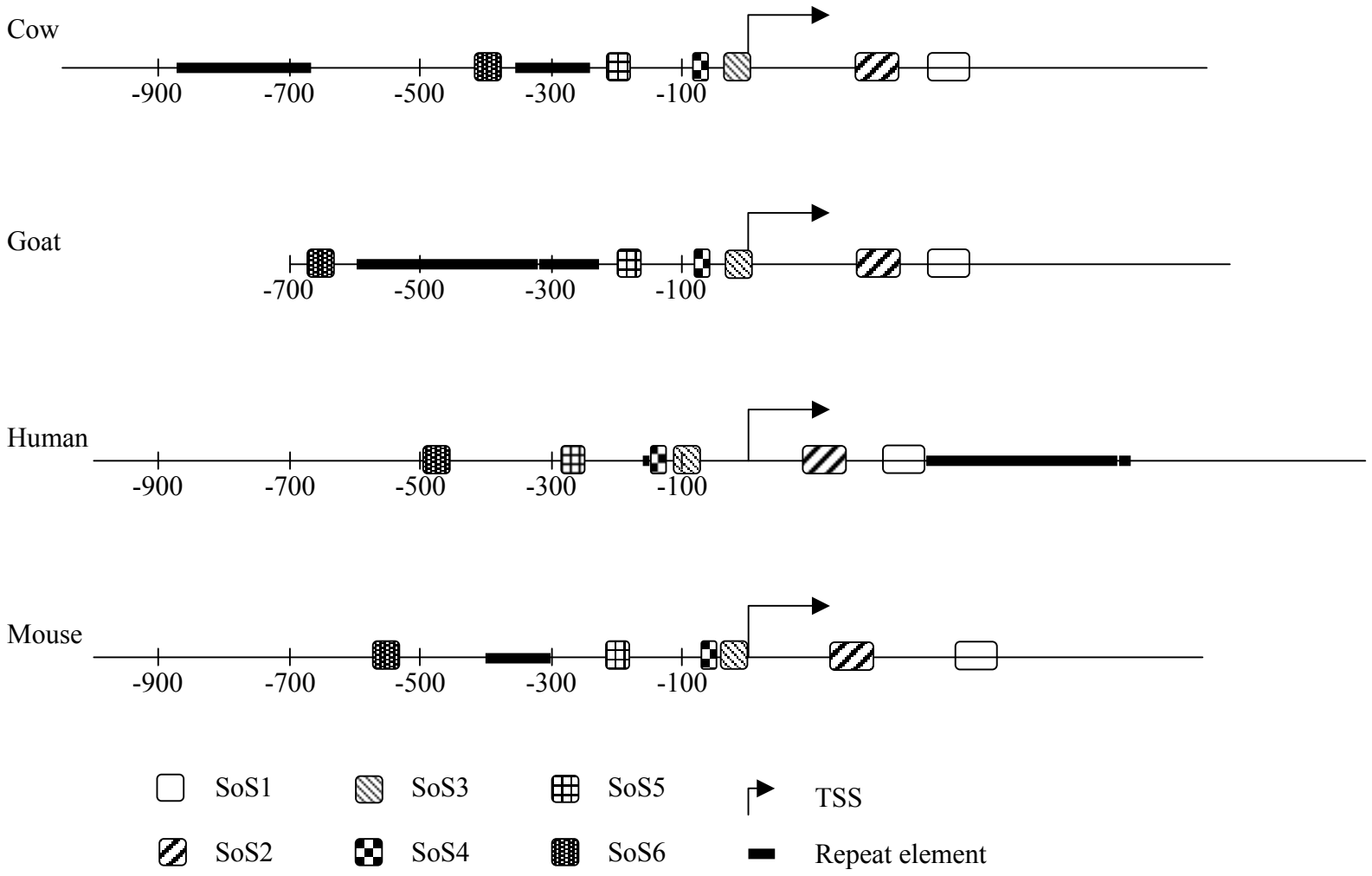


Figure 15: Location of SoSs and repeat elements in four *BTN* genes. The transcription start site and direction (arrow) was used as a reference point in this presentation. Boxes represent segments of similarity (SoS) and thick lines represent repeat elements. 3' end

One major and two minor repeat elements were identified in the most proximal 1kb of 5' flanking region in h*BTN*1A1 An *Alu* repeat covering 292bp and a contiguous $(aaat)_5$ tetramer were identified. Further, a simple monomer T repeat was located at -153 to -144. A relatively short 98bp repeat element was identified in m*Btn*1a1 (Table 7, Figure 15).

Pairwise and multiple alignments

The 5' prime flanking regions plus 5'UTR up to the CDS (see Table 7, Figure 15) sequences (minus their respective repeat elements indicated above) of the BTN gene of the four species were analyzed by pairwise sequence alignments. The overall similarity between cow and goat sequences is highest at 65%, with the mouse sequence being the most divergent from all the others (Table 8).

A multiple alignment of the region in the immediate vicinity of the TSS of *BTN* in the four species and subsequent search for widely described elements found in other genes, namely: Inr, TATA-box, DPE, TFIIB recognition element (BRE) and CpG islands yields none of the conserved core promoter elements in b*BTN*, g*BTN*, and m*Btn*1a1. However, a DPE is present in h*BTN*1A1 correct orientation and location relative to the TSS (Figure 16). No other core promoter elements were identified in h*BTN*1A1. Therefore, the promoter of *BTN* appears to be atypical of other commonly found genes and may possess as yet unidentified core promoter elements.

In a multiple alignment of all the sequences, the most similar region is located in the first introns and spans 22bp of which 18bp are identical and an additional 3 bases are conserved. Other homologous segments were also identified and given the designation segments-of-similarity (SoS) (Figure 15).

Table 8: Pairwise similarity index of orthologous *BTN* genes

	g*BTN*1 (1437bp)	h*BTN*1A1 (1944bp)	m*Btn*1a1 (1695bp)
b*BTN* (1750bp)	1.00[1] 65%[2]	0.259† 39%	0.040 19%
g*BTN* (1437bp)		0.205 36%	0.036 19%
h*BTN*1A1 (1944bp)			0.074 26%

[1] The similarity value of 1.00 marks the two most similar sequences, and does not necessarily denote perfect identity.
[2] Proportion of identical nucleic acid residues as a percent of the shorter sequence.
† The similarity coefficient is calculated relative to the maximum similarity.

A)

```
        1            15 16           30 31           45 46           60 61           75 76           90
1 bBTN\ GAGGAGGGAGTGTGT TGGGGTGGAAGGGTG TGGGGAGGCAGACTT TCCTGAGAGTACTTC CCCCTT---CCTTCT C-----CTTTAACTT
2 gBTN\ GAGGAGGGAGTGTGT TGGAGCGGAAGGGTG GGAGGAGGCAGGCTT TCCTGAGAGTACTTC CCCCTT---CCTTCT C-----CTTTAAGTT
3 hBtn\ --------------T TGCCCCGG--GGGCC ACAGCCAGCTTTCTC ACTTGGTAGCAGTGG CCTCTTGTGCCTTTT T-----CTCCAAGAT
4 mBTN\ ------------AAA TACACTTG---GTCA CCTGTGGGCAGGCTT CTCTAACAGCACA-C AGCCTTCTTCCTTCT GAAGAGCTCTCTCTT
 Consensus              TR   Y G   G Y     G  RGC   CTY  YYTRR AGYA       YCTT   CCTTYT       CTYY    T

           91          105 106         120 121
1 bBTN\TSS TTGCCAATGGGGCCA CAA------------ -----    100
2 gBTN\TSS TTGCCAATGGGGCCA CAA------------ -----    100
3 hBtn\TSS CACCCA----GGTCA GTATGTGTGGTTACT CTCAG    100
4 mBTN\TSS TGGCCCC-GGGGTGA CAAGCAGCCCTT--- -----    100
 Consensus Y  CC     GGY A   A
```

B)

```
         -60                                                          -1
1 bBTN\    GAGGAGGGAGTGTGTTGGGGTGGAAGGGTGTGGGGAGGCAGACTTTCCTGAGAGTACTTC
2 gBTN\    GAGGAGGGAGTGTGTTGGAGCGGAAGGGTGGGAGGAGGCAGGCTTTCCTGAGAGTACTTC
3 hBtn\    TTGCCCCGGGGGCCACAGCCAGCTTTCTCACTTGGTAGCAGTGGCCTCTTGTGCCTTTTT
4 mBTN\   AAATACACTTGGTCACCTGTGGGCAGGCTTCTCTAACAGCACACAGCCTTCTT-CCTTCTG
Consensus           G  Y  Y G   G      Y    RR RGCAG    YYYT     Y YYT
```

Figure 16: A) Multiple sequence alignment (ClustalW) of -60 to +40 region of bovine, caprine, human and murine *BTN* genes. TSSs are indicated by ***bold italics***, CT-motif in cow and goat *BTN* are boxed, and the DPE motif in h*BTN*1A1is underlined. Consensus of identical and conserved sequences are indicated in **bold** below the four sequences. B) Manual positional alignment of 60bp of the immediate 5' flanking region of *BTN* in the four species. Identical and conserved sequences are indicated.

Phylogenetic footprinting

Three of the SoSs are upstream of the TSS, one in the perimeter of the TSS and two downstream of the TSS. SoS3 is 41bp long and encompasses the TSS of cow, goat and mouse, but not of human *BTN*. SoS3 exhibits 56% identity across all four species with an additional 27% conserved bases.

A phylogenetic scan of the entire sequences of *BTN* from all four species identified 664 putative TF binding sites. Comparison of the putative TF binding sites in the six SoSs yielded several interesting features.

The region immediately around the TSS shows conservation across all species, except in mouse, some elements vicinal to the TSS are more preponderant than expected and appear in the same order of sequence in all species. ELK1 and NRF2, both members of the ETSF family of TFs, make-up the 5' boundary and includes GKLF1, BARBIE, and CP2. In cow and goat, these sets of TF footprints extend from about -10 to +39 and -11 to +39, while in human it covers the region -93 to -44. The mouse sequence shows only limited and spurious similarity sharing only the MZF1 site with b*BTN* and the E2F site with h*BTN*1A1 (Table 9).

SoS1 and SoS2 are of comparable size and reside in the first introns, with SoS2 being the one that is closer to the TSS. SoS2 proffered little similarity of putative TF binding sites across species. In all species, except humans, SoS2 possesses an EVI1 site, which in the case of cow and goat, overlaps with a GKLF1 site. On the other hand, SoS1 has a core similarity in its distribution of putative TF sites. The partially overlapping elements Hox1-3 and GATA (2 or 3) are present in all four species. In the cow and goat,

Table 9: Summary of putative TF binding sites in the six the SoS regions identified by multiple alignment of b*BTN*, g*BTN*, h*BTN1A1* and m*Btn1A1*.

Species	SoS1	SoS2	SoS3	SoS4	SoS5	SoS6
Cow	LMO2COM (+287) PLZF (+303) HOX1-3 (+313) GATA2 (+316) OCT1 (+316) CART1 (+317) CEBPB (+322) TALIBETAITF (+333) NEUROD1 (+333) TGIF (+337) CHR (+352)	HNF4 (+205) E2F (+207) GKLF (+227) EVI1 (+231)	MZF1 (-28) ELK1 (-2) GKLF (+7) BARBIE (+14) CP2 (+39)	STAT1 (-83) XFD (-74) CLTR_CAAT (-54)	ATATA (-202) MEIS_HOXA9 (-201) GKLF (-191) IRF3 (-189) E2F (-186) E2F (-184)	ATATA (-400) MEIS_HOXA9 (-395) FREAC7 (-394) SATB1 (-393) TATA (-393) BRN3 (-392) SATB (-390) MYT(-376) MZF (-370)
Goat	STAT3 (+280) ETS1 (+295) PLZF (+305) HOX1-3 (+315) GATA3 (+318) TAL1BETAITF2 (+335) NEUROD1 (+335) TGIF (+339) CHR (+354)	MOK2 (+201) GKLF (+229) EVI1 (+233)	ELK1 (-2) GKLF (+7) BARBIE (+14) CP2 (+39)	XFD (-74) CLTR_CAAT (-54) BTEB3 (-52)	HNF3B (-206) AMEF (-202) CRX (-200) GKLF (-190) IRF3 (-188) NMP4 (-187) E2F (-183) VDR_RXR (-179)	ATATA (-666) MEIS_HOXA9 (-661) FREAC7 (-660) SATB1 (-659) TATA (-659) BRN3 (-658) SATB (-656) HNF1 (-651) MZF (-636)
Human	TATA (+224) GATA1 (+235) GKLF (+236) CART1 (+239) HOX1-3 (+241) GATA2 (+244) OCT1 (+244) CART1 (+245) CEBPB (+250) COMP1 (+250) MTBF (+262)	MYOGNF (+107) ZID (+166) EVI1 (+177)	NEUROD1 (-120) NRF2 (-85) BARBIE (-83) E2F (-78) CP2 (-44)	OCT1P (-132) AHR (-124)	RFX1 (-281) GKLF (-271) EVI1 (-269) NMP4 (-265) EVI1 (-261) IRF2 (-260) EVI1 (-256)	IRF1 (-499) CDPCR3 (-479) SATB1 (-478) HNF (-473) TST1 (-468) SOX5 (-463)
Mouse	HIF1 (+331) GATA (+341) CART1 (+345) HOX1-3 (+347) GATA2 (+350) FREAC2 (+360) LEF1 (+372) RBPJK (+379)	EVI1 (+134) PAX5 (+168) MZF1 (+182)	MYOD (-45) ARE (-16)	MZF1 (-73) SP1 (-69)	TTF1 (-212) MYT1 (-207) SRF (-180)	TAL1ALPHAE47 (-557) GATA3 (-556) VBP (-556) PDX1 (-553) TCF11MAFG (-546)

these core elements are supplemented by PLZF, NEUROD1, TGIF and CHR. The human and cow sequence also have the same ordered series of putative TF binding sites for HOX1-3, GATA2, OCT1, CART1 and CEBP that fully or partially overlap. The mouse SoS1 however is the most dissimilar from the rest, but does possess a CART1 element binding site (Table 9).

