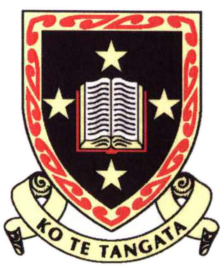
Clonal Cell Proliferation in the Mammary Gland



The University of Waikato
Te Whare Wānanga o Waikato

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy

in Biological Sciences

at

The University of Waikato

by

GREGORY MICHAEL JACOBSON

2004

Abstract

The primary aim of this thesis was to classify the nature of the clonal expansion of cells that gives rise to bovine and murine mammary epithelia. It is generally assumed that the mammary gland is derived from the outgrowth of mammary–determined stem cell progeny, and that this expansion produces a mixture of cellular clones throughout the gland. Molecular techniques that enable the 'visualization' of the size and distribution of these clones will lead to a better appreciation of mammary gland development and architecture.

In female mammals, the simplest 'mark' of clonality is provided by the inactivation of one or the other of the two X chromosomes. This event occurs early in embryogenesis and gives rise to stem cells that propagate this mark in progeny cells throughout the life of the animal. Two approaches, that utilized X-inactivation as a mark of clonality, were investigated, 1) naturally occurring X-linked gene polymorphisms in cattle, and 2) artificially introduced X-linked transgenes in mice.

The first part of this thesis deals with the isolation and utilization of polymorphic markers on the bovine X-chromosome as markers of clonality. Such expressed genetic markers are essential if one is to distinguish which of the two X-chromosomes is active in a given cell. Two X-linked bovine genes, glucose-6-phosphate dehydrogenase (*g6pd*) and the X-inactive specific transcript (*Xist*) were selected for sequencing and SNP detection. As no specific sequence data was available for these genes, a bovine lambda genomic DNA library was screened. The G6PD clone selected from a lambda library was shown to be a pseudogene and therefore not useful for the present study. Another clone, that hybridized with an *Xist* gene probe, was shown to actually represent a bovine Huntingtin-Associated Protein 1 like (*Hap1*-like) pseudogene, making it of no utility in this study. However, screening of regions of exon 1 of the *Xist* gene by direct PCR and sequencing, led to the discovery of an informative G↔T single nucleotide polymorphism (SNP). The variation, which was found in frequencies of 0.8 and 0.2 for 'G' and 'T' variants respectively (32% of females are heterozygous), was used in a number of different ways in order to assess the clonal profile of small

'blocks' of the gland (≤30 mm^3). In these preliminary experiments, RT-PCR analysis of *Xist* expression patterns within each of the small blocks (each comprising around 3 X 10^7 cells) showed general heterogeneity but the experiments were not conclusive. Limitations of the experimental methods used are discussed and suggestions of alternative approaches, that will offer higher resolution, are presented.

A second experimental approach, involving mice with a reporter transgene construct introduced into the X-chromosome (H253), was also used. The transgene consisted of 14 tandem repeats of a *lacZ* nuclear-localized reporter gene coupled with a mouse 3-hydroxyl-3-methylglutaryl coenzyme A reductase (*hmgcr*) promoter (Tam and Tan, 1992). A number of researchers had already utilized this transgenic model, and had confirmed that the cellular patterns of *lacZ* expression reflected actual X-inactivation patterns in a number of tissues. While carrying out initial investigations to ascertain the utility of this reporter system in investigations of the mammary gland it was discovered that the usual description of *hmgcr* as a 'house-keeping' gene is somewhat of a misnomer, as some tissues in the mouse show very strong patterns of *lacZ* expression while others gave minimal, or no, detectable expression. Furthermore, transgene expression was observed to increase during pregnancy and it was concluded that this effect was almost certainly a response to increased estrogen levels at this time.

Whole mount analysis of five mouse mammary glands from transgenic animals revealed heterogeneous patterns of *lacZ* staining, indicating that mammary epithelia is generally polyclonal. In total, 42% of terminal buds and alveoli displayed a single *lacZ* staining pattern. Analysis of these staining patterns in duct termini indicated that the majority of individual alveoli are derived from two or three division competent (stem) cells.

Acknowledgements

Firstly, thanks to my supervisor, Dick Wilkins, for constant availability and assistance throughout my many years of post-graduate study. Your ideas, and faith in me, were always helpful and encouraging. Ray Cursons deserves a special mention. Your technical advice was invaluable, and I will never forget your willingness to help everyone who approached you. Special thanks to Raewyn Towers for proof-reading, encouragement, and general assistance in the lab. Your heroic efforts, particularly when times were tough in the past few years, were very much appreciated. Bridget D. and Brett W., your proofreading efforts, and general encouragement, were helpful and appreciated.

Many thanks to the lab-rats, for support, humour, interesting gossip and for helping to generate a good working environment – Amanda, Anandan, Chris, Dave, Debbie, Geoff, Gina, Greg B, Gune, Henry, Marcel, The Marks, Peter, Sandy, Tao, You-Ying. Thanks again to Brett and Seamus for donations of enzyme and various reagents, and to Xedong for protocol suggestions. Barry, your instructions and direct assistance with much of the histological work was much appreciated. The assistance and patience of Ben Wilson and Goetz Labile made the transgenic mouse study possible. Thanks to Shinichi for advice with the statistics and to the staff at the Waikato Hospital histopathology lab for the H and E staining of bovine mammary sections. Thanks to Danni for encouragement and for general support.

A very special thanks to you Mum. This thesis would not have been possible without the unfailing support that I received from you. Your constant encouragement and interest in what I was doing was fantastic.

Table of Contents

List of tables

List of figures

Chapter 4

Chapter 5

Chapter 6

Appendix

Abbreviations:

aa	Amino acid
AL	Alveolar Lobules
APES	Amino-propyl-tri-ethoxy-silane
ASO	Allele Specific Oligonucleotide
β-gal	β-galactosidase
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
BLASTN	Basic Local Alignment Search Tool, nucleotide:nucleotide
BM	Basement Membrane
BMP	Bitmap image
bp	Base pair/s
BrdU	Bromo Deoxyuridine
cAMP	Cyclic Adenosine Mono Phosphate
CCD	Charge-coupled device
CMV	Cytomegalovirus
CNS	Central Nervous System
CSF	Colony stimulating factor
CT	Connective Tissue
CRE	cAMP response element
d	Day/s
DLLC	Differentiated Large Light Cells
DNA	Deoxyribose Nucleic Acid
dpc	Days post-conception
E2	estradiol (predominant physiological estrogen)
EDTA	Ethylenediamine-tetraacetic acid
EGTA	Ethyleneglycol-tetraacetic acid
ER	Estrogen receptor
ERE	Estrogen Response Element
ERMA	Environmental Risk Management Authority
EtBr	Ethidium bromide
FGH	Fetal Growth Hormone
gb	NCBI Genbank
G6PD	Glucose-6-Phosphate 1-dehydrogenase
GH	Growth Hormone
h	Hour/s
Hmgcr	3-hydroxy-3-methylglutaryl coenzyme A reductase gene
HGPRT/Hgprt	Hypoxanthine-guanine phosphoribosyl transferase gene
HSP	Heat shock protein
HUMARA	Human androgen receptor gene
IGF	Insulin-like growth factor
lacZ	Bacterial β-galactosidase gene
LINE	Long Interspersed Nuclear Element
LLC	Large Light Cells
LRC	Label-retaining cells
min	Minute/s
Mf3	Winged helix/forkhead transcription factor gene
MG	Mammary gland
MGF	Mammary Gland specific nuclear Factor
MMP	Matrix metalloproteinase

MMTV	Mouse mammary tumor virus
mRNA	Messenger RNA
NaDOC	Sodium deoxycholate
NCBI	National Centre for Biological Information
NLS	Nuclear Localization Signal
NP-40	Nonidet P-40
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pers. comm.	Personal communication
PFA	Paraformadehyde
PGK	Phosphoglycerate Kinase
PM	DNA Polymorphism
PNL	Peanut lectin
PR	Progesterone receptor
PRL	Prolactin
PRLR	Prolactin receptor
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
RT-PCR	Reverse-Transcriptase Polymerase Chain Reaction
SC	Stem Cell/s
sec	Second/s
SINE	Short Interspersed Nuclear Elements
SLC	Small Light Cells
SNP	Single Nucleotide Polymorphism
SRE	Sterol response element
SREBP	Sterol response element binding protein
STAT	Signal transducer and activator of transcription
STR	Short terminal repeat
TA	Transiently Activated cell/s
TDLU	Terminal Ductal lobular Units
TEB	Terminal End Bud
TED	Terminal End Duct
TGF- ß	Transforming Growth Factor-β isoforms
TIFF	Tag Image File Format
TU	Transitional units
ULLC	Undifferentiated Large Light Cells
VNTR	Variable number tandem repeat
WAP	Whey acidic protein
wks	Weeks
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
XIC	X-chromosome inactivation centre
XIST/Xist	X-inactive Specific Transcript
X_m	Maternally derived X chromosome
X_p	Paternally derived X chromosome

Preamble

Chapter 1 begins with a brief introduction and concludes with a literature review comprising:

1) A review of general mammary development, with particular emphasis on the importance of the Terminal End Bud (TEB) in murine and similar structures in ruminants.
2) The current understanding of mechanisms of cellular clonality.
3) A review of established knowledge of mammary ductal clonality.
4) Details of existing approaches for measurement of cellular clones.
5) Proposed experimental approaches for the study of cellular clonality in the bovine and murine glands.

Chapter 2 describes general methods and materials used during this project. More specific details of methods used can be found within the relevant results chapters. Chapter 3, the first results chapter, describes the detection of bovine Single Nucleotide Polymorphisms (SNPs) by DNA sequencing and analysis of X-linked genes. The second results chapter, Chapter 4, describes the development of an RT-PCR/RFLP approach to allow the visualization of the contribution of different stem cell clones to the mammary gland. In Chapter 5, experiments used to determine the suitability of an animal model (a *lacZ* transgenic mouse) for studies of cellular clonality in the murine mammary gland are presented. Chapter 6 details application of the techniques developed in Chapter 5, showing how the animal model is used to visualize clonal events within mammary tissue and to allow the prediction of precursor cell numbers contributing to individual terminal ductal structures.

Finally, in Chapter 7, a discussion of the relevance and importance of the findings in the present study is presented.

Chapter One

Cellular Clonality of Mammary Epithelial Tissue

1.1 General introduction

Mammary organogenesis follows a developmental program that occurs predominantly postnatally (Horseman, 1999). This special feature of the mammary gland makes it possible to observe changes in branching morphogenesis in adult tissue and to correlate these changes to different physiological events. In contrast, most other glandular tissues (for example, salivary glands) undergo branching morphogenesis that is near completion at birth, therefore making these tissues far less amenable to such study.

In virgin animals, ductal growth and lateral branching occurs to fill the mesenchymal space; in pregnant animals, further growth, with differentiation of milk-producing glandular components, occurs. At parturition the gland has been organised to allow for milk secretion and its efficient transfer to suckling young. After weaning, the gland undergoes radical restructuring, involution, to return it to an anatomical state similar to that of the virgin animal. Subsequent pregnancy requires the outgrowth of residual mammary tissue to produce a large mammary ductal network. Experimental characterisation of these stages of epithelial cell growth and reduction is essential for maximising profitability from dairy animals (Ellis, 1998).

Mammary epithelial tissue expansion occurs by proliferation from a pool of dedicated stem cells, with progeny of each of these stem cells forming small populations ('clones') of cells in the adult gland. This thesis examines the size and distribution of these stem cell clones within bovine and murine mammary glands by detecting randomly inactivated X-linked marker genes.

1.2 Stem cell theory

1.2.1 Introduction

The existence of stem cells, defined as cells that remain in the tissue and retain the ability to proliferate throughout life, was first suggested in 1901 (Adami, 1901). The stem cell has since been further defined as: an undifferentiated cell type, able to divide both symmetrically and asymmetrically, which is responsible for cellular replacement in a tissue. Stem cells must be able to switch between these different roles throughout life (Potten, 1986; Potten *et al.*, 1997). Consensus opinion now considers that such cells represent a very small part of any tissue population and are likely to exercise their proliferative function through division of intermediate cells, the so-called transiently activated (TA) cells (Potten and Morris, 1988). Self-renewal and production of a more differentiated progeny TA cell (so-called asymmetrical division) is described in the 1976 Oakberg-Huckins model of stem cell behaviour (Figure 1.1), which was based on studies of renewal of tritiated thymidine ([^{3}H]dThd)-labelled spermatogonial stem cells (Oakburg and Huckins, 1976).

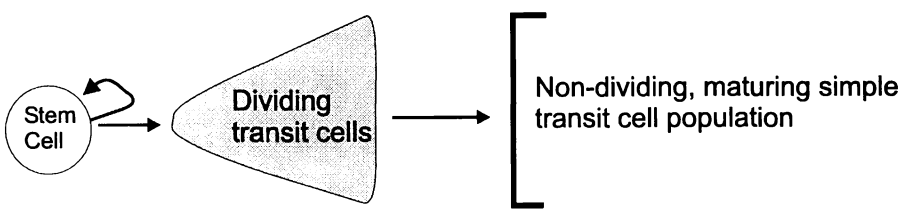


Figure 1.1 The Oakburg-Huckins model of stem cell growth. The stem cell divides to produce more terminally differentiated cells, the transit cells, although this is not at the expense of stem cell numbers as such cells are able to self-renew to preserve their numbers (adapted from Potten, 1986).

Gilbert and Lajtha (1965) proposed distinct stem cell growth behaviours for each of three broad tissue types identified, that is: 1) essentially nonreplacing – female germ-line, CNS and muscle; 2) nonrenewing but with regenerative capacity – liver, kidney, CT and apocrine glandular tissue; and 3) continually replacing tissues – haemopoietic, surface epithelia, the male germ line and holocrine glandular epithelia.

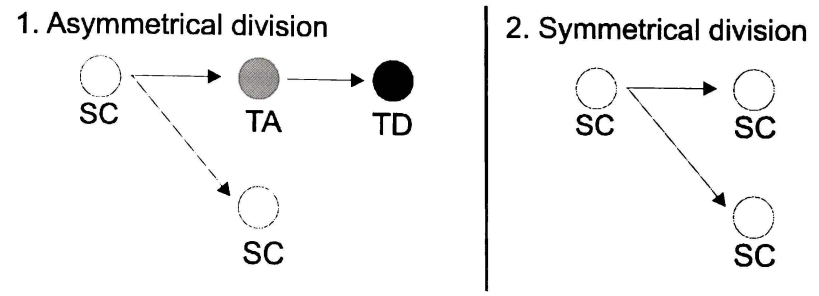


Figure 1.2 A summary of stem cell behaviour. SC = stem cell, TA = transiently activated cell, TD = terminally differentiated cell. Stem cells can divide 1) Asymmetrically to produce more differentiated progeny while self-renewing; or 2) Symmetrically to yield two parent-type stem cells. (Drawn from data in Potten (1986).

Division during normal growth is asymmetrical, which maintains numbers of proliferation capable cells. Stem cells are also capable of symmetrical division, particularly in response to disease or where massive accumulation of tissue is required quickly (Potten and Morris, 1988). For example, during mammary gland growth at pregnancy (see Figure 1.2, part 2). Therefore, stem cell division behaviour is dependent on tissue/niche location and developmental timing.

1.2.2 Stem cell hierarchies in Mouse gut: 'perpetual' regeneration

Few tissues have been subjected to rigorous study of the stem cell kinetics involved in their growth. An exception is mouse intestine, where cellular proliferation in the crypt-villus of the gut wall epithelia has been extensively studied (Potten, 1986). Such structures, and the ordered nature of the clusters or patches in which they form, provide models of self-renewing stem cells that produce a tissue 'subunit' through a series of increasingly differentiated daughter cells. The fact that each hierarchy is contained within a well defined space (the crypt), and that descendents of the hypothesized stem cells move out of the crypt in a well-organized fashion, facilitates observation throughout the depth of the crypt (Hermiston and Gordon, 1995).

The crypt-villus comprises a deep well-shaped pit, the crypt of Lieberkühn, with walls extending steeply to terminate at the apex of the villus. Studies of adult phenotypes have shown that an anulus of between 4 and 16 cells (the presumed stem cells), is located about four or five cells from the base of the crypt. In cellular kinetics studies, these cells have been found to exhibit slow cell cycle

times relative to other crypt cells (Potten *et al.*, 1997; Wright and Alison, 1984). Stem cell division results in a gradient of differentiation, with the most differentiated cells found at both the top and the bottom of the crypt. Cells moving down the axis differentiate into a paneth cell (involved in immunity and found at the base of the crypt) phenotype, while cells moving out of the crypt accumulate other markers of differentiation as they become increasingly terminally differentiated (Potten, 1986) (Figure 1.3). This stem cell activity produces four principal cell lineages, all of which differentiate to fulfil a terminal function as they move vertically up the crypt. In the mouse intestinal epithelium, they are eventually sloughed off by exfoliation at the tip of the villus (Hermiston and Gordon, 1995; Schmidt *et al.*, 1985).

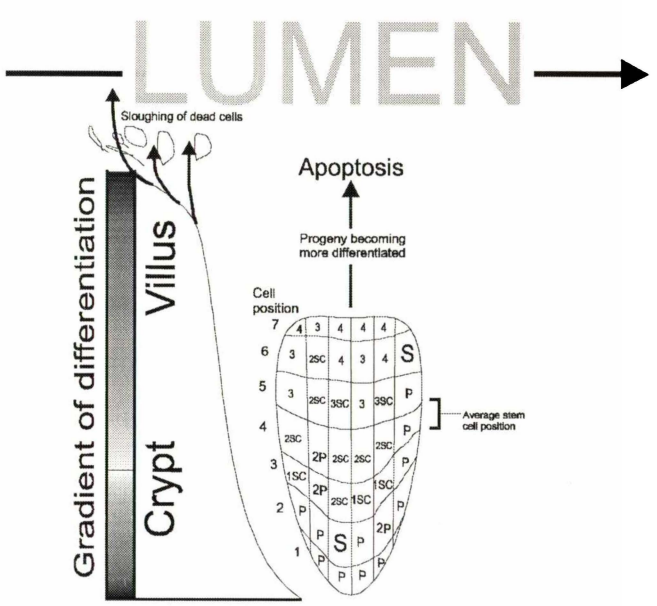


Figure 1.3. The mouse intestinal crypt-villus as a model of stem cell contribution to an organ structure. While it is thought that an anulus of stem cells (S) is located at an average depth of 4 cells from the base of the crypt, actual stem cells and their immediate progeny (1SC, 2SC, 3SC), that may have limited stem cell clonogenic capabilities, are found scattered throughout the lower and middle regions of the crypt. These cells produce the functional daughter cells, 1-4, that display increasing degrees of differentiation commitment as they enter the 'functional-zone', that is, the villus, as can be seen with differentiation markers applied to such populations. (Modified from Potten *et al.*, 1997).

In neonate chimeric animals crypts of mixed origin have been found, and this indicates polyclonality. However, this pattern is lost after a few weeks, presumably because of a "purificiation" by a predominance of progeny of a single anchored stem cell and its descendents in the crypt (Schmidt *et al.*, 1988).

Two theories of population and maintenance of the intestinal crypt by the stem cells have been suggested. In the first, the stem cell pedigree concept (SCPC), one master stem cell maintains the entire crypt through a population of daughter stem cells (Schmidt *et al.*, 1988). The second theory postulates that crypt population occurs via a small population of identical stem cells that arise simultaneously during gut morphogenesis (Hermiston and Gordon, 1995).

In support of the SCPC, stem cell organization within the crypt-villus has been suggested to result from a single primary progenitor that produces two lineages of progeny. The first, the enteroendocrine lineage, produces factors to stimulate the other lineage, the enterocyte lineage, by means of a enteric neural interface (Mills and Gordon, 2001).

1.2.3 Implications for studies of mammary clonality

As detailed above, mouse gut epithelial cellular populations have anatomical features that make them useful for study of epithelial cellular hierarchies/clonality in surface epithelia. However, as suggested by Gilbert and Lajtha (1965), glandular apocrine systems, because they do not require perpetual regeneration, are unlikely to show the exact same patterns of growth as seen in the mouse intestine.

Nevertheless, the production of an entire organ from a small number of progenitors is a presumed function of all stem cell populations in epithelial organs. Mills and Gordon (2001) imply that cellular growth patterns similar to that described in the crypt-villus system may exist in other organs, including the mammary gland, where single progenitors produce several different cell types (detailed below). Furthermore, it is thought that the crypt-villus model represents a standard model for the growth of all epithelial systems (Zajicek, 1995).

1.3 Organization and growth of mammary epithelia

The formation of the adult mammary gland, with ectodermal epithelia infiltrating a mesodermal fat pad to produce a ductal-alveolar 'tree', produces the cellular

structures to support lactogenesis. Several factors, including endocrine regulation, spatiotemporal control of cell division and death (apoptosis), regulated turnover of ECM components, cell-cell interactions, migration and differentiation, act to influence this development (Gumbiner, 1992).

The mammary gland is an extensively arborized epithelium with a fractal geometry (Guarini and Onofri, 1993) embedded within a fatty stroma. The adult gland comprises two cellular compartments, that is, the mesenchyme of fatty stroma invaded by blood vessels and nerves, and the milk-producing epithelial component, comprising ducts and lobules. A basement membrane of type IV collagen, laminin and glycosaminoglycans, separates these two tissue compartments. The epithelial compartment consists of three mature cell types, the myoepithelia, the ductal epithelia and the alveolar epithelia, all of which are distinguishable using epithelial specific cell markers (reviewed in Rudland, 1993).

The mammary ductal system has one or more layers of cuboidal epithelial cells, many of which border a lumen that is continuous throughout the gland. As noted above, these cuboidal cells are typically surrounded by a layer of elongate myoepithelial cells.

Classically, morphological features have been used to distinguish ductal and secretory epithelia from myoepithelia. The two mammary epithelia are distinguishable because ductal epithelia exhibits apical microvilli and specialized junctional complexes (as well as desmosomes) whereas myoepithelial cells typically exhibit smooth muscle-like myofilaments, pinocytotic vesicles and basement membrane on their basal surface (Rudland, 1987).

1.4 Mammogenesis – embryo to puberty

1.4.1 General features of rodent and ruminant embryonic mammary development

The seminal event in mouse mammary gland formation is lateral to midline ectoderm proliferation, resulting in the formation of a mammary band (or streak). In mouse embryos, this band is seen between d10 and d11 of the ~21 day murine

gestation (Sakakura *et al.*, 1987) whereas, in bovine, it is observed between d75 and d85 of a 280-day gestation (Ellis, 1998), that is, relatively earlier than in the mouse. Outgrowth of this band produces the mammary line that comprises a line of stratified cells representing three layers of epithelium - the surface epitrichium, the stratum intermedium and the malphigian layer or stratum germinatarium – as reviewed in Ellis (1998). After local thickening, the mammary 'hillock' and bud - mammae precursors - are observed. The mammary bud was originally thought to be the product of epithelial proliferation. However, its formation is now attributed to epithelial reorganization (Propper, 1978). Mammary bud formation coincides with mammary line degeneration.

The next step in development is proliferation of the mammary bud epithelia to form the primary sprout and teats. In ruminants, these result from the proliferation of the associated mesenchyme. Secondary sprouts soon emerge from the primary structure and, in rodents, both sprout structures (initially solid cords of cells) canalise soon after their formation to form open ductal structures. The mechanism of canalization is unclear, although in rodents it is assumed that cells at the centre of the cords either move outwards or die by apoptosis. This creates a hollow precursor ductal structure (Hogg *et al.*, 1983). At birth, the mouse gland consists of branching hollow tubes that open at the nipple.

Further embryonic mammary development is achieved by growth of epithelial elements into a primary undifferentiated embryonic mesodermal fat pad. Most of the fat pad precursors, primarily adipose tissue, fibroblasts, nerves and blood vessels, develop posterior to the mammary bud; however, compared to the other components, fibroblasts develop in close association with the mammary bud proper (Sakakura *et al.*, 1982).

1.4.2 Mammary development: birth to puberty

Isometric mammary gland growth, that is, proportional with increased animal size, occurs between birth and the onset of puberty at around 3 wks (reviewed in Sinha and Tucker, 1966).

1.4.3 Mammary development: puberty to involution

Late in puberty the immature gland grows rapidly to form a ductal system that extends to the borders of the fat pad. Between 3 and 12 weeks of age, this allometric growth is seen as intense mitotic activity at the mammary end buds, that have a specialized advancing edge for fatty stroma penetration (Silberstein and Daniel, 1982). Prolactin and progesterone are important inducers of lobular budding during this last stage of irreversible gland growth (Horseman, 1999) (Figure 1.4). This phase of growth is also dependent on the action of estrogen and growth hormone, probably through the action of insulin-like growth factor I (IGF-I), which acts to induce rapid ductal growth with extensive proliferation in the TEB – discussed below (Horseman, 1999).

During pregnancy the glandular epithelia undergoes extensive proliferation in response to prolactin, placental lactogens, progesterone and local growth factors (Horseman, 1999).

At parturition, levels of progesterone, estrogen, and placental lactogens, decrease sharply so that lactogenesis can proceed. Eventual involution occurs following milk stasis, in a signal transducer and activator of transcription (STAT)-regulated mechanism that involves a decrease in STAT5 activity and up-regulation of STAT3 expression (Philp *et al.*, 1996).

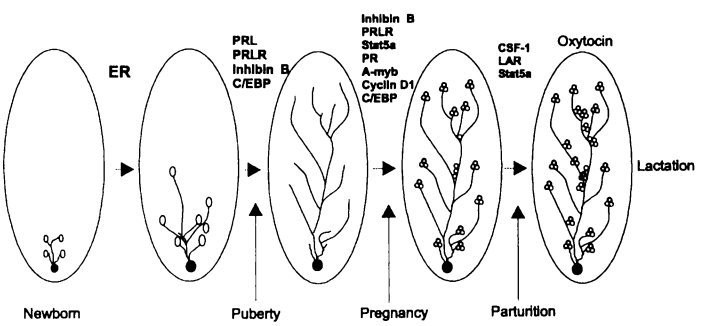


Figure 1.4. Diagram representing progressive steps in mouse mammary development under the influence of various hormones (receptor expression and transcription factor involvement is also indicated). ER – estrogen receptor, PRL – Prolactin, PRLR – prolactin receptor, inhibin ßB, C/EBPß, STAT5a – signal transducer and activator of transcription 5a, PR – progesterone receptor, A-myb, cyclin D1, CSF-1 – colony stimulating factor 1, LAR, Mf3 and oxytocin Involution, not pictured here, occurs after lactation ceases and involves upregulation of STAT3 and downregulation of STAT5 (redrawn from Hennighausen and Robinson, 1998).

The post-involution gland resembles the virgin gland morphologically. However, it has been shown to harbour a novel population of epithelial cells, located around the terminal end ducts, which survive apoptosis. This was discovered using measurements of longevity of cells expressing a Rosa-*lacZ* reporter construct activated by a second transgene (WAP-Cre) during lactogenesis II. These cells were also shown to be involved in glandular regrowth during subsequent pregnancy (Wagner *et al.*, 2002).

1.4.4 Human mammary development

Embryonic mammary development in humans is well characterized. Ectodermal thickening produces two lines (extending from the axilla to the groin) by the sixth week of embryonic development. Gland formation occurs by the development of more than 20 separate ectodermal cords that grow into the underlying mesoderm. At birth the gland consists of a primary lactiferous duct that branches into three to five secondary ducts that end in TEBs (reviewed in Sell and Pierce, 1994). Before the onset of puberty, proliferation of these simple branched ducts occurs. Monopodial (from the sides of existing ducts), and dichotomous (at the growing tips) branching at the termini of the primary ducts and of lateral end buds, also occurs around this time. This allows for the differentiation into small alveolar buds that undergo expansion at puberty to form terminal ductal lobular units (TDLU).

During pregnancy the TDLU differentiate rapidly to form alveolar lobules, some consisting of upwards of 200 ductules (Russo and Russo, 1987). Distension of these alveolar lobules produces the secretory alveoli seen to increase during the lactation phase. Post-lactational changes include alveolar involution but, overall, the "post-lactation" gland remains in a more differentiated state than that seen before pregnancy (reviewed in Rudland 1993).

Pockets of primitive epithelia exist within adult human glands. It has been suggested that these cells may represent either, virginal lobules that have remained refractory to endocrine stimulation or, reserves for the replacement of degenerating or exhausted glandular tissue (reviewed in Rudland 1993) as shown for similar cell populations in the mouse gland (Wagner *et al.*, 2002).

1.4.5 Ruminant mammary development

In cows, a 15-fold increase in the amount of DNA (that is, a cell number 'index') in the gland has been measured between 2-3 months of age, with the remaining prepubertal period characterized by linear growth (Sinha and Tucker, 1969). This is in contrast to the isometric growth seen in the rodent gland in the comparable period, that is, the bovine mammary gland grows allometrically (faster relative to general growth) between 3 and 9 months. However, between 9 months of age and the onset of pregnancy, the bovine mammary glandular epithelium is again observed to undergo isometric growth.

1.5 Comparing ruminant and mouse postnatal gland morphogenesis

Parenchymal morphogenesis in the ruminant mammary gland is thought to proceed in a distinct mechanism from that seen in the mouse gland (Hovey *et al.*, 1999) and this is particularly evident in histological observations (Ellis, 1998). Ruminant gland development is, however, similar to that of the human gland with ductal elongation by epithelial proliferation of terminal buds in TDLU (Russo and Russo, 1987).

There are significant species differences in the general structure of the mammary gland. Perhaps the most significant is that, while the virgin mouse has mainly primary and secondary ducts, a few end buds and few alveoli, the bovine mammary gland (and the mammary gland of mammals with a luteal phase in their estrous cycle) has a far more developed ductal and alveolar structure.

The mammary structure of newborn ruminants has a single primary duct extending from each teat. The base of this duct undergoes enlargement to produce a cavernous cistern while distal regions of the duct branch to form secondary and tertiary ducts. These secondary branches undergo further branching to form lobular clusters of ductules, similar to the TDLU seen in human glands (Hovey *et al.*, 1999). Hence, the bovine gland lacks TEBs; and ductal proliferation and advance is achieved via highly arborescent ductal structures embedded within a dense fibrous stroma (Ellis, 1998).

The mechanism of ductal lumen formation also differs greatly between ruminants and rodents. In the mouse, lumen formation occurs by apoptosis of anterior body cells within the TEB and is simultaneous with TEB penetration of the fat pad. Lumen formation in ruminants, where ductal apoptosis is rare (Molenaar *et al.*, 1996), is thought to occur by the coordinated and simultaneous separation of already formed (closed) ducts within an entire TDLU. The duct aperture size seems to be limited by constraints imposed by both neighbouring TDLU, that are undergoing similar separation, and by the surrounding dense connective tissue (Ellis, 1998). Expression of both Netrin-1 and its receptor, neogenin, has been shown to be important in the partitioning of mouse cap cells (discussed below) from the lumenal space, with elimination of either gene resulting in a disorganized TEB (Srinivasan *et al.*, 2003); the importance of this adhesion interaction in bovine lumen formation has not yet been examined.

Studies of ruminant mammogenesis, approached through examination of mouse models must, therefore, consider the fundamental, and sometimes dramatic, differences that exist between the two species.

1.5.1 Rodent and ruminant gland structural comparison

Different requirements for glandular support are thought to account for many of the anatomical differences observed in rodent and ruminant mammary glands. In ruminants, ductal growth is seen in "veins" throughout the network of fibroblasts within connective tissue (CT). These "veins" penetrate the adipose tissues of the mesenchyme, forming septa of rope-like cords. In contrast, ductal growth in the mouse occurs with the ducts directly abutting the adipose tissue. Hovey (1999) suggests that, in the ruminant, the extensive network of CT coupled with major ligaments provides structural support for the entire gland, a feature that is not required in the mouse, where the flattened glands lie directly adjacent to the underside of the animal. A table showing anatomical and developmental differences between the murine, bovine and human glands is presented below (Table 1.1).

Developmental stage	Rodent	Ruminant	Human
Embryonic	Mammary streak in embryo at d10-11. Around d20, fat pad is seen. A collection of hollow branching ductal tubes opening at the nipple	Mammary streak is first seen around d80. At birth a single primary duct extends from the teat	Development begins 7-8 wks post conception. Wks 12-16: groups of cells begin to branch, forming primitive milk ducts late in pregnancy
Neonate to puberty	Simple branched anlage of ductal components	Base of the primary duct enlarges to form a gland cistern while distal regions divide to form secondary and tertiary ducts. Becomes allometric immediately preceding puberty	Proliferation of simple ducts coupled with monopodial and dichotomous branching
Puberty	Between 3-12 wks allometric growth is seen. Ductal elongation continues to form sparsely branched ductal network	Clusters of ductules arising from larger ducts are formed. Pronounced lobule formation relative to rodent	3-4 years of growth form mature structure comprising 15-20 lobes (TDLU) that meet the nipple via milk ducts. Similar to structures seen in ruminants
Cycling virgins	Development continues and ducts fill the entire mammary fat pad	Allometric growth is maintained for the first few estrous cycles	A limited amount of ductal branching occurs
Pregnancy	Extensive branching and alveoli formation occurs to give secretory potential	Allometric growth returns	Terminal ductal regions differentiate rapidly to form alveolar lobules
Parturition	Terminal differentiation of ductal elements. Fully developed milk producing gland	Fully developed gland. Post-parturition allometry continues briefly	Alveolar lobule distension has produced secretory alveoli that allow for milk production
Postweaning	Involution, characterized by extensive apoptosis, returns the entire gland back to state of the mature virgin	Involution occurs. Gland is more differentiated than before pregnancy	Alveolar involution

Table 1.1. A table of comparisons between developmental and anatomical features of ruminant, human and rodent mammary glands. Sources: Tucker (1987), Sell and Pierce (1994), Hovey (1999).