SoS4 is the shortest region of similarity and the only similar binding sites are XFD and CLTR_CAAT which are present in cow and goat. SoS5 also shows little congruency in terms of putative TF binding sites. GKLF binding sites are present in all but mouse, and there are partially overlapping sites for IRF3 and E2F in cow and goat *BTN*.

SoS6 is the most similar in terms of putative binding sites in cow and goat *BTN*, with one of the total of nine binding sites alternately coding for either MYT1 or HNF1 in cow and goat respectively. The only elements in human *BTN* that are also present in cow and/or goat *BTN* are SATB1 and HNF1. As in the other regions, mouse *BTN* is most dissimilar from the other three species.

A wider scan of putative transcription sites that are present in all four species, regardless of whether they localize to the SoS, was also conducted. No patterns of collocation of the putative TF (or family of TF) binding sites were discernable. Furthermore, submission of the entire multipely aligned sequences to the Genomatix © website did not match with any previously described transcription modules.

DISCUSSION

Background

Coeval with the formation of the Holstein-Friesian Association of America in the 1880s, artificial selection in the high yielding Holstein-Friesian dairy breed has been particularly rigorous. With improvements in selection and breeding technologies, the selection pressure has been steadily growing.

Of the tools at the disposal of modern animal breeders, marker assisted selection conceivably affords the most direct control by which to hasten the rate of genetic gain of desirable traits in farm animals. The initial step towards this aim is to identify (or locate) genetic markers that will serve as flag posts along the length of the chromosome of interest. The first of such maps for the cow were published by Bishop *et al*. (1994) and Barendse *et al*. (1994) encompassing markers on all autosomal chromosomes as well as the X chromosome. Currently, an international collaboration to identify and map genetic markers in dairy animals hosts three websites listing more than 2,000 markers (http://www.cgd.csiro.au/cgd.html; http://locus.jouy.inra.fr/; http://sol.marc.usda.gov). Although the potential benefit from these markers is undeniable, without a higher density map, or markers specifically shown to tightly co-segregate with a gene of interest, they will not be of immediate practical use to breeders.

b*BTN* allelic distribution

In lieu of indiscriminate markers reposed in the database, we employed a PCR-RFLP approach to identify intragenic markers within the b*BTN* gene. Five PCR-RFLPs were detected; one in exon 8 and the rest were intronic (Table 2). Sons of four

informative grandsire families were genotyped for these markers and the resulting data used to construct an allelic frequency distribution (Table 5).

Based on the characterization of the marker sites in amplicons 1 and 2, b*BTN* appears to be bi-allelic with one variant (here dubbed allele A) making-up nearly 2/3 of the allelic pool in the half-sib families tested. In a similar study in Korean Holstein-Friesians, Lee *et al*. (2002) show that b*BTN* is bi-allelic with the allele distribution exhibiting the same skewed distribution. The largely skewed distribution favoring allele A is testimony to non-random selection pressures that forced b*BTN* out of conformity to the theoretical Hardy-Weinberg equilibrium (Hardy, 1908; Weinberg, 1909). It is to be noted however, there are several possible haplotypes as supported by the polymorphism detected in amplicon 1. Otherwise, this 5' region of the gene that comprises the two exons that code for the Ig-like folds, may exhibit a higher level of variability as evidenced by restriction patterns that were unexplainable from the available sequence (Figure 6).

As can be witnessed from the availability of data on hundreds of paternal half-sib daughters in the dairy bull DNA repository (DBDR) database, human interference has greatly impacted the level of genetic composition of genes relevant to many production and type traits. The interposition of human selection pressure on the natural progression of genetic drift and diversity has likely contributed to the skewing of the frequency of b*BTN* alleles, particularly those related to milk production and health. Extrapolating from this premise, it is also likely that genes that are in linkage disequilibrium to these highly selected loci were impacted by the selection pressure as well.

Allelic substitution effect of b*BTN*

Bovine butyrophilin is located on chromosome 23 in close proximity to the BoLA (bovine leukocyte antigen … cow MHC), PRL, and myelin oligodendrocyte glycoprotein genes (Ashwell *et al*., 1996a, Brunner *et al.*, 1996; McShane *et al*., 2001). The syntenic region in humans, p-arm of chromosome 6, codes for all these genes as well. However, the mouse genome has undergone rearrangements. Thus m*Btn*1a1 and PRL are found on chromosome 13, while the mouse H2 (MHC) genes are found on chromosome 17.

The first finding that linked the BoLA region to a single milk production trait, percent fat, was reported by Hines *et al*. (1986). However, even though predicted transmitting ability was detected, no differences in phenotype were observed. Despite the scanty density of mapped markers, it was the first study that conclusively established a discernible genetic influence of the BoLA region on a milk production trait.

To date, the only two genes located in the BoLA region that are candidates to directly influence milk production traits are b*BTN* (Ashwell, *et al*., 1996b) and PRL (Hallerman *et al*., 1988). The role of PRL in the physiological phenomenon of lactation is well known. Recent works by Horseman *et al*. (1997) and Ormandy *et al*. (1997) in which one or both alleles of PRL were ablated conclusively establishes this fact. On the other hand, the role of BTN was a matter of speculation until the recent generation of an m*Btn*1a1 knockout mouse that revealed its importance for milk-fat secretion (Ogg and Mather, unpublished).

The findings by Hines *et al*. (1986) could thus be due to the direct effect of b*BTN* and/or PRL, or an indirect effect of the BoLA or other vicinal genes. Prolactin, although predominantly expressed in the pituitary, is also found in other cell types and plays roles

unrelated to mammary gland functions and lactation. BTN, on the other hand, is a mammary-specific gene whose function in lactation was not past well-founded conjectures (for review see Mather, 2000; Mather and Keenan, 1998).

In order to elucidate if b*BTN* is a factor that influences milk production/health traits, the PCR-RFLP markers identified in b*BTN* were used in a QTL analysis. In a GDD model, a QTL analysis was conducted testing the hypothesis that substitution of one b*BTN* allele by another (as measured by its surrogate PCR-RFLP marker) will result in a significant difference in several economically important dairy traits, namely: milk yield, protein, fat, percent protein, percent fat, SCS and herd life.

The findings from this analysis reveal that in the four grandsire families in the study, none, but one of the markers tested showed a putative QTL effect. One marker in a single grandsire family showed a significant putative QTL effect on percent protein (Figure 8). However, it is difficult to draw any broader inferential conclusions from this sole statistically significant finding. None of the other grandsire families show significance for this particular trait among all the markers tested. In fact, there is not even a statistically non-significant but trend-setting pattern. Thus the significant effect observed could be the result of the effect of grandsire and not the marker site.

A similar QTL study, but using a genome wide scan of US Holstein DBDR families using 70 markers (Ashwell and Tassell, 1999) showed that chromosome 23 has a significant QTL effect on productive life and udder type trait. Another genome wide scan of North American Holstein-Friesian resource families with 174 markers (4 markers for chromosome 23) yielded no significant effect on the tested milk and health traits, a marker on BTA23 came closest to the threshold for SCS. Given that the BoLA resides

on this chromosome the results obtained are further evidence of the unreliability of the existing QTL detection methods. Moreover, Ashwell and Tassell (1999) did not report similar findings in the DBDR family; but did report significant QTL effect on SCS in an earlier publication (Ashwell, *et al*., 1996b).

In contrast to genome-wide scans, the chances of ascribing a putative QTL to a specific gene, b*BTN*, were greatly improved in this study by using intragenic markers. Our findings confirm the inconsistency that has plagued the use of QTL analysis methodologies to unequivocally establish a genetic link between a measurable production/health trait and a given genetic marker (or gene).

While the statistical outputs do not support an association between b*BTN* and the tested quantitative traits, this does not rule out the fact that b*BTN* could very well play a role in lactation. In fact, the work done by Ogg and Mather (unpublished) in a mouse model system unambiguously demonstrates that the expression of *BTN* is essential for the regulated secretion of milk-fat.

The findings from the QTL study and the experimentally validated role of BTN in milk-fat secretion appear discordant. However, there are three possible mutually non-exclusive explanations: 1) *BTN* is one of several possible quantitative genetic factors involved in milk-fat secretion, 2) the genetic differences between the two alleles are in regions of the gene that do not markedly alter its functional attributes, and 3) generations of selection pressure have limited the genetic variability, thus the scarcity of the BB genotype has increased the threshold required for the detection of statistically significant differences.

Complex traits, such as milk production and health, present a challenge in terms

of dissecting each genetic component that exerts influence on a phenotype. In addition to bBTN, other factors such as XO and ADRP conceivably contribute to the milk-fat secretion process (Banghart *et al*., 1998; Heid *et al*., 1998; Vorbach *et al*., 2002). bBTN maybe one-of-several equally important components in this process. Therefore, differences between the two alleles of b*BTN* might not be extricable from the effects of other players. Additionally, the heritability of these genetic factors is not large, so that the impact that environmental factors play in the expression of economically important phenotypes cannot be discounted.

The only exonic PCR-RFLP detected in b*BTN* resides in the last and largest exon coding for the cytoplasmic domain (Table 2). A substitution of nucleotide A to G translates to a conservative basic amino acid substitution of lysine to arginine. The remaining four PCR-RFLPs all reside in non-coding regions of the gene. Therefore, the two allelic forms of b*BTN* might not significantly differ in the discharge of their functions, thereby reducing the likelihood of detecting statistically discernable differences. Also, the power to detect a QTL effect by the PCR-RFLP markers were limited by the fact that the BB genotype was only represented in 9% of the tested half-sib sons, thereby making a sensitive comparison amongst the quantitative traits at best recalcitrant to current analysis methods. The rarity of genotype BB is a direct result of the skewing of the allelic frequency of b*BTN*. Artificial selection pressures have, by design or inadvertently, favored the accumulation of allele A. By design, because the selection for dairy production/health traits might have been more desirably influenced by allele A than allele B. Otherwise, b*BTN* might be closely linked to other sought-after genetic factor(s) that preferentially segregate with allele A, one credible candidate being

PRL.

The near genetic uniformity of a subpopulation of dairy breeds presents an obstacle in employing a QTL approach for detecting loci (gene) consequential to economically important traits. The only viable means by which to get around this limitation is to cross different breeds with maximal divergence in the phenotypic attribute of interest. This approach, although a theoretically viable option, cannot be practically applied to existing commercial populations without significant costs.

Expression of m*Btn*1a1 in HC11 cells

Several modes of regulation are ascribed to the control of 'milk genes' both *in vivo* and *in vitro*. Transcription is one level of control by which expression of most 'milk genes' is made sentient to the physiological status of the mammary gland. In addition to the contribution of the extracellular matrix (for example see Streuli *et al*, 1991), several hormones are known to play a significant role (for example see Stocklin *et al*, 1996). In the HC11 mouse mammary cell line system, a trio of lactogenic hormones, *viz* insulin, hydrocortisone, and PRL have been identified as major *trans*-acting factors required for the transcriptional induction of the major 'milk genes'.

The molecular processes by which PRL exerts its effect in mammary cells has been extensively worked out since its role in mammary gland development was initially recognized (Nandi, 1958). Schmitt-Ney *et al*. (1991) and Watson *et al*. (1991) were the first to independently identify STAT5 (signal transducer and activation of transcription) as the downstream effector of PRL signaling. Prolactin acts via the cell surface receptor PRL-R, which leads to the signaling cascade known as the JAK2/STAT pathway (Rui *et*

al., 1994: Pezet *et al*., 1997). The end point of this pathway is the binding of active STAT5 dimer to the γ-interferon activation sequence (GAS) element on target genes, such as the 'milk genes', which induces transcription (Gouilleux *et al*., 1994; Wakao *et al*. 1994).

The wide-ranging role of insulin in mammary gland physiology has been a field of interest since Nandi (1958) showed its importance for structural integrity in a mammary explant system. The effect of insulin depends on the physiological status of the mammary gland. During the developmental phase, insulin is primarily involved in stimulating DNA polymerase activity thus contributing to the progression of the cell cycle (Lockwood *et al*., 1967). With the onset of lactation, the functions of insulin shifts to maintaining mammary tissue (Kumaresan and Turner, 1965; Merlo, *et al*., 1996) and transcription of mammary specific genes such as α-lactalbumin (Prosser *et al*., 1987) and β-casein (Nicholas, *et al*., 1983). However, a direct requirement of insulin for lactation *in vivo* has not been definitively established, as it has been shown to be nonessential in lactogenesis (Kyriakou and Kuhn, 1973) and milk yield (Hove, 1978). The I required by mammary cells *in vitro* cultures likely functions as substitute for IGF-1 (reviewed by Neville et al., 2002).

Although hydrocortisone (H) exerts its influence via the glucocorticoid receptor (GR), its precise mode of action independent of other factors has yet to be fully elucidated (Mills and Topper, 1970). Studies done on the endogenous expression of 'milk genes' in rabbit mammary explants and HC11 cells, show that H (in the presence of I) appears to have a more direct and rapid effect on the expression of WAP, and little influence on the expression of β-casein (Doppler *et al*., 1991; Puissant and Houdebine,

1991). On the contrary, PRL by itself does not induce expression of WAP, but in conjunction with H induces triple the level of expression attained with either H or I alone (Hobbs *et al*. 1982; Doppler *et al*., 1991). In HC11 cells, it was found that maximal expression of β-casein expression was attained in the presence of both HC and PRL (Doppler, *et al*., 1989). These and other data (see also Stoecklin *et al*., 1996; Cella *et al*., 1998) suggest that the terminal effectors in the PRL and HC signaling pathways converge providing a combined regulatory mechanism.

In addition to the documented linear effects of each lactogenic hormone, significant progress has also been made in defining the various interactions among these hormones and their resultant effects. As briefly mentioned earlier, the interaction between the downstream effects of H and PRL affects the quantity of 'milk gene' expression. Immunoprecipitation studies show that GR and Stat5 are physically associated throughout the various stages of mammary gland development (Doppler *et al*., 1989; Cella *et al*., 1998; Stoecklin *et al*., 1996). Moreover, their interaction results in a functionally active transcription regulator that could either repress (Stoecklin *et al*., 1996) or enhance expression of target genes (Cella *et al*., 1998).

Several researchers have established that *BTN*, *in vivo*, is restricted to the mammary gland, and more specifically to the phases surrounding lactation (Ogg *et al*, 1996; Banghart *et al*, 1998). Aside from the qualitative examination of the temporal and tissue specificity of *BTN* expression, very little is known about the *cis*- and *trans*-acting factors involved in either the induction or the cessation of transcription. Notwithstanding the lack of a more incisive examination, at least *in vivo*, the expression profile of *BTN* supposes a well regulated control mechanism.

The only preliminary work done to characterize *trans*-acting factors on the expression profile of *BTN* in an established cell line was done by Aoki *et al*. (1997) in HC11 cells. In the study, no measurable enhancement of m*Btn*1a1 expression was attained by treating the cells with the accepted triad of lactogenic hormones, IHP. Further, the study was limited to the expression of endogenous *BTN* and therefore did not provide an insight into the identity of potential *cis*-acting elements, nor provide conclusive evidence as to the individual effects of the lactogenic hormones tested.

The second phase of the present study attempted to explicate the *trans*-acting factors influencing expression of endogenous m*Btn*1a1 in HC11 cells. HC11 cells were incubated in the presence or absence of I, H and PRL, and in the presence or absence of matrigel. By northern blot, m*Btn*1a1 message was only detected in the presence of all three lactogenic hormones, with or without matrigel (Figure 10). This result was also corroborated using mRNA samples isolated from HC11 cells treated with IHP on plastic (Figure 10).

Therefore, the expression of m*Btn*1a1 in HC11 cells is responsive to treatment with IHP, and is independent of a supplementary substratum. This result is in agreement with the widely held view that the expression of 'milk genes', in this case m*Btn*1a1, in HC11 cells requires the presence of IHP (Ashktorab *et al*., 1999). Moreover, when HC11 cells are cultured in an induction medium, the presence or absence of matrigel does not measurably alter the transcription of m*Btn*1a1. This is could either be because HC11 cells produce and deposit their own substratum (Chammas *et al*., 1994), or because the expression of m*Btn*1a1 in HC11 cells is not dictated by the presence or absence of a substratum.

TAZ: a subclone of HC11 cells

Although HC11 cells are the most widely used mammary cell lines to investigate hormone responsiveness and promoter assays, they are notoriously non-homogenous at times reverting to basal, as opposed to luminal, cells (Deugnier, *et al*., 1999). The original HC11 cell clone lost its m*Btn*1a1 expression in response to lactogenic hormone, IHP, which might be due to a growing heterogeneity in the cell population. To circumvent inconsistencies that might arise from this non-homogeneity, and in an attempt to obtain a more uniform subclone with an increased level of m*Btn*1a1 expression, single colonies of HC11 were propagated and screened. In the presence of IHP, subclone 9, hereto named TAZ (for tatsaechlicher ausdrückend zellen) was found to express an increased level of m*Btn*1a1 message as compared to contemporaneously isolated HC11 subclone (Figure 13). Interestingly, the tight hormonal control displayed by the original HC11 cells is diminished in TAZ cells (compare Figure 10 and Figure 13) where m*Btn*1a1 expression is detectable even in cells treated only with I. Aside from the novel m*Btn*1a1 expression profiled in TAZ cells, light microscopy revealed a more uniform morphology of the TAZ subclones as compared to the parental HC11 cells (Figure 12). TAZ subclones therefore represent a manifestly distinct form of a mouse mammary cell line in which expression of m*Btn*1a1 is not dependent on the presence of all three lactogenic hormones.

Phylogenetic footprinting

The transcription of BTN during and around lactation points to potential *cis*-acting elements that exert control over the regulated expression. A computational

approach was used to examine the putative promoter region of BTN in four species in which sequence data was available.

A region of the 5' flanking sequence of orthologous cow, goat, human and mouse *BTN* genes, starting from about -1kb and extending to the start of the CDS, minus their respective repeat elements (Tables 7, Figure 15), were analyzed for core promoter elements, previously described 'milk gene' elements, and shared transcription modules amongst the four species.

Prior to alignment and analysis of the orthologous sequences, the TSSs were delineated by either referring to experimentally determined annotations in the GenBank DNA sequence repository, or by exon mapping. This criterion was used to assess the TSS for cow, human and mouse sequences. The g*BTN* sequence archived in the GenBank data base (accession № AY036083) however was not experimentally established. Therefore, an arbitrary determination was made to map its TSS by exon mapping against b*BTN* (based on the high degree of homology) (see Tables 8). Using this method, an additional exon was revealed further 5' in g*BTN*. This is in agreement with the exon/intron organization and location of the start of the CDS observed in cow, human and mouse sequences.

The region of the orthologous genes extending from -60 to +40, with the TSS as a quasi-axisymmetrical center, was subjected to multiple alignment analysis (Figure 16). The alignment did not reveal a strikingly evident homology. Yet, an interesting feature in b*BTN* and g*BTN*, is a 21bp stretch of region straddling the TSS and exclusively comprising pyrimidines (Figure 16). Fifteen base pairs of this pyrimidine rich CT-motif is found immediately downstream of the TSS and shares a near perfect homology with

both the human and mouse sequences, the exception being a 3 bp insertion that interrupts the homology in both human and mouse sequences. In the mouse sequence, the additional inserted trimer is itself entirely made-up of pyrimidines.

The TSS region was then screened for core promoter elements widely described in both mammary and non-mammary genes as essential for the assembly of a preinitiation complex (PIC), *viz*. TATA-box, CAAT-box, pyrimidine-rich Inr element (for review see Roeder, 1996) and/or the recently identified DPE (Burke and Kadonaga, 1997; Burke and Kadonaga, 1996). None of the four genes possessed these elements, with the exception of h*BTN*1A1 which contained a DPE (Figure 16) at the prescribed position of +28 to +32. This is in stark contrast to other 'milk genes' such as cow κ-casein (Alexander *et al.*, 1988), cow lactoferrin (Teng, 2002), and mouse WAP (Campbell *et al.*, 1984) that possess canonical or noncanonical TATA- and CAAT-boxes.

A comparison of the pyrimidine-rich sequence straddling the TSS with the loose Inr element consensus sequence $YYA_{+1}N(T/A)YY$ (Roeder, 1996) reveals that the CT-motif of b*BTN* and g*BTN* diverge at two points: at position +1, both have C instead of the conserved A residue; and at position +3, both have a C instead of T/A. In addition to the ones described for the cow and goat sequence, h*BTN*1A1 has an additional divergence at position +5.

An inspection of other cow 'milk genes', namely all the caseins, α-lactalbumin, and β-lactoglobulin reveals that they do not posses such a pyrimidine rich region in the vicinity of their respective TSSs. Perhaps, the pyrimidine rich sequence is a novel element, or a variant Inr element that is required for basal expression of *BTN*.

Notwithstanding, the parallel functionality of the entire 21 bp CT-motif in all four

species is questionable since it is primarily located upstream of the TSS in the human and mouse genes, as contrasted to its straddling position in the cow and goat sequences. The differences in the TSS seen between the cow and goat on one hand and the mouse and human on the other, could plausibly be due to the trimer insertions of the latter. Further deletion/substitution studies could shed light on this issue.

In view of the absence of traditionally present core promoter elements, a manual positional-alignment was performed to determine if there are any conserved elements positioned equidistant from the TSS. A couple of short conserved stretches of sequences are observed (Figure 16). The first, GCAG/C, is at position -23 to -20 (cow, goat and human), and -22 to -19 (mouse). The second is a 4bp stretch of pyrimidines located at -15 to -12 (cow, goat and human), and -14 to -11 (mouse). The relevance of these tetramers is not clear, and no similar patterns have been previously described in other genes.

A phylogenetic footprinting was conducted on all four orthologous sequences of ~0.7kb (g*BTN*), ~1kb (b*BTN*, h*BTN*1A1 m*Btn*1a1) using GEMS Launcher© software (Genomatix Software GmbH) (Quandt *et al*., 1995) to detect and characterize putative TF binding sites. These sequences were devoid of repeat elements, and their 5' ends extended to their respective CDS start sites, as described above.

The GEMS Launcher identified a total of 664 putative TF binding sites, that is, on average, about one TF binding site per 10 bp. In order to systemize the search and maximize the chances of revealing regulatory networks of higher order transcription modules, six segments of the highest homology across all species were selected (Table 9, Figure 15) for comparison. These homologous locales were named segments of

similarity (SoS) and numbered 1 - 6.

The overall homology of the b*BTN* and g*BTN* is high (Table 8), and the level of homology within these islands of homology, *i.e*. SoSs is even higher. Therefore, there is an increased coincidence of the same putative TF binding sites occurring in the same SoS. As regards the putative TF sites in the human and mouse sequence, even though they share high sequence homology with their ruminant orthologues in the SoSs, the homology does not extend to coding for a large number of shared putative TF binding sites. A closer examination of each SoS is discussed.

SoS6 is the most 5' SoS relative to the TSS, representing a putative distal promoter module. The putative binding sites for SATB1 (Special AT-rich sequence-binding protein 1) (Dickinson *et al.*, 1997) is present in all but in mouse. SATB1 is a tissue-specific chromatin remodeling protein that effectively controls gene expression in thymocytes by affecting higher order architecture (Cai *et al.*., 2003). An HNF (hepatic nuclear factor 1) (Tronche and Yaniv, 1992) is also present in the goat and human sequences, but not in the cow and mouse sequence. Interestingly, in humans a member of the CLOX (cut-like homeodomian) family of transcription factors, CDPCR3 (Harada *et al.*., 1995), is present sharing the same footprint as SATB1. Liu *et al.* (1999) describe that the CLOX family member, CDP, and SATB1 interact *in vivo* via their DNA binding domains. This interaction was demonstrated to interfere with the DNA binding ability of both proteins. Moreover, overexpression of CDP was shown to override the transcriptionally repressive effect of SATB1 on MMTV promoter activity (Liu *et al.*., 1999). It is possible that both TFs work from the same overlapping putative binding site, where in the absence (or lower level) of CLOX, SATB1 negatively regulates *BTN*, which,

upon elevated expression of CLOX is depressed both by binding SATB1 directly as well as competitively binding to the shared *cis*-element. A computer based model construction of cow genes expressed during lactation found CLOX to be a common denominator in all the promoters examined (Malewski and Zwierzchowski, 2002).

The CLOX/SATB1 shared footprint is conspicuously absent throughout all the mouse SoSs examined -- although, this or a similar site maybe present at a more distal location. The absence of this putative *cis*-element could very well be one of the reasons why a transgene using m*Btn*1a1 5' flanking region did not drive the expression of a reporter gene (Mather and Ogg, unpublished work). It is also to be noted that SoS6 in the cow is found both in the b*BTN*1.7/CAT and b*BTN*1.2/CAT transfection vectors; therefore SATB1/CLOX can not by itself account for upregulating the core promoter. It is likely that another one or more modules located further upstream work cooperatively.

A member of the Hox family of TFs is also present in cow and goat SoS6. A triple ablation mouse model of Hoxa9 and two of its paralogs, Hoxb9 and Hoxd9, illustrates that this TF is required for mammary gland differentiation, post-pregnancy (Chen and Capecchi, 1999).

SoS5 lies along the notional proximal-distal promoter boundary. GKLF (Shields and Yang, 1998) putative binding sites are found in all but the mouse segment. GKFL is principally expressed in skin, oral and other epithelial cells where it is thought to be involved in terminal differentiation. (Garrett-Sinha *et al.*, 1996). The ablation mouse model of GKLF is lethal soon after birth and has been shown to be important for the barrier integrity of the epidermis (Segre *et al.*., 1999). The relevance of this TF towards expression of mammary genes has not been previously established, nonetheless, its

involvement in the terminal differentiation of other epithelial cells could be very well duplicated in the mammary gland. GKLF is demonstrated to be required for the expression of one of the components of laminin-5, the major extracellular matrix deposited by mammary epithelial cells (Miller *et al.*, 2001). From our results, expression of m*Btn*1a1 was neutral to the presence or absence of a matrigel substratum. Therefore, a direct or indirect effect of GKLF on the expression of *BTN* can not be inferred.

E2F (Kovesdi *et al.*., 1986) was traditionally perceived as a factor that, in partnership with pRb (retinoblastoma), is involved in controlling cell cycle progression. More recently the versatility of this group of TFs is becoming evident. A well-designed work by Müller *et al.* (2001), using human bone tissue derived U2OS cells, discovered that E2F may induce the expression of more than 1200 genes. Given the wide-ranging action of E2F, and its role in several distinct cellular processes, if it does contribute to the expression of *BTN*, this effect is unlikely to be tissue/lactation specific.

The human SoS5 has three putative binding EVI1 sites. No link to lactation and to the mammary gland has so far been established for the other putative sites occurring in SoS5 of more than one species, namely, IRF3 and NMP4.

The SoS4 is the region of high homology that most typifies the proximal promoter region in terms of proximity to the TSS. The cow and goat segments possess putative binding sites for BTEB3 (Martin *et al*., 2000) and STAT1; while the mouse sequence possesses a site for Sp1. BTEB3 and SP1 are members of the large SP1-like family of TFs. The Krueppel-like factors (KLF), of which GKLF is one, is a sub-group within the extended Sp1-like family. The zinc-finger motifs are the standard feature of all Sp1 family members, but their other domains are variable, and display contrariety in their

regulatory properties (for example, see Kwon *et al.*, 1999). One means by which this group of TFs display regulatory specificity is by the pattern of tissue distribution, with, for example, Sp1 present in most tissues (Hagen *et al.*., 1994), while Sp4 is restricted to nervous tissue (Hagen *et al.*, 1992). Another emerging means of regulation is by competition for the same binding sites, but by exerting opposing effects as was demonstrated by the inhibition of Sp1 activation by GKLF (Zhang *et al.*, 1998).

TATA-less promoters have been shown to require Sp1-binding sites (Dynan and Tjian, 1983), and Sp1 activation is enhanced in the context of TATA-less promoters (Colgan and Manley, 1995). Furthermore, Sp1 binding is implicated in establishing one or more TSSs (Kollmar *et al.*. 1994; Lu *et al.*., 1994; Boam *et al.*., 1995). The exact location does not seem to be strictly defined, but spans within the -50 to -150 range.