1.6 Endocrine regulation of mammary growth – stromal contribution

The importance of the stroma in directing epithelial branching morphogenesis has been demonstrated by tissue transplantation experiments, in particular those involving transplantation of salivary and mammary epithelia (interchanged between their native stromal environments). It has been shown that, regardless of the origin of the epithelial transplant, growth matched that expected for the associated stromal type present. This clearly showed that stroma had a major controlling affect on epithelial branching and general duct growth (Sakakura *et al.*, 1976; Sakakura *et al.*, 1987). The mechanism whereby undifferentiated epidermal cells are directed to differentiate and form mammary epithelia is thought to involve hormonal signals in a mesenchymal-epithelial communication. This communication has been implicated as having a significant influence during fetal glandular development (Cunha, 1994; Cunha and Hom, 1996).

Significantly, the size of the fat pad or, specifically, the amount of adipose in the fat pad, determines the number of epithelial cells found in the gland and, therefore, has a direct correlation with the potential for milk production (Hovey *et al.*, 1999).

The hormonal milieu involved in directing mammary development is complicated. For normal fetal mammary development there is a requirement for the metabolic hormones (insulin, growth hormone, the thyroid hormones, and glucocorticoids) and others such as the insulin-like growth factors: IGF-I (Kleinburg, 1998) and IGF-II (Manni *et al.*, 1994), as well as fetal growth hormone (FGF) (Barraclough *et al.*, 1990). This highlights the importance of mesenchymal-ectodermal communication in normal gland formation and maintenance (discussed below).

In the mouse, estrogen encourages ductal growth and TEB formation from ducts while progesterone causes the differentiation of alveolar structures. Implant experiments have shown that estrogen acts, both locally and systemically, to regulate epithelial growth via stromal epithelial interactions and/or growth factor production (Haslam, 1988). Associated experiments have shown that estrogen-responsiveness required the presence of stromal cells (Haslam, 1988) and that the

majority of TEBs and ductal epithelial cells lacked estrogen receptors (Zeps *et al.*, 1998). Locally, the basal 'B cells', which are considered to be mammary stem cells, in the resting gland differentiate into 'Clear' A cells under the influence of estrogen (Chepko and Smith, 1998).

Negative mediation of ductal growth is thought to occur via the action of a number of factors, including transforming growth factor-β isoforms (TGF-ß). Experiments using slow release implants releasing TGF-β_1 showed inhibition of ductal growth in a murine model – observed as reduced thymidine labelling (Silberstein and Daniel, 1987). A similar inhibitory affect has been observed in experiments in ruminants (Woodward *et al.*, 1995).

A model describing the contribution of various stromal, ovarian and pituitary-derived hormones to the progression of epithelial invasion of the fat pad has been presented by Hovey (1999) (see Figure 1.5).

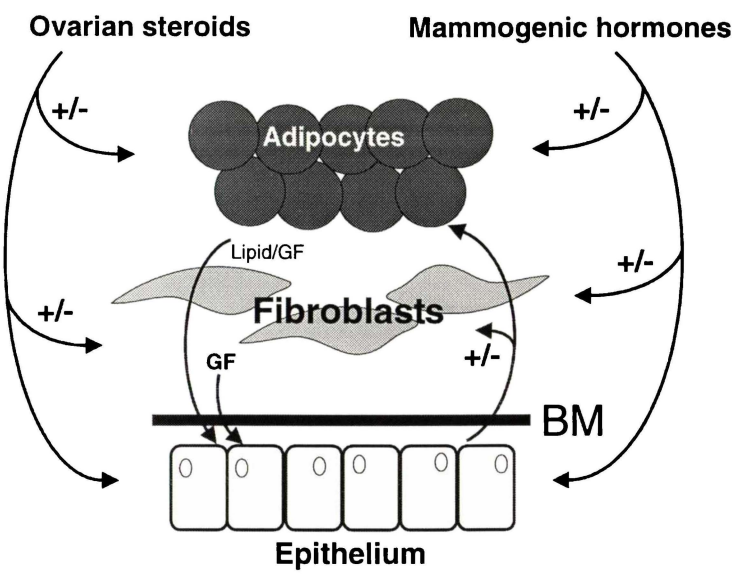


Figure 1.5. A schematic showing hormonal control of mammary epithelial growth. Three levels of regulation, from the pituitary, ovaries and the stroma itself are implicated in the regulation of mammary epithelial growth. GF – growth factors, BM – basement membrane (redrawn from Hovey *et al.*, 1999).

In summary, hypophyseal and ovarian hormones, coupled with local stromal hormones, interact in murine mammogenesis. Similar endocrine regulation is thought to be responsible for ruminant mammogenesis (Ellis, 1998). Studies of branching morphogenesis of the mammary cell line 1-7 HB2 have demonstrated

the importance of the fibroblast-derived morphogenic growth factors, including hepatocyte growth factor/scatter factor HSF/SF and various other fibroblast derived growth factors (Berdichevsky *et al.*, 1994), in this process. Furthermore, a role for matrix metalloproteinases (MMPs) has been demonstrated in collagen gel culture studies, where inhibitors of MMPs eliminated branching morphogenesis but did not directly affect epithelial proliferation (Simian *et al.*, 2001).

1.7 The Terminal End Bud (TEB)

The TEB is the generator of new epithelial tissue for mammary ductal formation and progression. The tip, flank and neck regions of the TEB are covered in an epithelial monolayer of 'cap' cells. These are stem cells that characteristically lack specialized features (such as intercellular junctions). Ductal penetration and the formation of a myoepithelial 'sleeve' around the duct, is effected by apical cap cell proliferation, while ductal elongation results from the formation of ductal epithelia achieved by the proliferation of 'body cells' within the luminal region of the TEB beneath the cap cell zone (Williams and Daniel, 1983). TEB structures are functional until the end of puberty, when the ductal network approaches the edges of the fat pad, at which time they are permanently replaced by terminal end ducts (TED) and alveolar buds (Topper and Freeman, 1980; Williams and Daniel, 1983).

The stem cell characteristics of cap cells has been demonstrated by histochemical and immunohistochemical staining that shows weak staining for markers of differentiation in cap cells and in a subset of body cells. Specific immunocytochemical staining for epithelia, using monoclonal antibodies (MAbs) to MFGM, anti-EMA and peanut lectin (PNL), produces a graded staining pattern with the loosely-packed apical peripheral cells showing weak staining that becomes stronger closer to the centre of the TEB. Myoepithelia specific markers, including antibodies to keratin, PKK2 and LP34, smooth muscle actin, myosin and vimentin (Rudland, 1991), stain the apical cells similarly to the epithelial markers described above. Staining becomes more intense in these peripheral cells more posterior to the apex of the TEB, and finally shows maximal intensity at the myoepithelia of the subtending duct of the TEB (Rudland, 1991). The cap cells at

the apex of the TEB stain positive, although weakly, for both myoepithelial and epithelial specific markers (Figure 1.6).

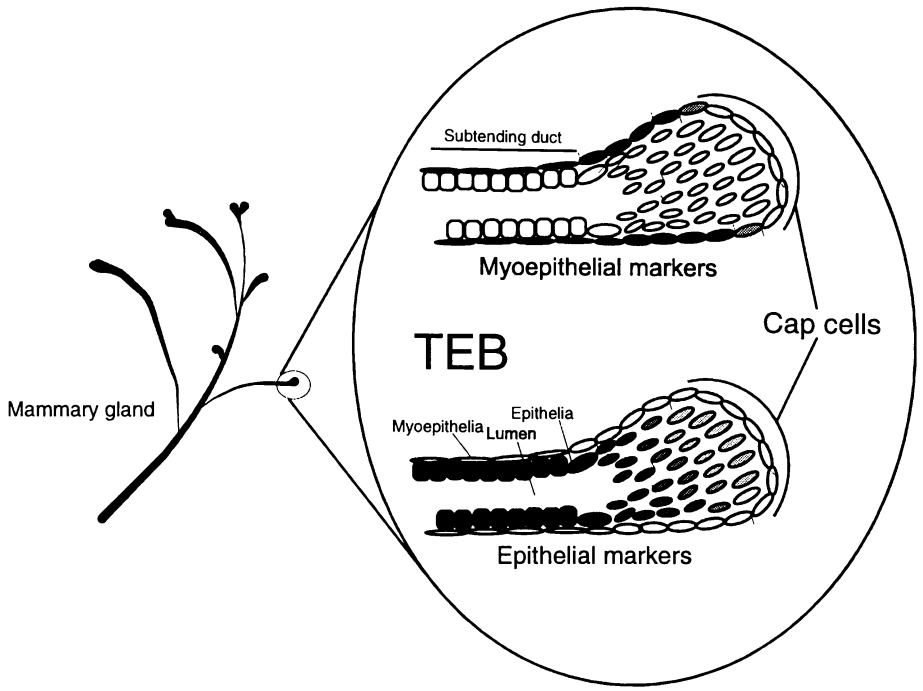


Figure 1.6. Schematic of a TEB in mouse mammary showing specific staining for myoepithelia and epithelia. Epithelial specific markers (MAbs to MFGM, anti-EMA and PNL) and myoepithelial specific markers (histochemical staining of Smooth muscle actin and BM proteins) (Rudland, 1993) reveal the relative location of different epithelia formed in the advancing duct by intensive mitosis of the cap cells. Although these apical peripheral cap cells stain weakly to moderately for both sets of markers, the epithelial specific markers stain the cortical cells most intensely while the myoepithelial markers stain peripheral cells with a gradual increasing intensity towards the subtending ductal myoepithelia (Lochter, 1998). (Drawn from sources in Lochter, 1998; Rudland, 1993).

Primary culture assays, involving immunomagnetic separation of epithelial and myoepithelial lineages, showed that while pure cultures of myoepithelial cells did not interconvert to ductal/secretory type epithelia, a portion of these cells did differentiate into myoepithelia cell types (Pechoux *et al.*, 1999). This demonstrates that the two lineages share a common progenitor stem cell, and that the epithelial type is more plastic than the myoepithelial type, presumably because myoepithelia represents a more terminally differentiated cell type.

Direct evidence for the existence of mammary stem cells in the TEB includes chemical carcinogen studies that revealed a correlation between the presence of TEB and predisposition to cancer in the mammary gland. The effective carcinogenicity of chemical carcinogens (for example, 7, 12-

dimethylbenz[α]anthracene [DMBA] and N-nitrosomethylurea) in rodents decreases 50 days postpuberty. In humans this developmental stage coincides with a natural decrease in TEBs (as they differentiate into alveolar bundles and terminal ducts) in the gland (Russo and Russo, 1978). McGregor (1977) found that prepubertal and adolescent women, ages characterized by high amounts of breast ductal formation, were the group most susceptible to radiation-induced breast carcinoma in Hiroshima and Nagasaki.

1.7.1 Apoptosis in the TEB

Apoptosis plays a critical role in ductal growth phase. In a study of apoptosis within the TEB during growth, Humphreys (1996) demonstrated that while only 7.9% of the rapidly dividing proliferative cap cells were undergoing apoptotic processes, up to 15% of cells in the three layers of body cells immediately adjacent to the lumen were apoptotic. Therefore, levels of apoptosis during development are higher than that observed during either involution or pregnancy, with 4% and 5% of cells in the TEB affected in each state, respectively (Li *et al.*, 1996; Quarrie *et al.*, 1995). Moreover, other developing tissues typically display far lower levels of apoptosis than that observed in the TEB; kidney displays 3% (Coles *et al.*, 1993), and optic nerve 0.25% (Barres *et al.*, 1992) apoptotic cells during normal development. Humphreys (1996) states that the significance of these relative levels of involvement must be viewed against the normally low background levels of apoptosis in quiescent cells (of only 0.1 to 1%).

In embryogenesis, rodent gastrulation is thought to be regulated by an apoptotic signal from the endoderm that communicates with cells within the developing embryo to cause cavitation. Cells that are anchored to the basement membrane escape the apoptotic response to this signal (Coucouvanis and Martin, 1995). Considering this work, Humphreys (1996) suggests that positional sensing, relative to both the TEB and to its environment, is critical to apoptotic signalling within the TEB and that tubular morphogenesis may occur by a mechanism similar to that seen in the early embryo.

Patterns of apoptosis in the ruminant mammary gland are thought to be substantially different from those in the rodent mammary gland. Molenaar (1996) reported low numbers of apoptotic cells in virgin ewe mammary glands and

emphasised that the observed numbers of apoptotic cells were much lower than that seen in comparable murine glands. Ellis (1998) suggests that this observation may reflect interspecies differences in lumen formation mechanics, as outlined in the gland comparison above (Table 1.1).

In summary, the terminal end bud is a dynamic structure via which ductal extension occurs by coordinated processes of proliferation and apoptosis.

1.8 Ductal branching of parenchyma

Branching parenchymal morphogenesis in the mouse has been divided into three distinct processes. First, the entire gland is populated by increased linear ductal length. Next, dichotomous branching occurs to expand the parenchymal area. And, finally, the growth of collateral duct segments fills any remaining areas (Ormerod and Rudland, 1984). Ellis (1998) stressed the importance of massive amounts of epithelial proliferation to provide the 'building blocks' for such growth, and this behaviour is examined below in a review of mammary stem cell biology.

1.9 A model of mammary development by epithelial outgrowth – stem cell candidates

1.9.1 Stem cells and division competence in mammary development

As in most mammalian tissues, a fine balance of cell death and proliferation in the mammary epithelial population is crucial for normal growth. Studies of the human gland showed that multiple rounds of oestrus do not lead to a huge accumulation of mammary tissue (Ferguson and Anderson, 1981). This implies that the activity of mammary stem cells, that is, those cells capable of responding to critical developmental timepoints by producing new epithelial cells, is balanced by apoptosis. Indeed, it is known that apoptosis removes much of the excess epithelial population produced during menses (Potten and Loeffler, 1990).

Triated thymidine ([^{3}H]dThd) labelling of virgin female rats (by intra-peritoneal injection), allows detection of actively dividing (DNA synthesizing) cells and has identified a population of cells (the cap cells) in mammary terminal end buds that produce separate lineages for luminal and possibly myoepithelial cells (Dulbecco *et al.*, 1982). This implies that stem cells are responsible for ductal elongation into the fat pad. The ubiquity of these stem cells in the gland was indicated by transplantation studies where it was shown that cells taken from any part of the mammary gland could repopulate the epithelia–free fat pad (Smith and Medina 1988). This represents clear evidence for the existence of a ubiquitous population of 'resting' cells that have the potential to act as stem cells within the mammary epithelial cell population.

Smith and Medina (1988a) have characterized the 'repopulating ability' of mammary epithelial components using a combination of autoradiographic, immunohistochemical and various morphological criteria, in conjunction with transplantation experiments involving different portions of the murine mammary gland. Mammary stem cells were defined as 'uncommitted mammary cells' that have the potential to divide by mitosis to produce committed cell progeny without the total loss of their own uncommitted status. The work showed that indigenous mammary epithelia become division-defunct after a finite number of cell divisions. The same authors found 'pale staining' mammary cells that: 1) were present at all stages of development, 2) divided soon after being placed in culture, 3) were connected by desmosomes and hemidesmosomes to their neighbouring cells. This was important as it indicated that they were not simply "wandering lymphocytes", and 4) produced daughter cells that functionally differentiate in the presence of lactogenic hormones. Therefore, the cells displayed the characteristics expected of stem cells (Smith and Medina, 1988a).

Christov (1993) used bromodeoxyuridine (BrdU) labelling in virgin BALB/C mice to identify three division-competent cells in each gland that contributed to epithelial growth. Smith (1996) observed that, in the progeny of a pool of 5 X 10^3 transplanted β-galactosidase expressing mammary epithelial cells, lobular, ductal and combined lobulo-ductal populations could be observed in an epithelial-free fat pad. These experiments suggest that, while some division-competent mammary

stem cells are committed to ductal growth and some to lobular growth, a third cell type, which is able to produce both ductal and lobular progeny, may exist.

A more precise identification of these stem cells was carried out by Christov (1993) who used the presence of condensed chromosomes, as markers of division competence, to establish a model of epithelial cell hierarchies in the murine mammary gland. This model describes three division-competent epithelial cell populations and states that the Small Light Cell (SLC) – synonymous with the cap cell, the suspected mammary stem cell - differentiates into the terminally differentiated mammary cells (both secretory and myoepithelial) after passing through two precursor stages. SLC were seen to lose their division competence as they acquired signs of differentiation (such as lipid droplets and desmosomes). These SLC represent about 3% of the total mammary epithelial population and are characterized by naked chromosomes, that is, chromosomes that lack extensive transcriptional machinery. These cells do not exhibit secretory potential due, in part, to absence of luminal contact (Chepko and Smith, 1997).

According to the proposed model, a subset of the Undifferentiated Large Light Cells (ULLC) secondary progenitor cells eventually differentiate into Differentiated Large Light Cells (DLLC) and finally to functional secretory epithelia while the remaining ULLC differentiate into pre-myoepithelial cells for ultimate terminal differentiation into functional myoepithelia (Figure 1.7).

The proposed model for the production of functional differentiated cells in the mouse is supplemented by results obtained in [^{3}H]dThd studies of human breast tissue. Such studies indicate that basal clear cells (the human SLC analogs) divide frequently, and that the myoepithelia originates from a precursor within the mammary epithelium after a number of cellular divisions of the precursor cell (Joshi *et al.*, 1986).

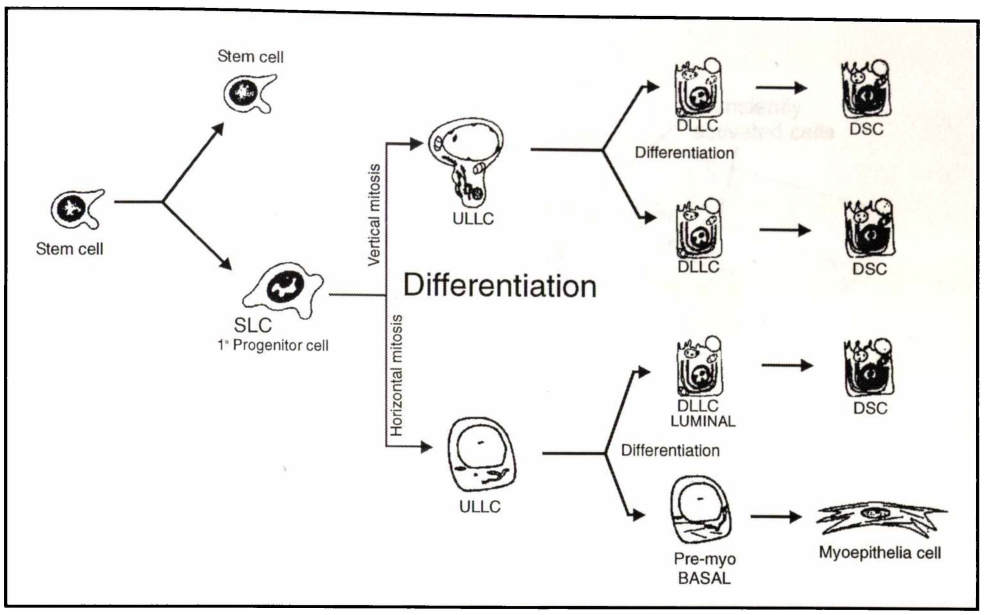


Figure 1.7. Schematic of suggested hierarchies of division competent stem cell to functionally differentiated cell in mammary tissue. SLC – small light cells, ULLC – undifferentiated large light cell, DLLC – differentiated large light cell. DSC – differentiated secretory cell (Adapted from Chepko and Smith, 1997).

This work is confirmed by other [^{3}H]dThd studies. Zeps (1996) found that, while 50% of mouse mammary cells were labelled after three injections, most of this signal was lost after two weeks (by halving kinetics) but a small portion of cells (varying regionally between 1 in 50 and 1 in 1000 cells) retained a [^{3}H]dThd signal. The authors suggest that this population of cells represents the slowly dividing functional mammary precursors that were first described by Potten and Loeffler (1990).

The primary progenitor (SLC) cells, proposed in the Chepko and Smith model, were shown to be capable of multiple cellular divisions before becoming committed to a particular differentiation pathway. These division-competent (and differentiation-committed) primary progenitor SLC divide by cytokinesis (perpendicular or parallel to the basement membrane) to produce secondary progenitor ULLC. Typically, perpendicular division is symmetrical and produces two secretion committed ULLC (or two SLC). In contrast, parallel division is asymmetrical and produces both a pre-myoepithelial cell and a progenitor ULLC. The outcome of these divisions seems to hinge on the orientation of the cleavage furrow and the timing of division (Figure 1.8).

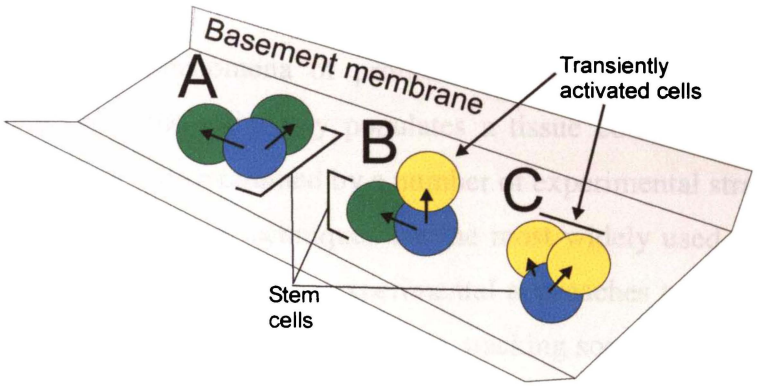


Figure 1.8. Possible stem cell division strategies in mammary organogenesis – predicted by the model described by Chepko and Smith (1997). **A)** Symmetric division of a stem cell (SC; blue) occurs to produce two identical SC (green) in the population of a tissue compartment, **B)** Asymmetric division of a SC to give a more terminally differentiated daughter cell or TA cell (yellow) and another SC, and finally **C)** symmetrical division to give two TA cells with exhaustion of stem cell potential. Each of these division strategies are thought to occur during mammary organogenesis.

Cells resembling SLCs and LLCs have been described in human (Ferguson, 1985), bovine (Nickerson *et al.*, 1982), and in ovine mammary tissue (Ellis *et al.*, 1995). The fact that these undifferentiated cells are found in a number of species lends support to the notion that they possibly exist as stem cells in the mammary gland.

Recently, three distinct progenitor populations - luminal-restricted, myoepithelial-restricted and bipotent - have been identified in human mammary tissue (Stingl *et al.*, 2001). The authors used single cell culture system to show that a portion of cells in the human mammary gland population were bipotent and capable of generating each of the other lineages (luminal and myoepithelia) in culture. Furthermore, they proposed that each of the cell types observed in the human population may be directly analogous to the cells observed in the mouse mammary, as has been proposed previously (Kordon and Smith, 1998; Smith and Medina, 1988a).

1.10 Clonality studies

Cellular clonality, the phenomena of proliferation by a subset of tissue-type determined stem cells that ultimately populates a tissue compartment within a mammalian organ, has been examined by a number of experimental strategies. Inactive X chromosome based techniques are the most widely used methods for the study of cellular clonality. Other experimental approaches to the assessment of clonal structures include cytogenetic analysis, tracking somatic mutants *in vivo*, and retroviral integration (reviewed in Wainscoat and Fey, 1990). X chromosome inactivation based assays have aso been used in tumour clonality studies where they are used to distinguish clonal tumour tissue from normal (presumed polyclonal) tissue in adjacent tissues (Busque *et al.*, 1996).

1.10.1 Types of clonality assays

There are two primary methods employed to observe clonality in the tissues of female eutherian mammals, that is: 1) examination of pre-existing variation (for example differences between alleles) on the X chromosome, and 2) examination of introduced variation. The latter includes the observation of tissues within chimeric and transgenic animals, as well as animals manipulated by retroviral-mediated introduction of cellular markers.

1.10.1.1 Protein histochemistry method

The protein or histochemical method, developed by Fialkow (1973), is used *in vitro* (gel-based) analysis of glucose-6-phosphate dehydrogenase (G6PD) isozymes to distinguish particular patterns of X chromosome inactivation throughout a tissue. While a simple technique, because of the infrequency of the G6PD variant - with only 1% of the Caucasian and 30% of the African population informative (Beutler *et al.*, 1967) – its application is limited to just a small subset of the general population.

1.10.1.2 DNA and RNA based methods

DNA and RNA based methods rely on naturally occurring allelic differences in X-linked genes. Vogelstein (1985), devised an RFLP strategy based on the detection of a variable number tandem repeat (VNTR) in the hypoxanthine

phosphoribosyltransferase (*HPRT*) gene for the assessment of clonality in human tumours. This technique makes use of the fact that the inactive X chromosome is modified by methylation whereas the active chromosome remains unmethylated. By using methylation-sensitive restriction endonucleases, the paternal and maternal chromosomes can be distinguished.

Since their first description by Vogelstein (1985), polymorphisms within the human androgen receptor (*HUMARA*) gene have been widely used as cellular clonality markers. This gene exhibits an 11-31 unit VNTR. In order to assess clonality using this loci, regions of interest are first micro-dissected from tissue sections on glass microscope slides and the DNA extracted and digested using a methylation-sensitive restriction enzyme (usually *Hpa* II). Next, the region containing the VNTR is amplified by PCR and the products electrophoresed by polyacrylamide gel electrophoresis (PAGE). A densitometric analysis of the bands is performed and the relative amounts of each allele determined (Figure 1.9). This assay is widely used because the majority (86%) of individuals are heterozygous for the VNTR (Gale *et al.*, 1996).

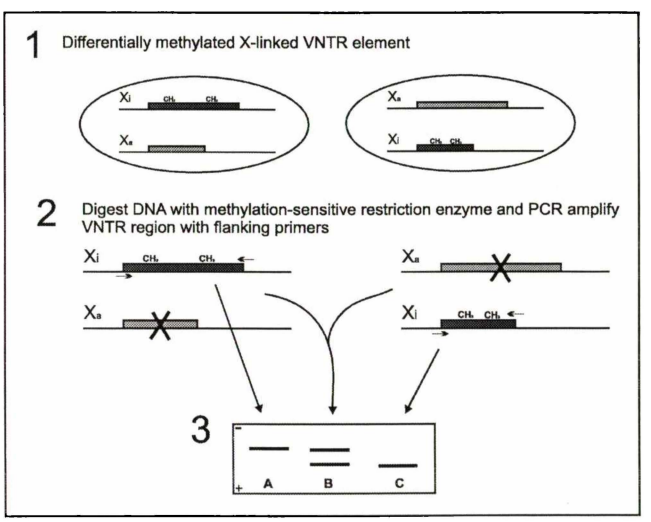


Figure 1.9. Showing cellular clonality measurement by discrimination of an X-linked Variable Number Tandem Repeat (VNTR). 1) Different choice of methylated inactive (X_i) and unmethylated active (X_a) X chromosomes in two different cells within the same tissue, 2) X_a is unmethylated and, therefore, digested with a methylation-sensitive restriction enzyme (such as *Hpa* II). PCR is carried out to amplify the intact Xi VNTR region which, because it is methylated, remains undigested. 3) The resultant PCR products are electrophoresed on an acrylamide gel in order to discriminate the products by size. Monoclonality is evident when one band is seen while polyclonality is assumed when two bands are observed in a single lane.

The X-linked phosphoglycerate kinase (*pgk*) gene has often been used in X-inactivation assays. PGK has 8 *Hpa* II sites located close to a *Bst* XI restriction

enzyme polymorphism (PM) at the 5′ of the gene. Assays based on the *pgk* gene allow for the analysis of very small (≈100) populations of cells (Gilliland *et al.*, 1991).

1.10.1.4 Retroviral incorporation

Engelhardt (1995) used retroviral incorporation as a marker of cell lineages involved in sub-mucosal salivary gland development (in constructs comprising *lacZ,* or alkaline phosphatase, under the control of cytomegalovirus [CMV] enhancer). These methods allowed the resolution of X-inactivation profiles and, therefore, clonality within a transfected tissue. Such studies have limited informative regions in a tissue due to the small areas infected by an introduced retrovirus.

1.10.1.5 General chimeric animal studies

Manipulation of mouse blastocysts can also yield useful animal models within which clonality can be examined. Specifically, the examination of a combination of two lineages of cells throughout a chimeric animal provides investigators with a tool by which to examine the clonal profile of a tissue. A requirement for such studies is that cells derived from each parental lineage can be distinguished by histochemical techniques.

By utilizing mice that are chimeric for electrophoretic variants of NADP-dependent isocitrate dehydrogenase, the developmental history of skeletal muscle myotubes was shown to be via cell fusion rather than by acytokinesis (Mintz and Baker, 1967). Furthermore, coat colour markers have been used to show that skin epithelia has its origin in 64 randomly arranged progenitor cells which 'stream' from the neural crest to populate the nascent skin epithelial zone (Wolpert and Gingell, 1970).

The timing of paternal chromosome inactivation in extraembryonic tissue has been studied using an elegant assay with chimeric mice produced by the injection into blastocysts (of a sandy coat coloured strain) of cells exhibiting an X-linked translocation (Cattanach's translocation) 3 or 4 days post-coitum. An

examination of the resulting coat colour patterns revealed that X-inactivation had not taken place by embryonic d4 (Gardner and Lyon, 1971).

The experiments described above show the utility of chimeric animals in clonality assays. Similarly, one would expect transgenic animals that are heterozygous for an X-linked marker gene (and accordingly displaying mosaic patterns of expression throughout a model animal) to be equally informative.

1.10.1.6 Limitations of existing techniques for clonality assessment

X chromosome based clonality assessment takes advantage of the inactive-X choice mosaicism throughout an animal. As the choice of inactive X chromosome is random and early (occurring in the committed stem cell), and because daughter cells from a single stem cell maintain the same inactive X chromosome as their parent cell, different populations (clones) of cells in female animals have either X chromosome inactivated, that is, are mosaic for chromosome choice. This random choosing results in different clones within a tissue that each exhibit a particular inactive X chromosome.

The single most important requirement in an X-based clonality assay is that each parental chromosome, that is, paternally-derived X (X_p) and maternally-derived X (X_m), can be distinguished. As detailed above (section 1.10.1), this can be achieved either through examination of the methylation (inactivation) status of variable length loci on the chromosome, by distinguishing differential expression of variable regions (at the transcript level), or by examining isozymes of proteins produced by the expression of genes containing polymorphisms.

In the STR (or VNTR) clonality assessment method (described in section 1.10.1.2), monoclonality for either allele is assumed when a single product is observed after electrophoresis and polyclonality when double bands (that is, both allele amplification products) are seen (Figure 1.9). A prerequisite for such experiments is that both parental alleles have been genotyped and that the animal under study is heterozygous for the polymorphism examined.

This technique, while elegant in its simplicity, has several failings. The most important is the potential for sampling error caused by the inclusion of non-target cells in the tissue isolated for DNA extraction. In the literature, such introduced error is seen in contrasting assessments of clonality of human mammary tissue, such as that detailed in the human mammary clonality studies below, where measurements of the clonality of the same tissue type give different clonal profiles.

Another failing of existing approaches is that such techniques are exclusively *in vitro* and, therefore, provide little direct anatomical framework by which to determine the significance of clonal regions within a tissue. Accordingly, Garcia *et al.* (2000) has identified *in situ* assessment as being a crucial 'next step' required for further understanding of tumour cellular clonality.

For the present study a novel alternative approach, that examines an X-linked single nucleotide polymorphism (SNP) in heterozygous animals as a means of discriminating allelic differences, and therefore of determining clonality within tissue niche, was designed and applied. As detailed in Chapter 3, the bovine experiments used a polymorphism of the bovine X-inactive Specific Transcript (*XIST*). Accordingly, details of *XIST*, its expression and transcriptional regulation, and its involvement in establishing the inactive X chromosome are discussed below.

1.11 X-inactive Specific Transcript (*XIST*)

1.11.1 Inactivation of an X chromosome as a mechanism for dosage compensation

In observations of mature nerve cells in female cats, Barr (1949) observed a well developed 'nucleolar satellite' that was almost always absent in male animals. This morphological marker was found to be a reliable indicator of sex which, the author determined, could be readily ascertained from observation of any suitably stained tissue section, where the condensed chromosomal body (Barr body) was seen on the inner cell membrane.

In her seminal work, Lyon (1961) proposed that random inactivation of one X chromosome in the stem cells of female mammals is responsible for dosage compensation of X-linked genes between the sexes. This inactivation effect would produce a mosaicism of inactive X chromosome choice throughout an animal.

According to the Lyon hypothesis, X inactivation involves random inactivation of one X chromosome in the stem cell, and this inactivation decision is maintained throughout a stem cell clone (that is, a stem cell and its daughter cells). This effect, known as Lyonization, produces a heterochromatinized, hypoacetylated, hypermethylated and condensed chromosome that is almost exclusively transcriptionally inactive (Lyon, 1961, 1992).