The Sp1 (in mouse) and BTEB3 (in cow and goat), anchored at -69, -54 and -52 respectively could serve as the elements that define the TSS and probably interact with a masked Inr sequence to establish a PIC. It is also of significance that there is a positional disparity in the actual start of transcription downstream of the Sp1/BTEB3 elements. Boam *et al.* (1995) demonstrated that mutating the native Inr sequence results in Sp1 directing transcription from several other sites. Comparison of the TSS locale (Figure 16) of all four species depicts the mouse TSS to be 15bp downstream of the comparatively homologous section of the cow and goat TSS. The cow/goat + 1CCC positionally correspond to the mouse -16 AGC, with the additional trimer CTT at -10 creating a significant divergence from its coordinate ruminant sequence. The Inr sequence disparity seen in m*Btn*1a1 conceivably contributes to its promiscuous transcriptional start sites as observed by Mather and Ogg (unpublished).

The human SoS4 region does not possess any of the SP1/BTEB/KLF family putative binding sites. Since SoS4 in humans lies over 100bp upstream of the TSS, an extended search over the -50 to -150 region relative to the TSS was conducted, but did not yield any putative binding sites to the Sp1/BTEB/KLF binding sites. However, Oct1P and Ahr putative binding sites were detected. Oct1P (Verrijzer *et al*., 1992) is a member of the larger POU domain TFs that share a similar DNA binding domain. Oct-1 is ubiquitously expressed and has not to date been implicated in regulating mammary specific genes. AhR (aryl hydrocarbon/dioxin receptor) heterodimerizes to the AhR-nuclear translocator (Arnt) and is responsible for responses to seemingly dissimilar synthetic and naturally occurring xenobiotic insults (Gu *et al.*, 2000). Thus far, no high affinity endogenous ligand has been identified for Ahr (for review see Denison and Nagy, 2003). Despite a close association observed between AhR-Arnt and the estrogen receptor in mammary cells (Safe *et al.*, 2000), Le Provost *et al.* (2002) have conclusively determined that AhR-Arnt is not essential. So the relevance of the presence of Oct1P and AhR putative binding sites is not evident.

h*BTN*1A1 lacks the Sp1/BTEB present in the cow, goat and mouse sequence to potentially serve as a yardstick for the TSS. However, the DPE located at about +30 could very well serve an important role. Burke and Kadonaga (1996, 1997) have shown that DPE acts in cooperation with an Inr to bind TFIID and direct PIC assembly in TATA-less promoters. However, even though h*BTN*1A1 possesses a DPE element located at the exact distance from the TSS (discussed earlier), it lacks a classical Inr. Therefore, it can be hypothesized that, in h*BTN*1A1, the DPE serves as the sole element that orients the assembly of the PolII transcription machinery, or else the Inr found in

h*BTN*1A1, and its homologues in cow, goat and mouse define a new class of Inr sequences.

SoS3, in terms of position relative to the TSS is located straddling the TSS in cow and goat, but resides in the proximal promoter in human and mouse *BTN*. Therefore interpretation of the significance of putative TF binding sites, will have to be taken in the context of their co-location (Table 9). A putative TF binding site common to all species in SoS3 is BARBIE (so named to reflect that it is a barbiturate responsive element), a site that is widely described as a response element associated with the cytochrome P450 family of genes. Moreover, AhR, has also been described in relation to its role in cytochrome P450 expression. No previous work has attempted to define the role of the BARBIE box on expression of 'milk genes'. The descriptions of E2F and GKLF have been provided above in relation to their occurrence in both cow and goat SoS5. It is to be noted, that the segments of similarity are only a reflection of the degree of local homology and do not show location relative to TSS. Thus SoS3 in h*BTN*1A1, unlike the cow and goat, spans from -120 to -44, and thus is still 5' of its TSS. In this context, the presence of E2F may have a similar and as yet undefined significance. A computational approach to define regulatory elements important for genes in the mammary gland has identified E2F as one of a number candidates that potentially play a dual role in regulation both the 'milk genes' as well as the 'involution genes' (Malewski and Zwierzchowski, 2002) .

CP2 putative TF binding site is present in all but the mouse SoS3. CP2 TF is expressed in many types of tissues (Jane *et al.*, 1995; Murata *et al*., 1998) but so far described as binding elements further 5' of the TSSs in α-globin (Kim *et al*., 1990) and

cytochrome P450 genes (Sueyoshi *et al*., 1995).

The singular distinction of SoS2 is that the EVI1 putative *cis*-element is present in the first intron in all the species examined. EVI1 (ectopic virus integration site 1) is a contributor to retrovirally induced myeloid leukemia (Mucenski *et al*., 1988) and a known transcription repressor (Bartholomew *et al.*, 1997) in coordination with co-repressors (Chakraborty *et al*.., 2001; Palmer *et al*., 2001). EVI1's significance towards mammary specific genes is not apparent from its tissue distribution and regulatory functions ascribed to it in relation to hematopoietic differentiation.

The SoS phylogenetic comparisons indicate that the *BTN* promoter does not conform with the promoters of other well characterized 'milk genes'. TF-binding sites that are well characterized in other 'milk genes' are not found in the SoSs. Moreover, at least in TAZ cells, mRNA synthesis appears to be unaffected by treatment with the lactogenic hormones, IHP. This of course, is in line with the absence of the classical response elements for PRL and HC. Therefore, the *BTN* promoter most likely presents a nouvelle network of transcription factor binding sites that are capable of directing mammary gland and lactation specificity. The SATB1-CLOX and Sp1/BTEB/KLF putative binding sites offer the best candidates to have biologically relevant effects on the expression of b*BTN*.

Further, the lack of a previously described Inr in the TATA-less promoters of all four species of butyrophilin points to separate a mechanism by which core promoter elements can direct the PIC to the proper TSS, or else these TSS possess a new as yet uncharacterized Inr element.

Promoter analysis of b*BTN*

To identify potential regulatory elements in b*BTN* a number of CAT reporter constructs that encoded either 5' only, or 5' and 3' flanking regions of b*BTN* (Table 1) were transfected into the original HC11 cell line and the subcloned TAZ cells. For maximal inductive response, the cells were all cultured in IHP-containing induction media and the CAT activity subsequently quantified as a measure of promoter activity.

A region of the promoter that spans -453 to -1042 (construct pb*BTN*1.7/CAT) was found to be sufficient to drive expression of the reporter gene significantly above the background (pBasic/CAT) control. This elevated expression was observed in both the original HC11 cells (Figure 10) and TAZ subclone (Figure 13) cells. The construct pb*BTN*1.2/CAT, encoding a shorter 5' flanking region, registered a slight and statistically insignificant CAT activity when compared to the basal control.

The addition of the 3' flanking region of b*BTN* does not alter the promoter activity of the 5' flanking region. This is illustrated when comparing pb*BTN*1.7/CAT vs. pb*BTN*1.7/CAT/+1.7 and pb*BTN*1.7/CAT/-1.7, and pb*BTN*1.2 vs. pb*BTN*1.2/CAT/+1.7, pb*BTN*1.2/CAT/-1.7 as assayed in the original HC11 cells. A similar result was obtained when comparing the pb*BTN*1.7/CAT vs. pb*BTN*1.7/CAT/+1.7 in the TAZ cells.

The data thus indicate that while the core and proximal promoter regions are sufficient for basal transcription, the distal promoter region is essential for an elevated expression of b*BTN*. Furthermore, the 3' end flanking region makes little or no contribution to controlling transcription.

It is therefore plausible that the -453 to -1042, region of b*BTN* representing the distal promoter sequence, possess one or a few TF *cis*-elements that exert a positive

effect on the core promoter, either as an activator or derepressor. These putative *cis*-acting elements are thus thought to provide a regulatory platform from whence *trans*-acting factors can convey mammary specificity and impart lactation dependency. In order to shed light on the putative transcription factor binding sites and their modular organization, a phylogenetic footprinting of the cow, goat, human and mouse gene sequences was carried out.

Future directions

This paper investigates two questions about the b*BTN*. First, we analyzed the allelic substitution effects of b*BTN* on quantitative production and health traits in dairy cattle. Secondly, we utilized two approaches to decipher the putative proximal promoter region most important for the expression of b*BTN*. On the one hand, we employed reporter constructs in a transient transfection assay to delimit a segment important for its expression. On a parallel approach, we applied a computational phylogenetic comparison to identify conserved regions across four species and the corresponding putative transcription factor binding sites.

Efforts to select for an economically more favorable allele of *BTN* in Holstein-Friesians are not supported by the findings in this study. The skewed allelic distribution points to some selective pressure that favored allele A. This skewed allelic distribution maybe accounted for by one or more QTLs that are closely linked to *BTN*. However, without direct evidence that relates the allelic difference to a phenotypic difference, there is no significant advantage that can be gained by using one of the alleles of *BTN* as a factor in a MAS program. A more conclusive result could feasibly be obtained in a

future study by using phenotypically divergent subpopulation of animals.

On the other hand, the use of the putative promoter of b*BTN*, specifically the region extending to -1kb, could likely serve to drive exogenous gene expression in the mammary gland system. Given that the b*BTN*1.7/CAT construct was able to drive expression of a reporter construct in a cross-species cell line (HC11 cells), *in vitro*, it may perform better in a ruminant *in vivo* system. However, the computational promoter analysis raises more questions than it answers. The Inr and putative promoter regions of *BTN* largely do not conform to known and conserved elements of 'milk' and other widely described genes. It is unlikely that the proximal putative region of lactation specific genes contain the much sought-after 'master switch' that activates all 'milk genes'. It still remains to be determined which narrowly delimited regions (elements) of b*BTN* are crucial for its expression. A more incisive study utilizing either point mutations or extensive deletion constructs could reveal important elements and lead to the description of the modules that regulate the expression of b*BTN*. Further, comparing results obtained here with that obtained from a transient transfection assay in non-mammary cells could help establish which regions confer mammary specificity.