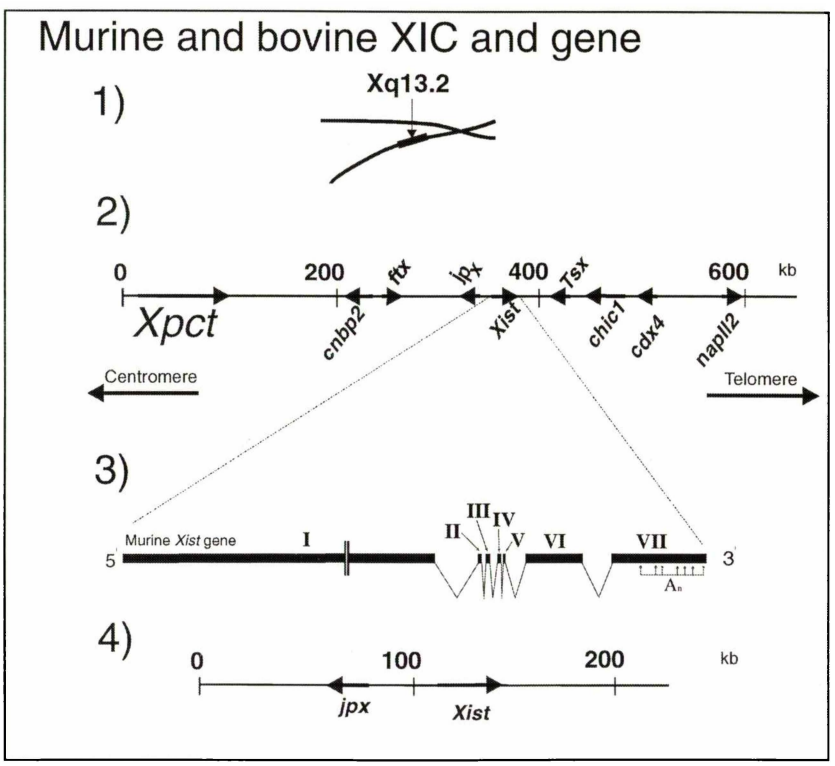


Figure 1.10. A diagram showing **1)** the approximate location of the *Xist* gene on the mouse X chromosome, **2)** a graphic of the murine XIC and **3)** a schematic representation of the *Xist* gene itself. The gene was initially thought to include 6 exons encoding a 15 kb transcript (Brockdorff *et al.*, 1992) but was described in 1999 as comprising 7 exons (I-VII) with 7 possible polyadenylation signals (A_n) having a transcription product of 17.4 kb (Hong *et al.*, 1999). An eighth exon in the mouse has also been putatively identified but was not included in the diagram due to lack of experimental proof of its existence (Chureau *et al.*, 2002). Finally, **4)** the recently sequenced bovine XIC, a 233 kb Bacterial Artificial Chromosome (BAC), including the two genes discovered in this region (*Xist* and *Jpx*) is included for comparison. (This figure was drawn from the three sources referenced within the caption: Brockdorff *et al.*, 1992; Chureau *et al.*, 2002; Hong *et al.*, 1999).

Such chromosome-wide inactivation is mediated by the X-inactive Specific Transcript (*Xist*) (Figure 1.10). During X-inactivation, *Xist* is expressed only from the chromosome destined for inactivation, and the gene product (RNA) coats the same chromosome in a cis-acting manner that contributes to the silencing of X chromosome gene expression.

1.11.2 Timing and nature of X-inactivation

X inactivation timing studies have measured the relative amounts of X-linked gene products produced at different embryonic stages in male and female heterozygous mice. Kratzer and Gartler (1978) used mouse litters (with an approximate male to female ratio of 1:1) to show bimodality of hypoxanthine guanine phosphoribosyltransferase (*HGPRT*) activity until the eight cell morula stage. This was followed by X inactivation, indicated by a gradual movement to unimodal activity (that is, single gene expression), occurring at blastulation.

Tissue-specific differences in the onset of X-inactivation are known to occur in mouse extraembryonic tissues. Observations of female mouse embryos that are heterozygous for Cattanach's translocation (In7;X) indicated that mouse extraembryonic tissues show an inactivation bias for the paternal X chromosome (Takagi and Sasaki, 1975). Experiments with mouse embryos lacking paternal X chromosome showed that one of the two 'maternal-type' X chromosomes was inactivated in extraembryonic tissues (Rastan *et al.* 1980). This suggests that while the choice mechanism of an inactive chromosome in this tissue normally selects the paternal chromosome for inactivation, the presence of this chromosome is not an absolute requirement for successful inactivation and chromosome counting.

Xist expression has been detected at very early stages in embryonic development. Avner (2000) used RT-PCR to show the presence of *Xist* mRNA at day 4 of embryonic development. This correlates with non-random X-inactivation in extraembryonic tissues, that is, preferential extraembryonic paternal X inactivation, which is completed by 3.5–4.5 days post conception (dpc). The same authors also showed that *Xist* expression was maternally imprinted until the blastocyst stage, at which time expression became random (Avner *et al.*, 2000).

1.11.3 Levels of *XIST/Xist* transcript

Quantitative RT-PCR has shown that, in the kidney of both the adult mouse and the 7.5 dpc mouse embryo, there are less than 2 X 10^3 *Xist* transcripts per cell (Buzin *et al.*, 1994). Such low expression levels are intriguing as they indicate that the *Xist* transcript is not expressed in sufficient quantities to coat the entire X chromosome (the mouse X chromosome is approximately 1.5 X 10^8 bp so at least 7000 15 kb transcript copies would be required for complete linear coverage). This low copy number suggested that *Xist* expression may be the consequence of X chromosome inactivation rather than being an integral part of its causative mechanism (Brockdorff *et al.*, 1992; Brown *et al.*, 1992). Subsequently, however, this theory has been disproved by experiments with *Xist* knock-out animals that demonstrated that *Xist* is crucial to normal X chromosome inactivation (Marahrens *et al.*, 1997; Penny *et al.*, 1996).

1.11.4 *XIST/Xist* mode of action

The *Xist* transcript associates with and coats the inactive X chromosome, and it has been suspected that the transcript may interact via certain consensus sequences along its length (Clemson *et al.*, 1996). There are several features of the *Xist* transcript that suggest functional domains within its sequence. Both the human and mouse *XIST/Xist* cDNAs are long (19.3 and 17.8kb respectively, Hong *et al.*, 1999; Hong *et al.*, 2000) and no evidence for conserved open reading frames (ORF) within the gene has been found (human - Brown *et al.*, 1992; mouse - Brockdorff *et al.*, 1992). Furthermore, peptide nucleic acid (PNA) studies that eliminated normal *Xist* transcript function in mouse cell culture, have demonstrated that the target *Xist* transcript contains sites whose disruption results in loss of X-inactivation. While having no effect on steady state levels of *Xist* in treated cells, the addition of one particular PNA (as a PNA-transportan conjugate) resulted in the complete loss of X inactivation. The binding site of the effective PNA-transportan was found in 'region C' (antisense PNA sequence: aaattccatgactctagaa), the third of four repetitive stretches corresponding to a region at the 5′ end of exon 1 of the gene (Beletskii *et al.*, 2001). Binding of the PNA-transportan within region C was shown to eliminate histone mH2A association with the inactive X chromosome, and caused the loss of localized *Xist*

RNA and the loss of an observable Barr body. This study was the first to demonstrate that the *Xist* transcript is organized into functional domains.

Immunoprecipitation experiments have shown that the *Xist* transcript associates specifically with hypoacetylated chromatin, often co-localizing with the hypoacetylated histone H4 at the promoters of X-inactivated genes. This suggests that the transcript acts directly in silencing X-linked genes (Gilbert *et al.*, 2000).

1.11.5 Nonrandom X chromosome inactivation and skewing

Clonality experiments assume fidelity of X inactivation, that is, that X chromosome inactivation has taken place randomly throughout all tissues. However, some experimental evidence suggests that this pattern may not always be present. Busque (1996) tested X-inactivation ratios of 295 human females using a human androgen receptor gene (*HUMARA)* based assay on blood cells and found that significantly increased skewing of chromosomal choice was correlated with advancing age. Some 37.9% of individuals aged 60 years and older showed chromosome-choice skewing, while only 8.6% of neonates exhibited some degree of skewing. In this same study, the total pool of blood stem cell progenitor number was calculated to be between 8 and 16 cells. This estimate, reached by stoichiometric calculation using skewed inactivation patterns in neonates, agrees with those for other tissues and suggests that X chromosome inactivation occurs at an early stage in embryogenesis before determination (Fialkow, 1973).

Analyses of organ morphogenesis using X-linked markers for inactivation are more informative if the size of the stem cell pool present at the time of inactivation is known. Estimates of the size of this progenitor population have been made. Puck (1992) examined skewing in neonatal mice and estimated that a pool of between 10 and 20 cells were present at the time of inactivation, while Montiero (1998), using simulation studies, estimated that between 4 and 16 cells were present at the time of X chromosome inactivation. Both of these suggestions reflect the findings of X-inactivation timing experiments (Kratzer and Gartler, 1978), where bimodal expression of the X-linked gene *hgprt* was shown to occur until the 8-cell (morula) stage.

Ohno (1967) predicted that each adult tissue in mammals must have arisen from a small number of progenitors (reviewed in Migeon, 1998). Accordingly, inactivation events (even if truly random) acting on this small population of cells may lead to stochastic variation resulting in unequal distributions of inactivation choice. In this scenario, it is stochastic selection, not non-random X chromosome inactivation, which produces the uneven ratio of inactive X chromosomes.

1.12 Using Single Nucleotide Polymorphisms (SNPs) to measure clonal regions

An alternative to the methods discussed above is to detect and distinguish SNPs in transcripts from X-linked genes as a direct measure of chromosome activity. As individual clones of cells express the same allelic variant of an SNP, this method would allow a measurement of the size, and relative contribution, of such cell populations (clones) in a tissue.

Single Nucleotide Polymorphisms (SNPs) are naturally occurring single base changes in DNA sequence that result in polymorphic alleles at a particular loci. SNPs, most of which are presumed to be functionally silent, are defined as a frequent and naturally occurring single nucleotide sequence variation (Weaver and Hedrick, 1997).

With the recent completion of the first major phase of the human genome sequencing project, a partial description of SNPs found in humans (that is, from the limited number of individuals sequenced in the project) has been made available. With some 1.42 million SNPs mapped in the human genome (by The SNP Consortium and The International Human Genome Sequencing Consortium), such polymorphic variation represents the majority of known human sequence variation (the remainder being due to rearrangements, repeat length polymorphism and insertion/deletion events) (Sachidanandam *et al.*, 2001). The identification and characterization of SNPs has been accelerated by the completion of the working draft of the human genome and huge financial investment by

pharmaceutical companies, whose particular interest is in disease-gene linked SNPs.

While only 4.2% of polymorphisms fell within exons, 85% of all exons were within 5 kb of an SNP (Sachidanandam *et al.*, 2001). The human X chromosome exhibits a low frequency of SNPs, with only 1 SNP per 3.77 kb compared with the genome-wide average frequency of 1 SNP per 1.9 kb. Only one other chromosome, the Y chromosome, has a lower SNP frequency (one SNP per 5.19 kb).

The exploitation of single nucleotide polymorphisms as markers of different X-linked alleles has the advantage over other clonality measurement approaches in that it is not complicated by variable methylation that could affect DNA-based methods. X chromosome inactivation assays based on SNP discrimination have used *G6PD* (Beutler and Kuhl, 1990), Iduronate-2-sulfatase (*IDS*) (Wilson *et al.*, 1993) and palmitoylated membrane protein (*p55*) (Luhovy *et al.*, 1995a) gene – all in *in vitro* experiments. Heterozygotes for these SNPs occur at frequencies of 27%, 41% and 48% respectively. In practice such SNPs could be distinguished by 'marker cocktails', with approximately 90% of females informative when using a combination of *PGK* and *HUMARA* DNA analysis and 68% informative for *IDS*, *p55* and *G6PD* SNPs (Gale, 1999).

1.13 Studies of mammary clonality

1.13.1 Terminology

A cellular clone ("clone") represents a family of cells deriving from a common progenitor, that is, a single stem cell and its daughter cells. So, "clonality" is a broad term that refers to the number and organization of clones involved in the formation and maintenance of the particular tissue being examined. A "patch" is a group of cells that share a common genotype contiguous at the moment of consideration (Nesbitt, 1974).

1.13.2 History

The first attempts at elucidation of the clonality of mammary tissue involved assessment of X-linked enzyme histochemical activity of the housekeeping gene *g6pd* (Thomas *et al.*, 1988). Cells from normal (C3H) animals were introduced to blastocysts of GPDX animals, to produce chimeric mice (C3H-GPDX) that were mosaic for *G6PD* isoforms. Thomas (1988) examined d15 lactation C3H-GPDX chimeric animals for *G6PD* expression (lactation had been previously identified as an enhancer of *g6pd* expression in rats (Richards and Hilf, 1972) and non-lactating animals were shown not to be informative). Heterozygotes showed the expected dual population of G6PD variants with most individual acinar exclusively showing one or the other parental type. Some acinar and all ducts display a mosaic (or mixed) parental lineage, suggesting a polyclonal origin. Furthermore, the distribution of patches in the mammary tissue appeared non-random, which indicated that 'coherent clonal growth' was producing a relatively large patch size for each parental-type inactive X in lactating breast tissue.

The *in situ* results suggested a polyclonal origin for the majority of the larger ducts, with only some smaller ducts and most individual alveoli displaying a "monophenotypic" (homogeneous) staining pattern (Thomas *et al.*, 1988). These experiments also showed that, for the marker measured, physiological state mediates gene expression. Moreover, it was demonstrated that X-linked gene products could be exploited in the wholemount measurement of mammary cellular clonality. Most importantly, the experiments showed that cell mingling and

clonal growth occurred during histogenesis, and resulted in demonstrable cellular clones in the mouse gland.

1.13.3 Stem cell number in the mouse mammary

Kordon (1998) transplanted mouse mammary tumour virus (MMTV) infected epithelial cells into epithelial-free mouse fat pads and estimated that a single stem cell is theoretically able to produce 10^{12} to 10^{13} multipotent offspring (that is, 11 doublings through 4 transplants were observed) before reaching senescence. Smith (1996) estimated that there are around 2500 stem cells in the adult mouse inguinal mammary gland. The authors also showed that, in Czech II mice, any portion of the mammary gland was sufficient to produce an entire functional gland after transplantation into a cleared (epithelia-free) mammary fat pad.

Other studies have shown that between 0.2 and 0.45% mammary epithelial cells exhibit the ability to efflux Hoechst 33342 dye, a behaviour that is considered a universal stem cell marker (Alvi *et al.*, 2002). Limiting dilution cell transplant experiments, where these undifferentiated 'side population' cells were introduced to cleared mammary fat pad, showed that such cells functioned to generate epithelial structures.

1.13.4 Calculating tissue progenitor cell number

Observations of ratios of each of the two possible X-inactivation choices (X_p- or X_m-derived inactive X chromosome) in an adult tissue can be used to estimate numbers of precursor cells that have contributed to that tissue. Typically, such approaches make use of the fact that determinative events of early embryo development are random binominal sampling events and, therefore, that statistical analyses of differentiated mosaic tissues can give information on these determinative events themselves (Nesbitt, 1971). Such analyses are performed by analysis of variation (ANCOVA analysis) in mosaic compostion between mosaic individuals, and within individual tissues in mosaic animals. These statistical manipulations provide data regarding both the number of cells present in the embryo at the time of inactivation (n_1) and the number of primordial precursors that were present in each tissue (n_2) (McMahon *et al.*, 1983; Nesbitt, 1971; Stone, 1983).

A more direct method of assessment, which can be applied to the assessment of clonality in individual structures, has been demonstrated. Stone (2002) produced an estimate of progenitor cell numbers in individual taste-buds by determining the relative amounts of blue to non-staining cells after *lacZ* histochemistry of tissues from a hemizygous H253 transgenic animal (using the method detailed in Chapter 6, section 6.3.3).

There has been some doubt cast on the accuracy of such methods as, in order to perform the analysis, two assumptions, namely: 1) that cells are completely mixed at 'sampling', and 2) that proliferation occurs equally from all progenitors, must be made (Mead *et al.*, 1987). In practice, these assumptions are difficult to test.

1.13.5 Cellular clonality in the human mammary gland

It is generally accepted that cancers arise after a set of mutations accumulate within a stem cell and that these mutations cause unchecked proliferation (reviewed for bowel cancer in Fearon and Vogelstein, 1990). Because, it would be unlikely that identical mutation patterns would occur in adjacent cells, leading to the formation of a homogenous tumour, individual tumours are assumed monoclonal (that is, originating from a single cell), while hyperplasic outgrowth is considered a polyclonal process comprising cells that are still responsive to both endogenous and exogenous stimuli (Noguchi *et al.*, 1994a). It would be valuable to assess complete tumours for clonality to determine the relative contribution of their constituent cellular clone/s.

There are conflicting reports regarding the clonality of breast carcinoma. In contrasting studies, Smith (1971) used *G6PD* variants to show that mammary carcinomas were monoclonal, while Fialkow (1976), also using isozyme analysis, concluded that mammary carcinomas were polyclonal. Noguchi (1994a) suggested that such discrepancies may be due to the inadvertent inclusion of stromal tissue with the parenchymal tissue analysed.

Noguchi *et al.* (1992) reassessed the clonality of human mammary carcinoma using RFLP analysis of *PGK* variants. A monoclonal pattern was the predominant pattern observed in the tumours examined. This was in contrast with

the polyclonal patterns of clonality observed in normal breast tissue. The authors used this method to determine that predominantly intraductal carcinoma (specifically when found dispersed throughout the breast) was monoclonal (Noguchi *et al.*, 1994a).

Recent re-evaluations cast doubt on the findings of Noguchi *et al.* (1992) and produced a different measurement of patterns of clonality in human mammary tissue. Using a *HUMARA* assay, Tsai *et al.* (1996) found that entire lobules and large ducts in normal human mammary were monoclonal. While the patch size was found to be between 0.08 and 0.45 mm^2, measurements of homogeneous expression of one *HUMARA* variant in an entire block from a human reduction mastectomy (from a *HUMARA* heterozygote), demonstrated that the amount of tissue derived from a particular stem cell (that is, the patch or clone size) may, in some circumstances, be quite large. In further support of this evidence for large clone size in human breast tissue, G6PD histochemical studies of breast tissue from Sardinian women heterozygous for a temperature sensitive G6PD variant showed suspected monoclonality within all 111 small ducts and 57 lobules examined (Novelli *et al.*, 2003).

1.14 Aims and hypotheses

It has been suggested that the clonal mechanics of all epithelial tissue systems resemble those observed in the intestinal crypt-villus (Zajicek, 1995). If so, then one would expect individual mammary ducts to display a profile of clonality that is roughly similar to that found in mouse large intestine. As explained above (section 1.2.2), mouse gut progenitor stem cells are located near the base of the intestinal crypts. The progeny from these cells become more differentiated as they move towards the villi apex; where they ultimately undergo natural cell death and sloughing. Mouse mammogenesis differs in some important mechanistic aspects as it involves the TEB penetrating the fat pad with intense mitotic division by stem cells (cap cells) at the tip of the duct. Accordingly, progeny of these progenitor cells form end bud components, which ultimately contribute to the mammary ducts and alveoli.

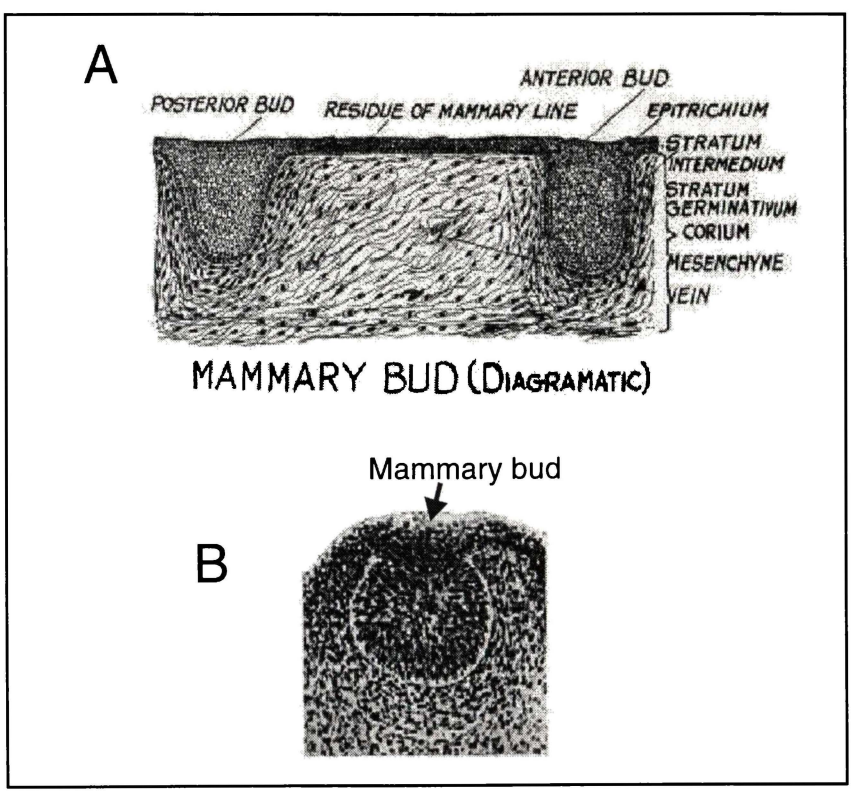


Figure 1.11. Diagrams showing A) mouse mammary buds and B) a ruminant mammary bud, originally from Turner (1930). The ruminant bud is characterized by clear separation from the surrounding mesenchyme while, in contrast, the mouse mammary bud directly abuts the mesenchyme.

It is presumed that the entire population of stem cells that contribute to this ductal/alveolar epithelium are found within the mammary bud (Figure 1.11).

Because of the assumed large number of such stem cells in this structure and, assuming an equal mitotic contribution from these cells, it was predicted that early outgrowth would form polyclonal mammary sprouts. Further ductal extension could result from either: 1) a large population of stem cells contributing randomly to new lateral branches by chance orientation to a branch point after penetration of the primary epithelial duct through the fat pad. This would produce a heterogeneous mixture of small cellular clones throughout the gland or, 2) stem cells being deposited as 'reservoirs' along the ductal basement membrane during growth. Hence, with ductal extension, there would be a reduction in total stem cell numbers available for contribution to new ductal outgrowth. This may lead to skewing where large terminal regions of the gland are produced from single stem cell progeny or, 3) a very early selection of one stem cell in the mammary bud for all subsequent ductal formation, resulting in an almost completely monoclonal gland.

This thesis had the following aims:

- To develop methods for measuring the size of individual stem cell clones in animal models
- To classify the distribution of individual stem cell clones in murine and bovine mammary epithelia using *in situ* and *in vitro* techniques, and to propose a model of cellular clonality in the mammary gland.
- To compare observations from each model animal, bovine and murine, to draw new conclusions as to the suitability of each animal for the study of human breast disease.
- To test the hypotheses that: 1) mammary tissue is predominantly polyclonal, that is, does not display extensive regions of homogeneity for X-linked markers (which would indicate monoclonality), 2) homogeneous expression of X-linked markers is limited to ductal termini, and 3) cellular clonality patterns in murine and bovine mammary tissue are fundamentally similar.

To achieve these aims, and to test the hypotheses, a series of experiments were designed to determine patterns of clone size and distribution in mammary tissue.

These experiments assume that branching decisions occur along the ductal axis where stem cell progeny contribute to new ductal structures and that these structures can be measured using *in vitro* experiments or wholemount (and/or thin section) tissue preparations.

Initially, an *in vitro* method using X-linked SNPs was applied to distinguish alleles of selected X-linked bovine genes and, therefore, the cellular clone profile of the tissue. This approach assumes that the stem cells contributing to a gland all display a random choice of an inactive X chromosome; so that females that are heterozygous for the expressed X-linked polymorphism will be mosaic for each allele throughout the gland. The bovine house-keeping gene X-inactive specific transcript (*XIST*) was screened for polymorphism and an expressed SNP was identified. RT-PCR/RFLP analysis was used to measure the expressed SNP allele ratio in heterozygous animals and, hence, to determine if the gland progenitor pool is small; which would result in an over-representation of one allele (assuming no chromosome selection causes a skewed ratio) in 30 mm^3 blocks of tissue (representing around 3 X 10^7 cells).

It was decided that study of a transgenic mouse model would be the best way to examine cellular clonality in the murine gland. A mouse with an X-linked transgenic modification (H253), a transgene construct that includes the bacterial *lacZ* gene, was chosen for use in the study. Entire whole mount glands from hemizygous female mice were stained using a histochemical reaction that detected the transgene product (bacterial β-galactosidase). Examination of patch size and distribution, in whole mount preparations and in thin sections, revealed the clonality of the entire mammary ductal network in individual glands. These data were used to make conclusions about the numbers of progenitor stem cells that contribute to individual TEB/alveoli.

Chapter Two

General Materials and Methods

2.1 Screening Lambda bovine genomic DNA library

2.1.1 Preparing host strain for lambda sensitivity

A loop of an XL1-blue MRA (P2) host strain glycerol stock was streaked onto an LB-agar plate and incubated overnight at 37°C. An individual colony was then used to inoculate 5 ml of LB media (see appendix) supplemented with 0.2% (w/v) maltose. The media was cultured in a shaking incubator until OD_{600} 0.5 was reached (assuming that 1 OD_{600} = 5 X 10^8 cells/ml, this represents a cell concentration of 2.5 X 10^8 cells/ml). This culture was then supplemented with $MgSO_4$ to a final concentration of 10mM and incubated for an additional 5 min before transfection with the lambda virus.

2.1.2 Titre of pre-made Lambda bovine DNA library

A lab stock of a proprietary lambda bovine genomic DNA library (Stratagene cat. # 945701), created using DNA extracted from the liver of a cow (of undisclosed breed), was titred to determine phage concentration. This library consists of fragments of the *B. taurus* genome (between 9 and 23 kb) within DASH® II lambda library vector arms (Figure 2.1) and was constructed by ligation of partially digested *Bam*HI fragments of the *B. taurus* genomic sequences into *Bam*HI digested DASH® II lambda library arms.

The phage titre was assumed to be approximately 1 X 10^4 phage forming units per µl (pfu/µl). A 2 µl aliquot of the library phage was used to infect 2 ml of host strain (OD_{600}=0.5) prepared as detailed above. After 20 min incubation at 37°C a series of aliquots (200 µl, 20 µl and 2 µl) were each added to 5 ml of lambda top agarose (50°C), aliquoted into 10 ml falcon tubes and thoroughly mixed by inversion. Immediately following mixing, each of the phage-infected host strain

suspensions was poured onto separate pre-poured LB-agar plates (10 cm X 10 cm) and an even distribution of media ensured by tilting.

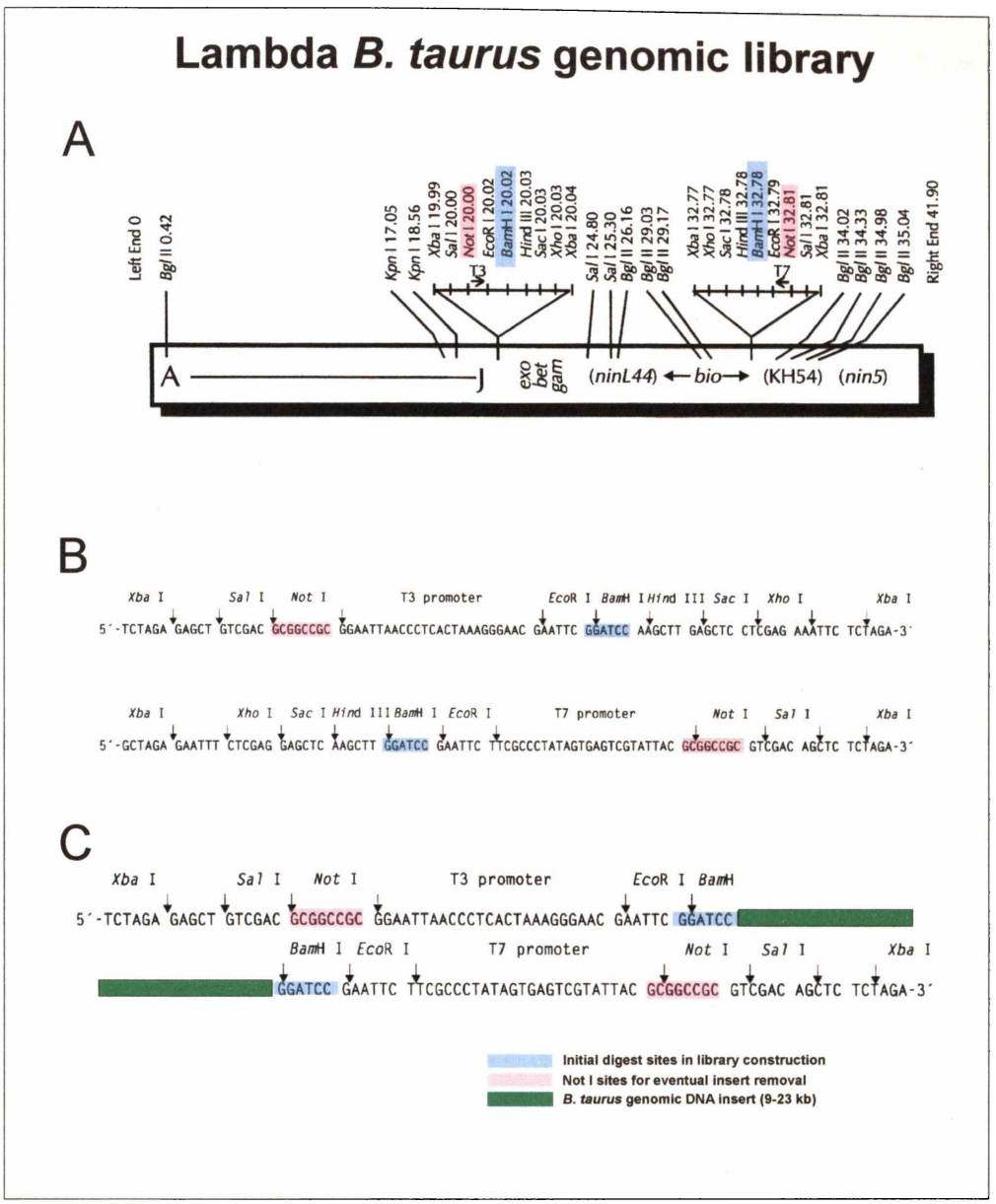


Figure 2.1. A diagram of the DASH® II lambda library vector showing A) unmodified commercial DASH® II lambda vector, B) the vectors Multiple Cloning Site (MCS) and C) a representation of Stratagene # 945701 *B. taurus* genomic DNA library (drawn from manufacturer product information sheets with modification).

After the top agarose had solidified the plates were incubated overnight at 37°C. The next day, plaques on each plate were counted to determine the concentration of pfu. The phage titre for the bovine genomic DNA library used was determined to be 1.4 X 10^4 pfu/µl.

2.1.3 Library screen: transfection of host strain XL1-blue MRA (P2)

To infect the host strain with lambda, a 10 µl aliquot of the phage library (1.4 X 10^5 pfu/µl) was inoculated into 2 ml of host strain prepared as detailed above and allowed to transfect for 20 min, after which time 35 ml of LB top agarose (50°C) was added and mixed thoroughly by inversion. Next, a pre-poured 22.5 cm square LB-agar plate was warmed to 37°C and the top agarose/infected host strain poured, spread by tilting and allowed to solidify. Finally, the plates were placed in a 37°C incubator overnight to allow for plaque formation.

2.1.4 Plaque lifts

After the overnight incubation described above, the plates were chilled at 4°C for 3 h to decrease the likelihood of top agarose peeling during the plaque lift. Sheets of Hybond™-N+ membrane were gently laid onto the top agarose, so that all plaque-bearing portions were completely covered. A syringe needle was used to punch holes through both the Hybond membrane (Amersham) and the underlying agarose and agar layers. These reference holes were later used to position autoradiographs, membranes and plates, so that plaques that corresponded to a strong probe hybridization signal on the membrane could be individually selected from residual viral plaques on the top agarose layer on the plate. To avoid contamination between membranes, needles were flamed for 5 sec before each reference hole was punched.

The first of each of the duplicate membranes used for plaque lifting on each plate was laid down for 1 min. Subsequent membranes (usually one duplicate set were produced per plate) were laid down for a longer period (> 2 min) in an attempt to ensure that approximately equal numbers of phage were transferred in each lift. After the appropriate application time, the filters were carefully lifted from the plate with a single and even motion (using two forceps). The duplicate membranes were placed (plaque-side up) onto a near-saturated filter paper containing denaturing solution (0.5 M NaOH, 1.5 M NaCl) for two minutes, and the membranes placed onto neutralising solution (0.5 M Tris/1.5 M NaCl) saturated filter paper for 5 min. Following these treatments, the membranes were

transferred onto dry filter paper, dried in an 80°C oven for 5 min, and the DNA from the plaques cross-linked to the membranes by exposure to 120 mJ of UV light (at 254 nm) using a BLX-254 crosslinker (GibcoBRL). The membranes were stored in plastic bags at -20°C until use in hybridization experiments.

2.1.5 Phage-library Probe preparation

PCR products were labelled using Rediprime™ kit (Amersham) chemistry according to the manufacturer's instructions. Briefly, 20 ng of gel-purified PCR product was placed in a 0.6 ml microfuge tube and the solution volume brought to 45 µl by the addition of TE (pH 8). The sample was denatured at 96°C for 5 min and the solution allowed to cool briefly, after which it was added directly to the Rediprime reagents in a 1.7 ml tube (the proprietary reagent mixture includes Klenow, dNTPs, and random hexamer primers). The solution was mixed thoroughly by gentle finger-flicking after the addition of 5 µl (50 µ Ci) of Redivue α-[^{32}P]dCTP (Amersham). The labelling reaction was then allowed to continue at 37°C for 30 min and then the labelled probe stored on ice until required.

2.1.6 Probe Hybridization to plaque lifts

The library membranes lifted from the primary screen plates were placed into hybridization tubes (Hybaid HB-OV-BXL, 7.25 cm Internal Diameter, 30 cm Length) and prehybridized for 2 h in Church and Gilbert hybridization solution (see Appendix 5) at 64°C. Ten ml of fresh Church and Gilbert hybridization solution was poured into a 50 ml tube and the labelled probe, denatured by the addition of 15 µl of 4M NaOH (final concentration ~1 M), added directly to the solution.

This solution was added to the Hybaid™ tubes containing the prehybridized membranes and the probe hybridized overnight at 64°C, with rotation, in a Hybaid™ oven (Midi Dual 14). After hybridization the probe solution was collected into a Falcon tube and stored at 4°C. The membrane was washed with 3X SSC, 0.1% SDS for 10 min, then with 0.1X SSC, 0.1% SDS for a further 5 min. The membranes were removed from the Hybaid™ tubes and washed in a tray for another 5 min in the same solution. When the counts had decreased

sufficiently (as assessed by Geiger counter), the washed membranes were sealed in plastic bags and immediately placed in an autoradiography cassette for overnight exposure to autoradiographic film (XR-OMAT, Kodak).

2.1.7 Selecting positive clones

Plaques that gave a good hybridization signal were identified on their original LB plates by reorienting the membrane with the autoradiograph and the LB plate. Plaques were collected by inserting the large bore end of a pipette tip through the upper (agarose) layer of the plating media and selecting the region of media that included the plaque of interest. This agarose plug was placed into 1 ml of SM buffer (see Appendix 5) and allowed to elute for 1 h at room temperature on a flat-bed shaker. Due to the proximity of plaques on the plates, most of the clones selected in primary screens were mixed with their non-target neighbours; consequently secondary, tertiary (and sometimes quaternary) screens were performed to ensure the isolation of a pure lambda clone.

2.1.8 Amplifying Lambda phage – DNA extraction

A 100 µl inoculum of an overnight culture of XL1-blue MRA (P2) host strain (supplemented with 10 mM $MgCl_2$) was inoculated into 15ml of LB media and grown to OD_{600}=0.5. Phage aliquots of 150 µl were inoculated into 800 µl aliquots of these plating cells, which were then incubated at 37°C for 20 min. After this incubation the infected host strain (~950 µl) was added directly to 100 ml of NZY broth (Appendix 5) and incubated for 16-20 h with shaking. Next, 1 ml of chloroform was added and each of the cultures shaken for a further 5 min to lyse any remaining bacterial cells. The vessels were placed in the 4°C cold-room for 30 min to allow the chloroform to settle, then the aqueous phase elements (phage and bacterial debris) were poured into 50 ml Beckman centrifuge tubes. The tubes were balanced and centrifuged for 10 min at 10 000 rpm (10 000 RCF) at 4°C using a Beckman JA20.1 rotor. The supernatant was collected and the pellet discarded. This centrifugation was repeated once to ensure that the cellular debris had been completely removed. Filtered 1.25% glycerine was added to a final ratio of 0.01 % by volume. After thorough mixing the samples were centrifuged at 20 000 rpm (40 000 RCF) in a Beckman JA20.1 rotor for 2.5 h.

The supernatant was discarded and the pellets, representing identical and largely intact phage, were pooled into the same 1.5 ml eppendorf tube and re-suspended in 400 µl of TE buffer (pH 8). DNAse and RNAse were added, to final concentrations of 37.5 µg/ml and 12.5 µg/ml respectively, and the tubes incubated at 37°C for 10 min. Following this incubation, 20 µl of 10% SDS (0.5% final concentration) was added, mixed into the sample by vigorous flicking of the tube and an incubation carried out at 60°C for 5 min. Finally, the suspended DNA was purified by three extractions in Tris-buffered phenol (AR grade) and two extractions in chloroform.

After these purification steps the DNA was precipitated by the addition of a $1/10^{th}$ volume of 3 M sodium acetate (pH 5.2) and two volumes of ethanol. After five min at RT the DNA was pelleted by a 10 min centrifugation at 13 000 rpm (12 000 *g*) in a Heraeus (Biofuge Pico) bench-top centrifuge. The pellet was washed once with 70% ethanol, air-dried and resuspended in 100 µl of TE buffer (pH 8).

2.1.9 Removal of arms from the lambda clone

The DASH II lambda vector used in the construction of the library has *Not*I sites within the multiple cloning site (MCS) which flank its bovine DNA insert (Figure 2.2). For each candidate clone, a 3 µl (~750 ng) sample of the extracted Lambda DNA was digested overnight at 37°C in 20 µl of 1X 'H' digestion buffer with 10 U of *Not*I (Boehringer). Following this digestion, the fragments were ethanol precipitated, washed in 70% ethanol and the DNA resuspended in 50 µl of TE buffer (pH 8). The restriction pattern was analysed by 1% TBE gel electrophoresis.

The gel was placed on a mylar sheet and bands corresponding to the insert DNA of the lambda vector, that is, bands of lower molecular weight than those corresponding to the lambda arms, were excised from the gel using #10 scalpel blades. For a number of clones, *Not*I digestion gave more than three bands, indicating that the restriction endonuclease cuts within the insert itself (in such cases these bands were collected separately).

The DNA fragments within each of the gel slices were extracted using a Qiaex II Gel extraction kit (Qiagen cat #20021) according to the manufacturer's instructions. Briefly this required incubation of a gel slice for 10 min at 50°C in 'QX1 buffer' with 1 µl/ml 'Qiaex II DNA binding solution' resin. Following centrifugation and serial washing steps, the DNA was eluted into 20 µl of TE buffer (pH 8). After purification, this extract was used directly as a cloning insert.

2.1.10 Cloning of insert DNA into phagemid Bluescript KS+

Cloning of the insert from the selected lambda clone was carried out using a pBluescript II KS+ (pBS KS+) cloning vector (a 2961 bp vector; Stratagene cat. # 212207) with XL1-blue host strain. For each cloning experiment 200 ng of vector (in 10 µl) was made linear by digestion with 10 U each of the restriction enzymes *Not*I (gclggccgc) and *Sal*I (gltcgac) in the Boehringer digestion buffer 'H'. These enzymes both cut once within the pBS KS+ multiple cloning site (MCS). Following digestion the vector was precipitated with two volumes of ethanol, washed in 70% ethanol, and the vector pellet resuspended in 4 µl of deionised water.

The insert DNA, digested from lambda with *Not*I as described above, was ligated into the pBluescript KS+ vector using 1 U T4 DNA ligase in a 10 h reaction at 16°C in T4 DNA ligase buffer. Vector to insert 5′ end ratios of 1:1, 1:2 and 1:3 were trialled for each ligation experiment.

After ligation of the insert to the appropriate docking *Not*I 5′ sticky end overhang, the linear vector-insert fragment was purified and the remaining sticky ends were blunted in a 10 µl reaction with 1 U Klenow in the presence of 5 mM dNTPs. The blunt-ended fragment was again purified by precipitation and another ligation performed (as above) to produce a closed circular DNA molecule (Figure 2.2) for transformation.

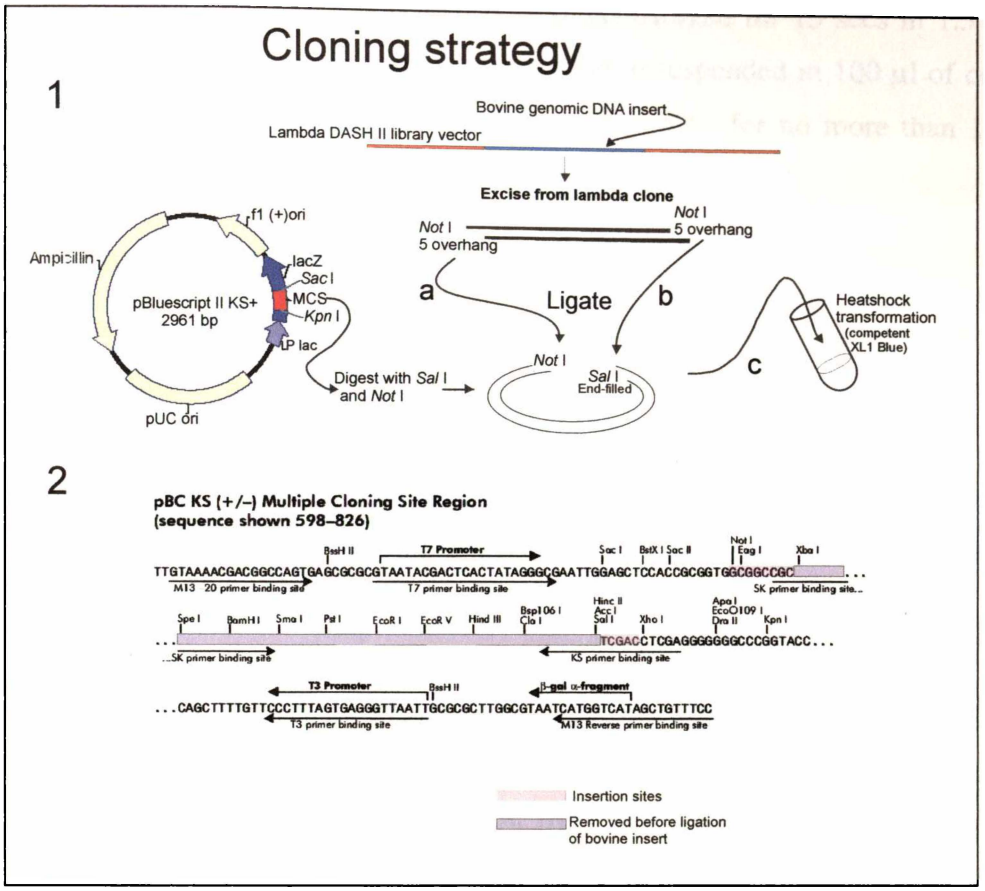


Figure 2.2. Diagram showing 1) the cloning strategy used in library clone purification and 2) annotated cloning vector Multiple Cloning Site (MCS) sequence. After excision of the bovine genomic DNA insert from the arms of the lambda library vector the bovine fragment was 1a) sticky-end ligated on a pBS KS+ vector and 1b) blunt end filled and ligated onto the complementary end of the vector under construction. 1c) Finally the vector was purified and aliquots used directly for transformation of the host strain used (XL1-blue) (see below) and after a short incubation plated onto LB agar supplemented with ampicillin. The vector preparation is detailed by a diagram of the vector Multiple Cloning Site (MCS) (2). The diagram shows *Not*I and *Sal*I states post insertion, with the *Sal*I site modified by end-filling (drawn from Stratagene product information sheets).

2.1.11 Preparation of competent host strain

Host strains used in cloning experiments were prepared immediately before they were required, as such cells give an improved transformation frequency compared to competent cells stored for longer periods of time at -80°C.

A 0.1 ml aliquot of XL1-blue host strain from a 10 ml overnight liquid culture (10 ml of LB broth inoculated with cells from a single colony cultured on an LB plate at 37°C overnight, then sealed and stored at 4°C for up to a week before use) was used to inoculate 9.9 ml (final ratio: 1 part in 100 parts) of LB broth. The strain was cultured until OD600 ≈ 0.4, indicating that cells were in the exponential/log phase of growth.

One ml aliquots of the host strain culture were centrifuged for 15 secs in 1.5 ml tubes, the supernatant was discarded and the pellet resuspended in 100 µl of cold 1X TSS (see appendix 5). The cells were stored at 4°C, for no more than 1 h, until required in transformation.

2.1.12 Transformation

Transformations of total plasmids produced from the above ligations were performed using a heat-shock protocol. Between 1 and 3 µl of the ligation reaction mixture was pipetted into each 100 µl aliquot of host strain in 1X TSS (prepared as above) on ice. After 30 min the tubes were placed into a 42°C water bath for 1.5 min, and then immediately into an ice bucket for 5 min. One ml of LB broth (without antibiotic) was added and the tubes incubated at 37°C for 30 min (to allow the newly transformed cells to express plasmid encoded antibiotic resistance factors before being plated out). After the tubes were centrifuged for 20 sec at 11 000 rpm and the supernatant discarded, the cells were finally resuspended in 50 µl LB supplemented with 10 mg/ml ampicillin.

2.1.13 Plating and selection

A 20 µl aliquot of each transformation, produced as described above, was plated onto LB agar (supplemented with 100 µg/ml ampicillin, 0.5 mM IPTG and 40 µg/ml X-gal) and distributed evenly over the plate using a sterile glass spreader. The plates were allowed to partially dry in a laminar flow hood for 10 min, then incubated overnight at 37°C and assessed for potential insert-positive clones by eye. White colonies (that is, those potentially exhibiting an insert disrupting the vector lacZ gene) were selected using a sterile toothpick and streaked across a square on a grid labelled LB-Amp agar plate. The toothpick was then immersed into 1 ml of LB broth AMP media for 10 sec. Both the liquid media and these 'selection' plates were incubated overnight at 37°C, the former with shaking. Re-streaked plates were again assessed for the presence of white colonies and the inoculated LB broth tubes that corresponded to positive (white) streaked colonies, subjected to a plasmid extraction as below.

2.1.14 Alkaline lysis plasmid extraction

For initial clone screening purposes a rapid alkaline lysis technique, essentially as detailed by Cormack and Somssich (1997), was used. In brief this involved centrifuging 1 ml overnight cultures of candidate clone-containing XL1 host cells for 30 sec at 10 000 rpm (12000 g), removal of supernatant, and re-suspension of the cellular pellet in 0.2 ml of Solution II (1% SDS, 0.2 N NaOH). After gentle mixing for a few minutes, 0.2 ml of Solution III (3M potassium acetate pH 5.5) was added and the samples were mixed by inversion. Next, the samples were centrifuged at 13 000 rpm (12 000 *g*) for 2 min to remove cellular debris and the supernatant collected and transferred into clean 1.5 ml plastic tubes. An equal volume of 100% isopropanol was added and, after thorough mixing, the samples were centrifuged for an additional 2 min to pellet the plasmid DNA. The supernatant was discarded and the pellet air-dried for 10 min. The pellets were resuspended in 50 µl of TE buffer with 10 µg/ml of DNAase-free RNAase A (Sigma) at 37°C for 10 min. Finally, a phenol/chloroform extraction was performed to remove the RNase and the plasmid was resuspended in TE buffer.

2.2 Other protocols used

2.2.1 DNA sequencing

Sequencing reactions were carried out using ¼ reaction volumes (with a final volume of 5 µl) using a protocol provided by The University of Waikato DNA Sequencing Facility. The reaction consisted of 2 µl of Bigdye II™ (Applied Biosystems 377), 1 µl of primer (0.8 pmol/µl; 20mers were usually used), 0.5 µl of template (100 ng of plasmid; or 1-10 ng of PCR product) and 1.5 µl of deionised water. The cycling conditions used were: 18 to 25 extensions using an MJ Research PTC-100 thermocycler (MJ Research, Inc) program of 96°C/10 sec and 60°C/4 min. Products were purified by precipitation, in a final concentration of 0.12 mM sodium acetate (pH 4.6) in 2 volumes of ethanol, for 20 min at RT. The precipitated DNA was collected by centrifugation for 30 min at 13 000 rpm. The pellet was washed in 200 µl of 70% ethanol, the tube centrifuged for another 10 min, the ethanol discarded and the pellet dried at 80°C for a few seconds in a thermo-cycler. The samples were supplied in this form directly to the sequencing facility for electrophoresis.

2.2.2 RNA extraction from bovine and murine tissues

RNA was extracted using Trizol™ (Stratagene) reagent according to the manufacturer's instructions with minor modifications. Briefly, this involved placing a piece of frozen tissue (50-120 mg) into a folded sheet of mylar plastic sheet, smashing the tissue using a light hammer and tipping the resulting powder into a 1 ml aliquot of Trizol™ in a 1.7 ml tube. The tube was then capped and shaken to mix the contents and to ensure a consistent homogenate. After incubation for 20 min at RT, 0.2 ml of chloroform was added and, following thorough mixing by vortexing, the samples were centrifuged at 13 000 rpm for 10 min at 4°C and the aqueous phase collected. Isopropanol (0.5 ml) was added and the samples mixed and stored at 4°C for 30 min. Finally the samples were centrifuged for 10 min at 13 000 rpm and the resultant pellet washed in 1 ml of 75% ethanol, air-dried and, after re-suspension in 50 µl of DEPC-treated deionised H_2O, stored at -80°C.

2.2.3 DNA extraction from bovine milk for mass SNP screening

For the extractions, 50 ml milk samples were collected either directly from an animal, or from retail milk (Anchor™ Milk). The milk sample was centrifuged for 5 min at 3 500 rpm (2 200 *g*), and the top 45 ml of supernatant (and the upper fat layer) discarded. The remaining 5 ml was transferred to a 10 ml tube and after the addition of 5 ml of TE buffer (pH 7.5, 4°C) the tube was inverted a few times to ensure mixing and then centrifuged at 500 rpm (45 *g*) for 5 min. Next, the upper 9 ml was removed by aspiration and discarded and the remaining 1 ml of solution carefully removed (avoiding disruption to the debris pellet). This 1 ml of solution was transferred to a 1.5 ml tube and centrifuged for 15 sec at 13 000 rpm (12 000 *g*) and the majority of the supernatant discarded (leaving about 200 µl of solution). Following the addition of 1 ml of TE buffer (as above) to this 200 µl solution and centrifugation for 15 sec at 13 000 rpm, the supernatant was again discarded, this time retaining only the lowest 100 µl of solution (that is, the fraction containing the debris-free somatic cells). Finally, 100 µl of Proteinase K buffer (see Appendix 5) - supplemented with 40 µg of Proteinase K - was added and the solution mixed by pipetting through a medium bore tip. The cell protein components were digested by incubation at 55°C for 2 h. Genomic DNA was

obtained by standard phenol/chloroform extractions and ethanol precipitation. The numbers of cells present in the original milk sample were estimated using the method shown in Appendix 6 (pers. comm. R.J. Wilkins, The University of Waikato).

2.2.4 End-labelling oligonucleotide probes for use in RT-PCR/RFLP experiments

Deoxyoligonucleotide probe (20 pmol of 5′ ends) was brought to a volume of 17 µl by the addition of deionised water. End-labelling was carried out in a 20 µl volume with 1 µl (10 U) T4 polynucleotide kinase (PNK; Roche), 2 µl of 10X PNK buffer and 1 µl of Redivue [γ^{33}P]dATP (10 µ Ci; Amersham), for 30 min at 37°C. The reaction was stopped by the addition of 1 µl of 2 M EDTA (final concentration of 0.2 M EDTA) and the tubes stored in an ice-bath before direct addition to the hybridization solution.

2.2.5 Hybridization

UV cross-linked membranes were pre-hybridized in 5 ml of Church and Gilbert hybridization buffer (see Appendix 5) for 2 h before the direct addition of 20 pmol of labelled oligonucleotide probe. Hybridization was carried out for between 6 and 12 h, then the excess probe was removed by washing in 2X SSC/0.1% SDS for 10 min at the hybridization temperature. Finally, the membranes were dried and exposed to autoradiographic film overnight (or to a Fuji phosphorimaging plate 2 h to overnight).

2.2.6 General PCR strategy

PCR was carried out as detailed in the individual results chapters. Typically, annealing was carried out at a temperature 5°C lower than the average of the primer T_m. Where this approach failed to yield products, annealing temperature was lowered and/or cycle number increased in an attempt to amplify the difficult target molecule.

2.2.7 Developing autoradiographs

Autoradiographs were taken from their cassettes in the dark and placed in a fresh developer (Kodak) until hybridization-positive regions were well-developed (usually after ~2 min). The film was lifted from the developer solution, allowed to drain briefly and placed in a fixative solution until the cloudiness on the surface of the film had cleared. The film was then washed thoroughly in tap water and allowed to drip dry.

2.3 Image acquisition:

2.3.1 BAS-1800 II Phosphorimager (Fuji)

Phosphorimaging with the BAS-1800 uses 'photostimulable phosphor layer (BaFX:Eu)' coated imaging plates (23 X 25 cm). These phosphor plates store charge from radioactive isotopes and this charge is later read by a laser-coupled photomultiplier tube within the instrument. The distributor claims that the system offers 10 to 100 times the sensitivity of film based autoradiography and hence increases membrane throughput times as well as allowing direct image capture to a computer for image quantification.

A Fuji BAS-1800II phosphorimager was used for much of the autoradiographic 'development'. For the development of blots probed with ^{32}P-labelled probes, membranes were dried in a 37°C oven, placed in a plastic bag, and then taped, signal-side up, to the positioning grid within a BAS cassette. Next, a phosphor plate that had been cleared of residual signal by exposure to light for 20 min, was placed into the cassette with the dry membranes. The cassette was laid flat for the required exposure time – usually 2 h, although this was dependent on approximate signal intensity on the membrane, measured using a Geiger counter.

Following exposure the plates were removed and placed face up into a BAS-1800 reader instrument and an image read using the associated software package 'BAS - Image Read'. The resultant image was cropped and its contrast/brightness adjusted with 'BAS – Image Gauge' software.

2.3.2 Photographic image acquisition

2.3.2.1 Nikon Coolpix995

Low magnification images were collected using a Nikon Coolpix995 via a modified 10x eyepiece attachment. Wholemount images between 100x and 200x magnification were collected in 'manual mode' as TIFF images (setting: 'Hi': without compression) and full size ('3:2') to ensure maximum resolution upon subsequent image manipulation and enlargement.

2.3.2.2 Zeiss AxioHR digital camera

Both higher magnification and/or resolution images for the *in situ* hybridization (Chapter 4) and histochemical (chapter 5) studies were collected using a Zeiss Axiocam HR on a fluorescent microscope with digital imaging software capability. Images were collected at magnifications between 10x (for wholemount glands) and 400x (for *in situ* hybridization results). As above, images were saved as TIFF images for later assessment and manipulation.

2.3.2.3 Agarose gel images

Ethidium bromide (EtBr) stained agarose gels were observed using a TFX-35M transilluminator (LifeTech) and images collected using a Cohu™ CCD camera with Scion 'Image' frame capture software.

2.4 Histology

2.4.1 APES treatment of glass slides

In order to treat slides to allow cross-linking adherence of tissue sections, pre-cleaned glass slides (Semi-frosted; Erie Scientific) were soaked for 60 sec in 1% 3-aminopropyltriethoxy-silane (APES), rinsed in running tap water, and washed 3 X 10 min in DEPC H_2O. The slides were subsequently allowed to air-dry, and were stored in an airtight container until required.

2.4.2 Fixation of animal tissues

The bovine tissues used in this study were fixed in a freshly prepared 4% paraformaldehyde solution. This was made by dissolving 4 g

of paraformaldehyde in 66.5 ml of DEPC-high quality water at 60°C with 3 drops of 2M NaOH added to assist dissolving. Finally, 33 ml of 3X PBS (final 1x) was added and pH adjusted to 7.4 with 1 M HCl after the solution had cooled.

2.4.3 Paraffin-embedding murine and bovine mammary tissues

Paraformaldehyde fixed bovine mammary glands were cut into thin (~4 mm thick) tissue blocks before processing, whereas fixed mouse glands were processed whole. The tissues were treated for 1 h each in 30%, 50%, 70%, 90% ethanol and for 3 X 1 h in 100% ethanol (final in absolute ~98% ETOH). After dehydration, the tissues were cleared in xylene for 3 X 1h and placed in histology cassettes. The cassettes were submerged in baths of molten paraffin wax at 58°C for 3 X 1 h. Next, the tissue cassettes were taken from the final paraffin bath and the tissue block removed and placed into a metal pouring mould (pre-filled with molten paraffin wax). The plastic cassette insert was placed over the top of the mould, ensuring that the liquid wax covered it, and the mould placed on an ice-cooled tray for 20 min. Next, the moulds were placed in a beaker into which cold tap-water was run for 20 min. The moulds were finally cooled in a fridge at 4°C for 20 min, after which time the paraffin blocks were removed from the moulds. The excess wax was trimmed in preparation for microtome sectioning.

2.4.4 Microtome sectioning

Serial thin sections (10-15 μm) were produced using a microtome blade at a 6 degree angle to the block. These sections were floated in a water bath and collected onto coated glass slides. Finally, the slides were dried overnight in a 42°C oven.

2.4.5 Frozen sections

Blocks of tissue from sampled glands were frozen in liquid N_2 immediately after dissection. Before sectioning, the tissues were mounted onto a block using OCT (Tissue-Tek™) mounting solution. Frozen sections (10 μm) were produced, lifted onto silane-coated slides and allowed to air-dry for 1 h. The slides were then fixed in 4% paraformaldehyde for 5 min, and dried in an ethanol series (70%

5 min, then 100%) after which they were stored at 4°C in 100% ethanol before use (up to 1 month).

2.4.6 Mouse mammary wholemounts

The *lacZ* histochemically-stained mammary tissues were dehydrated in an ethanol series and cleared in acetone for 30 min. After clearing, the wholemounts were re-hydrated, stained overnight in carmine alum and dehydrated in an ethanol series to 100% ETOH. Following this dehydration, the mounts were soaked in xylol for 1 h (or until cleared) and then mounted with a xylene-based mounting media (D.P.X) using another glass slide as a cover-slip.

2.5 Ethics and ERMA approvals

The bovine tissue used in this study was sourced from animals at a research abattoir. As they were not slaughtered specifically for use in the experiments described in this thesis, no special ethics approval was required for this work. The transgenic animal tissues used in the murine studies were from culled animals bred for an unrelated study conducted at another facility. This study had the appropriate approval (Agresearch Ruakura AEC number 3499).

Cloning of bovine DNA fragments was carried out with approval from ERMA. The internal biological safety committee approval reference number is: GM000/UOW1.

2.6 Oligonucleotides used in this study

The oligonucleotides used as primers or probes in the present study are listed in Table 2.1

For bovine *XIST* sequencing	
BovXistF:	ATTGTGGTATCATGAGGTGGGAAA
BovXistRev:	AGCGAGGTGCTATGCTAACTCAT
Xist2Forward	GCCAGTTAGTGGAGGATGGAATTA
Xist2Rev	ATCAAAATCCGACCCCAGCATTAG
GregX	GGGTAGGTGTTCCTCTTGAGGAA
BXist.Rev.2.1	CCAGCATTAGCCACGGGGTAGGT

XistR1	AGGGATTCCTCTTCTGCCAACAG
XistR2	AGTTGGCCAACGATCATCTGCGA
XistF2	AGATGAGTTAGCATAGCACCTCG
XbackForw	GCCCATCGGGGCTGCGGATACCTG
Bov.Xist.Prom.Rev	CCCAACACTTTTCCAACCATATAA
X1F	ATTGTGGTATCATGAGGTGGGAAA
Bov.XF1.back	TTTTGAACACTCTAACAAAGCAGA
XF2.2	TCAGAAAGGGTGGTAGAATCGG
XR1	CGAGGTGCTATGCTAACTCAT
Bov.PM.specific.ALUI	GGTGCCGTGGCCGAAGCCCTAGC
X2R.outer	TCAAAAATCCGACCCCAGCATTAG
Bov.Xist.Prom.F	WAAGATRTMNGRMTTGCAATCWTC
PromForw1	TTCTTAAAGCGCTGCAATTTGCTG
Promforw2	TATTTTGCCCATCGGGGCTGCGGA
BXPM.F	AATTCATAAAATATTTTTAAACAGCT
BXPM.R	GGCAACCAAAAATCATTCCA
AluPM-MnlIIUpst.	GAACGTCACGCCATTGCTCTGCATTG
XF2	GCCAGTTAGTGGAGGATGGAATTA
For amplifying a probe for screening of the λ library for G6PD:	
Bov.G6PD.Probe.F	RGTGTTGAARTGATCTCAGAGGT
Bov.G6PD.Probe.R	TTCTCYMGMTCAATCTKGTGCAGC
Bov.G6PD.2ndF	ATGCAGAACCACCTMCTGCAGATG
Bov.G6PD.2ndR	TAGGTAAGGTCCAGCTCCGACTCCT
For sequencing the pXIST λ library insert:	
pXIST-T7-2	CCTCTCTGTAGCGAAGCTCCTATC
reverse primer	GGAAACAGCTATGACCATG
pXIST-T3-2	TCCTCAGGCACCAACTCTTCCACA
pXIST.3	CATCCCAGGTCTAGTTGCCATAGG
pXIST.4	GGTGTCCCAAGGCAATGGGAAGCT
pXIST.M13.2	GGAAGTGCTAGTGGTGGTTGGCAA
pXIST.5	GAAGGGACTGCAGATTAGACACAGG
pXIST.5.b	CAGAGATGGGGGACATTCCAGAA
pXIST.6	CCCCAACTGAACTGCTACCTGCCA
pXIST.7	GAGAGGCCTCTGAGCTGAGT
pXIST.8	TGGCAAGATCTATGGAGCTA
pXIST.9	GGGGAAGAGGAAGGAGACTT
pXIST.9.rev.2	TCCACCATGACCCGTCTGTC
pXIST.9.M13-21.2	GCCTCCTGCATCCAGATGTT
pXIST.D.Rev.2	CTTGTGTCCAGAAGCCCAATG
pXIST.10	CCTGGTCTCTGGCAGCTCAT
pXIST.11	CCACACAATCCAGGATGAAT
pXIST.12	TGATGTGGTGGCCTCAATTA
For sequencing the pG6PD λ library insert:	
pG6PD.m13.2	GTACAAGCTGGATTTAGAAGAGGC

pG6PD.Rev.2	ACCAAGATGATGACCAAGAAGCCT
pG6PD.B.M13.3	CTAAAGAGCCTCTTGATGAGGCTG
pG6PD.C.Rev.2	TGAATGGCAGTCTCCGTGGCATT
pG6PD.B.M13.4	TGCCAAAGAATTGATGCTTTCAA
pG6PD.C.M13-21.2	GCTGGGAAGGGTGTGAATGCCATT
pG6PD.B.m13.5	TGATAGTTGAGCATCCTCACCAGG
pG6PD.B.Rev.3	AAGAAGGCGGCATCCATACT
pG6PD.rev.3	GTGTACGGCAGGTACGTCTA
pG6PD.R.4	GGCACCTACAAGTGGGTGAA
pG6PD.rev.5	CAGGGGCAAGGTCCTCAGAA
pG6PD.6	TGGGTGGGCACGTGGATTAC
For bovine G6PD exon 10 screen	
Bov.Ex10.screen.F	CCAAAGCCTGCTGTGTCCATCC
Bov.Ex10.screen.R	CAGGGTGTCACCAAGGTGCTGA
For bovine G6PD Exon 11 screen	
pG6PD.B.screen.F	CCACACTGCTCCTTCTCTGTAG
pG6PD.B.screen.R	AGGCTGCCCTTTCCGCCACG
pG6PD.B.Ex11.Screen.F	GCACCTTGGTGACACCCTGT
pG6PD.B.ex11.Screen.R	ACGAAGAGGGGGATGGAGAG
Others:	
T27-XF2-c/t-PM	CCTCTCACTGGCTCCAGCT
T27-g/t-pm.2	CCTCTCACTGGCTCCAGC
Prom1-g/t-PM	CCCCAAGTATCGTAAAATACTGA
T27-PM.rev	CTCACCGCCAAGTTCACACT
Prom.PvuII.R.new	CAACCAAAAATCATTCCAAA
26mer-G	ACTAGAAGAAAGTTCAGGTTTTTAAA
26mer-T	ACTAGAAGAAAGTTAAGGTTTTTAAA
G6PD-HindIII-cloning	GCTCAAAGCTTGTGGGGGTTCAC
G6PD-SacI-cloning	GTGGAGGCAGACGAGCTCTGAA
GAPDH Forward	ACTCACTCAAGATTGTCAGCAATG
GAPDH Reverse	GTCATGAGCCCTTCCACAATGCCA
For mouse RT-PCR:	
HMGCoA Reverse	CACCGCGTTATCGTCAGGATGAT
HMGCoA Forward	CCCTGGGAAGTTATTGTGGGAACA
GFP reverse	CTGCTTGTCGGGCCATGATATAGAC
GFP forward	CTTCAAGGACGACGGCAACTACAA
LacZ reverse	CGTTTCGTCAGTATCCCCGTTTA
LacZ forward	GCAAAGACCAGACCGTTCATACAG
HMGCoA probe reverse	CTTCGTCCAGACCCAAGGAAACC
LacZ probe forward	TCGCTCGCCACTTCAACATC

Table 2.1. Oligonucleotides used in this study.