APPENDIX

Multiple alignment of sequence of the 5' flanking region and 5' UTR of *BTN* paralogous genes from four species, namely, cow, goat, human and mouse.

```
bBTN     1  gaattcagca gaaagtgtat cagtaagtta acaaaaatgt ggacAAAGAC TTGTGTATAA ACTGCACTAT
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN     1  at-------- ---------- ---------- ---------- ----AAAGAC TTATATATAA AGTGTATTAT
mBtn     1  ccataccctc cccccg---- ---------- ---------- ---------- ---------A AGTGCAGTCT

bBTN    71  TTATAATAGG CAAAAAACCC AAGATGTGCC ATGATCTCAG GAATTAtagg aaccttcaat atgatataaT
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN    29  TTACAATAGG AAAAAAATCC CAAATGTCCC ACAATCATGG CAATTA---- ---------- ---------T
mBtn    28  TTATACTAGA AAAAGAACTA GAAATct--C ATAATCTTCG CAAATA---- ---------- ---------T

bBTN   141  ATAGCTATTA ACAATGTTTC AAATTACATA Annnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN    76  ACAACTATTA ACAATATTTT GAGTTACACA Aaaagtctaa ttga------ ---------- ----------
mBtn    73  ATGCGTATTA GCTATGCTAT GAACTAtgca ggaaaactta ctatgaactt atcactatga actgatatat

bBTN   211  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN   120  ---------- ---TTGCTTT TATGATATAC ATAAGC---- ---------- ---------- ----------
mBtn   143  attgttctta aatTTTATTT TATATTTATG TACAGC---- ---------- ---------- ----------

bBTN   281  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN   143  ---------- ---------- ---------- ---------- ---------- ---------- ----------
mBtn   179  ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN   351  nnnnnnnnnn nnnnnnnnnn nnnnnnnnat tatataaaag tatatcaaat tATATAAATC ATCAGTCCTA
gBTN     1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN   143  ---------- ---------- ---------- ---------- ---------- -ATAAAATAA ATCTTGGAag
mBtn   179  ---------- ---------- ---------- ---------- ---------- -ATAGAAACA ATCATTGATA

bBTN   421 TTAC------ ---------T GGAGGCTATC ATTACTTgct gacattatga gtgatttcta cctt----AA
gBTN     1 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN   162 gactatgcca aaatgtgaaT TGAGGCTATC ATTAGTTagt agattaagtg tgacctccat tttataaaAA
mBtn   198 AAAC------ ---------T Gttt------ ---------- ---------- ---------- ----------
```

```
bBTN 436 AATATACATA TCTTTAAAAT GTTTTCTATA ACAGTAAGg- ---GGTTTAA CCCAAAACAG GGAGTTGTAG
gBTN   1 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 192 TATTTTTCTG TATTTTTAAA TTTTTTCATA ACAATGAGag gtaGGTTTAG TATAAATTAA GGAATTGTAG
mBtn 207 --TTTTTCTT TATCTTTGCA TTTTTTCAgt aataaatgaa aa---TTCAA AACCAAATAA GAAATTGCTG

bBTN 538  GCCTCATAAA Ttgctt---- ---------- ---------- ---------- ---------- ----------
gBTN   1  ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 302  GCCTCATGAt gagttgcttc aaaagaaata attgt----- ---------- ---------- ----------
mBtn 272  ATCTCATGAC Tgatggcagg gtgaagcgcc aggtccttgt gcagttatac cttgaaggtg gacatccagt

bBTN 554 ---------- ---------- ---------- ---------- ---------- TTAGGGAGGT AGCTTTGCAG
gBTN   1 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 337 ---------- ---------- ---------- ---------- ---------- ---------- -------CAG
mBtn 342 ggactcctgc cacccacacc cacattcctg aaggtgtctc atggaaaaga TCAGGGAGGG AGAGCTGCAG

bBTN 574 AGAGGCAACT c-TGGGCCCT GAAAACTGAA TcATACCCAT GCATTCTCAG CCTATCTTTA G---------
gBTN   1 ---------- ---------- ---------- --ATACCCAT GCATTCTCAG CCTATCTTTA G---------
hBTN 340 GGAGGCAACT ttTGGGCCCT GTAAACAGAA T---------- ----TCACAG CCTACCCTTA agactggcct
mBtn 412 ccattgtgga ctcactc--- ---------- ---------- ---------- ---------- ----------

bBTN 634  ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 20   ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 397  ttaatggagt tgttccaaca gacacagcac ctaaggtgga cggacgctcc aataacgaag atggctatag
mBtn 429  ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN 634 ---------- ---------- ---------- -------TTT TCAGCTCATA TAAATAATAA TAGAATAAAA
gBTN  30 ---------- ---------- ---------- -------TTT TtAGCTCATA TAAATAATAA TAGAATAACA
hBTN 467 acagaggcta ctgctacctt ggcctctggg ctttcaTTTT TCAACCCGCA GAAATAATAA TAGAATAACA
mBtn 429 ---------- ---------- ---------- ------TTTA GCTATTCACA GATGTAATGA CAAAGTAAtt
                                                  SoS6

bBTN 667 ACTTCCCCTC TTCTAAATCC CTATTAnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
gBTN  63 ACTTCCCCTC TTCTAAATCC CTATTAtatg ctaagcannn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
hBTN 537 ATTTATCCTC cgcnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 463 tactttctgg gctcctattc tcttgcctgt tttgtttcca atactgtttg tgtctaatac ttttccaact

bBTN 737 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
gBTN 133 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
hBTN 607 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 533 tggcataatt caaacaaggt attagtaaca ttagtctttt tcttaaaagt aacaaacacc ccactctcnn
```

```
bBTN 807 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 203 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
hBTN 677 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnn--- ---------- ---------- ----------
mBtn 603 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn

bBTN 807 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 273 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
hBTN 714 ---------- ---------- ---------- ---------- ---------- ---------- ----------
mBtn 673 nnnnnnnnnn nnnnnnnnnn nnnnnncctc ccccccccatg aacttgggtt aaaagaactg aagccacaga

bBTN 807 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 343 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
hBTN 714 ---------- ---------- ---------- ---------- ---------- ---------- ----------
mBtn 743 gttaaattca caggctgatg gcctcatga- ---------- ---------- ---------- ----------

bBTN 807  ---------- ---------- ---------- ---------- ---------- -------GCC CAAGACTTTG
gBTN 413  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnGCC CAAGGCTTTG
hBTN 714  ---------- ---------- ---------- ---------- ---------- ---------- ----------
mBtn 782  ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN 820 CATCTTATTT TAGTGTTTTC TTAAATCCTC TTTCTTTTTC GCTCTGTCCC CTTtTGTAA- --ACAGGTTA
gBTN 473 CATCTTATTT TAGTGTTTTC TTTAATCCTC TTTCTTTTTC CCTCTGTCCC CTTcTGTAA- --ACAGGTTA
hBTN 714 ---------- ---TGTTTCC TTGCCTCTCT TCTCTTTTTC TCTTTTCTGC CTTTcATTTG GTCCAGCTCa
mBtn 782 ---CTCATTT CAG---TTGC TCAAGTCTTC TTTCTTTTTG TCCCCATTCC CTAT-ATTCG GTACAGCTCT
                                        SoS5

bBTN 887 TCAATGCATA GCGTAATTTC TAAGCTGAAG ACCACAAATC CACAGACTCT TCTCCTCTCT GGAACTTtTT
gBTN 550 TCAATGCATA GCCTAATTTC TAAGCTGAAG ACCACAAATC CACAGACCCT TCTCCTCTCT GGAACTT-TT
hBTN 771 tcannnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnttca
mBtn 835 TTAATGCATA TATCGTTCTC TTAGGGGAGG Aggatgaacc caaactacct gaccactaat ctgtagtcca

bBTN 957 T--------- ---------- ------CCCC CAAAGTAAAT ACATCAGGGA TGAGGAGGGA GTGTGTTGGG
gBTN 620 T--------- ---------- ------CCCC CAAAGTAAAT ACATCAGGGA TGAGGAGGGA GTGTGTTGGA
hBTN 841 tctcccnnnn nnnnnn---- -------CCT CAGAGTGAAT ATACCAGGCC ACCT------ ----------
mBtn 905 catgtttaaa aggctgctcc tcccccCACC CCGAATAAAT ACACTTGGTC ACCT------ ----------
                                        SoS4

bBTN  992 GTGGAAGGGT GTGGGGAGGC AGACTTTCCT GAGAGTACTT CCCCCTTCCT TCTCC----- ---TTTAACT
gBTN  664 GCGGAAGGGT GGGAGGAGGC AGGCTTTCCT GAGAGTACTT CCCCCTTCCT TCTCC----- ---TTTAAGT
hBTN  884 ---------- ----GGAGGC AGGCTTTCCT GAGAGCACCT CCCGCTTCCT TTTCCTGCTG GCTTTTAACT
mBtn  959 ---------- ----GTGGGC AGGCTTCTCT AACAGCACac agc-CTTCTT CCTTCTGAAG AGCTCTCTCT
                                        SoS3
```

```
bBTN 1064 TTTGCCAATG GGGCCACAAC CAGCTTTCTC ATTTGGTAGC AGAAGCTTGT TGGTGCCTTT TCCTCCAACA
gBTN  726 TTTGCCAATG GGGCCACAAC CAGCTTTCTC ATTTGGTAGC AGAAGCTTGT TGGTGCCTTT TCTTCCAACA
hBTN  940 TTTGCCCCGG GGGCCACAGC CAGCTTTCTC ACTTGGTAGC AGTGGC-CTC TTGTGCCTTT TTCTCCAAGA
mBtn 1014 TTGGCCCCGG GGTGACAAGC AGCCCTTTTC ACTTGATCAC TGTGGC-TCT GGCTCCCTTT TCCTCtgggt

bBTN 1134 TCAG--AGGT CAGTATCTGG GACTGTTCTC AC--CCCTGA GAGAGCCAGC AGCC--AGTC ATT-------
gBTN  796 TCAG--AGGT CAGTATCTGG AACTGTTCTC AC--CCCTGA GAGAGCCAGC AGCCAGAGTC ATT-------
hBTN 1009 TCAcccAGGT CAGTATGTGT GGTTACTCTC AGCTCCCTGA GGGAGTGAGC AGCCAGAGgg gcatagcata
mBtn 1083 ctgtcgaaat cggtaggtgc ttc-ACTCTC AGCTCagctc tctttg---- ---------- ----------


bBTN 1191 -TTCTCTTTC CTGGGCTCCT TCCTTCCTAG TCCCAGGCCC TCTATTCTTC CAGCCCTTTG GCTCTCTTCA
gBTN  855 -TTCTCTTTC CCGGGCTCCT TCCTCCCTAG TCCCAGGCCC TCTACTCTTC AAGCCCTTTA GCTCCCTTCA
hBTN 1079 cTTCTCTCCC TCCAGCTCCT TTCTCCTTGG TCCTAGGCCC TCTACTCTTC AAGCCCTTTT GCTTGCTTCA
mBtn 1128 ---------- ---------T CTCTTCTCTG TACTAGGCTT TCTGTTCCTC AAGCTCTTCA GCTCtgcct-
                                                                  SoS2
bBTN 1260 GCTCCCCTTT CTTGT----- ---------- ---------- ---------- ---------- ----------
gBTN  924 GCTCCCCTTT CTTGT----- ---------- ---------- ---------- ---------- ----------
hBTN 1149 GCTCCtgcTT CTTGC----- ---------- ---------- ---------- ---------- ----------
mBtn 1178 -CTCCCCTCT CTctcagact ttgtcaagac tgtatgtacc tcacggtgta actcccagag atcaccctcc

bBTN 1275 ---------- ---------- ---------- ---------- TCTGTCTCCT CCAGCCTTTC TTCTTTCCTG
gBTN  939 ---------- ---------- ---------- ---------- TCTGTCTCCT CCAGCCTTTC TTCTTTCCTG
hBTN 1164 ---------- ---------- ---------- ---------- TCCATCTCCT CCCATCTTTT CTCTCTTCtc
mBtn 1247 tgagagctgc tgggcttaca gttgagaaac acaccttgtc TTTCTCTCCT CCTTCGTTTC ATTTcatgt-

bBTN 1305 CTCCTGCTTA TTTCCCTAGT CTCTATCTCT GCCTCCATCA CTGTATCTTT ATTATCATTG CTAAAAGCAG
gBTN  969 TTCCTGCTTA TTTCCCTGAT CTCCATCTCT TCCTCCATCA CTGTATCTTT ATTATCACTG CTAAAAGCAG
hBTN 1194 caattc---A TTTCCTTGGA CTCCATTTCT GCCTTTATAC CCTTATCTTT ATTATCATTG CCAAAAGCAA
mBtn 1316 ---------- ---------T CTCCATTTCT ACCTCCGTGG CTTTATCTTC ATTATCACTT CTAAACACGA
                                                                  SoS1
bBTN 1375 ATGTCAAATT ATTGTA---- ---------- ---------- ---------- -TTTGAAAGA GGAGA-----
gBTN 1039 ATGTCAAATT ATTGTA---- ---------- ---------- ---------- -TTTGAAAGA AGAGA-----
hBTN 1261 ATACCAAATT CTTGTA---- ---------- ---------- ---------- -TTTgnnnnn nnnnnnnnnn
mBtn 1367 ATAACAAAGT ATcccactcg attcgatttt actttattgt tttattgtta tTGTAAATGA GGAGA-----

bBTN 1405 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 1069 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 1296 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 1432 ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

```
bBTN 1405 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 1069 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 1366 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 1432 ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN 1405 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 1069 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 1436 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 1432 ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN 1405 ---------- ---------- ---------- ---------- ---------- ---------- ----------
gBTN 1069 ---------- ---------- ---------- ---------- ---------- ---------- ----------
hBTN 1506 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
mBtn 1432 ---------- ---------- ---------- ---------- ---------- ---------- ----------

bBTN 1405 ---------- ---------- ---------- TTTTTTTGAC GCAGTAGTTC CAAGACTTGT CTATTTATAA
gBTN 1069 ---------- ---------- ---------- TCTTTTTGAC ACAGTAGTTC TAAGACTTGT CTATTTATAA
hBTN 1576 nnnnnnnnnn nnnnnnnaaa aaggagactt TTTTTTTAGT GCAGTGATTC CAAGATTTGT TCATCGATAt
mBtn 1432 ---------- ---------- ---------- TTTCTTcatt atctacaact g--------- ----------

bBTN 1445 TACA-GAAAC CCGTTAATGG GTACTATGCA TCACAATTGC CTTCTCCTAA GACTCTCTTG GGGGGTTAGA
gBTN 1109 TACAtGAAAC CCATAAATGG GTACTATACA TCACAATTGC CTTCTCTTAA GACTCTCTTA GGGGTTTAGG
hBTN 1646 gcaaccaac- ----TAATGA GTGTTATGCC TCACAGTTCC CTTCTCCTAA Gttttagag- ----------
mBtn 1453 ---------- ---------- ------TGCC TCGCGGCTCC ATTCTggagg cagtcgaggg ctggaggacc

bBTN 1514 GCCCATTTTC TGtTTTGTAC AGACAGGATT GACTAACCTT AGGGTGGTAG GTGGATGTGT GCTGACGGCA
gBTN 1179 GCCAATTTGC TGgTTTGTAC CGACAAGATT GACTAACCTT AGGGTGTTAG GTGGGATGGc agccaga---
hBTN 1700 ---------- ---------- ---------- --------TT AGAGTAGTGG GTGGGAAGGT GATAACCACA
mBtn 1497 ag-------- ---------- ----------

bBTN 1584 GccgggACTG AAAGTCTACC CTT-GGCAGA AGGCAAGGAG GAAGCTTAAG GCGCAAATGA TGCTTTCCA-
gBTN 1249 ------ACTG AAAGTCTACC CTT-GGCAGA AGGCAAGGAG GAAGCTTAAG GTGCAAATGA TGCTTTCCA-
hBTN 1700 Gtgtaaacta aaaatccata ctggGGATGG AGGCTGAGAG GAGGTTTCAG GGGCAAATGa ccagaacact
mBtn 1499 ---------- ---------- ---------- ACGTACAGAG GAAGGGTATG GGGCAGGCGC TGTTGTAAA-

bBTN 1653 ---A------ ---------- ---------- ---------- -------GCT GAAAAGAATT GCTGGGAGGG
gBTN 1309 ---A------ ---------- ---------- ---------- -------GCT GAAAAGAACT GCTGGGAGGG
hBTN 1802 tgcA------ ---------- ---------- ---------- -------GCT GGAAAGAACT GTAGAGAGGA
mBtn 1538 ---Atggact gaaaatgacc ctgtagggga aatacagagc cctccagGTT GGAAGAAACT GGTGGa----

bBTN 1677 CACAGGAGCA CGGTGTCTCA ATAACTCC-- TAGTTATTGC CTATTCTTTC ATCTCTCCTA GGCTGAAGTT
gBTN 1333 CACAGGAGCA TGGTGTCTCA ACAACTCT-- TAGTTATTGC CTATTCTTTC ATCTCTCCTA GGCTGAcaaa
```