Chapter Three

Screening Bovine X-linked Genes for Polymorphism

3.1 Introduction

One of the aims of this thesis was to measure the size and distribution of cellular clones in bovine mammary tissue. To achieve this, a marker that discriminates between neighbouring cellular clones is required and, for this purpose, most studies use polymorphic alleles of X-linked genes.

X-linked gene polymorphisms have been used for analysis of the cellular clonal profile in studies of a number of tissues and disease states. For example, glucose-6-phosphate dehydrogenase (*G6PD*) polymorphisms in myeloproliferative disorders (Luhovy et al., 1995), hypoxanthine phosphoribosyltransferase (*HPRT*) polymorphisms in haematological samples (Browett *et al.*, 1988), and human androgen receptor (*HUMARA*) polymorphisms in studies of thyroid gland clonal composition (Jovanovic *et al.*, 2003). PCR-based *in vitro* analysis of the mosaic expression of these markers showed that the clonal profile of a tissue could be found by simply determining the presence of the different alleles by polyacrylamide gel electrophoresis (PAGE) (also reviewed in Chapter One).

At the beginning of this study there was little available genome sequence data for X-linked bovine genes. Accordingly, it was thought necessary to screen a genomic library for bovine homologues of known X-linked genes. The candidate genes chosen were:

1) *G6PD:* An X-linked housekeeping gene used previously in studies of cellular clonality in human specimens (Peng *et al.*, 2000). The human *G6PD* gene is known to be reasonably polymorphic, especially for SNPs, with 18 SNPs reported in human *G6PD* (Genbank: NM_000402). It was assumed that the

bovine homologue will also display a high level of variation thereby making it a good candidate for this study.

2) X inactive-specific transcript *XIST*: Bovine *XIST* was thought likely to contain useful polymorphism for two primary reasons. Firstly, there was previously published polymorphism for the human *XIST* gene (Rupert *et al.*, 1995), and secondly, evolutionary studies had shown that the gene lacked conserved coding potential (Hendrich *et al.*, 1993). Both of these findings indicated that *XIST* sequence constraints were not likely to be rigid. Fluorescence *in situ* hybridization (FISH) studies had shown that the *XIST* transcript localized to a single X chromosome *in cis* in female somatic cells (Duthie *et al.*, 1999), meaning that it would provide a focal nuclear-localized marker of X inactivation pattern and therefore the origin of particular groups of cells derived from the same progenitor (that is, an individual clone).

It was thought that the best markers of clonality for use in this study were single nucleotide polymorphisms (SNPs). SNPs give an advantage over variable number tandem repeat (VNTR) polymorphisms in this work as they can be readily detected by both PCR-based and eventually by *in situ* methods. Going (2003) anticipates that such *in situ* methods applied to study of the human breast will reveal clonal profiles in normal and diseased tissues without the potential for contamination from neighbouring tissues (discussed in Chapter 4).

This Chapter describes: 1) the isolation of putative *G6PD* and *XIST* clones from a bovine genomic DNA λ library, and 2) the direct sequencing of bovine *XIST* using consensus primers designed from known *XIST/Xist* gene sequences. The annotation of the *G6PD* and the *XIST* clones identified in the library screening is described in appendices A.2.1 and A.2.2, respectively. The *XIST* sequence obtained by direct sequencing was subjected to RT-PCR to confirm sex-linkage, and Northern blot analysis to determine whether the gene was expressed at reasonable levels in mammary tissue. Subsequently, the *XIST* fragment obtained by direct sequencing was screened for SNPs by detection of sequence heterogeneity at the DNA level. The potential utility of each of the identified

SNPs in the proposed bovine mammary gland cellular clonality studies (Chapter 4) is discussed.

3.2 Methods and Materials

3.2.1 Library screening for the *Bos taurus G6PD* gene

A probe was produced by PCR amplification with consensus primers designed from known mammalian *G6PD* sequences. The primers were:

Bov.*G6PD*.2ndF (forward)	atgcagaaccacctmctgcagatg
Bov.*G6PD*.2ndR (reverse)	taggtaaggtccagctccgactcct

This primer pair amplified a ~1 kb fragment that included exons 9 to 12 of the gene, a region shown to be polymorphic in other species (see Beutler and Kuhl, 1990).

The λ genomic DNA library was screened using an identical method as used for the *XIST* gene screening, which is detailed below in section 3.2.3 and in Chapter 2, section 2.1.5.

3.2.2 Detection of SNPs in the *Bos taurus G6PD* gene

A direct PCR-based approach was used to amplify regions of the bovine *G6PD* clone isolated from the λ library. Primers were designed from the pG*6PD*.B clone sequence (as described in section 3.3.1) and these were used to amplify the presumed exons 10 and 11 from genomic DNA samples.

Exon 10 specific primers:

Bov.Ex10.screen.F	ccaaagcctgctgtgtccatcc (92-114)
Bov.Ex10.screen.R	cagggtgtcaccaaggtgctga (420-442)
	Product: 350 bp

Exon 11 specific primers:

p*g6pd*.B.Ex11.Screen.F	gcaccttggtgacaccctgt (423-443)

p*g6pd*.B.ex11.Screen.R acgaagaggggatggagag (575-595)

Product: 172 bp

(The numbers in parentheses indicate primer binding position within the *G6PD* sequence gb: AF531754 published from this study).

Exons 10 and 11 of the '*G6PD* gene' from ten animals were sequenced in both forward and reverse directions using the general PCR protocol described in Chapter 2 (section 2.3.7).

3.2.3 Library screening for *Bos taurus XIST*

A lambda genomic library was screened for a *XIST* clone. A 980 bp probe fragment was designed to the known bovine sequence (within exon 1) using the primers shown below:

X1F (forward) attgtggtatcatgaggtgggaaa

XistR2 (reverse) agttggccaacgatcatctgcga

Product: 979 bp

These primers were designed from a genomic *Bos taurus XIST* gene fragment DNA sequence deposited previously into Genbank by the author (gb: AF104906, position 1125-2104). PCR products were electrophoresed on a 1% TBE agarose gel, purified using a gel extraction kit (Qiagen) and end-labelled as described in Chapter 2, section 2.1.5.

A commercial phage library was screened as detailed in Chapter 2, section 2.1. library clones that gave a strong signal after hybridization with the *XIST* probe were selected, amplified (see Chapter 2, section 2.1.8) and the 5′ and 3′ λ arms of the vector removed by digestion with *Not*I. The inserts were isolated in a 1% TBE gel, purified using a gel extraction kit (Qiagen), cloned into a Bluescript KS+ vector (see Chapter 2, section 2.1.10) and sequenced using ABI dRhodamine terminator chemistry with a ABI 377 DNA sequencer (see Chapter 2, section 2.2.1 and below).

Standard T3 and T7 primers (T7: gtaatacgactcactatagggcg, T3: aattaaccctcactaaaggg) or M13-20 and M13 reverse primer (M13-21: gtaaaacgacggccagtg and M13_reverse: tcatggtcatagctgtttcc) were used to obtain the initial sequence data. Subsequently, clones that showed similarity to those database sequences with a known coding potential were sequenced further using internal primers designed as more sequence information was obtained.

3.2.4 *XIST* DNA sequencing and screening strategy for SNPs

DNA sequencing of PCR products of bovine *XIST* was carried out as outlined in Chapter 2, section 2.2.1. SNPs were detected by direct inspection of heterogeneity on electropherograms or by ClustalX-facilitated comparison of the sequence data files from a number of different animals. Confirmation of sequence heterogeneity for SNP2 was carried out by performing replicate sequencing runs (with at least three-fold total sequence coverage).

This sequencing project focused on the 5′ end of the *XIST* exon 1, as sequence data for this region had been obtained previously by comparative genetic studies in other species (Hendrich *et al.*, 1993).

A portion of the bovine *XIST* gene was sequenced from PCR products amplified from a *B. taurus* genomic template using universal consensus primers. Universal primers (BovXistF: attgtggtatcatgaggtgggaaa, BovXistRev: agcgaggtgctatgctaactcat), designed to a known 650 bp bison *XIST* gene fragment (Bison bonasus, gb: L10727) but with consideration of sequence data from - in order of descending weighting because of different relative similarities to bison *XIST* - human, equine and leporine, were used to amplify and sequence a portion of the B. taurus gene. Regions flanking this fragment were amplified by sequential 'walking'; each PCR used one known bovine sequence (found in earlier sequences) and one primer designed from the *XIST* sequence data of other species.

A gapped BLASTN search of the bovine *XIST* fragment was performed against sequences stored on the NCBI databases (http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi). ClustalX software was used

for the alignment of *XIST* sequences - human (gb:HUMXIST), murine (gb:LO4961), equine (gb:U50911), leporine (gb:OCU50910) and bison (gb:BIS*XIST*AA) - to confirm sequence similarity.

3.2.5 *XIST* expression analysis by RT- PCR

The RT-PCR/RFLP method described in detail in section 3.2.7, using primers (BXPM.F and BXPM.R) that incorporate a *Pvu*II site in the reaction products, was used to amplify a 142 bp fragment corresponding to bovine *XIST*.

Beta-actin (*β-actin*) was chosen as a house-keeping gene to determine the quality of the RNA used and to provide a semi-quantitative measure of relative *XIST* levels in the mammary and liver tissues examined. The primer sequences used for amplification of a 251 bp fragment from the bovine *β-actin* transcript were:

Forward: GGCATCCACGAGACCACCTWCAA

Reverse: CACATCTGCTGGAAGGTGGACAG

3.2.6 Northern blotting

3.2.6.1 mRNA enrichment from total RNA for Northern blotting

mRNA was enriched from a 100 µl aliquot of pooled mammary gland total RNA (50 µl from each of two animals, representing around 3.2 X 10^7 cells) using the PolyATtract® mRNA Isolation system (Promega), and resuspended in 20 µl of DEPC-treated deionised water. This protocol selects mRNA by binding biotinylated oligo(dT) beads to polyA RNA and then selecting these beads paramagnetically using streptavidin-coated magnetic beads. A 5 µl aliquot of this mRNA (~0.5 µg) and enriched mRNA prepared from male bovine liver using the same technique were used in Northern blotting as described below.

3.2.6.2 Gel composition and electrophoresis conditions

Northern blotting was carried out according to Ausubel (1997) at conditions recommended for successful blotting of large transcripts (the *XIST* transcript is 19.3 kb and 17.4 kb in the human and mouse, respectively (Hong *et al.*, 1999; Hong *et al.*, 2000) and it is likely to be of a similar length in *Bos taurus*).

Specifically, gels with a low agarose concentration (0.7%) were used, ethidium bromide (EtBr) was omitted from the loading buffer, the gel was pre-treated in NaOH before blotting, and blotting was carried using 10 X SSC transfer buffer.

All gel apparatus was pre-treated by soaking in 0.2 M NaOH. A 0.7% agarose MOPS gel (14 X 10 cm) was prepared by dissolving 0.7 g of agarose in 72 ml of DEPC water (see Appendix 5) and boiling. The solution was allowed to cool to 65°C and 10 ml of 10 X MOPS (see Appendix 5) added with continual stirring. Finally, 18 ml of 12.3 M formaldehyde was added and the solution mixed and poured into a gel tray.

For each loading, 0.5 µg of an enriched mRNA sample (section 4.2.2.1) was bought to a volume of 11 µl with DEPC-H_20 in a 200 µl PCR tube and 5 µl of 10X MOPS, 9 µl of 12.3 M formaldehyde, and 25 µl of formamide was added. The sample was denatured by incubation at 65°C for 15 min. After denaturation, 10 µl of a formaldehyde loading buffer (see Appendix 5) was added to each sample, the contents mixed by gentle flicking, and the samples loaded directly on the gel. Electrophoresis was carried out for 4 h at 5 V/cm (50 V) in 1X MOPS buffer until the bromophenol blue (BPB) dye front had migrated approximately half of the length of the gel. A single lane was removed from the gel, washed for 20 min in 0.5 M ammonium acetate, and then stained for 40 min in the same solution after the addition of 0.2 µg/µl EtBr. Staining of this lane showed diffuse staining with no clear bands, as would be expected of an mRNA enriched specimen.

3.2.6.3 Blotting

The gel was washed 2X 5 min in DEPC-H_2O, soaked in 50 mM NaOH/1.5 M NaCl for 30 min and neutralized in 0.5 M Tris (pH 7.4)/1.5 M NaCl for 20 min. Finally, the gel was soaked in 10X SSC for 45 min before blotting overnight onto Hybond N+ nylon membrane using a standard upward capillary transfer (Ausubel *et al.*, 1997). Following blotting, the membrane was washed briefly in 2X SSC to remove residual agarose and allowed to air-dry. The RNA was then cross-linked to the membrane in a UV cross-linker (Stratagene, 254 nm, 120 mJ/cm^2).

3.2.6.4 *XIST* DNA probe hybridization

The membranes were hybridized overnight at 60°C with the probe produced as described below, washed in 3X SSC/0.1% SDS for 10 min at 60°C and for 5 min in 1X SSC/0.1% SDS, and exposed to autoradiographic film (XR-MAT) overnight.

A 303 bp *XIST* DNA probe encompassing the SNP-containing region, was PCR amplified from a SNP2 heterozygote DNA template (animal 696) using the primers PF1 and BXPR (see section 3.3.5), gel-purified, and labelled using random priming (Rediprime™ kit, see Chapter 2, section 2.1.5).

3.2.7 Measuring the frequency of SNP2 by PCR-RFLP

A PCR/RFLP test was designed to allow the routine and reliable detection of SNP2. This diagnostic test utilises mismatches incorporated into a forward primer (BXPM.F) to produce a *Pvu*II restriction site (CAG|CTG) in products with a guanine residue immediately 3′ of the primer binding site (G-allele). The PCR product from the other allele gives a PCR product (T-allele) that is indigestible with *Pvu*II. The PCR product represents positions 161 to 303 of Genbank AF104906 (Appendix A.3.2); SNP2 is at position 188 of this sequence.

A partial nested-PCR reaction, minimised the co-amplification of a SINE sequence (submitted to Genbank as gb:AF181665) that was sometimes seen as a spurious product on the gel at about 430 bp. This reaction used the following first round primer pair:

PF1	TTCTTAAAGCGCTGCAATTTGCTG
BXPM.R	GGCAACCAAAAATCATTCCA

The PCR reaction was carried out using 50 ng of genomic DNA in a final reaction volume of 25 µl (comprising 200 µM dNTPs, 0.7 U of *Taq* DNA polymerase, and 0.8 pmol of each primer, for 94°C/2 min and 20 cycles of 94°C/30 sec, 60°C/20 sec, 72°C/30 sec), and gave a 303 bp product.

A second round of PCR was carried out using the forward primer:

BXPM.F AATTCATAAAATATTTTTAAA**CAG**CT

TARGET TTAAGTATTTTATAAAAATTT**TTG**GA Actual bovine *XIST* sequence

The forward primer BXPM.F annealed to its target with first cycle mismatches (bolded) to produce a *Pvu*II site. BXPM.R was used as the reverse primer in this second reaction.

Five to ten ng of PCR product from the first reaction was used as template in a reaction with a final volume of 25 µl (with identical amounts of PCR reagents to those used above). The PCR conditions used were 94°C/3 min followed by 25 cycles of 94°C/15 sec, 50°C/30 sec, 72°C/30 sec with a final extension of 2 min at 72°C.

Following PCR, 10U of *Pvu*II was added directly to the reaction and, after 3 h of digestion at 37°C, the resultant fragments were ethanol-precipitated, washed and re-suspended in TE buffer before final electrophoretic separation on a 2% TBE agarose gel (Figure 3.1).

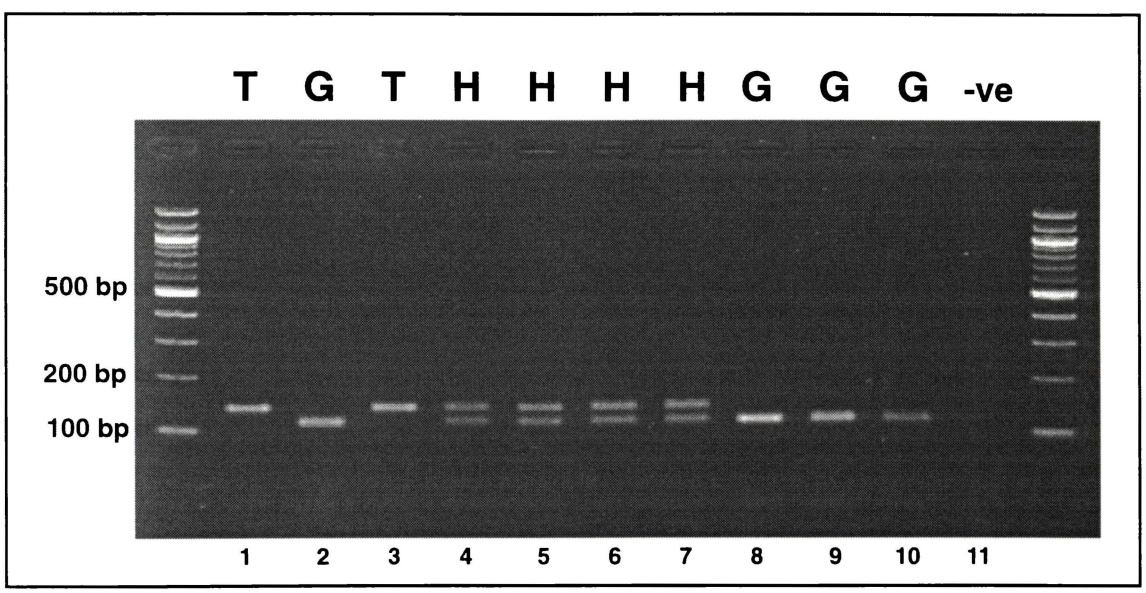


Figure 3.1. Sample TBE agarose gel (2%) showing a range of digested PCR products from homozygous (lanes 2, 8-10 showing 'G' allele homozygotes, lanes 1 and 3 showing the 'T' Allele homozygotes), and heterozygous (lanes 4-7) animals. Lane 11 is a negative control from the PCR and the lanes on the far left and right of the gel are 100 bp DNA ladder (New England Biolabs).

3.2.8 SNP2 frequency calculated on pooled DNA from a large number of animals

A population-wide assessment of allelic ratios was measured using a protocol described previously (Uejima *et al.*, 2000) with minor adaptations. A 142 bp fragment was amplified from a 1 µl aliquot (approximately 10 ng) of the first round template using the *Pvu*II restriction site primers BXPM.F and BXPM.R, as described above using 26 cycles of PCR. Finally, 0.2 pmol of a ^{33}P end-labelled reverse primer was added before the final cycle of PCR and an extension at 72°C was carried out for 5 min, after which the tubes were placed on ice (Figure 3.2).

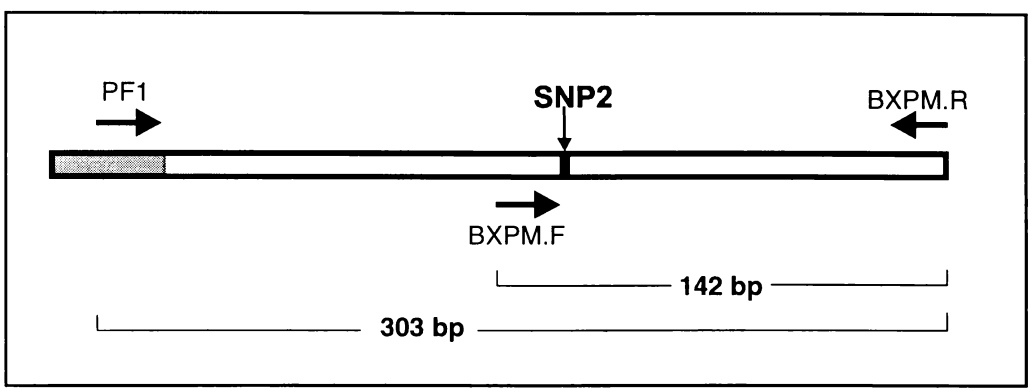


Figure 3.2. Diagram showing PCR strategy for both individual animal and mass-screen allelic ratio determination experiments. Screening of individuals was carried out using nested PCR with the primers: PF1 and BXPM.R, followed by an 'internal' PCR with the primers: BXPM.F and BXPM.R. The mass-screening experiment used a ^{33}P-5' labelled reverse primer (BXPM.R) in the final cycle of PCR before final digestion with the restriction enzyme *Pvu*II.

Templates used included DNA extracted from a retail consumer milk sample (see Chapter 2, section 2.2.3). The control for this experiment was a heterozygous DNA sample from animal #696.

After PCR, fragments were digested with *Pvu*II (as described above) and separated on a 0.75 x 80 x 80 mm 15% acrylamide gel (see Appendix 5) run at 20 V/cm for 1.5 h. After electrophoresis the gels were dried and exposed overnight to a phosphor-imaging plate. Digital images of the phosphor plates were collected and relative quantities of cut to uncut fragments determined for each of the samples using gel analysis software.

3.3 Results

3.3.1 pG6PD.B clone: AF531754

Preliminary sequencing of putative *G6PD* clones using M13 sequencing primers identified one clone that potentially contained the *G6PD* target region of interest. This clone, which was inserted in the 5'-3' orientation relative to the vector's Reverse primer site, showed similarity to exons 9-12 of previously published G6PD sequences. *Not*I restriction digest of this sub-cloned fragment and gel electrophoresis for fragment size analysis on a 0.7% TBE agarose gel, showed that the clone corresponding to the putative *G6PD* gene insert size was around 5 kb (Figure 3.3).

The clone was sequenced from the reverse primer end of the vector using a walking approach until the putative end of the gene was reached (that is, only the region similar to *G6PD* was sequenced). Sequence analysis was carried out using published Genbank sequences for reference. The sequence analysis is detailed in Appendix A.2.1.

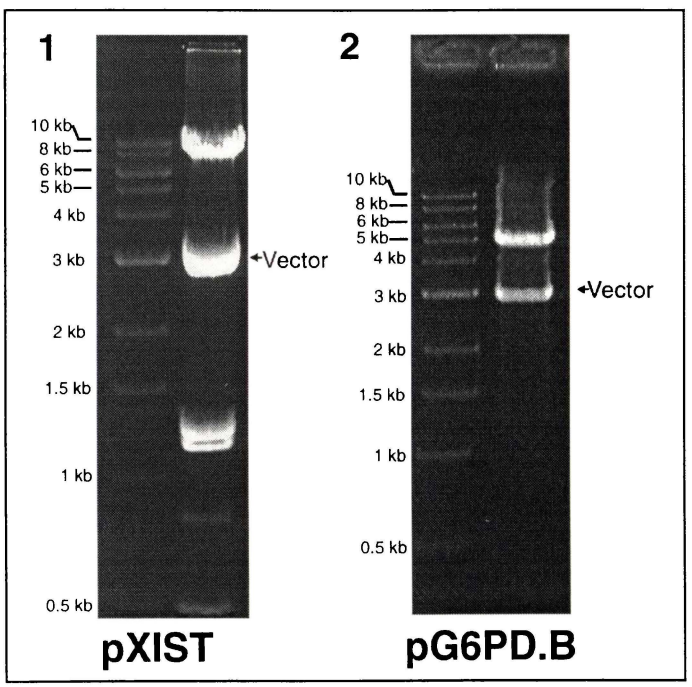


Figure 3.3. Photographs of 0.7% TBE agarose gels showing the selected λ clones after sub-cloning into pBluescript, plasmid purification and restriction digest with *Not*I (plus *Sal*I for pXIST). **1)** pXIST a fragment with total length of around 12 kb (represented as fragments of approximately 8 kb, 1.2 kb, 1.3 kb, 700 bp and 500 bp) and **2)** a bovine *G6PD*-like clone (p*G6PD*.B) comprising a 5 kb insert (digested just with *Not*I). The band at 3 kb is the linearized vector.

As detailed in the Appendix, this clone was later shown to be a likely pseudogene sequence, with similarity to published authentic *G6PD* sequences, and was annotated accordingly before being submitted to Genbank (reference gb: AF531754).

3.3.2 Bovine *XIST*

3.3.2.1 Sequencing *XIST*

In total, 2182 bp of *XIST* was sequenced from PCR products and submitted to Genbank (gb: AF104906) (See Appendix A.2.3.2 for sequence data and Chapter 2, section 2.6, for primer list).

3.3.2.2 BLAST confirmation of *XIST* sequence

BLASTN searching (http://www.ncbi.nlm.nih.gov/BLAST) returned 410 hits for the bovine *XIST* sequence (in section 3.3.1.1). These gapped results were predominantly short sequences but also included two *H.sapiens* (gb:AL353804, M97168), one leporine (gb:U50910) and an equine (gb:U50911) sequence, all of which have at least moderate overlap with the query sequence and some of which encompass its entire length. Similarity was also found to several rodent sequences including *Mus musculus* (Mouse; gb:MMU29341) and *Microtus rossiaemeridionalis* (Southern vole; gb:AJ310130).

The *Bison bonasus* (gb: L10727) sequence (that had the most weighting placed on it for the initial primer design) had 98% identity to *Bos taurus XIST* in the 650 bp overlap between the two sequences.

3.3.2.3 RT-PCR to confirm *XIST* sex-linkage

As shown in Figure 3.4, the *XIST* transcript was detected in the mammary and female liver samples but not in the male liver sample.

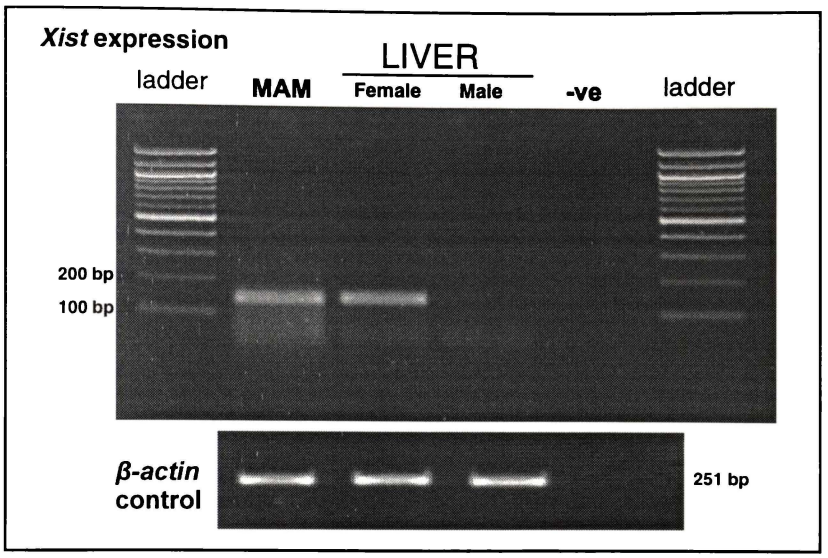


Figure 3.4. RT-PCR to confirm female-specific *XIST* expression and lack of expression in the male liver control tissue. A bovine *β-actin* product was amplified as a control and the expected product of this house-keeping gene (251 bp) is shown.

This RT-PCR result showed X-linkage of the *XIST* transcript as no *XIST* mRNA was detected in RNA extracted from male liver. Furthermore, expression was detected in the mammary specimen, which indicated that *XIST* would be detectable in the mammary specimens analysed in Chapter 4.

3.3.2.4 Northern blot

A Northern blot was carried out according to the protocol in section 4.2.2. The negative control (male liver enriched mRNA) was chosen, as male somatic cells are not expected to express *XIST* and an *XIST* probe should not give an autoradiographic signal in this sample. The suitability of the male liver mRNA negative control was previously confirmed by an RT-PCR reaction using RNA samples from the livers of both female (+ve) and male (-ve) animals that showed the expected female specific expression *XIST* expression profile and that the house-keeping gene control gene β-actin was detectable in all tissues examined (section 3.3.2.3). The blot was probed with a 303 bp *XIST* probe and the resultant autoradiograph is shown in Figure 3.5.

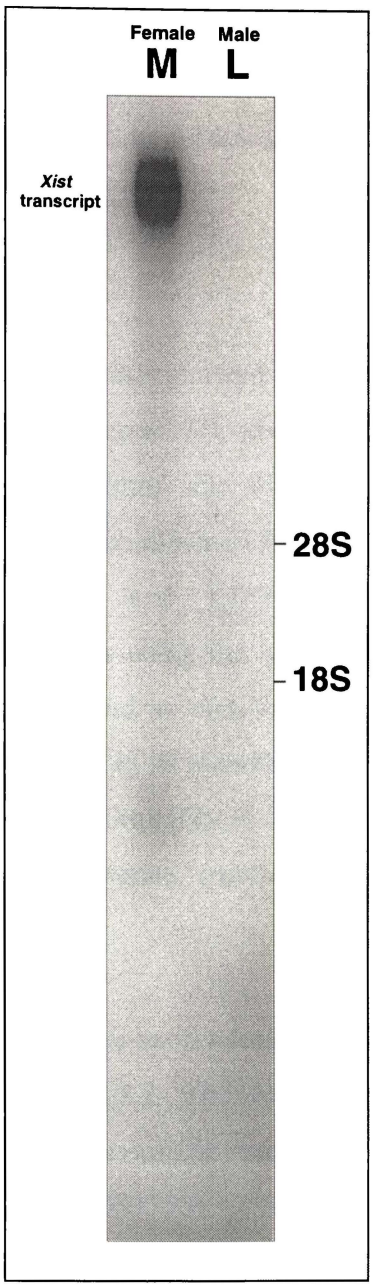


Figure 3.5. Northern blot of mammary and liver (male) enriched mRNA (0.5 μg loaded). A 303 bp α-[^{32}P]dCTP labelled DNA corresponding to a 5′ region of exon 1 of the *XIST* gene was used as a probe.

This Northern blot confirmed the findings of the RT-PCR experiment presented in section 3.3.2.3. Specifically, this blot showed that the *XIST* gene is expressed in a sex-specific pattern, with expression detectable in mammary tissue (female) but not in liver tissue (male). We endeavoured to repeat this experiment but had insufficient enriched RNA to do so. Ideally, this experiment should be repeated with: 1) a greater amount of enriched mRNA, 2) detection of a housekeeping gene to confirm the quality of the specimens loaded, and 3) an additional specimen of

enriched mRNA from female liver. As an alternative, real-time quantitative PCR (QPCR) could be used as, like Northern blotting, this would show relative levels of *XIST* expression and would confirm the sex-specific expression pattern.

3.3.2.5 pXIST clone: AF531755

0.7% TBE agarose gel electrophoresis showed that the putative *XIST* clone insert had a length of around 12 kb (Figure 3.3, above). The selected subclone was sequenced from the 'Reverse primer' site towards the insert to yield a total fragment size of 3562 bp. Sequence obtained from the M13 -21 direction showed database similarity to SINE and LINE-rich sequences. Somewhat disappointingly, Genbank searches using the sequence obtained showed that the presumed '*XIST*-positive' clone had no similarity to *XIST* gene sequences from other animals (or to the probe used in its identification). An *in silico* translation of this sequence showed significant similarity to the huntingtin-associated protein 1 (*HAP-1*) gene in mouse and human (mouse gb: NM_010404; human gb: NM_003949).

A complete sequence analysis, with exon/intron boundary prediction, was carried out and is described in Appendix 2.2.2. The clone was annotated as an HAP1-like pseudogene (from the limited available sequence data available) and was submitted to Genbank (gb: AF531755). X-linkage of this fragment was not confirmed.

3.3.3 SNPs in the *Bos taurus XIST* gene

Analysis of the 2182 bp of *XIST* gene sequence (gb:AF104906) obtained by direct sequencing revealed two SNPs.

3.3.3.1 Candidate *XIST* SNP 1 (SNP1)

Ten cows were screened for polymorphisms in the 2182 bp fragment of *XIST* exon 1 by direct sequencing, using ClustalX to compare the obtained sequences. A putative SNP designated SNP1, resulting in an *Mnl*I RFLP, was detected in some sequences (Figure 3.6). Unfortunately, this polymorphism could not be reliably detected. This possibly could be due to allelic dropout during PCR of the

region. However, this explanation is not likely, given that optimal amounts of template was used in the PCR wheras allelic dropout usually only occurs under suboptimal conditions when the total template is present at a low copy number and/or is of poor quality, for example, where PCR is performed on single cells. A number of alternative strategies were used in an attempt to eliminate this problem.

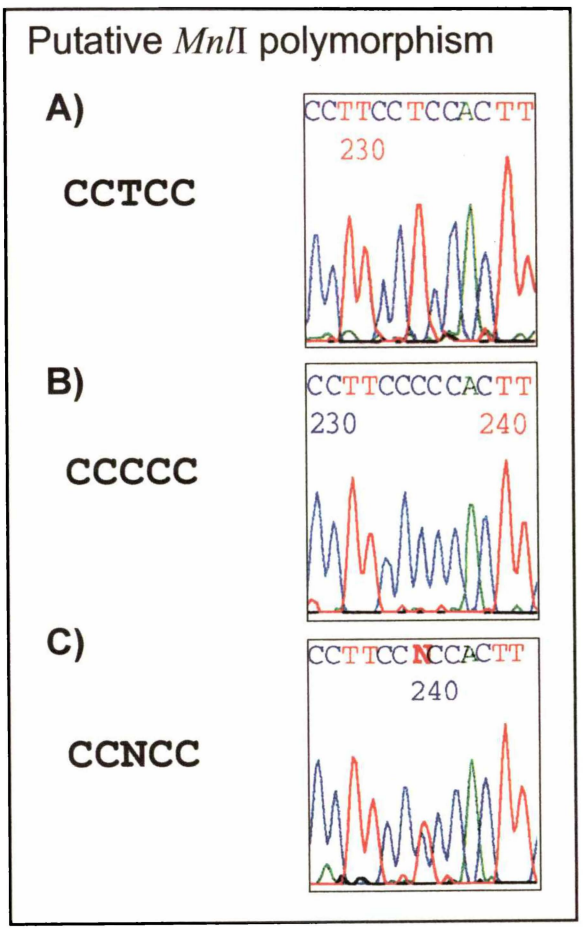


Figure 3.6. Partial electropherograms showing sequences from animal homozygous for the: a) 'C' variant b) 'T' variant and c) a heterozygous animal. While early sequencing efforts gave promising results, this polymorphism was not reliably detectable (images were cropped from complete electropherograms).

However, none of these methods, including the use of a proof-reading polymerase (*Pwo*) in equal proportion to *Taq* polymerase or PCR using an allele-specific PCR forward primer that bound immediately 5′ to the SNP, produced a reliable detection of both alleles of the SNP.

3.3.3.2 Candidate *XIST* SNP 2 (SNP2)

A second putative SNP was identified by direct examination of the sequence electropherograms obtained in 3.3.4.1 (Figure 3.7). This polymorphism has two alleles, designated as 'G allele' and 'T allele' to indicate that one variant has guanine while the other variant has thymine at position 188 of Genbank sequence AF104906 (see Appendix A.3.2).

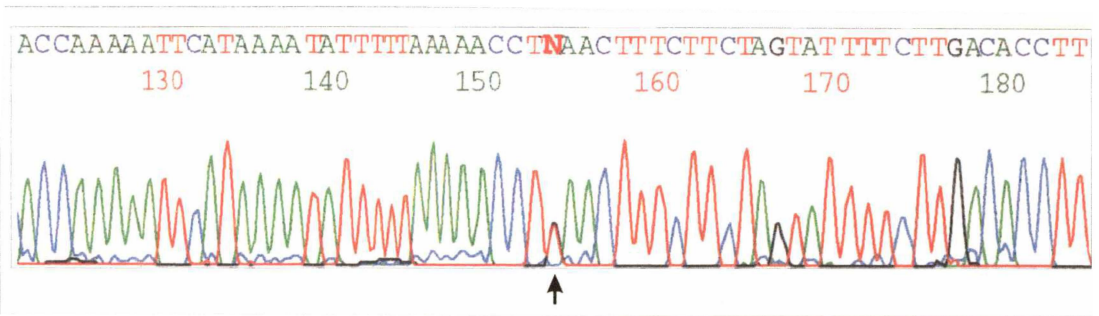


Figure 3.7. A partial electropherogram showing both the position of SNP2 and general sequence quality obtained for this region of the bovine *XIST* gene. The SNP is indicated by an arrowhead and heterogeneity is seen at this position as the template DNA for this sequencing reaction was sourced from a heterozygous animal.

Initial sequencing of the putative SNP-containing region was carried out at least 3 times to ensure high sequence accuracy and to confirm the authenticity of the polymorphism (a complete electropherogram for this region of *XIST* in a heterozygous animal is presented in Appendix 1).

3.3.4 Screening a population for SNP2 by PCR-RFLP

The PCR-RFLP protocol described in section 3.2.7 was applied to the polymorphic *XIST* loci of 88 female animals (representing 176 total alleles). The 'G allele':'T allele' ratio was found to be about 4:1, with frequencies of 0.8 and 0.2 for the G and T alleles, respectively.

While the numbers of animals screened were reasonably low (n=88), differences in the allelic frequency were apparent in the three different breeds examined. Friesian animals showed a frequency of 0.74 to 0.26 (n=58) while Jersey cows exhibited a frequency of 0.65 to 0.35 (n=20) for the G and T alleles, respectively. None of the Sahiwal animals examined in the screen (n=10) exhibited the T allele (Table 3.1).

Breed	No. of cows	Frequency of G allele	Frequency of T allele
Fresian	58	0.74	0.26
Jersey	20	0.65	0.35
Sahiwal	10	1	0
	88 total	Avg. 0.80	0.20

Table 3.1. Allelic frequencies for SNP2. SNP2 Allelic ratio calculation from a population of animals (n=88).

Total number of individuals with genotype data:	88
Population data sample size (number of chromosomes):	176
Average estimated heterozygosity:	32%

Average Allele Frequency:		Average Genotype Frequency:	
G	0.8	GG	0.64
T	0.2	GT	0.32
		TT	0.04

3.3.5 SNP2 Allelic ratio calculation on pooled DNA from a large number of animals

Samples of commercial milk were assessed for total allelic ratios (section 3.2.5) and, after comparison of the relative amounts of cut to uncut PCR product, were found to display an allelic ratio of approximately 1:5 for the T and G alleles of SNP2 respectively. This result is similar to that obtained by large scale screening of individual animals (Figure 3.8) but may overestimate the frequency of the G allele.

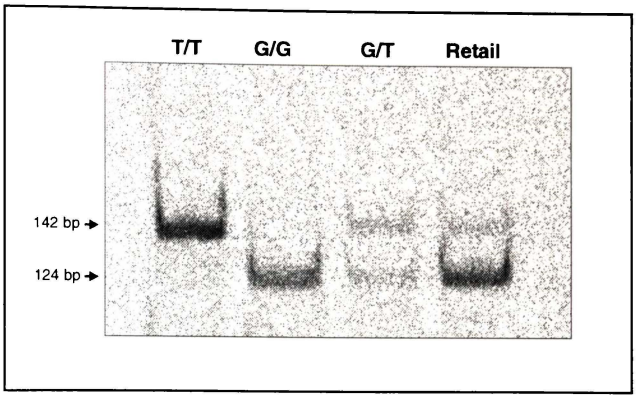


Figure 3.8. A 15% acrylamide gel phosphor image showing separation of digested PCR products in the allelic ratio experiment. The homozygous control samples (**T/T and G/G**) gave a single band after PCR, restriction digest and electrophoresis. The individual heterozygote sample gave a ratio of around 2:1 (G:T alleles) which deviates from the expected 1:1 ratio for this sample. The pooled retail milk sample (**Retail**) gave a ratio of approximately 5:1 for the G and T alleles, respectively. This showed a higher G allele than the allelic average determined by direct PCR-RFLP (with agarose gel analysis). The higher molecular weight fragment (142 bp) is uncut product (all T allele product) and the lighter fragment (124 bp) is completely cut (just G allele product). Also, see Figure 4.3 in Chapter 4, which shows the amplification bias in the G/T lane more clearly than in this figure.

3.3.6 Other potential bovine *XIST* SNPs

The bovine X-inactivation centre (XIC), which contains the *XIST* gene, was recently sequenced and has been published in Genbank in 2002 (gb:AJ421481) (Chureau *et al.*, 2002). This sequence was determined from a number of bovine genomic bacterial artificial chromosomes (BAC) (determined by 'shotgun' sequencing of sub-cloned sheared BAC fragments to 10-fold coverage) to produce a 233 kb XIC fragment. This fragment was subjected to a two-sequence BLAST search using AF104906 *XIST* (the partial bovine *XIST* sequence obtained in the present study) as a query. This search revealed an additional putative SNP. The putative A↔C SNP is located at position 116103 of AJ421481; position 17 of AF104906.

AJ421481: 116087 ttcttaaagcgctgcactttgctgcgaccgcca 116119 (Chureau BAC sequence)
AF104906: 1 ttcttaaagcgctgcaatttgctgcgaccgcca 33/EXON 1 start (our sequence)

A portion of the original electropherogram is shown in Figure 3.9.

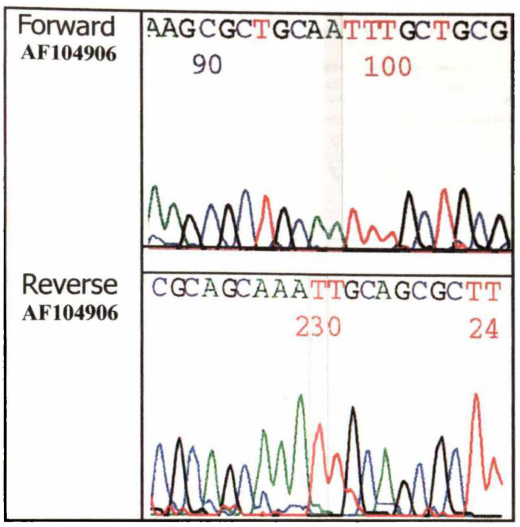


Figure 3.9. Electropherogram, in both forward and reverse directions, showing sequence data for AF104906 corresponding to SNP-containing region in the BAC sequence (AJ421481) and confirming a putative polymorphism A↔C SNP3 (Grey bars).

This corresponds to the promoter region of the bovine *XIST* gene located 16 bases 5′ to the start of the exon and therefore, while possibly evidence for an authentic SNP, is not useful in the proposed study (Chapter 4) as it would not be present in transcripts. The frequency of this SNP (designated SNP3) remains to be assessed at the population level.

3.3.7 *Bos taurus XIST* transcript features

This bovine *XIST* sequence was compared with similar regions in human, mouse and bovine (Figure 3.10). The comparison revealed marked sequence conservation, particularly of potential AG/CT splice site acceptor/donor sites,

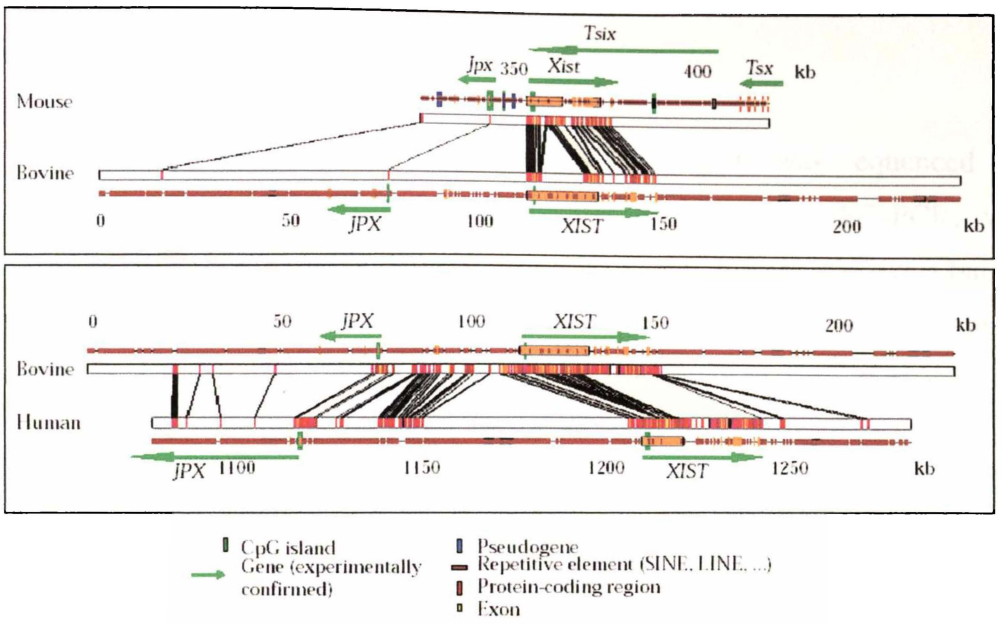


Figure 3.10. Showing conservation of gene orientation in the bovine, murine and human XIC. Used with permission from Chureau (2002).

between all three species. Bovine *XIST* is likely to share mouse exons 1-7, and possibly exon 8, which is immediately adjacent to the polyA signal region. While there has been no experimental confirmation of these exons in the bovine gene, their presence was confirmed in the human gene (with the exception of a predicted homolog of mouse exon 3). The bovine gene sequence and its predicted transcript, most closely matches the murine homolog and lacks human *XIST* exon 2 and 7 equivalents, as predicted by Chureau *et al.* (2002).

In summary, these sequence data indicate that the bovine *XIST* transcript size is likely to be between 16-19 kb and is more similar, structurally, to the murine than the human *XIST/Xist* gene.

3.4 Discussion

A 2182 bp (2.2 kb) fragment of the bovine *XIST* gene was sequenced and confirmed to be expressed in a sex-specific pattern using RT-PCR, with expression detectable in tissues from female animals (mammary and liver) but not in male liver tissue. The confirmation of *XIST* expression in bovine mammary tissue indicated that, if a suitable polymorphism was found to distinguish different alleles of *XIST*, the gene transcript could be used as a marker of inactive X chromosome choice in our study of cellular clonality in the bovine mammary gland (Chapter 4).

Three SNPs were detected within the 2.2 kb *Bos taurus XIST* fragment making this sequence highly polymorphic when compared to the human X chromosome in general; where about 1 SNP is found every 3.8 kb (Sachidanandam et al., 2001). Of the three putative SNPs identified in bovine *XIST*, one was confirmed to be authentic by PCR/RFLP-based screening of a large population of animals. This SNP (SNP2) was selected for use in our bovine mammary gland clonality experiments (Chapter 4). Assuming that the 1x sequencing coverage used in this study gave accurate results, the SNP frequency, at about 1 SNP per 700 bp, may reflect weak sequence constraints within the analysed sequence. It is interesting to note that none of the SNPs identified in this study are located in the *XIST* 5′ repeat region, which has been shown to be centrally important to transcript function (Beletskii et al., 2001). While this is likely to mean that the SNPs identified here are functionally silent, it is possible that SNPs within the *XIST* gene or promoter affect the initiation and establishment of X inactivation by somehow affecting either the "randomness" of X chromosome choice or transcript function. While it has not been fully characterized, SNP3 may be particularly interesting in this respect, as it is found in the promoter region of the *XIST* gene.

In a screen of 88 animals of a variety of breeds, the *XIST* G allele was found at 80% of SNP2 loci with the T allele accounting for the remaining 20% of loci. This is, therefore, a reasonably common polymorphism and, as around 32% of animals are heterozygous for the SNP, the sourcing of informative heterozygotes is very simple.

The mass-screening allelic screening experiment (section 3.3.4), which effectively averages out of differences in allelic frequency between breeds, revealed a similar ratio of T to G alleles to that found in the individual animal screening for SNP2 (presented in section 3.3.5). One assumes that retail milk represents the pooled somatic cells from the thousands of animals commercially milked in a large collection area; maybe 2-10 X 10^3 animals. The results of this screening indicated that in this large female population, the G and T alleles are present at a ratio of approximately 4:1. Unfortunately, however, the PCR approach used was found to have an amplification bias to the G allele and this placed limitations on the usefulness of this technique at the population level. This bias is examined further and discussed in the next chapter.

Future investigations

The experiments in this chapter identified a highly heterogeneous SNP for use in the clonality studies presented in the next chapter. However, more direct approaches would have improved the efficiency of screening for both the targeted gene sequences and for SNPs within these sequences.

The titre of the λ library used in this study was low (1.4 X 10^5 pfu/μl) and would not have been fully representative. In fact, considering the low titre, screening of this library should probably not have been attempted. However, by using a library of suitably high titre, an *XIST* exon 1 probe applied to a bovine genomic DNA library is likely to have revealed an authentic *XIST* clone. Exon 1 of *XIST* in both human and mouse is an extremely long exon (about 10kb), so a positive clone identified in this way is likely to contain a large stretch of *XIST* "coding" sequence.

Alternatively, a bovine mammary cDNA library or EST database could have been used. This would have almost certainly prevented the isolation of pseudogenes, which was found to be a problem when using the genomic library. The fact that a library sourced from mammary tissue would also immediately confirm the expression of these genes in the mammary gland is an additional advantage. After

their identification in a library, the X-linkage of these genes could be confirmed using expression analysis in male and female animals. Then, techniques such as single strand conformation polymorphism (SSCP) analysis, originally used by Orita et al. (1989), could be used to identify polymorphic DNA, and SNPs could be subsequently identified by sequencing these polymorphic templates. Finally, the frequency of specific SNPs could be determined by allele-specific PCR or single nucleotide primer extension (SNuPE), and suitable SNPs selected for use in clonality studies.

Chapter 4

The size and distribution of cellular clones in the bovine mammary gland

4.1 Introduction

The measurement of the size and distribution of cellular clones in bovine mammary tissue is important for a number of reasons. First, there is potential for economic gain through improvements to gland development and the generation of more persistent lactation by manipulation of growth from mammary stem cells. Second, close anatomical similarities between the bovine and human glands make the bovine gland a possible model for human breast cancer (Ellis and Capuco, 2002). Third, and although this may seem paradoxical given the statement just made above, the fact that bovine mammary tumours are extremely rare (Povey and Osborne, 1969; Swett *et al.*, 1940) suggests that there may be critical differences between human and bovine stem cell regulation, and the elucidation of these may aid research into the prevention and treatment of human breast disease (Ellis and Capuco, 2002).

While cellular clonality in the bovine gland has never been directly studied, Novelli *et al.* (2003) used histochemistry to identify the extent of homogeneity in sections of human breast tissue by determining the extent of expression of a temperature sensitive G6PD variant (the Sardinian mutant). This *in situ* study showed homogeneity in all 111 small ducts and 57 lobules examined with heterogeneity (polyclonality) detected in just a single large duct. This study supported the idea that many terminal ductal alveolar units (TDLU) have a monoclonal origin. However, *in vitro* studies using PCR based assessment of X-linked markers have been inconclusive, with Tsai *et al.* (1996) and Noguchi *et al.*

(1992) producing data suggesting that the small ducts and TDLU in the human gland were, respectively, monoclonal and polyclonal in origin. Peng *et al.* (2000) obtained data that suggested general clonal heterogeneity in the mammary gland, using one-step RT-PCR and *Pvu*I discrimination of a G6PD polymorphism to show both heterogeneous and homogeneous patterns of clonality in microdissected samples collected from normal and carcinoma tissue, respectively. Considered together, these data indicate that monoclonality exists in individual TDLU but that individual cellular clones or patches do not extend over large portions of the gland.

In the light of these human data, the working hypothesis being explored in the current work is that the bovine MG will display a similar clonal profile to that observed in the human MG.

In order to measure the size and distribution of cellular clones in the bovine MG and to draw comparisons between the clonal profiles in the two glands, we chose to use an X chromosome inactivation (XCI) based approach. Methods based on XCI are commonly used in clonality studies and reveal cellular clonal profile *in vitro* by comparison of ratios of expressed alleles of X-linked genetic polymorphic in heterozygotes (Busque and Gilliland, 1996; Shibata *et al.*, 1996; Tsai *et al.*, 1996).

Small blocks of bovine mammary tissue were examined to ascertain whether they were clonal in nature using a direct RT-PCR/RFLP protocol.

This approach, first described by Uejima *et al.* (2000), measures ratios of expression of two allelic variants of an expressed SNP by incorporation of a radioactively-labelled primer in the last cycle of PCR followed by allele-specific digestion of the reaction products and measurement of ratios of the two alleles by polyacrylamide gel electrophoresis (PAGE). The bovine *XIST* SNP used in the current study, SNP2, is an expressed G/T difference close to the 5′ end of exon 1, as described in Chapter 3.

4.2 Methods

All RNA extractions and enzymatic manipulations were carried out in solutions made using diethylpyrocarbonate (DEPC) treated deionized double-distilled water.

4.2.1 Sample collection and storage

Fresh mammary and liver tissue samples were dissected into ~2 cm^3 blocks, snap-frozen in liquid nitrogen, and stored at -70°C until required.

4.2.2 RT-PCR/RFLP – Hot Stop method

In this method, total RNA was extracted from a series of small mammary tissue blocks; each was 30 mm^3 and comprised around $3X10^7$ cells. First strand cDNA synthesis with random hexamer primers was used to produce cDNA (section 4.2.2.3). A fragment of the *XIST* cDNA was amplified by PCR with primers designed to introduce a *Pvu*II restriction site in products from one of the expressed variants of the gene (allele 'G'; as described in Chapter 3, section 3.2.7). A 5′ ^{33}P[γdATP] labelled reverse primer was added before the last cycle of PCR to label all reaction products. RFLP was performed using the restriction enzyme *Pvu*II and this revealed either heterogeneity (indicating polyclonality) or homogeneity (indicating possible monoclonality) by examination of digested products on a 15% PAGE gel (Figure 4.1).

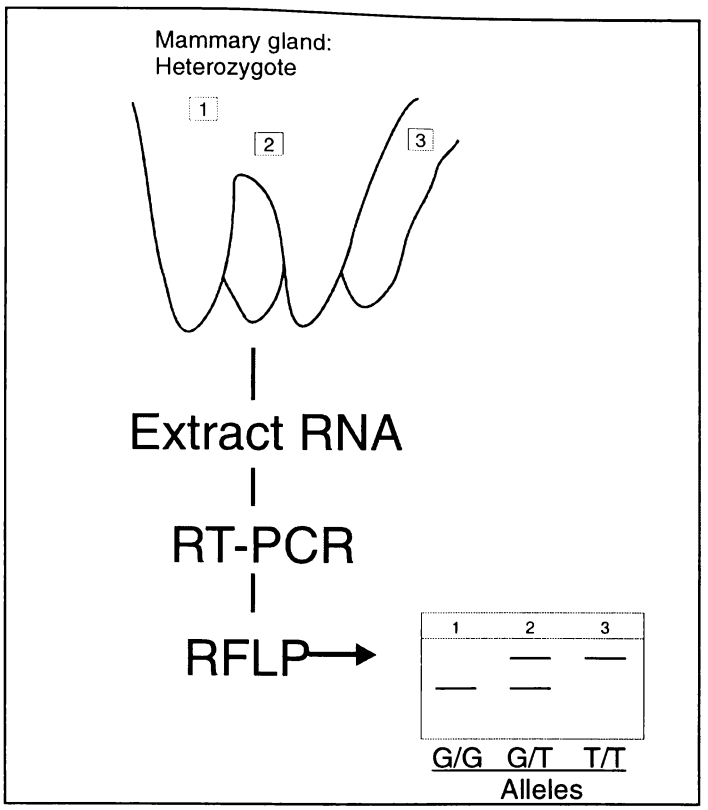


Figure 4.1. Strategy used to measure allelic ratios in MG samples. Total RNA was extracted from bovine mammary tissue and RT-PCR used to amplify a diagnostic *XIST* fragment. Restriction enzyme digest of the fragment and PAGE shows a banding pattern that reflects the ratio of alleles in the tissue. RFLP produce three possible banding patterns: 1) just the 'G'-allele, indicating homogeneity, 2) heterogeneity, indicating polyclonality, and 3) homogeneity for the 'T' allele.

4.2.2.1 RNA extraction from tissue blocks – estimated cell number

Total RNA was extracted from approximately 30 mm^3 cubes of frozen tissue (see Chapter 2, section 2.2.2) and was resuspended in 50 µl of DEPC-H_2O. Each of these tissue samples weighed around 30 mg. Assuming that 1 g of tissue contains around 10^9 cells, total RNA from approximately 3.0 X 10^7 cells (10^6 cells/mm^3) would be present in each resuspended total RNA sample.

In addition, 10 µm thin sections of mammary tissue were collected from paraffin blocks. Each of these 1 cm^2 thin sections contains a total around 1 mm^3 of tissue, which corresponds to approximately 10^6 cells (section 4.2.2.1).

4.2.2.2 DNase treatment of total RNA

DNase treatment of extracted RNA was carried out to ensure that no contaminating DNA was present in the sample used in the RT-PCR experiments.

One microlitre of a 10x Mn^{2+} buffer (6.6 mM $MnCl_2$, 100 mM Tris pH 7.8) and 1 µl (1U) of RNase-free DNase (Promega; cat.# M610A) was added to 8 µl of total RNA and the reaction incubated at 37°C for 30 min. The reaction was stopped by adding 1.8 µl of 25 mM EGTA (final concentration of 2 mM) and incubation at 65°C for 10 min. An aliquot of this RNA was used in first strand cDNA synthesis.

4.2.2.3 First strand cDNA synthesis

A 2 µl aliquot of DNase-treated RNA was added to a PCR tube with 0.5 µl of random hexamer (50 ng/µl) and 7.5 µl of DEPC water, mixed thoroughly and incubated at 75°C for 5 min before being placed on ice. After 5 min, 2 µl of 10X synthesis buffer, 4 µl of 25 mM $MgCl_2$, 2 µl of 0.1M DTT, 1 µl of 10 mM dNTPs and 0.5 µl of SuperScript™III Reverse Transcriptase (200 U/µl) was added to each tube, the tubes incubated at RT for 10 min then at 50°C for 1 h. The reaction was stopped by incubation at 70°C for 5 min.

4.2.2.4 PCR of *XIST* SNP diagnostic fragment

A 1 µl aliquot of the cDNA produced by first strand synthesis was added to a 25 µl PCR reaction with 2.5 µl of 10X PCR buffer (Roche), 2 µl of dNTPs (2.5 µM), 17.5 µl of high quality deionised water, 0.625 µl of Taq Polymerase (0.625 U), and primers for the desired *XIST* product, and amplified using 35 cycles of PCR: 94°C/30 sec, 50°C/30 sec, 72°C/30 sec. The primers used, BXPM.F and BXPM.R, are described in Chapter 3, section 3.3.5.

The 'hot-stop' PCR reaction was completed by including end-labelled (see Chapter 2, section 2.2.4) reverse primer (BXPM.R) in each reaction tube, to a final concentration of 1 pmol, and carrying out a final cycle of PCR. This prevents the production of labelled heteroduplex products, which could be produced if the labelled reverse primer was incorporated at an earlier cycle of PCR.

4.2.2.5 RFLP

PCR products were digested by the direct addition to a 25 µl PCR reaction of 5 U of the restriction enzyme *Pvu*II followed by brief mixing and incubation at 37°C overnight. After digestion, the PCR products were loaded onto a 15% TBE acrylamide gel (see Appendix 5) and electrophoresed at 180 V (18 V/cm) for 1.5 h.

Gels were dried using a flat bed gel drier (Biorad) and exposed to phosphor plates for 2-12 h (as described in Chapter 2, section 2.3.1).

4.2.2.6 Direct extraction of RNA from 10 µm tissue sections for RT-PCR

Multiple 10 µm sections were collected from paraffin blocks of MG from each of three animals, two *XIST* G/T heterozygotes and a homozygous *XIST* G allele animal. Next, total nucleic acids in the samples were extracted according to the protocol described by Frank (1996) who compared nine different extraction protocols and found that the following protocol gave the best yield of RT-PCR amplifiable RNA.

The sections were placed in 1.7 ml tubes, 100 µl of 0.5% Tween-20 (DEPC) was added and the tubes agitated briefly before incubation at 90°C for 10 min. After the samples had cooled, 100 µl of 5% Chelex (in TE buffer) was added to each sample and the tubes incubated at 99°C for 10 min. Next, the tubes were briefly vortexed and centrifuged at 11 000 *g* for 15 min. Following centrifugation the residual wax was removed from the surface of the solutions and, after heating the samples to 45°C, 100 µl of chloroform was added. The solutions were gently mixed by finger-flicking and were again centrifuged at 11 000 *g* for 10 min. The upper phase was collected and the nucleic acids concentrated by precipitation in a $1/10^{th}$ volume of 3 M sodium acetate (pH 5.5) and 3 volumes of 100% ethanol (AR). This RNA was used as template in a first strand extension reaction as detailed in section 4.2.2.3.

4.2.2.7 PCR controls

In Chapter 3, an amplification preference for the 'G'-allele was noted. To measure this effect, and to determine the influence of contaminant polyclonal tissue on the ability to detect clone homogeneity in tissue blocks, two control experiments were performed:

1) Firstly, a heterozygous DNA template was amplified using 25 cycles of PCR so that the unequal amplification of the alleles of SNP2, which was seen in earlier experiments, could be quantified.

A genomic DNA template was diluted 1/10 (200 ng/µl), 1/20 (100 ng/µl), 1/30 (66 ng/µl) with water, and a 1 µl aliquot of each dilution was PCR amplified in a 25 µl reaction as described in section 4.2.2.4. The PCR products were digested and electrophoresed using 15% PAGE to determine allelic ratios. The results of this experiment are shown in section 4.3.4.1.

2) To determine whether contamination from non-epithelial polyclonal regions within the sampled tissue would affect detection of clone homogeneity in the glandular tissue a further control experiment was carried out.

A PCR product was amplified from two male animals, hemizygous for the T and G alleles, respectively, using the primers BXPM.F and BXPM.R. The product was diluted to the ratios indicated in Table 4.1. As the detected allele was also present in the "contaminant", the actual basic ratios of "contaminant" to "sample epithelia", of between 1:3 and 1:7, were effectively equivalent to ratios of between 1:3.5 and 1:7.5. Fifty nanograms of this mixed "contaminant:epithelia" template was reamplified using 20 cycles of PCR (as detailed in section 4.2.2.4) and digested with *Pvu*II (5 U) directly in the PCR reaction buffer. The effect of the inclusion of the contaminant mock "polyclonal tissue" was determined by comparison of the ratios of products obtained from each PCR reaction. These are presented in section 4.3.3.2.

Sample	Ratio	Spiked Contaminant	'Epithelial'
1	1:3	0.5 G/0.5 T	3 G
2	1:3	0.5 G/0.5 T	3 T
3	1:4	0.5 G/0.5 T	4 G
4	1:4	0.5 G/0.5 T	4 T
5	1:5	0.5 G/0.5 T	5 G
6	1:5	0.5 G/0.5 T	5 T
7	1:6	0.5 G/0.5 T	6 G
8	1:6	0.5 G/0.5 T	6 T
9	1:7	0.5 G/0.5 T	7 G
10	1:7	0.5 G/0.5 T	7 T

Table 4.1. Table showing the ratios of template used in control experiments. These experiments were designed to determine whether inclusion of connective tissue in the epithelial sample would affect the resolution of the proposed RT-PCR/RFLP experiments.

4.2.2.8 Calculation of percentage epithelia in mammary blocks

Photographs were collected from haematoxylin/eosin stained mammary gland sections from each of the animals in the study. Epithelial cell numbers were then measured relative to cells within the connective tissue component of the sections by counting nuclei of each cell type in five 400 μm^2 frames within each photograph. This count was then used to calculate the average percentage contribution for both stromal and epithelial cells in the sections.

4.2.2.9 Animals used in the study

The primary study used mammary tissue from cull animals (animal/genotype: 696 G/T, 5780 G/G and 6244 G/T) and liver samples from a heifer and a male animal. In an extension to this study, mammary samples from a further three SNP2 heterozygous heifers were analysed by the same RT-PCR/RFLP method applied to the samples from the above animals. The heterozygous control used in PCR optimisation was DNA from animal 696 (a heterozygote), and the two hemizygous control samples for the 'T' and 'G' alleles were from male animals.

4.3 Results

4.3.1 Calculation of percentage of epithelia in mammary blocks

Analysis of cell nuclei in haematoxylin and eosin stained mammary tissue sections revealed that only around 80% of the cells in mammary gland sections were epithelial cells. Because there were considerable variations from animal to animal, analyses were carried out on a number of glands from which total RNA was to be extracted for the RT-PCR experiments (Table 4.2). The table shows a 'corrected ratio' determined by measuring the actual areas of lobular tissue, comprised of epithelial cells, and of connective tissue (CT) in each sample. This correction factors in relative density of cells in epithelia and CT with the area of the sampled sections in which each cell type is found. These estimates were used when interpreting the results of the control experiments, which are relevant to the potential resolution of the Hot-Stop method.

Animal Sampled	Ratio of cell density (CT:Epithelial nuclei)	% Area epithelia	Corrected ratio of CT:epithelia
696	1: 4.8	66.3%	1:5.6
5780	1: 4	79.5%	1:5.18
6244	1: 3.25	83%	1:4.3

Table 4.2. Table showing ratios of nuclei found within connective tissue and epithelial components of mammary gland sections that correspond to the blocks sampled for the RT-PCR experiments. CT: Connective tissue.

Images collected from thin sections of each of the glands are shown below (Figure 4.2). These images show both the extent of individual lobules (in two dimensions) and the relative numbers of epithelial and stromal cells.