```
hBTN 1829 CTTTGGAAAG CGGagggttg acagagccgg TAGTTGTCTC CTGTCCATT- -CATCTCCTA GGCTGAAGCT
mBtn 1601 -----GAACA GGGCGCTTGC GGAACCCA-- TAGTTACCTC CTGACTGTT- -TCTCTCCCA GCCTGAAGCT

bBTN 1744 CCCGAca--- ---------- ---------- ---------- --------
gBTN 1400 ctcgagccca ccagc----- --ATCTTGCT GCCCAGAAAG GTTGG---
hBTN 1897 CCTGAgggga ctcacatcag ttATCTTGCT GCTCCAGAAG GGTGGgag
mBtn 1662 Cttggcgggc ttcattgccc cagttagctc agag------ --------
```

Legend:
Region of highest homology (gray shading).
Segment-of-similarity (double underlined).
TSS (***bold italics***).
Repeat element represented with 'n.'

REFERENCES

Aggeler J, Park CS and Bissell MJ (1988). Regulation of milk protein and basement membrane gene expression: the influence of the extracellular matrix. *Journal of Dairy Science*. 71:2830-2842.

Alexander LJ, Stewart AF, Mackinlay AG, Kapelinskaya TV, Tkach TM and Gorodetsky SI (1988). Isolation and characterization of the bovine kappa-casein gene. *European Journal of Biochemistry*. 178:395-401.

Amadou C, Ribouchon MT, Matei MG, Jenkins NA, Gilbert DJ, Copeland NG, Avoustin P, and Pontarotti P (1995). Localization of New Genes and Markers to the Distal part of the human major histocompatibility complex (MHC) region and comparison with the mouse: New insights into the evolution of mammalian genomes. *Genomics* 26:9-20.

Andres AC, Zuercher G, Djonov V, Flueck M and Ziemiecki A (1995). Protein-tyrosine kinase expression during the estrous-cycle and carcinogenesis of the mammary-gland. *International Journal of Cancer*. 63:288-296.

Aoki N, Ishii T, Ohira S, Yamaguchi Y, Negi M, Adachi T, Nakamura R and Matsuda T. (1997). Stage specific expression of milk fat globule membrane glycoproteins in mouse mammary gland: comparison of MFG-E8, butyrophilin, and CD36 with a major milk protein, beta-casein. *Biochimica Biophysica Acta*. 1334:182-190.

Ashktorab H, Thonabulsombat C, Reed WA and White KL (1999). Bovine β-casein gene promoter activity and hormonal induction of its expression in a mammary epithelial cell line. *Transgenics: Biological Analysis Through DNA Transfer*. 3:23-29.

Ashwell MS, Ogg SL, Mather IH (1996a). The bovine butyrophilin gene maps to chromosome 23. *Animal Genetics*. 27:171-173.

Ashwell MS, Rexroad, CE, Miller RH, and VanRaden PM (1996b). Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. *Animal Genetics*. 27:235-242.

Ashwell MS and Van Tassell CP (1999). Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. *Journal of Dairy Science*. 82:2497-2502.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, and Struhl K, editors (1988). *Current Protocols in Molecular Biology*. John Wiley and Sons Inc., New York.

Baguisi A, Behboodi E, Melican DT, Pollock JS, Destrempes MM, Cammuso C, Williams JL, Nims SD, Porter CA, Midura P, Palacios MJ, Ayres SL, Denniston RS, Hayes ML, Ziomek CA, Meade HM, Godke RA, Gavin WG, Overstrom EW, Echelard Y (1999). Production of goats by somatic cell nuclear transfer. *Nature Biotechnology*. 17: 456-461.

Ball RK, Friis RR, Schoenenberger CA, 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 Journal*. 7:.2089-2095.

Banghart LR, Chamberlain CW, Velarde J, Korobko IV, Ogg SL, Jack LJ, Vakharia VN, and Mather IH (1998). Butyrophilin is expressed in mammary epithelial cells from a single-sized messenger RNA as a type I membrane glycoprotein. *Journal of Biological Chemistry*. 273:4171-4179.

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick B W, Ryan, AM, Clayton D, Li L, Neibergs HL, Zhang N, Grosse WM, Weiss J, Creighton P, McCarthy F, Ron M, Teale AJ, Fries R, Mcgraw RA, Moore SS, Georges M, Soller M, Womack JE and Hetzel DJS (1994). A genetic linkage map of the bovine genome. *Nature Genetics*. 6:227-235.

Bartholomew C, Kilbey A, Clark AM, and Walker M (1997). The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation. *Oncogene*. 14:569-577.

Bauman DE and Griinari JM (2003). Nutritional regulation of milk fat synthesis. *Annual Review of Nutrition*. 23:203-227.

Bellini M, Lacroix JC, and Gall JG. (1993). A putative zinc-binding protein on lampbrush chromosome loops. *EMBO Journal*. 12:107-114.

Berger S L and Kimmel AR, Editors (1987). *Guide to Molecular Cloning Techniques*. Academic Press, Orlando.

Birkenfeld J, Kartmann B, Anliker B, Ono K, Schlötcke B, Betz H and Roth D (2003). Characterization of zetin 1/rBSPRY, a novel binding partner of 14-3-3 proteins, *Biochemical Biophysical Research Communications*. 302:526-533.

Birnboim HC and Doly J (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. 7:1513-1523.

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL, Hawkins GA, Toldo SS, Fries R, Grosz, MD, Yoo J and Beattie CW (1994). A genetic linkage map for cattle. *Genetics* 136:619-639.

Blanchette-Mackie EJ, Dwyer NK, Barber T, Coxey RA, Takeda T, Rondinone CM, Theodorakis JL, Greenberg AS, and Londos C (1995). Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *Journal of Lipid Research*. 36:1211-1226.

Blayney, DP (2002). The changing face of US milk production. *USDA Statistical Bulletin 978*.

Boam DS, Davidson I, and Chambon P (1995). A TATA-less promoter containing binding sites for ubiquitous transcription factors mediates cell type-specific regulation of the gene for transcription enhancer factor-1 (TEF-1). *Journal of Biological Chemistry*. 270:19487-19494.

Bohmer F-D, Kraft R, Otto A, Wernstedt C., Hellman U, Kurtz A, Muller T, Rohde K, Etzold G, Lehmann W, Langen P, Heldin C-H, and Grosse R (1987). Identification of a polypeptide growth inhibitor from bovine mammary gland. Sequence homology to fatty acid- and retinoid-binding proteins. *Journal of Biological Chemistry*. 262:15137-15143.

Botstein, D, White, RL, Skolnick, M and Davis, RW (1980). Construction of a genetic map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*. 32:314-331.

Bovenhuis H and Weller JI (1994). Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. *Genetics*. 137:267-280.

Brunner RM, Guerin G, Goldammer T, Seyfert H, and Schwerin M (1996)**.** The bovine butyrophilin encoding gene (BTN) maps to chromosome 23. *Mammalian Genome*. 7:635-636.

Burke TW and Kadonaga JT (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes and Development*. 10:711-724.

Burke TW and Kadonaga JT (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes and Development*. 11:3020-3031.

Burkhard P, Stetefeld J and Strelkov SV (2001). Coiled-coils: a highly versatile protein folding motif. *Trends in Biochemical Sciences*. 11:82-88.

Cai ST, Han HJ, Kohwi-Shigematsu T (2003).Tissue-specific nuclear architecture and gene expession regulated by SATB1. *Nature Genetics*. 34:42-51.

Cainarca S, Messali S, Ballabio A and Meroni G (1999). Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. *Human Molecular Genetics*. 8:1387-1396.

Campbell KHS, McWhir J, Ritchie WA and Wilmut I (1996). Sheep cloned by nuclear transfer from cultured cell line. *Nature*. 380: 64-66.

Campbell SM, Rosen JM, Hennighausen LG, Strech-Jurk U, and Sippel AE (1984). Comparison of the whey acidic protein genes of the rat and mouse. *Nucleic Acids Research*. 12:8685-8697.

Cavaletto M, Giuffrida MG, Fortunato D, Gardano L, Dellavalle G, Napolitano L, Giunta C, Bertino E, Fabris C, and Conti A (2002). A proteomic approach to evaluate the butyrophilin gene family expression in human milk fat globule membrane. *Proteomics*. 2:850-856.

Cella N, Groner B, and Hynes NE (1998). Characterization of Stat5a and Stat5b homodimers and heterodimers and their association with the glucocortiocoid receptor in mammary cells. *Molecular and Cellular Biology*. 18:1783-1792.

Chakraborty S, Senyuk V, Sitailo S, Chi Y, and Nucifora G (2001). Interaction of EVI1 with cAMP-responsive Element-binding Protein-binding Protein (CBP) and p300/CBP-associated Factor (P/CAF) results in reversible acetylation of EVI1 and in Co-localization in Nuclear Speckles. *Journal of Biological Chemistry*. 276:44936-44943.

Chammas R, Taverna D, Cella N, Santos C and Hynes NE (1994). Laminin and tenascin assembly and expression regulate HC11 mouse mammary cell differentiation. *Journal of Cell Science*. 107:1031-1040.

Chan EKL, Hamel JC, Buyon JP and Tan EM (1991). Molecular definition and sequence motifs of the 52-kD component of human SS-A/Ro autoantigen. *Journal of Clinical Investigation*. 87:68-76.

Chen F, and Capecchi MR (1999). Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proceedings of the National Academy of Sciences, U S A*. 96:541-546.

Chomczynski P and Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*. 162:156-159.

Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce de León FA and Robl JM (1998). Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science*. 280, 1256-1258.

Colgan J and Manley JL (1995). Cooperation between core promoter elements influences transcriptional activity in vivo. *Proceedings of the National Academy of Sciences U.S.A*. 92:1955-1999.

Cox TC, Allen LR, Cox LL, Hopwood B, Goodwin B, Haan E and Suthers, GK (2000). New mutations in the MID1 provide support for loss of function as the cause of X-linked Optiz syndrome. *Human Molecular Genetics*. 9:2553-2562.

Danielson KG, Oborn CJ, Durban EM, Butel JS and Medina D (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. *Proceedings of the National Academy of Sciences*. 81:3756-3760.

Denison MS and Nagy SR (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annual Review of Pharmacology and Toxicology*. 43:309-334.

Deugnier MA, Faraldo MM, Rousselle P, Thiery JP, and Glukhova MA (1999). Cell-extracellular matrix interactions and EGF are important regulators of the basal mammary epithelial cell phenotype. *Journal of Cell Science*. 112:1035-1044.

Diaz E and Pfeffer SR (1998). TIP47: A cargo selection device for mannose-d-phosphate receptor trafficking. *Cell* 93:433-443.

Dickinson LA, Dickinson CD and Kohwi-Shigematsu T(1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural

element in a matrix attachment region. *Journal of Biological Chemistry*. 272:11463-11470.

Dinarello CA, Wolff SM, Goldfinger SE, Dale DC and Alling DW (1974). Colchicine therapy for familial mediterranean fever. A double-blind trial. *New England Journal of Medicine*. 291:934-937.

Doppler W, Groner B and Ball RK (1989). Prolactin and glucocorticoid hormones synergistically induce expression of transfected rat beta-casein gene promoter constructs in a mammary epithelial cell line. *Proceedings of the National Academy of Sciences U.S.A.* 86:104-108.

Doppler W, Villunger A, Jennewein P, Brduscha K, Groner and B, Ball RK (1991). Lactogenic hormone and cell type-specific control of the whey acidic protein gene promoter in transfected mouse cells**.** *Molecular Endocrinology*. 5: 1624-1632

Dorit RL and Gilbert W (1991). The limited universe of exons. *Current Opinion in Genetics and Development*. 1:464-469.

Dosogne H, Arendt J, Gabriel A and Burvenich C (2000). Physiological aspects of bovine milk secretion. Annales de Medecine Veterinaire. 144:357-382.

Dynan WS and Tjian R (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*. 35:79-87.

Febbraio M, Abumrad NA, Hajjar DP, Sharma K, Cheng W, Pearce SF, Silverstein RL (1999). A null mutation in murine CD36 reveals an important role in fatty acid and lipoprotein metabolism. *Journal of Biological Chemistry*. 274:19055-19062.

Feltus FA, Groner B and Melner MH (1999). Stat5-Mediated Regulation of the Human Type II 3ß-Hydroxysteroid Dehydrogenase/Δ^5-Δ^4 Isomerase Gene: Activation by Prolactin. *Molecular Endocrinology*. 13:1084-1093.

Miller KA, Eklund EA, Peddinghaus ML, Cao Z, Fernandes N, Turk PW, Thimmapaya B, Weitzman SA (2001). Kruppel-like factor 4 regulates laminin alpha 3A expression in mammary epithelial cells. *Journal of Biological Chemistry*. 276:42863-42868.

Franke WW, Heid HW, Grund C, Winter S, Freudenstein C, Schmid E, Jarasch ED, and Keenan TW (1981). Antibodies to the major insoluble milk fat globule membrane-associated protein: specific location in apical regions of lactating epithelial cells. *Journal of Cell Biology*. 89:485-494.

Franke WW, Hergt M, and Grund C (1987). Rearrangement of the vimentin cytoskeleton during adipose conversion: formation of an intermediate filament cage around lipid globules. *Cell*. 49:131-141.

Garrett-Sinha LA, Eberspaecher H, Seldin MF, and de Crombrugghe B (1996). A gene for a novel zinc-finger protein expressed in differentiated epithelial cells and transiently in certain mesenchymal cells. *Journal of Biological Chemistry*. 271: 31384-31390.

Ghosal D, Shappell NW, and Keenan TW (1994). Endoplasmic reticulum lumenal proteins of rat mammary gland. Potential involvement in lipid droplet assembly during lactation. *Biochim Biophysica Acta*. 1200:175-81.

Goldfarb, M (1997). Two-dimensional electrophoretic analysis of human milk-fat-globule membrane proteins with attention to apolipoprotein E patterns. *Electrophoresis*. 18:511-515.

Gouilleux F, Wakao H, Mundt M and Groner B (1994). Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. *EMBO Journal*. 13:4361-4369.

Greenwalt DE and Mather IH (1985). Characterization of an apically derived epithelial membrane glycoprotein from bovine milk, which is expressed in capillary endothelia in diverse tissues. *The Journal of Cell Biology*. 100:397-408.

Grønning LM, Dahle MK, Taskén KA, Enerbäck S, Hedin L, Taskén K and Knutsen HK (1999). Isoform-Specific Regulation of the CCAAT/Enhancer-Binding Protein Family of Transcription Factors by 3',5'-Cyclic Adenosine Monophosphate in Sertoli Cells. *Endocrinology.* 140:835-843.