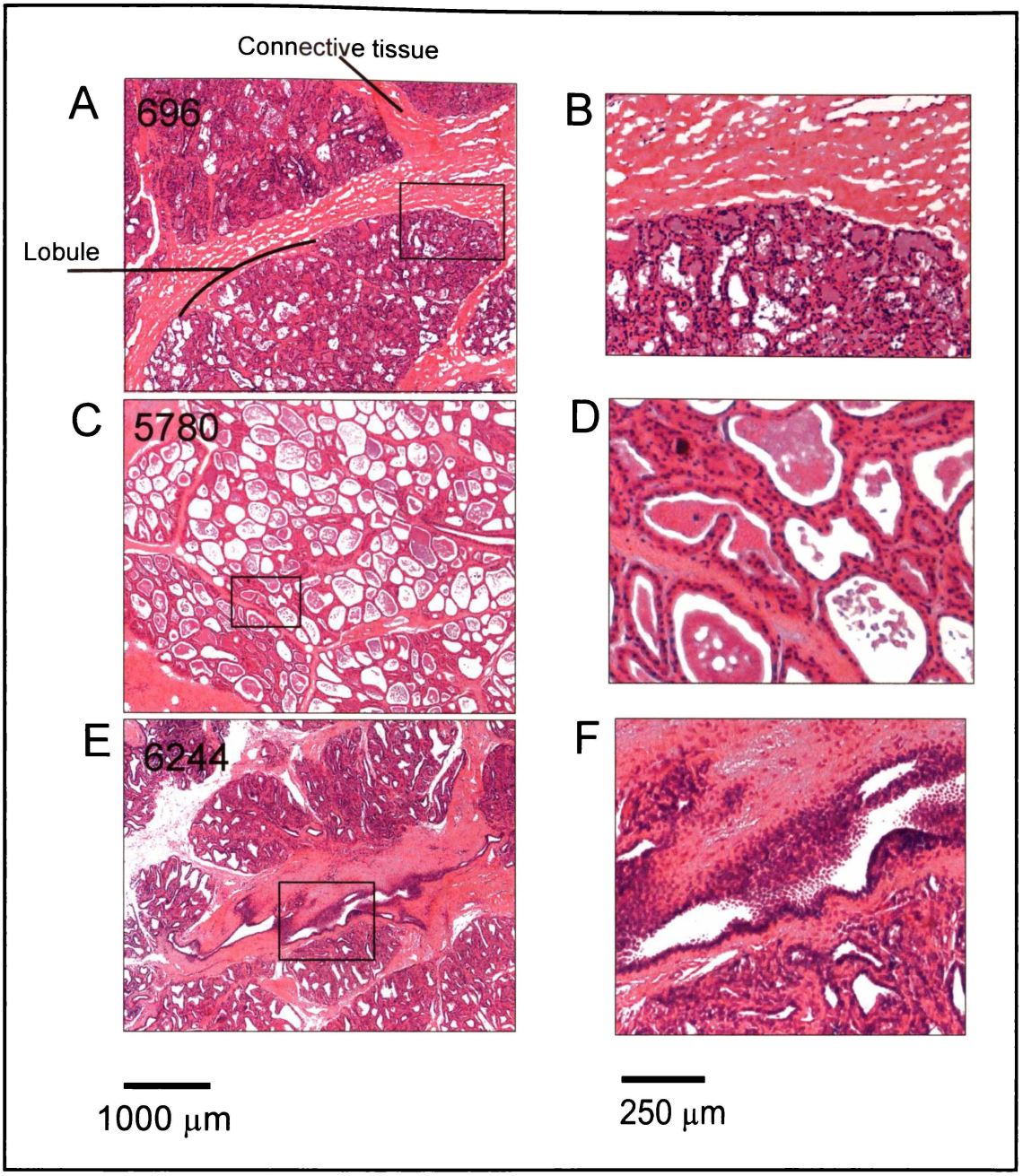


Figure 4.2. Bovine Mammary Gland Sections. The extent of lobular bundles and distribution of nuclei within the mammary glands of the animals can be seen. Images A, C and E are from glands 696, 5780 and 6244 respectively. Images B, D and F are higher magnification images from the same glands, 696, 5780 and 6244, respectively.

4.3.2 Control experiments for RT-PCR/RFLP method

4.3.2.1 Xist SNP2 Heterozygote DNA PCR/RFLP

In earlier work, preferential G allele amplification was observed. This was quantified by PCR (Figure 4.3) and, after 25 cycles, twice as much G as T allele

product was found using Gelstar 4™ software quantification. Although different PCR conditions were tried (annealing temperature, cycle time etc.) it was not possible to correct for this preferential amplification (data not shown).

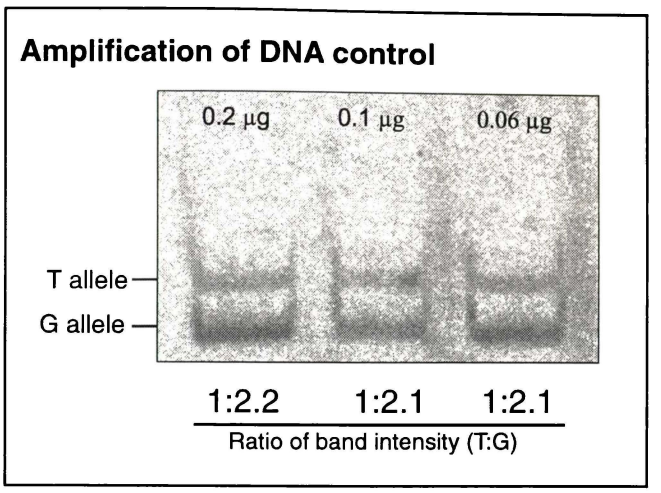


Figure 4.3. Amplification of a heterozygous DNA (from animal #696) as a control. After 25 cycles of PCR, twice as much 'G'-allele product as 'T'-allele product was observed. The amount of genomic DNA loaded to each reaction is indicated on the figure as is the ratio of band intensity.

4.3.2.2 Contaminant spiking experiment

The methodology used in the contaminant spiking experiment, using different ratios of 'contaminant' to 'target' is described in section 4.2.2.7. The results of this experiment are shown in Figure 4.4.

The results of this experiment were difficult to interpret, especially in light of the known amplification bias, which favours the G allele. Where the G-allele predominated, co-amplification of a T-allele product was always observed. Furthermore, amounts of this co-amplified product were not seen to decrease with increasing relative concentration of the G allele. It is perhaps possible that incomplete digestion of the products meant that the G allele product band was underrepresented in these lanes. However, this seems unlikely as this RFLP method had previously been used to successfully genotype animals (including males) in the individual animal screening experiments in Chapter 3.

Where the T-allele was predominant, at a ratio of between 1:5.5 and 1:7.5 of "stromal (contaminant) cells" to "epithelial cells", a product corresponding to the

T allele was found to predominate, with only minimal amplification of the contaminating G allele. Even when the contaminant was present at a ratio of 1:3 to the target T allele, effectively meaning that the T allele template was 3.5 fold the amount of the G allele template, there was clearly more T allele product than G allele product. The T allele product was found in increasing amounts with decreasing amounts of contaminant. As noted above, this was not true of the reactions where the G allele was predominant.

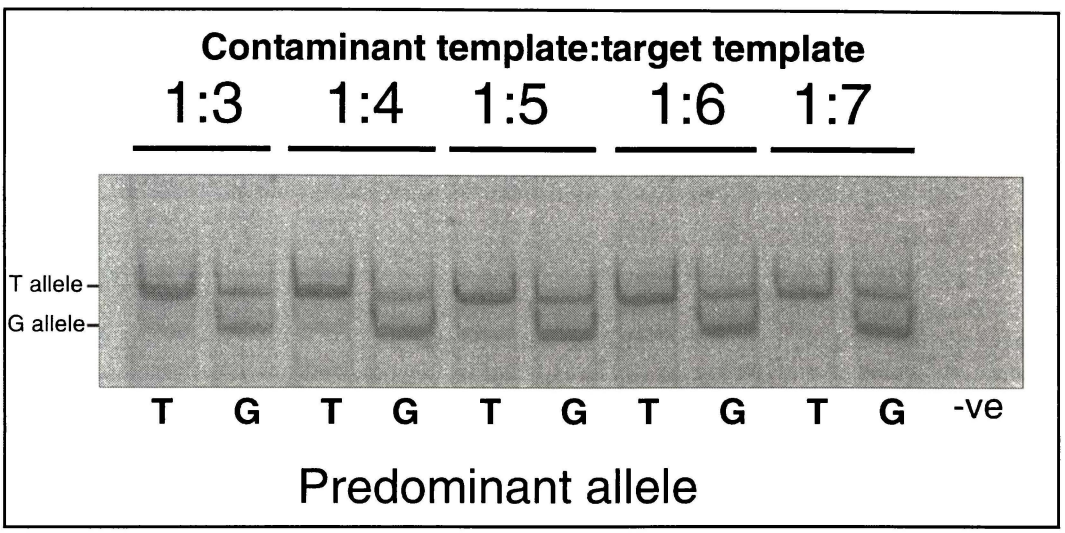


Figure 4.4. Allele ratio banding patterns observed using different amounts of a mixed 'contaminant' template with 'desired' template. The ratio of contaminant to target template is shown above the gel picture while the predominant allele is indicated below the picture. A PCR negative control is shown at the extreme right of the gel picture.

If the same allelic ratio as that observed in PCR amplification from a heterozygous genomic DNA template (see 4.3.2.1) was found in this experiment, it would be expected that the amplification preference should have doubled the amount of G allele product relative to the T allele product. Clearly, this was not the case.

This contaminant spiking experiment showed that inclusion of stromal tissue with epithelial tissue in the mammary blocks sampled may affect the resolution of homogeneously 'G'-allele expressing samples (Figure 4.4). The apparent preferential amplification of the T allele, which is the opposite result from that expected in this PCR, is difficult to explain.

4.3.3 RT-PCR/RFLP showed differential allelic expression

Figures 4.5 and 4.6 show gel pictures of alleles expressed in older cull animals and heifers, respectively. Tables 4.3 and 4.4 show this data in tabulated form with an indication of relative amounts of each allele detected in heterozygous animals.

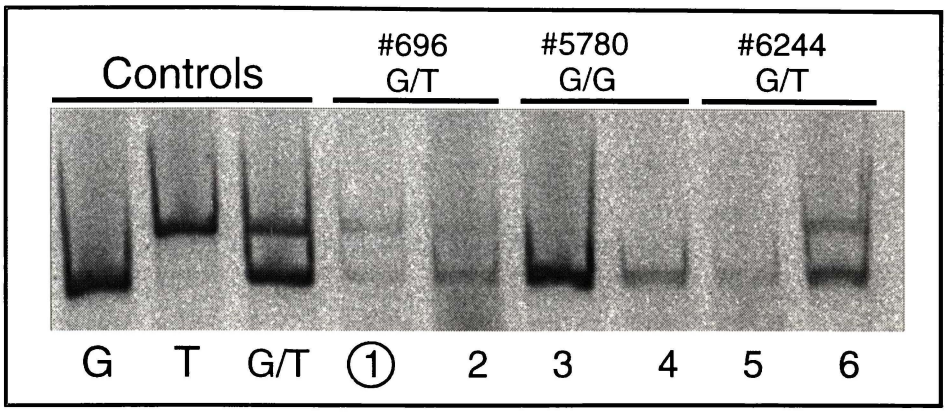


Figure 4.5. Alleles expressed in cull animals. Three controls are included: homozygotes for the 'G'-allele and the 'T' allele, and an amplification product from a heterozygous sample. Actual tested mammary RNA samples are sourced from: 1) and 2) animal #696, a heterozygote, samples 3) and 4) animal #5780, a homozygote for the 'G' allele, and 5) and 6) animal #6244, a heterozygote. (See Figure 4.8 for controls). The circled number corresponds to the only suspected homogeneous specimen from the 16 blocks sampled.

Animal	Genotype	Sample 1	Sample 2
696	G/T	g/T	G/t
5780	G/G	G	G
6244	G/T	G/t	G/T

Table 4.3. Table showing expressed alleles in random mammary gland samples in heterozygous and homozygous animals. The relative intensity of each amplified band is shown by the case of the letter used in the results boxes. The letter case used indicates relative band intensity.

Patterns of clonality in MG of heifers

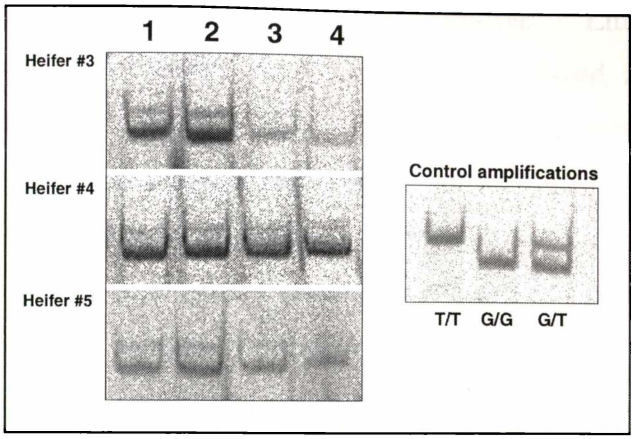


Figure 4.6. SNP2 Alleles expressed in five samples from three individual heifers. Control amplification/RFLP from a DNA template is shown in the inset panel.

Animal	Genotype	1	2	3	4
3	G/T	G/t	G/t	G/t	G/t
4	G/T	G/t	G/t	G/t	G/t
5	G/T	G/t	G/t	G/t	G/t

Table 4.4. Table showing detected alleles in samples collected from three heifers. In all samples, the G allele preferentially amplified but due to limitations of the PCR strategy used it was impossible to determine if this digestion pattern reflected genuine homogeneity (for exclusively G allele expressing regions) in the sample.

Analysis of the 30 mm^3 samples from 6 animals gave largely equivocal results. In nearly all cases, the G allele predominated, but it could not be determined whether this was due to preferential amplification during the PCR or actual homogeneity. In only one sample, sample 1 from animal #696, was evidence for homogeneity of the T allele found. If this result is viewed alongside the preferential amplification for the G allele found in the first control experiment (section 4.3.2.1), then it may be significant, as it would indicate that there was a predominance of cells expressing the T allele in this specimen.

At least some of these technical problems would have been overcome if it had been possible to RT-PCR from much smaller tissue samples. However, efforts to use template extracted from 1 mm^3 tissue samples failed to give reliable PCR amplification (data not shown).

4.4. Discussion

The results obtained here are clearly not definitive. Little evidence for homogeneity (suggesting possible monoclonality) was found in samples taken from bovine mammary glands. However, the fact that these samples were relatively large (30mm^3, corresponding to approximately 3 x 10^7 cells, containing an estimated 27000 TLDUs, at approximately 1 TDLU per 1 mm^3 and 1000 cells per TDLU) might mean that smaller areas of homogeneity, <30 mm^3, would not be detected. Interesting though, if one was to accept that one sample out of the sixteen analyzed (that is, specimen 1 for animal #696; see Figure 4.5) was actually displaying homogeneity, then probability considerations, with 1 of 16 30 mm^3 regions shows homogeneity so 1 of 8 15mm^3 regions will show homogeneity etc., would predict that, on average, regions of around 2 mm^3 in the bovine gland must be homogeneous. If this is the case, then these regions would be larger than those shown to be homogeneous in the mouse (Thomas *et al.*, 1988). However, given that the bovine gland has a weight of approximately 10 kg, compared with approximately 0.7 g for the mouse gland, it is perhaps likely that much larger areas of the bovine gland have a monoclonal origin (the standard formula for calculating mammary weight [kg] is: 0.045 X W $^{0.82}$; where W=maternal bodyweight). This would particularly likely if the numbers of progenitors in both glands are approximately equal. If, however, the bovine gland is derived from a larger progenitor population, say 1000 fold that found in the mouse gland, then it would be unlikely that extensive regions of homogeneity would be found in the bovine gland. If the size of the progenitor population for both glands is found to be similar, then a transit-amplifying cell population may be particularly important in the bovine gland. Future studies on glands from these species are likely to examine the relative size of the stem cell progenitor population and possible differences in involvement of a transit-amplifying cell population in gland growth.

It is interesting to note that, in the 0.7 g murine mammary gland, analysis of epithelial patch sizes show that the gland is predominantly heterogeneous, with homogeneity only observed in some terminal ducts (Kisseberth and Sandgren, 2004; Thomas *et al.*, 1988), whereas, in the ≈1500 g human gland, homogeneity is reported in individual TDLU (Novelli *et al.*, 2003; Tsai *et al.*, 1996). In these

studies of the human breast, heterogeneity, indicating polyclonality, was seen only in larger ducts.

Clearly, it would be preferable to measure clonality in individual TDLU or, at the very least, just a few TDLU, in a sample of ≤ 1 mm^3 (the approximate size of a single TDLU). Attempts were made to analyze the clonal profile in single 10 mm^2 sections, comprising around 1 mm^3 of tissue (10^6 cells), but these were not successful, as it proved impossible to extract sufficient RNA from these tissues for RT-PCR. If amplification had been possible, then such specimens are more likely than larger blocks to reveal possible homogeneity in expression of the SNP marker.

Two technical problems appeared to limit the RT-PCR analyses, the amplification method and the low level of *XIST* transcripts. The RT-PCR/RFLP technique used in this study was adapted from the Hot Stop technique originally described by Uejima (2000). In our hands, this method was found to have a number of limitations, especially when applied to discrimination of alleles of the *XIST* SNP. A requirement of the Uejima method was that very few cycles of PCR were performed. This is to limit the potential for amplification bias and possible inaccurate assessment of clonality. We found that the *XIST* transcript was difficult to amplify directly from cDNA. We believe that this may be due to either of two possible reasons: 1) the expression of the gene was too low for reliable detection (as seen in the requirement for an enriched mRNA for detection of the *XIST* transcript by Northern blotting in Chapter 3), or 2) the mismatches in the forward primer that were required to produce a *Pvu*II site in products of the G allele. This may have prevented the primer from annealing correctly during early cycles of PCR. The PCR amplification bias was an added complication, as it resulted in a two-fold preference for the G allele after 25 cycles of PCR. Although the PCR conditions were manipulated in an attempt to counteract this bias, no real improvement was produced. In summary, the analyses lacked both sensitivity and discrimination.

Future investigations

- A full characterization of cellular clonality in the bovine MG will benefit our basic understanding of MG biology. In particular, deducing the

'where and how' of mammary stem cell proliferation and cellular clone formation will reveal ways in which the bovine gland can be manipulated to achieve more sustained lactation and greater output. It is also possible that identification of differences between the human and bovine glands may give clues as to why cancer is so rare in the bovine gland (that is, beyond the obvious difference in lifespan).

- *XIST* SNP2 has potential as a marker of clonality as heterozygotes are found at a high frequency (32% heterozygosity) and this makes the selection of informative animals simple. Furthermore, many studies of *XIST* expression have shown that it is expressed exclusively from a single X-chromosome so it should be a very reliable indicator of the inactive X chromosome.
- More sensitive RT-PCR techniques should be able to extend the detection sensitivity enormously, for example, making it possible to analyse the clonal profile in a sample of just 10 to 100 cells from a microdissected TDLU.
- As an extension to our experiments, an *in situ* hybridization approach could be applied to thin sections to discriminate alleles of the *XIST* SNP. Allele specific *in situ* hybridization (ASISH) had been used previously to distinguish single nucleotide differences in *H19* and *IGF2* transcripts with a ^{35}S labelled 30mer oligoprobe (Adam *et al.*, 1996). Alternative *in situ* methods, such as allele-specific primed *in situ* synthesis (PRINS) or allele-specific 'padlock' probe hybridization (Nilsson *et al.*, 2000) may also be trialled to achieve the desired resolution and sensitivity for discrimination of an SNP. We believe that these techniques may be applicable to mammary clonality studies, particularly when a highly expressed gene is chosen for analysis. Such an approach would allow measurement of clonal profiles by direct observation of individual TDLU and ducts on 10 μm sections and profiles would not be affected by the presence of surrounding stromal tissue.
- With the publication of bovine mammary EST databases it has now become simple to look for SNP in other (more highly expressed) X-linked genes. As outlined in the discussion of the previous chapter, this would

allow the rapid identification of alternatives to *XIST* for both *in vitro* and *in situ* experimental assessment of clonality.

- *In situ* approaches using a well-expressed X-linked gene polymorphism as a marker of clonality will be useful as it eliminates the potential effect of contaminating stromal (which is presumably polyclonally-derived) tissue. Such methods will reveal the clonality of individual structures in the bovine and comparison of these profiles with those of human and mouse may reveal differences that must be considered when using particular animal models of human breast disease.
- Serial sections through a portion of gland could be reconstructed using 3D imaging software after *in situ* hybridization on these sections to detect each allele of the marker. This would give a better indication of general patterns of cellular clone size and distribution throughout the gland than would be possible from observations of 2D sections.

Chapter Five

The H253 Mouse as a Model of Cellular Clonality

5.1 Introduction

The bovine study in presented in Chapters 3 and 4 did not give a reliable measurement of cellular clonality in the mammary gland. Therefore, it was decided that an alternative approach, using observations of patterns of mosaic expression of an X-linked transgenic marker in whole mount mouse mammary glands, would allow a better description of clonality in the mammary gland and aid future attempts to determine the cellular clonal profile of the bovine gland.

5.1.1 Mouse models of cellular clonality

Both whole mount and serially-sectioned mouse organs have been used in studies of cellular clonality, for example, during kidney ductulogenesis (Lipschutz *et al.*, 1999). While very few studies have directly examined clonality in the mouse mammary gland, the use of *G6pd* isozymes as markers of different cellular clones in C3H-GPDX "normal" mice enabled Thomas *et al.* (1988) to score the clonality of individual structures in mouse mammary glands. They found staining patterns that indicated polyclonality in both ducts and alveoli. However, the marker could only be reliably detected in lactating animals, so this animal was not useful in investigations of cellular clonality at other developmental time-points.

The best model systems for study of cellular clonality allow for the unambiguous detection of the reporter gene activity without interference from high background levels of endogenous gene activity (Weiss *et al.*, 1997). Obviously, for transgenic mouse models of cellular clonality, transgene expression must also

reflect the X-inactivation status of the cell. However, it is a well-established that anomalies of transgene expression occur quite frequently. Krumlauf (1986) observed tissue variation in the expression of an X-linked transgene for an α-fetoprotein minigene that, when present on the inactive X chromosome, underwent transcriptional repression in the foetal liver but was not inactivated in the visceral endoderm. Furthermore, other X-linked transgenes, such as a concatemeric chicken transferrin gene insert (Goldman *et al.*, 1987) and a human α1(I) collagen gene (Wu *et al.*, 1992) insert, have been found to either totally escape normal inactivation, due to lack of methylation, or are incompletely inactivated, and hence expressed at low levels from the, supposedly silent, inactive X chromosome.

Despite the problems that are sometimes encountered with transgenic animals, some X-linked mouse transgenes have been found to be expressed in a similar pattern to endogenous genes; examples include CAT-32 (Collick *et al.*, 1998) and M-TK*neo*1 (Dandolo *et al.*, 1993) transgenes, which are methylated and inactivated similarly to endogenous X-linked genes.

5.1.2 Selection of mouse strain for this study

The inherent weaknesses in many of the existing X-linked transgenic mouse lines led Tam (1992) to develop the H253 transgenic mouse line for study of organogenesis. H253 animals exhibit an X-linked *lacZ* reporter gene construct that is reliably inactivated in a similar manner to the endogenous genes on the inactive X chromosome. This construct is 8.9 kb in size and contains the *E.coli* *lacZ* gene encoding β-galactosidase (β-gal) under the regulation of the 'housekeeping gene' promoter from the mouse 3-hydroxy-3-methylglutaryl coenzyme A reductase (*Hmgcr*) gene. The *SV40* large T-antigen nuclear localization signal (nls) is present at the 5′ end of the *lacZ* gene, and the transgene product is a fusion of β-galactosidase and NLS (Figure 5.1). The authors showed that 14 tandem constructs were incorporated in the X chromosome at position A6 (Figure 5.2), producing a 120 kb transgenic region, and that expression of the *lacZ* gene gave detectable amounts of gene product in the tissues examined (Tam and Tan 1992).

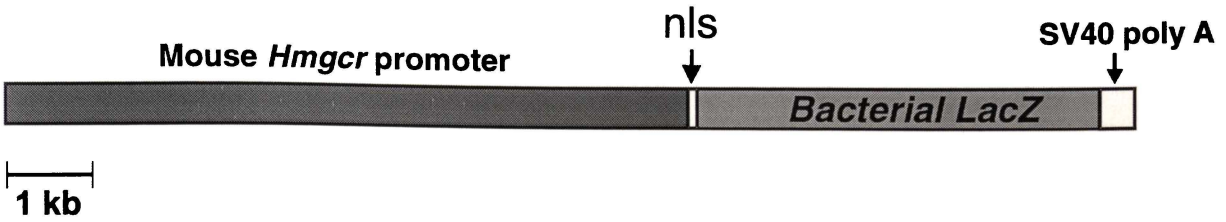


Figure 5.1. Diagram of the 8.9 kb transgene *lacZ*-NLS fusion construct used in the H253 mouse. The construct includes 5.6 kb of *Hmgcr* promoter elements and a 3.3 kb region including a 60 bp fragment of the *SV40* large T antigen (nls) for nuclear localization, the bacterial *lacZ* gene and the *SV40* polyadenylation signal (redrawn from Tam and Tan, 1992).

Furthermore, it has been shown that, in H253 animals, the transgene: 1) was expressed in all cells examined in homozygotes and in a mosaic pattern in hemizygotes, 2) showed preferential inactivation in visceral yolk sac when present

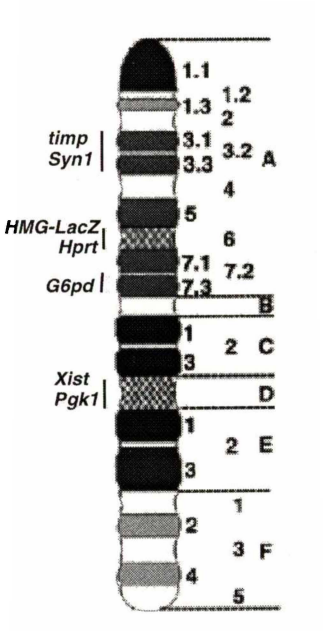


Figure 5.2. Schematic representation of the location of *HMG-lacZ*, *Hprt*, *G6pd*, *Xist* and *Pgk1* genes relative to the chromosomal bands on the murine X-chromosome. (adapted from Hadjantonakis *et al.*, 2001).

on chromosome X_p, and 3) was present on an intact chromosome, and therefore did not accelerate chromosomal deterioration by translocation (Tam and Tan, 1992; Tan *et al.*, 1993).

H253 mice have been used to study the timing of X-inactivation during embryogenesis. Using these mice, Tan (1993) showed that pronuclei of newly fertilized oocytes exhibit histochemically detectable β-galactosidase and that this marker was detectable until d3.5 when the blastocyst began to show mosaic

expression, with preferential expression of genes on X_p that indicated the beginning of X inactivation.

5.1.3 The HMG-CoA reductase (*Hmgcr*) gene promoter

Three regulatory elements have been found within the rat *Hmgcr* promoter: 1) a sterol response element (SRE) for SRE-binding protein (SREBP) mediated regulation of expression, 2) a cAMP-response element (CRE) and 3) an estrogen-response element (ERE). These elements are located at, respectively, -160, -100 and -90 nucleotides relative to the transcription start site in the endogenous gene and show significant sequence similarity among human, rat, mouse and hamster *Hmgcr* genes (Di Croce *et al.*, 1999). All of these sites could modulate transgene expression levels.

So, in tissues or developmental stages when the concentration of estrogen is low, it could be predicted that clonality assays that rely on expression from the HMG-*lacZ* transgene construct (described in section 5.1.2) may not be informative. This is likely to be a particular problem in tissues that lack other strong up-regulatory signals, such as mammary tissue and possibly the endometrial components of the uterus, where expression of both the endogenous gene and any transgene employing the *Hmgcr* promoter may be too low for reliable detection (Figure 5.3).

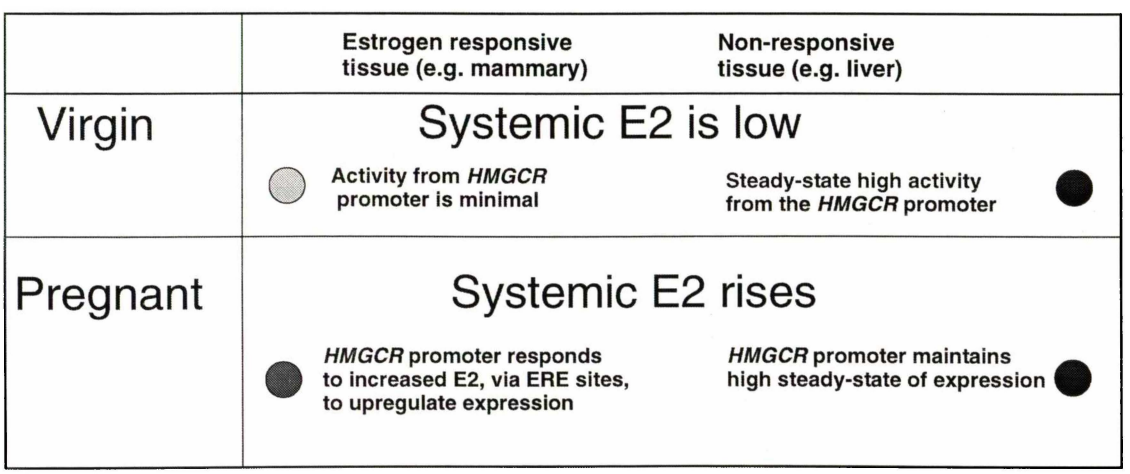


Figure 5.3. Diagram showing *Hmgcr* promoter activity in relation to physiological state and systemic estrogen levels. A low level of expression from genes with the *Hmgcr* promoter occurs in the estrogen-responsive tissues of virgin animals (in the absence of other upregulatory signals). In contrast, expression levels in the liver, the site of cholesterol biosynthesis (and therefore exhibiting steady-state requirement for the endogenous gene product from *Hmgcr*), are not dramatically affected by increased systemic estrogen. Expression in mammary tissue is predicted to increase during pregnancy. (Drawn from data in Di Croce *et al.*, 1999).

These potential effects on the expression of gene constructs that use the *Hmgcr* promoter can be best predicted by examining systemic maternal levels of estrogen (specifically, estradiol [E2]). As indicated in Figure 5.4, estradiol levels are raised in pregnant animals and peak at around d18. Progesterone levels show an even more dramatic increase in response to pregnancy, but fall sharply after peaking at around d15 of gestation.

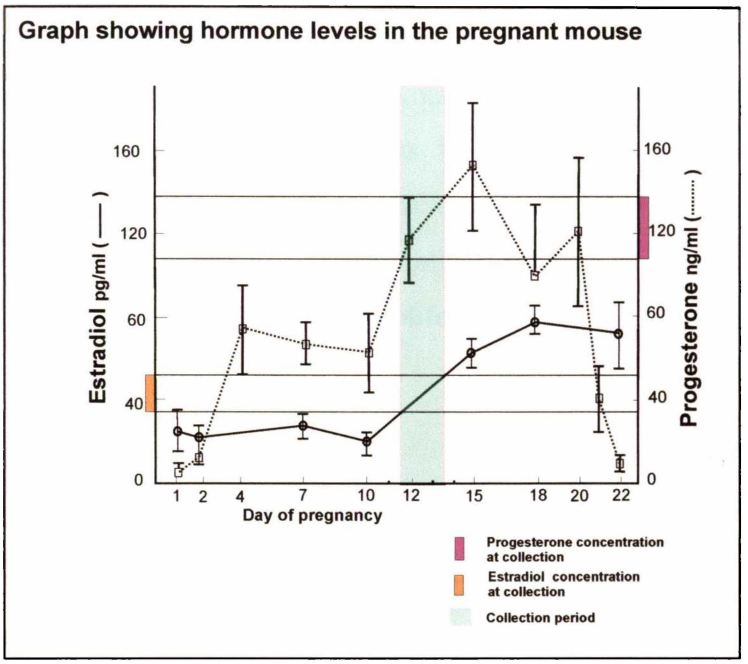


Figure 5.4. Graph showing systemic hormone levels in the pregnant mouse. Pregnant mice were sacrificed and the mammary glands removed for *lacZ* staining within the 2-day period between d12.5 and d13.5 of gestation (green bar). Estradiol (E2) levels begin to increase at around d10 of gestation and peak at around d18. In contrast, progesterone levels rise sharply at the beginning of gestation and peak earlier than estrogen levels (around d15) at which time they decrease sharply until parturition where progesterone levels approximate those seen in virgin animals – the decline in level of systemic progesterone is necessary for lactogenesis. (Adapted from: http://mammary.nih.gov/reviews/lactation/McCarthy001/index.html).

Based on the graph reproduced in Figure 5.4, the pregnant animals used in the present study were estimated to have estrogen levels of between 35 and 46 pg/ml and progesterone levels of between 100 ng/ml and 137 ng/ml. Nonpregnant animals were estimated to have approximate levels of estrogen and progesterone of 25 pg/ml and 5 ng/ml, respectively.

5.1.4 Estrogen: Mode of action via nuclear receptors

Upon ligand (estradiol) binding, estrogen receptors (ER) dimerise and associate with specific 15 bp inverted palindromic repeats (the ERE - discussed below) in the promoters of target genes. After association with the ERE, the receptor complex interacts with other components of the cellular machinery to effect expression of the target gene (Couse and Korach, 1999).

Estrogen receptor knock-out (ERKO) animals display a very rudimentary mammary ductal system relative to controls and this structure is devoid of terminal end buds (TEBs) (Couse and Korach, 1999). Tissue recombination experiments using subrenal capsule grafts in nude mice, which lack ER on mammary epithelial or stromal cells, have demonstrated that the estrogen dependency of mammary growth is stromally-mediated (Cunha *et al.*, 1997). That is, the signal for epithelial cells to proliferate appears to come from adjacent stromal cells "activated" by estrogen.