Gu YZ, Hogenesch JB, Bradfield and CA (2000). The PAS superfamily: Scensors of environmental and developmental signals. *Annual Review of Pharmacology and Toxicology*. 40:519-561.

Guggenmos J, Schubart AS, Ogg S, Andersson M, Olsson T, Mather IH and Linington C. (2004). Antibody cross-reactivity between myelin oligodendrocyte glycoprotein and the milk protein butyrophilin in multiple sclerosis. Journal of Immunology. In publication.

Hagen G, Muller S, Beato M, and Suske G (1992). Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic Acids Research*. 20:5519-5525.

Hagen G, Muller S, Beato M, and Suske G (1994). Sp1-mediated transcriptional activation is repressed by Sp3. *EMBO J*. 13:3843-3851.

Hallerman EM**,** Theilmann JL**,** Beckmann JS**,** Soller M**,** and Womack JE (1988). Mapping of bovine prolactin and rhodopsin genes in hybrid somatic cells. *Animal Genetics*. 19:123-131.

Harada R, Berube G, Tamplin O J, Denis-Larose C, and Nepveu A (1995). DNA-binding specificity of the cut repeats from the human cut-like protein. *Molecular and Cellular Biology*. 15:129-140.

Hardy, G.H (1908). Mendelian Proportions in a Mixed Population. *Science*, NS. XXVIII:49-50.

Hare MP, Cipriano F, and Palumbi SR (2002). Genetic evidence on the demography of speciation in allopatric dolphin species. *Evolution*. 56:804-816.

Heid HW, Schnolzer M, and Keenan TW (1996). Adipocyte differentiation-related protein is secreted into milk as a constituent of milk lipid globule membrane. *Biochemical Journal*. 320: 1025-1030.

Henry J, Miller MM, and Pontarotti P (1999). Structure and evolution of the extended B7 family. *Immunology Today*. 20:285-288.

Henry J, Ribouchon M, Depetris D, Mattei M, Offer C, Tazi-Ahnini R, and Pontarotti P (1997a). Cloning, structural analysis, and mapping of the B30 and B7 multigenic families to the major histocompatibility complex (MHC) and other chromosomal regions. *Immunogenetics*. 46:383-395.

Henry J, Ribouchon M-T, Offer C and Ponarotti P (1997b). B30.2-like domain proteins: A growing family. *Biological and Biophysical Research Communications*. 235:162-165

Hines HC, Allaire FR, and Michalak MM (1986). Association of bovine lymphocyte antigens with milk fat percentage differences. *Journal of Dairy Science*. 69:3148-3150.

Hobbs AA, Richards DA, Kessler DJ and Rosen JM (1982). Complex hormonal regualtion of rat casein gene expression. *Journal of Biological Chemistry*. 257:3598-3605.

Horseman ND, Zhao W, Montecino-Rodriguez E, Tanaka M, Nakashima K, Engle SJ, Smith F, Markoff E and Dorshkind K (1997). Defective mammopoiesis, but

normal hematopoiesis, in mice with a targeted disruption of the prolactin gene. *EMBO Journal*. 16:6926-2935.

Hove K (1978). Maintenance of lactose secretion during acute insulin deficiency in lactating goats. *Acta physiologica Scandinavica*. 103:173-179.

Husaini,Y., Wilkins,R.J. and Davey,H.W. (2001) *Bos taurus* butyrophilin (BTN) gene, partial CDS. GenBank Accession No. AF037402.

Huynh HT, Robitaille G and Turner JD (1991). Establishment of bovine mammary epithelial cells (MAC-T): an in vitro model for bovine lactation. *Experimental Cell Research*. 197:191-199.

Inoue S, Orimo A, Hosoi T, Kondo S, Toyoshima H, Kondo T, Ikegami A, Ouchi Y, Orimo H and Muramatsu M (1993). Genomic binding-site cloning reveals an estrogen-responsive gene that encodes a ring finger protein. *Proceedings of The National Academy of Sciences of The United States Of America.* 90:11117-11121.

Ishii T, Aoki N, Noda A, Adachi T, Nakamura R, and Matsuda T. (1995). Carboxy-terminal cytoplasmic domain of mouse butyrophilin specifically associates with a 150-kDa protein of mammary epithelial cells and milk fat globule membrane. *Biochimica Biophysica Acta*. 1245:285-292.

Itoh K, Itoh Y and Frank MB (1991). Protein heterogeneity in the human Ro/SSA ribonucleoproteins. The 52- and 60-kD Ro/SSA autoantigens are encoded by separate genes. *Journal of Clinical Investigation*. 87:177-186.

Jack LJ and Mather IH (1990). Cloning and analysis of cDNA encoding bovine butyrophilin, an apical glycoprotein expressed in mammary tissue and secreted in association with the milk-fat globule membrane during lactation. *Journal of Biological Chemistry*. 265:14481-14486.

Jane SM, Nienhuis AW, and Cunningham J M (1995). Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. *EMBO Journal*. 14: 97-105. [Erratum in: *EMBO J*. 1995. 14:854].

Jarvi SI, Goto RM, Gee GF, Briles WE and Miller MM (1999). Identification, inheritance and linkage of B-G-like and MHC I genes in crane. *The Journal of Heredity*. 90:152-159.

Johnson VG and Mather IH (1985). Monoclonal antibodies prepared against PAS-I butyrophilin and GP-55 from guinea-pig milk-fat-globule membrane bind

specifically to the apical pole of secretory-epithelial cells in lactating mammary tissue. *Experimental Cell Research*. 158:144-158.

Jolivet G, L'Hotte C, Peirre S, Tourkine N and Houdebine LM (1996). A MFG/STAT5 binding site is necessary in the distal enhancer for high prolactin induction of transfected rabbit as1-casein-CAT gene transcription. *FEBs Letters*. 389:257-262.

Karall C, Looft C, Barendse W and Kalm E (1997). Detection and mapping of a point mutation in the bovine butyrophilin gene using F-SSCP-analysis. *Animal Genetics*. 28:66.

Kato Y, Tani T, Sotomaru Y, Kurokawa K, Kato JY, Doguchi H, Yasue H, Tsunoda Y, et al. (1998). Eight calves cloned from somatic cells of a single adult. *Science*. 262: 2095-2098.

Keele JW, Wray JE, Behrens DW, Rohrer GA, Sunden SL, Kappes SM, Bishop MD, Stone RT, Alexander LJ and Beattie CW (1994). A conceptual database model for genomic research. *Journal of Computational Biology*. 1:65-76.

Keenan TW, Dylewski DP, Ghosal D, and Keon BH (1992). Milk lipid globule precursor release from endoplasmic-reticulum reconstituted in a cell-free system. *European Journal of Cell Biology*. 57:21-29.

Kennel De M A, De Bouwerie M, Kolopp-Sarda MN, Faure GC, Bene MC and Bernard CC (2003). Anti-myelin oligodendrocyte glycoprotein B-cell responses in multiple sclerosis. *Journal of Neuroimmunology*. 135:117-125.

Kim C G, Swendeman SL, Barnhart KM, and Sheffery M (1990). Promoter elements and erythroid cell nuclear factors that regulate alpha-globin gene transcription in vitro. *Molecular and Cellular Biology*. 10:5958-5966.

Kollmar R, Sukow KA, Sponagle SK, and Farnham PJ (1994). Start site selection at the TATA-less carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase promoter. *Journal of Biological Chemistry*. 269:2252-2257.

Kovesdi, I., Reichel, R., Nevins, J. R. (1986) Identification of a cellular transcription factor involved in E1A transactivation. *Cell* 45:219-228.

Kralj M and Pipan N (1992). The role of exocytosis in the apocrine secretion of milk lipid globules in mouse mammary gland during lactogenesis. *Biology of the Cell*. 75:211-216 [Erratum in: *Biol. Cell* 1992. 76:288. Metka K(corrected to Kralj M);Nada P(corrected to Pipan N)].

Kubota C, Yamakuchi H, Todoroki J, Mizoshita K, Tabara N, Barber M, and Yang X (2000). Six cloned calves produced from adult fibroblast cells after long-term culture. *Proceedings of the National Academy of Sciences*. 97: 990-995.

Kumaresan P and Turner CW (1965). Effect of insulin and alloxan on mammary gland growth in rats. *Journal of Dairy Science* 48:1378-1381.

Kwon HS, Kim MS, Edenberg HJ, and Hur MW (1999). Sp3 and Sp4 can repress transcription by competing with Sp1 for the core cis-elements on the human ADH5/FDH minimal promoter. *Journal of Biological Chemistry*. 274:20-28.

Kyriakou SY and Kuhn NJ (1973). Lactogenesis in the diabetic rat. *Journal Endocrinology*. 59:199-200.

Lagziel A, Lipkin E, Ezra E, Soller M, and Weller JI (1999). An MspI polymorphism at the bovine growth hormone (bGH) gene is linked to a locus affecting milk protein percentage. *Animal Genetics*. 30:296-299.

Lander ES and Botstein D (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 121:185-199. [Erratum in: *Genetics* 1994. 136:705].

Lanza RP, Cibelli JB, Blackwell C, Cristofalo VJ, Francis MK, Baerlocher GM, Mak, J, Schertzer M, Chavez EA, Sawyer N, Lansdorp PM and West MD (2000). Extension of Cell Life-Span and Telomere Length in Animals Cloned from Senescent Somatic Cells. *Science*. 288:665-669.

Le Provost F, Riedlinger G, Hee Yim S, Benedict J, Gonzalez FJ, Flaws J and Hennighausen L (2002). The aryl hydrocarbon receptor (AhR) and its nuclear translocator (Arnt) are dispensable for normal mammary gland development but are required for fertility. *Genesis*. 32:231-239.

Lee KH, Chang KW, Cho KH, and Lee KJ (2002). Genetic polymorphisms of BTN and STAT5a genes in Korean proven and young bulls. *Asian-Australasian Journal of Animal Sciences*. 15: 938-943.

Lehrach H, Diamond D, Wozney JM and Boedtker H (1977). RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry*. 16:4743-4751.

Levieux D and Ollier A (1999). Bovine immunoglobulin G, beta-lactoglobulin, alpha-lactalbumin and serum albumin in colostrum and milk during the early post partum period. *Journal Of Dairy Research*. 66:421-430.

Lieber JG, and Evans RM (1996). Disruption of the vimentin intermediate filament system during adipose conversion of 3T3-L1 cells inhibits lipid droplet accumulation. *Cell Science*. 109:3047-3058.

Linsley PS, Peach R, Gladstone P and Bajorath, J (1994). Extending the B-7 (CD80) gene family. *Protein Science*. 3:1341-1343.

Liu J, Barnett A, Neufeld EJ, and Dudley JP (1999). Homeoproteins CDP and SATB1 interact: potential for tissue-specific regulation. *Molecular and Cellular Biology*. 19: 4918-4926.

Liu J, Prickett TD, Elliott E, Meroni G, Brautigan DL. (2001). Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proceedings of the National Academy of Sciences U S A*. 98:6650-6655.

Lockwood DH, Voytovich AE, Stockdale FE and Topper YJ (1967). Insulin dependent DNA polymerase and DNA synthesis in mammary epithelial cells in vitro. *Proceedings of the National Academy of Sciences USA*. 58:658-664.

Lu J, Lee W, Jiang C, and Keller EB (1994). Start site selection by Sp1 in the TATA-less human Ha-ras promoter. *Journal of Biological Chemistry*. 269:5391-5402.

Malewski T and Zwierzchowski (2002). Genes expressed in cattle mammary gland - computational analysis of the 5'-upstream sequences in search for factors conferring a tissue-and stage-specific transcription. *Animal Science Papers and Reports*. 20:5-20.

Manabe R, Whitmore L, Weiss JM and Horwitz AR (2002). Identification of a novel micro-tubule associated protein that regulates microtubule organization and cytokinesis by using a GFP-screening strategy. *Current Biology*. 12:1946-1951.

Mansfield E, Chae JJ, Komarow HD, Brotz TM, Frucht DM, Aksentijevich I and Kastner DL (2001). The familial Mediterranean fever protein, pyrin, associates with microtubules and colocalizes with actin filaments. *Blood*. 98:851-859.

Martin KM, Cooper WN, Metcalfe JC and Kemp PR (2000). Mouse BTEB3, a new member of the basic transcription element binding protein (BTEB) family, activates expression from GC-rich minimal promoter regions. *Biochemical Journal* 345:529–533.

Mather IH, Tamplin CB and Irving MG (1980). Separation of the proteins of bovine milk-fat-globule membrane by electrofocusing with retention of enzymatic and immunological activity. *European Journal of Biochemistry*. 110:327-336.

Mather IH and Keenan TW (1998). Origin and secretion of milk lipids. *Journal of Mammary Gland Biology and Neoplasia*. 3:259-273

Mather, IH (2000). A review and proposed nomenclature for major proteins of the milk fat globule membrane. *Journal of Dairy Science*. 83:203-247.

McKenzie HA (Ed.) (1970 & 1971). *Milk Proteins Vol I & II*. Academic Press, New York.

McShane RD, Gallagher DS Jr, Newkirk H, Taylor JF, Burzlaff JD, Davis SK, and Skow LC. (2001). Physical localization and order of genes in the class I region of the bovine MHC. *Animal Genetics*. 32:235-239.

Mendel, G (1865). Versuch uber pflanzen-hybridenVerh. *Natforsche Ver. Brunn*. 4:3-47. [English translation from the website: http://www.mendelweb.org/Mendel.html].

Merlo GR, Graus-Porta D, Cella N, Marte BM, Taverna D and Hynes NE (1996). Growth, differentiation and survival of HC11 mammary epithelial cells: diverse effects of receptor tyrosine kinase-activating peptide growth factors. *European. Journal of Cell Biology*. 70:97–105.

Miller KA, Eklund EA, Peddinghaus ML, Cao Z, Fernandes N, Turk PW, Thimmapaya B and Weitzman SA (2001). Kruppel-like factor 4 regulates laminin alpha 3A expression in mammary epithelial cells. Journal of Biological Chemistry. 276:42863-42868.

Mills ES and Topper YJ (1970). Some ultrastructural effects of insulin, hydrocortisone, and prolactin on mammary gland explants. *Journal of Cell Biology*. 44:310-328.

Molenaar AJ, Davis SR, Jack LJ and Wilkins RJ (1995). Expression of the butyrophilin gene, a milk fat globule membrane protein, is associated with the expression of the alpha S1casein gene. *Histochemical Journal*. 27:388-394.

Mondy BL and Keenan TW (1993). Butyrophilin and xanthine oxidase occur in constant molar proportions in milk lipid globule membrane but vary in amount with breed and stage of lactation. *Protoplasma*. 177:32-36.

Mucenski ML, Taylor BA, Ihle JN, Hartley JW, Marse HC III, Jenkins NA, and Copeland NG (1988). Identification of a common ecotropic viral integration site,

Evi-1, in the DNA of AKXD murine myeloid tumors. *Molecular and Cellular Biology*. 8:301-308.

Müller H, Braken AP, Vernell R, Moroni CM, Christians F, Grassilli E, Prosperini, E, Vigo E, Oliner J D, and Helin K (2001). E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes and Development*. 15:267-285.

Murata, T., Nitta, M., and Yasuda, K. (1998). Transcription factor CP2 is essential for lens-specific expression of the chicken alphaA-crystallin gene. *Genes to Cells*. 3:443-457.

Nandi, S (1958). Endocrine control of mammary gland development and function in C3 11/HE Crgl mouse. *Journal of the National Cancer Institute*. 21:1039-1062.

Neville MC, McFadden TB and Forsyth I (2002). Hormonal regulation of mammary differentiation and milk secretion. *Journal of Mammary Gland Biology and Neoplasia*. 7:49-66.

Nicholas KR, Sankaran L and Topper YJ (1983). A unique and essential role for insulin in the phenotypic expression of rat mammary epithelial cells unrelated to its function in cell maintenance. Biochimica et Biophysica Acta. 25:763:309-314.

Oftedal, OT (2002). The mammary gland and its origin during synapsid evolution. *Journal of Mammary Gland Biology and Neoplasia*. 7:225-252.

Ogg SL, Komaragiri MV and Mather IH (1996). Structural organization and mammary-specific expression of the butyrophilin gene. *Mammamlian Genome*. 7:900-905.

Ogg SL and Mather IH, unpublished.

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

Palmer S, Brouillet JP, Kilbey A, Fulton R, Walker M, Crossley M, and Bartholomew C (2001). Evi-1 transforming and repressor activities are mediated by CtBP co-repressor proteins. *Journal of Biology of Chemsitry*. 276:25834-25840.

Peterson JA, Patton S, and Hamosh M (1998). Glycoproteins of the human milk fat globule in the protection of the breast-fed infant against infections. *Biology of the Neonate*. 74:143-162.

Peterson JA, Scallan CD, Ceriani RL, and Hamosh M. (2001). Structural and functional aspects of three major glycoproteins of the human milk fat globule membrane. *Advances in Experimental Medicine and Biology*. 501:179-87.

Pezet A, Ferrag F, Kelly PA and Edery M (1997). Tyrosine docking sites of the rat prolactin receptor required for association and activation of stat5. *Journal of Biological Chemistry*. 272:25403-25050.

Ponting C, Schultz J and Bork P (1997). SPRY domains in ryanodine receptors (Ca^{2+}-release channels). *Trends in Biochemical Sciences*. 22:193-194.

Prattes S, Horl G, Hammer A, Blaschitz A, Graier WF, Sattler W, Zechner R, and Steyrer E (2000). Intracellular distribution and mobilization of unesterified cholesterol in adipocytes: triglyceride droplets are surrounded by cholesterol-rich ER-like surface layer structures. *Journal of Cell Science*. 113:2977-2989.

Prosser CG, Sankaran L, Hennighausen L and Topper YJ (1987). Comparison of the roles of insulin and insulin-like growth factor I in casein gene expression and in the development of α-lactalbumin and glucose transport activities in the mouse mammary epithelial cell. Endocrinology. 120:1411-1416.

Puissant C and Houdebine LM (1991). Cortisol induces rapid accumulation of whey acidic protein mRNA but not of αs1 and β-casein mRNA in rabbit mammary explants. *Cell Biology International Reports*. 15:121-129. [Erratum: *Cell Biol Int Rep* 1991. 15:457].

Quaderi NA, Schweiger S, Gaudenz K, Franco B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S, Marion RW, Hennekam RC, Opitz JM, Muenke M, Ropers HH and Ballabio A. (1997). Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nature Genetics*. 17:285-291.

Quandt K, Frech K, Karas H, Wingender E and Werner T (1995). MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research 23:4878-4884 (1995).

Reddy BA, Kloc M, and Etkin L (1991). The cloning and characterization of a maternally expressed novel zinc finger nuclear phosphoprotein (xnf7) in Xenopus laevis. *Developmental Biolology*. 148:107-116.

Rhodes DA, Stammers M, Malcherek G, Beck G and Trowsdale J (2001). The Cluster of BTN genes in the extended major histcompatability complex. *Genomic*. 71: 351-362.

Robinson GW, McKnight RA, Smith GH and Hennighausen H (1995). Mammary epithelial-cells undergo secretory differentiation in cycling virgins but require pregnancy for the establishment of terminal differentiation. *Development*. 121:2079-2090.

Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochemical Sciences*. 21: 327-335.

Rosen KM, Lamperti ED and Villa-Komaroff (1990). Optimizing the northern blot procedure. BioTechniques. 8: 398-403.

Ruddy DA, Kronmal GS, Lee VK, Mintier GA, Quintana L, Domingo R Jr, Meyer NC, Irrinki A, McClelland EE, Fullan A, Mapa FA, Moore T, Thomas W, Loeb DB, Harmon C, Tsuchihashi Z, Wolff RK, Schatzman RC, and Feder JN (1997). A 1.1-Mb transcript map of the hereditary hemochromatosis locus. *Genome Research*. 7:441-456.

Rui H, Lebrun JJ, Kirken RA, Kelly PA and Farrar WL (1994). JAK2 activation and cell proliferation induced by antibody-mediated prolactin receptor dimerization. *Endocrinology*. 135:1299-1306.

Safe S, Wormke M and Samudio I (2000). Mechanisms of inhibitory aryl hydrocarbon receptor-estrogen receptor crosstalk in human breast cancer cells. *Journal of Mammary Gland Biology and Neoplasia*. 5:295-306.

Sambrook J, Fritsch EF and Maniatis T (1989). *Molecular cloning: A laboratory manual, 2nd edition*. Cold Spring Harbor Laboratory Press.

Sasaki M, Eigel WN and Keenan TW (1978). Lactose and major milk proteins are present in secretory vesicle-rich fractions from lactating mammary gland. *Proceedings of the National Academy of Sciences U S A*. 75:5020-5024.

Saurin AJ, Borden KLB, Boddy MN and Freemont PS (1996). Does this have a familiar RING? *Trends in Biochemical Sciences*. 21:*208-214.*

Schenker T and Trueb B (2000). BSPRY, a novel protein of the Ro-Ret family. *Biochimica Biophysica Acta*.1493:255-258.

Schmitt-Ney M, Doppler W, Ball RK, and Groner B (1991). Beta-casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. *Molecular and Cellular Biology*. 11:3745-3755.

Schweiger S, Foerster J, Lehmann T, Suckow V, Muller YA, Walter G, Davies T, Porter H, van Bokhoven H, Lunt PW, Traub P and Ropers HH (1999). The Opitz syndrome gene product, MID1, associates with microtubules. *Proceedings of the National Academy of Sciences U S A*. 96:2794-2799.

Segre JA, Bauer C and Fuchs E (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nature Genetics*. 22:356-400.

Seto, MH, Liu H-L C, Zajchowski DA and Whitlow M (1999). Protein fold analysis of the B30.2-like domain. *Proteins:Structure, Function and Genetics*. 35:235-249.

Shields J M and Yang VW (1998). Identification of the DNA sequence that interacts with the gut-enriched Krueppel-like factor. *Nucleic Acids Research*. 26:796-802.

Smale ST and Kadonaga JT (2003). The RNA Polymerase II core promoter. Annual *Review of Biochemistry*. 72: 449-479.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ and Klenk DC (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemsitry*. 150:76-85.

Smith SJ, Cases S, Jensen DR, Chen HC, Sande E, Tow B, Sanan DA, Raber J, Eckel RH, and Farese RV Jr. (2000). Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. Nature Genetics. 25:87-90.

Southern EM (1975). Detection of specific sequences among DNA fragments separated by gel-electrophoresis. *Journal of Molecular Biology*. 98:503-517.

Spitsberg VL, Matitashvili E and Gorewit RC (1995). Association and coexpression of fatty-acid-binding protein and glycoprotein CD36 in the bovine mammary gland. *European Journal of Biochemistry*. 230:872-878.

Stefferl A, Schubart A, Storch M, Amini A, Mather I, Lassmann H, Linington C (2000). Butyrophilin, a milk protein, modulates the encephalitogenic T cell response to myelin oligodendrocyte glycoprotein in experimental autoimmune encephalomyelitis. *Journal of Immunology*. 165:2859-2865.

Stein PA, Toret CP, Salic AN, Rolls MM, and Rapoport TA (2002). A novel centrosome-associated protein with affinity for microtubules. *Journal of Cell Science*. 115:3389-3402.

Stoecklin E. Wissler M, Gouilleux F and Groner B (1996). Functional interactions between Stat5 and the glucocorticoid receptor. *Nature* 383:726-728.

Streuli CH, Bailey N and Bissell MJ (1991). Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. *Journal of Cell Biology*. 115:1383-1395.

Sueyoshi T, Kobayashi R, Nishio K, Aida K, Moore R, Wada T, Handa H, and Negishi M. (1995). A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp 2d-9) gene in mouse liver. *Molecular and Cellular Biology*. 15:4158-4166.

Syu LJ, and Saltiel AR (1999). Lipotransin: a novel docking protein for hormone-sensitive lipase. *Molecular Cell*. 4:109-115.

Takahashi M, Inaguma Y, Hiai H, and Hirose F (1988). Developmentally regulated expression of a human "finger"-containing gene encoded by the 5' half of the ret transforming gene. *Molecular and Cellular Biology*. 8:1853-1856.

Tankslet D, Medina H and Rick CM (1981). The effect of isoenzyme selection on metric characters in an interspecific backcross of tomato - Basis of an early screening-procedure. *Theoretical and Applied Genetics*. 60: 291-296.

Taylor C, Everest M and Smith C. (1996). Restriction fragment length polymorphism in amplification products of the bovine butyrophilin gene: assignment of bovine butyrophilin to bovine chromosome 23. *Animal Genetics*. 27:183-185.

Teng CT (2002). Comparative analysis on the structural features of the 5' flanking region of kappa-casein genes from six different species. *Genetique, Selection, Evolution*. 34:117-128.

The French FMF Consortium (1997). A candidate gene for familial Mediterranean fever. *Nature Genetics*. 17:25-31.

The International FMF Consortium (1997). Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. *Cell*. 90:797-807.

Tissot C and Mechti, N (1995). Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus I long terminal repeat expression. *Journal of Biological Chemistry*. 270:14891-14898.

Tronche F, and Yaniv M (1992). HNF1, a homeoprotein member of the hepatic transcription reulatory network. *BioEssays* 14:579-587.

Valivullah HM, Bevan DR, Peat A, and Keenan TW (1998). In vitro fusion of microlipid droplets from mammary gland. *Progress in Clinical and Biological Research*. 270:421-425.

Vernet C, Boretto J, Mattei MG, Takahashi M, Jack LJ, Mather IH, Rouquier S, and Pontarotti P (1993). Evolutionary study of multigenic families mapping close to the human MHC class I region. *Journal of Molecular Evolution*. 37:600-612.

Verrijzer CP, Alkema MJ, van Weperen WW, van Leeuwen HC, Strating MJJ, van der Vliet PC (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO Journal*. 11:4993-5003.

Vilotte JL and Soulier S (1992). Isolation and characterization of the mouse alpha-lactalbumin-encoding gene: interspecies comparison, tissue- and stage-specific expression. *Gene*. 119:287-292.

Vonderhaar BK and Ziska SE (1989). Hormonal regulation of milk protein gene expression. *Annual Review of Physiology*. 51:641-52.

Vorbach C, Scriven A, and Capecchi MR (2002). The housekeeping gene xanthine oxidoreductase is necessary for milk fat droplet enveloping and secretion: gene sharing in the lactating mammary gland. *Genes and Development*. 16:3223-3235.

Wakao H, Gouilleux F and Groner B (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription gene family and confers prolactin response. *EMBO Journal*. 13:2182-2191. [Erratum in: *EMBO Journal* 1995. 14:854-855].

Watson CJ, Gorden KE, Robertson M and Clark AJ (1991). Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Research*. 19: 6603-6610.

Weinberg, W. 1909, Über Vererbungsgesetze beim Menschen: *Zeitschrift Abstammung und Vererblichkeit*. 1 377–393. 440–460.

Weller JI, Kashi Y and Soller M (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science*. 73:2525-2537.

Welsch U, Schumacher U, Buchheim W, Schinko I, Jenness P and Patton S (1990). Histochemical and Biochemical Observations on milk-fat-globule membranes from several mammalian-species. *Acta Histochemica*. Suppl. 40: 59-64.

Wilmut I, Schnieke, AE, McWhir J, Kind AJ and Campbell, KHS (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature*. 385: 810-813.

Wooding FB (1971). The mechanism of secretion of the milk fat globule. *Journal of Cell Science*. 9: 805-821.

Wooding FBP (1973). Formation of milk-fat globule membrane without participation of plasmalemma. *Journal of Cell Science*. 13:221-235.

Wu CC, Howell KE, Neville MC, Yates JR and McManaman JL (2000). Proteomics reveal a link between the endoplasmic reticulum and lipid secretory mechanisms in mammary epithelial cells. *Electrophoresis*. 21:3470-3482.

Ye H, and Serrero G (1998). Stimulation of adipose differentiation related protein (ADRP) expression by ibuprofen and indomethacin in adipocyte precursors and in adipocytes. *Biochemical Journal*. 330:803-809.

Ye T-Z, Gordon CT, Lai Y-H, Fujiwere Y, Peters LL, Perkins AC and Chui DHK (2000). Ermap, a gene coding for a novel erythroid specific adhesion/receptor membrane protein. *Gene*. 242:337-345.

Zaczek M and Keenan TW (1990). Morphological evidence for an endoplasmic-reticulum origin of milk lipid globules obtained using lipid-selective staining procedures. *Protoplasma*. 159: 179-182.

Zavizion B, Gorewit RC and Politis I (1995). Subcloning the MAC-T bovine mammary epithelial cell line: morphology, growth properties, and cytogenetic analysis of clonal cells. *Journal of Dairy Science.* 78:515-527.

Zemer D, Revach M, Pras M, Modan B, Schor S, Sohar E and Gafni J (1974). A controlled trial of colchicine in preventing attacks of familial mediterranean fever. *New England Journal of Medicine*. 291:932-934.

Zhang W, Shields JM, Sogawa K, Fujii-Kuriyama Y, and Yang VW (1998). The Gut-enriched Krüppel-like Factor Suppresses the Activity of the CYP1A1 Promoter in an Sp1-dependent Fashion. *Journal of Biological Chemistry*. 73: 17917 - 17925.