5.1.5 Aims

Our experimental design focussed on determining the utility of this transgenic animal as a model for assessment of cellular clonality, specifically of glandular development in mammary tissue. Specifically, we aimed to:

1) Confirm that tissues in the animal model show normal X inactivation in epithelial cellular populations and that exogenous (transgene product) histochemical staining can be distinguished from endogenous staining.

2) Assess the mammary gland as an environment for efficient expression of the transgene by determining the relative levels of expression of the endogenous 3-hydroxy-3-methylglutaryl coenzyme A reductase (*Hmgcr*) gene in various tissues by RT-PCR.

3) To determine relative expression levels of the *lacZ* reporter construct in the mammary and liver tissues from both virgin and pregnant transgenic mice.

4) To optimise a staining method for histochemical detection of the transgene product in mammary tissue.

5) To determine whether the H253 transgenic mice are suitable for study of cellular clonality in the mammary gland.

5.2 Methods

5.2.1 Dissection of mammary glands from transgenic animals

Glands were dissected immediately after CO_2 euthanasia. A throat to pubis incision was made and extended down each of the legs. The skin was drawn back from the underlying muscle and pinned to the dissection board. Glands were visible as fatty tissues on the underside of the skin. Thoracic and inguinal glands were prepared for removal by dissection of skeletal muscle remnants, particularly of those muscles that overlay the thoracic glands, and by cutting through the thin layer of connective tissue surrounding each of the glands (Figure 5.5).

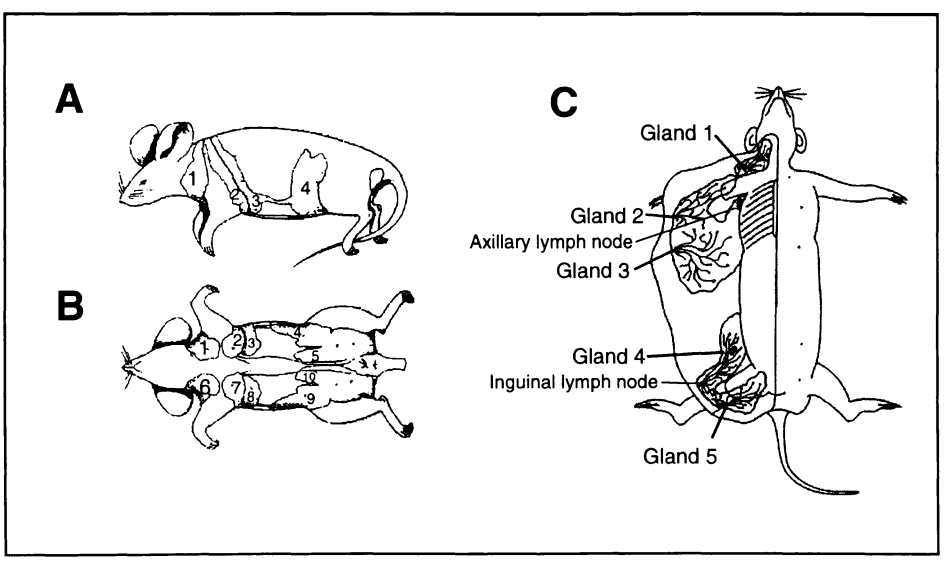


Figure 5.5. Diagram showing mouse mammary gland dissection. A) and B) The location of the murine mammary glands (labelled 1-10) and C) the presentation of the glands after ventral-side dissection (Ball, 1998).

Forceps were used to hold the most dorsal parts of the glands while they were carefully lifted away from the skin by gentle strokes with a #10 scalpel blade. Next, the glands were placed onto glass microscope slides and spread out to their

original *in vivo* dimensions. Finally, the gland was placed into a glutaraldehyde-based fixative solution (as detailed in section 5.2.3) for 1-2 h.

Most other tissues examined, including intestine and liver, were dissected from the animal after another throat to pubis incision through the ventrally-located muscle and connective tissues surrounding the peritoneum. These tissues were fixed as above, but thin-walled structures such as intestine were found to stain more successfully when the fixative was supplemented with 2% paraformaldehyde.

5.2.2 Determination of relative expression levels of endogenous *Hmgcr* gene and of the transgene by RT-PCR

A comparative assessment of expression of the endogenous mouse gene *Hmgcr* in both pregnant and virgin animals was carried out to examine the possibility for tissue-specific expression of the *lacZ* transgene, which is under the control of the same promoter.

Mouse mammary gland and liver were collected from three-month-old pregnant and virgin animals immediately following euthanasia. Total RNA was extracted from a portion of these tissues using Trizol™ reagent (as detailed in Chapter 2, section 2.2.2), an aliquot of the total RNA diluted in 10 mM Tris (pH 7), and quantified using a spectrophotometer (Smartspec, Biorad). A one-step protocol was used (Superscript II kit - Invitrogen) with a 'Taq polymerase positive, Reverse transcriptase negative' reaction included as a control.

RT-PCR was performed on 1 ng of template cDNA to detect the *Hmgcr* gene product using following primers:

Forward: CCCTGGGAAGTTATTGTGGGAACA

Reverse: CACCGCGTTATCGTCAGGATGAT

Product size: 161 bp from cDNA (457 bp from DNA)

5.2.3 RT-PCR to confirm *lacZ* expression in mouse tissue

RT-PCR was performed as above to measure *lacZ* expression in liver tissue from H253 hemizygous animals using the following primers:

Forward: GCAAAGACCAGACCGTTCATACAG

Reverse: CGTTTCGTCAGTATCCCCGTTTA

Product size: 163 bp from cDNA (also 163 bp from DNA)

5.2.4 Histochemical detection of ß-galactosidase presence

5.2.4.1 Staining protocol selection

Many different tissue methods for *lacZ* reporter staining have been reported. In order to select a method that was appropriate for general and mammary whole mount staining, two different staining protocols were assessed using a selection of different tissues.

Method 1. Fixation in 2% paraformaldehyde (PFA), 0.25% glutaraldehyde, and 0.01% Nonidet-P40 (NP-40) in 1X PBS for 2 h at RT.

Incubation in: 2 mM $MgCl_2$, 0.01% sodium deoxycholate (NaDOC), 0.02% NP-40 for 2 h with rocking. Staining in a solution comprising 1 mg/ml X-gal (in dimethyl formamide - DMF), 30 mM $K_4Fe(CN)_6$, 30 mM $K_3Fe(CN)_6.3H_2O$, 2 mM $MgCl_2$, 0.01% NaDOC, 0.02% NP-40, in 1X PBS overnight at 30°C.
(This method can be found at http://mammary.nih.gov/tools/histological/Histology).

Method 2. Fixation in 0.2% glutaraldehyde, 0.5 mM EGTA (pH 7.3), 2 mM $MgCl_2$ in 1X PBS at 4°C for 1 h.

Staining overnight at 30°C in a solution comprising 1 mg/ml X-gal (in DMF, as above), 2 mM $MgCl_2$, 0.01% NaDOC, 0.02% NP-40 (final concentration 0.02%), 5 mM $K_4Fe(CN)_6$, 5 mM $K_3Fe(CN)_6.3H_2O$, in 1X PBS (Lobe *et al.*, 1999).

NaDOC and Nonident-P40 are detergents included to aid penetration of the tissue by the staining reagents. $MgCl_2$ is included as a β-galactosidase cofactor. The EGTA in the fixative acts by chelating excess divalent cations (exhibiting a higher affinity for calcium, log K_{app} 6.68 at pH 7, than magnesium, log K_{app} 1.61 at pH 7, so as not to deplete enzyme cofactor; EGTA acts similarly to EDTA).

After overnight staining by either method, tissues were post-fixed in 4% PFA and stored in 2% PFA. The relative amounts of staining were determined by visual assessment. It was found that the standardised Method 2 (Lobe *et al.*, 1999) gave superior mammary gland staining compared to other published methods.

5.2.4.2 Technical notes

A number of technical factors critical to the success of these experiments were identified. These were choice of fixative, length of fixation, staining pH, composition of the staining solution and method of storage of stained tissues.

It was found that to ensure complete staining throughout the entire gland, a short fixation in 0.25% glutaraldehyde, rather than the frequently recommended fixation in 2-4% paraformaldehyde, ensures adequate penetration of staining reactants at a later stage of the experiment. Furthermore, this fixative seemed to preserve enzyme activity more efficiently than a more thorough fixation in paraformaldehyde. A caveat is that tissues 'pre-fixed' in such a manner also required 'post-fixing', in 4% paraformaldehyde overnight, to prevent tissue deterioration and subsequent diffusion of the histochemical stain. It is relevant to note that Shimohama *et al.* (1989) recommended that longer term storage of stained tissues should be in a fixative solution, to preserve both staining patterns and general morphology.

As detailed in the introduction, the pH at which the staining reaction was carried out was critical to the success of the histochemical staining. Ideally, one would like to use a pH near to the transgenic β-galactosidase optima of pH 7.3, but a higher pH is necessary to decrease the background resulting from the endogenous mammalian enzyme, which has a pH optima of around pH 7. In order to find an optimal pH, which gave acceptable β-galactosidase activity with minimal

background, trials with Tris-HCl buffered staining mixtures were carried out. The inclusion of 100 mM Tris (pH 8), recommended in some studies, was initially thought to negatively affect the staining efficiency of mammary tissues. It was later realised that this result was misleading, as the very weak staining found in mammary tissue from virgin animals was later attributed to low estrogen levels in the animals causing a low transgene expression (as discussed below). Accordingly, most subsequent experiments utilised low concentrations (usually 20 mM) of Tris-HCl (pH 7.8) buffering, as this was found to be sufficient to buffer the staining reaction at the desired pH overnight.

5.2.5 Whole mount mammary staining: controls and assumptions

5.2.5.1 Controls

If random inactivation of an X chromosome within each stem cell in a large pool of mammary progenitors occurs, then mammary tissue in transgenic animals has a certain number of 'in-built' staining controls. That is, when both the histochemical staining reaction and random X-inactivation are 100% efficient throughout an organ in a hemizygous animal, one expects to find an approximately equal number of blue- and non-staining epithelial elements. In the mammary gland, this would be best seen at the level of individual terminal end buds and/or ducts.

Direct examination of whole mount tissues permits an approximate measurement of the penetration efficiency of the staining reagents to be made. A gland displaying random X-inactivation is presumably a mosaic of cells representing both inactive X chromosome 'choices'. Hence, blue-stained tissues should be represented at all depths throughout the mammary gland. If not, then one assumes that staining did not take place evenly and that only thin, well-stained, portions of the gland should be assessed. The unavailability of homozygous H253 animal as a positive control was disappointing, as this control would have showed whether histochemical staining could be detected throughout the gland.

5.2.5.2 Assumptions regarding detection of the transgene product

It was assumed that the *lacZ* transgene construct in the H253 mice: 1) is expressed like other X-linked genes in the mouse, and 2) does not cause uneven inactivation profiles to the X chromosome in the mouse tissues examined. If statistically useful results are to be obtained, then chimeric and/or transgenic animals used in such studies must: 1) exhibit random mixing of the parental progenitor cell types, or of cells exhibiting each inactive X chromosome "choice", for chimeric and transgenic animals, respectively, and 2) exhibit approximately equal numbers of descendents from each of the two possible progenitor 'lines'. That is, for the transgenic animal, equal numbers of cells exhibiting either an inactive X_p or an inactive X_m (Hermiston and Gordon, 1995).

5.2.6 Assessment of X-inactivation fidelity in transgenic animals

The present study used both H253 transgenic and C57 lab-strain mouse lines. To confirm that H253 transgenic mouse showed normal patterns of X chromosome inactivation when bred with the C57 mice, the percentage of positive histochemical staining of crypts or follicles, in intestine and skin, respectively, was measured in both a hemizygous H253 transgenic mouse and in a C57 mouse. Intestine was removed, cut longitudinally, stained for β-galactosidase, placed between two glass microscope slides, and flattened. Skin was removed from the animal, fixed, and stained. A series of images were obtained at 100x magnification using a light microscope and a Nikon 995 'Coolpix' camera.

The following digital image analysis methods were devised using advice from contributors at various online imaging discussion forums. The digital JPEG images were imported into the computer program "Photoshop 7™" (Adobe) and their brightness/contrast adjusted to ensure that stained regions could be clearly distinguished from the background. The colour images were then converted to greyscale (IMAGE>MODE>GREYSCALE) and the threshold command (ADJUST>THRESHOLD; with a greyscale midpoint of 128) used to convert the images to black and white. Finally a histogram of greyscale (IMAGE>HISTOGRAM) was obtained (effectively all pixels were converted to black or white) and the percentage of the frame that was black (corresponding to stained

regions) was ascertained (Figure 5.6). This technique relied on extensive manipulation of the original image, but simple trials on frames with a known percentage blue-staining area showed that an accurate measurement of such areas could be obtained using these methods.

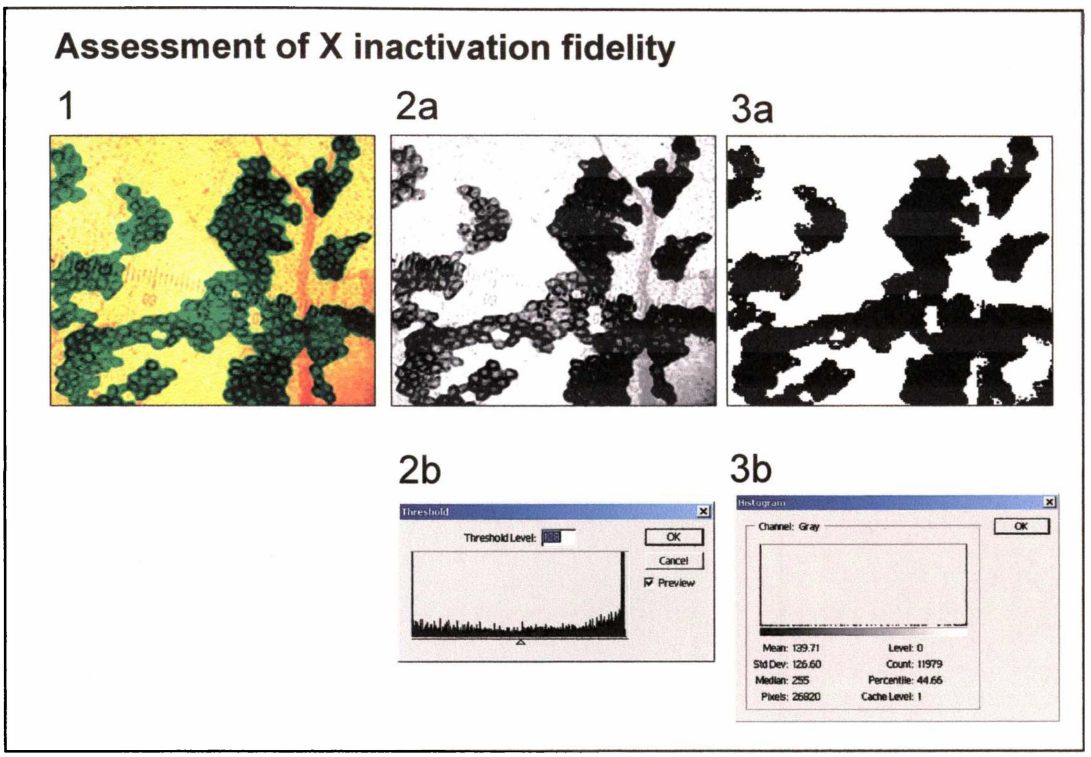


Figure 5.6. Image analysis steps completed using Photoshop 7™. A raw TIFF image at 3.3 mega-pixels was: 1) captured using a Nikon 995 camera (fine quality mode) and converted to greyscale (2a and b). Then, 3 a) the image was "bitmapped" with a threshold of 50%, resulting in a black and white image and finally, 3 b) a histogram of greyscale (effectively of just black and white channels) was obtained. The stained (β-gal positive) area in this sample frame is 44.66%.

The percentage area of stained intestinal crypts or hair follicles was measured in each of the examined frames. The combined average of stained regions for each tissue type was determined and is presented in Table 5.1.

5.2.7 Whole mount mammary staining: experimental

Entire mouse mammary glands were stained according to the protocol detailed in "Method 2" (section 5.2.4). The relative amounts of histochemical staining in whole glands from transgenic virgin and pregnant animals, as well as from non-transgenic animals, are shown in results section 5.3.8.

5.3 Results

5.3.1 RT-PCR indicates that the H253 mouse expresses the *lacZ* transgene in liver

As detailed in section 5.2.3, a preliminary RT-PCR experiment was performed to ensure that the model animal was expressing the transgenic construct. The electrophoresed product of this RT-PCR reaction is shown in Figure 5.7.

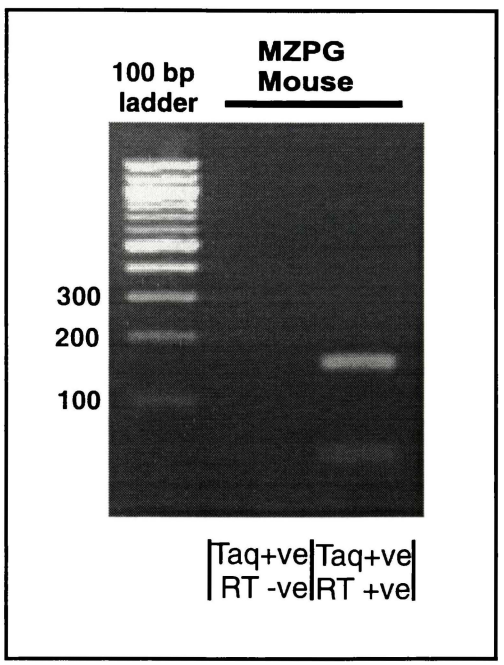


Figure 5.7. RT-PCR for *lacZ* in liver tissue from H253 mice. The products were electrophoresed on a 2% TBE agarose gel (5 V/cm for 50 min). Taq+ve/RT-ve: Taq Polymerase only, no reverse transcriptase (control); Taq+ve/RT+ve: has both Taq Polymerase and reverse transcriptase.

The reverse transcriptase (RT) positive lane shows a product of the size expected for the *lacZ* gene (163 bp). The control reaction gave no products. This simple RT-PCR experiment showed that bacterial *lacZ* was expressed in a H253 mouse tissue in the breeding background used in this study.

5.3.2 RT-PCR indicates that endogenous *Hmgcr* expression in the mammary tissues of pregnant animals is higher than that in virgin animals

A product of the size expected for the *Hmgcr* gene (161 bp) was observed in both mammary and liver tissues. Levels of the fragment in tissue collected from 3-month-old hemizygous H253 animals at d12.5-13.5 of gestation were 5 to 10 fold

higher than levels in mammary tissue from their virgin littermates (Figure 5.8). However, the liver sample controls showed no obvious differences between expression levels in pregnant and virgin animals.

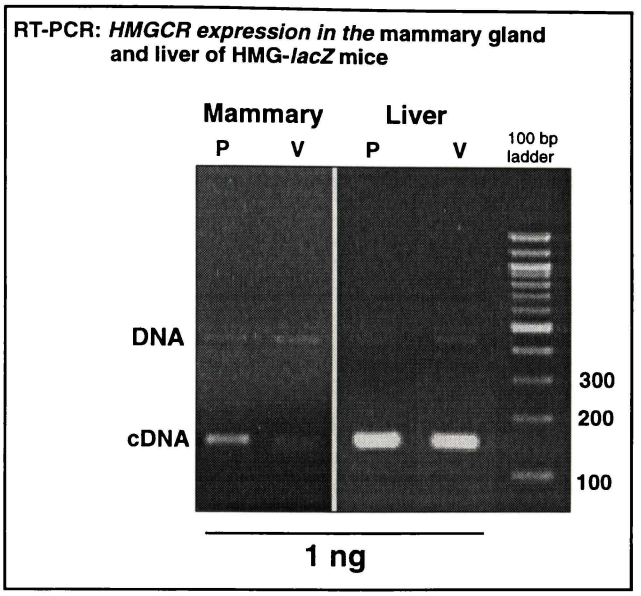


Figure 5.8. RT-PCR products showing relative levels of expression of the *Hmgcr* gene product in the mammary gland and liver of transgenic animals. While the mammary sample from the pregnant animal showed detectable levels of Hmgcr gene product (with 1 ng of total RNA template), a similar sample from a virgin animal showed very low levels of expression, as seen by the faint product band after 35 cycles of PCR. In contrast to the general measurement of expression from the mammary glands, both liver samples showed strong expression. This was expected, as hepatic cell lines have been shown to exhibit up to 10-fold the expression relative to other cell lines examined (Tam et al., 1994). The RT-PCR results from a series of *Gapdh* control amplifications were as shown below in Figure 5.9.

The difference in relative expression of the endogenous *Hmgcr* gene was measured using gel analysis software (GelPro 4™). A 5-fold higher *Hmgcr* expression was found in mammary tissue from the pregnant animal compared to levels in tissues from the virgin animal.

5.3.3 RT-PCR for the *lacZ* transgene showed a marked reduction in expression in mammary tissue from virgin animals

Relative expression levels of the construct were measured. The level of *lacZ* reporter gene transcript in the tissues from the pregnant animal was between two and five-fold higher than the level found in the virgin animal (measured using GelPro 4™ software). This is similar to the five-fold difference found in endogenous *Hmgcr* expression (see section 5.3.2). There appears to be a slightly lower transgene expression in the liver tissue from the virgin animal when compared with liver tissue from a pregnant animal.

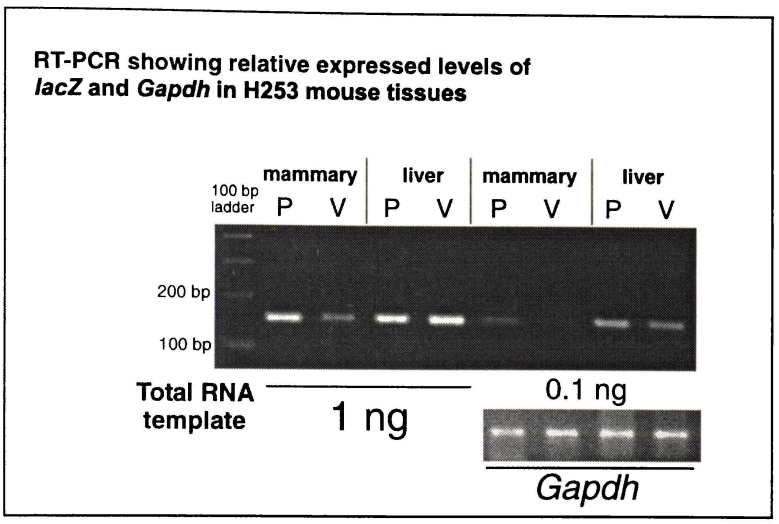


Figure 5.9. RT-PCR showing relative levels of expression the *lacZ* transgene and *Gapdh* in mammary and liver samples from MZPG mice. Tissues from pregnant (P) and virgin (V) animals were sampled, total RNA extracted, and RT-PCR performed using 1 ng and 0.1 ng of total RNA as template. The *Gapdh* control was amplified using the primers shown in Chapter 2 using only 0.1 ng of RNA as template.

5.3.4 Optimization of staining protocol

Staining trials were carried out using two basic protocols, as detailed above in section 5.2.4. Firstly, a *lacZ* staining technique published online, involving fixation in 2% PFA, and histochemical staining in a reagent mixture that included 30 mM ferri- and ferro-cyanide (see http://mammary.nih.gov/tools/histological/Histology), was tested. The second protocol tested (Lobe *et al.*, 1999) required a short (1 h) fixation in 0.2% glutaraldehyde and staining in 1 mg/ml X-gal, in a reaction solution that included 5 mM of each of ferro- and ferri-cyanide with standard β-gal/X-gal staining reagents.

Although neither method resulted in complete penetration of dense tissues, such as intact liver, the latter technique (Method 2) yielded superior results, as determined by direct observation of relative depths of staining. Using this method, entire liver stained superficially to a depth of approximately 1 mm. However, the entire mammary gland showed thorough staining, as indicated by an observation of blue-stained cells at all depths of the tissue.

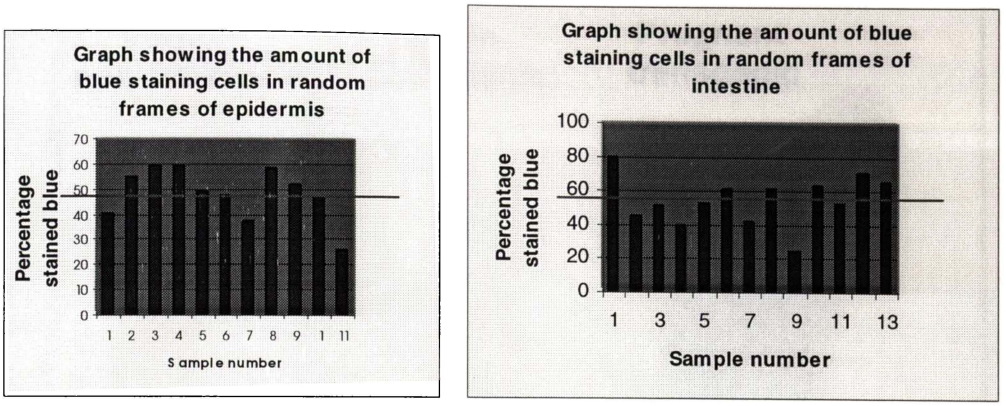
5.3.5 Assessment of X-inactivation fidelity

Random digital images were collected from both intestinal epithelia and skin epidermis and image analysis carried out to measure the percentage blue staining, as detailed in the methods section (section 5.2.6).

Data from a total of 11 x 0.8 mm^2 of epidermis and 13 x 0.8 mm^2 of intestine was collected and the average percentage staining determined (Table 5.1 and Figure 5.10).

Epidermal frames			Intestinal Epithelial frames		
SAMPLE	Frame #	Percentage	SAMPLE	Frame #	Percentage
1	660	40	1	652	80
2	666	55	2	653	45
3	679	59	3	654	51
4	676	59	4	655	40
5	774	49	5	656	53
6	775	48	6	657	61
7	777	38	7	658	42
8	778	58	8	651	61
9	781	52	9	659	25
10	783	47	10	660	63
11	784	26	11	666	52
			12	676	71
AVG % stained		48	13	677	66
			AVG% stained		55

Table 5.1. Showing tabulated data from the assessment of percentage-stained regions of both epithelial and epidermal tissues examined as detailed in section 5.3.5. Averages of 48% and 55% of stained tissue were obtained for epidermis and intestine respectively.



A **B**

Figure 5.10. Graphs showing the dataset as represented in table 5.1. **A)** epidermal staining and **B)** intestinal staining, each result is shown in a bar-graph with the average result indicated with a horizontal red bar.

The data shows that, when measurements are made from a large area of an epithelial tissue, approximately 50% of cells are found to be positive for *lacZ* expression. The percentage staining results obtained ranged from 26 % to 59% of imaged area for the epidermis, and from 25% to 80% of the imaged area for the intestinal epithelium. The average area of blue staining cells was 48% for epidermis and 55% for intestine.

5.3.6 Histochemical staining of gross whole mount tissues

The inguinal mammary gland from a pregnant animal was physically larger, specifically 2.2x the size of the control gland from a virgin animal of the same age, and was more intensely blue-stained than the gland from a virgin animal (Figure 5.11). The relative levels of staining detected were consistent with the hypothesis that elevated estrogen levels during gestation enhance the expression of the *lacZ* transgene. Interestingly, weak staining was seen in the gland from the virgin animal immediately after X-gal treatment but this was no longer detectable after clearing and counterstaining.

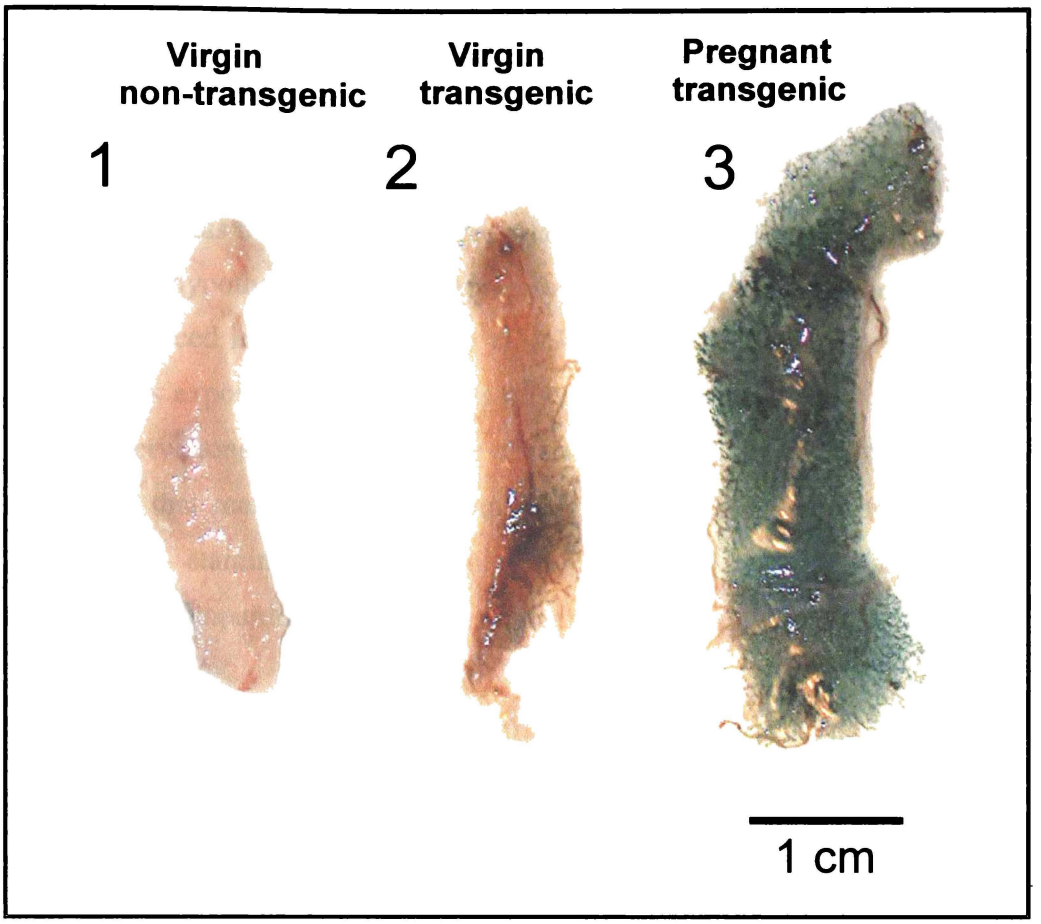


Figure 5.11. Digital images showing side-by-side comparison of whole mount mammary glands before counterstaining using Method 2 (section 5.2.4). 1) Number 4 inguinal mammary gland from a virgin non-transgenic H253 animal, 2) Number 4 inguinal mammary gland from a virgin transgenic animal (hemizygous H253), and 3) Number 4 inguinal mammary gland from a d12.5 pregnant transgenic (hemizygous H253) animal. All animals were three months old at sacrifice. Gland sizes are about 2 cm^2 for the virgin animal and 4 cm^2 for the pregnant animal.

5.4 Discussion

5.4.1. General comments

The mouse *Hmgcr* promoter, that mediates expression of the transgene in the H253 animal, has the potential to upregulate transgene activity via EREs. Hence, our working hypothesis was that expression mediated by this promoter would be severely reduced in tissues in which *HMGCR* was not normally required and/or those tissues exposed to low amounts of systemic estrogen. This was supported by our observations and explained the observation of the group who developed the H253 line, that mammary tissue from virgin animals did not show appreciable *lacZ*-positive staining (pers. comm. S.S. Tan, Children's Medical Research Foundation, Camperdown, NSW, Australia).

In contrast to mammary glands from virgin H253 animals, glands from pregnant animals (measured at d12.5 of gestation or later) showed considerable *lacZ* staining (evidenced in Figure 5.11 of the results section). We believe that, in the mammary gland, increased levels of the ovarian hormone estrogen causes upregulation of expression from the transgenic construct. This up-regulation results in increased β-galactosidase that produces a more intense histochemical stain in mammary glands from gestational versus virgin animals.

Our RT-PCR experiment showed that transgene expression in the mammary was affected by physiological state, with pregnant animals exhibiting *lacZ* levels that were 2 to 5-fold higher than in their virgin litter-mates. In contrast to the levels seen in mammary tissue, *lacZ* reporter expression levels in liver tissue were not found to vary significantly between pregnant and virgin animals.

The hypothesis regarding the effect of estrogen on expression from the HMG-*lacZ* reporter gene is strongly supported by our observations of appreciable *Hmgcr* promoter activity, and detectable *lacZ* transcripts, in the mammary tissue from a pregnant H253 animal but minimal expression in same tissue from a virgin animal. Unfortunately, because only a few animals were tested in the RT-PCR, further evaluation of this would be required for us to conclude that this result is real. This is because stochastic effects could occur during gland development to produce a gland that is derived from just a few progenitors, making the possibility

of an over-representation of one X-chromosome choice more likely (Takagi, 2001). If the underrepresented chromosome was the transgene-bearing chromosome, then histochemical staining would be deficient or absent, regardless of high levels of systemic estrogen.

5.4.2 RT-PCR analysis of endogenous levels of *Hmgcr* gene transcripts

Although, in some cases, transgene artefacts may suppress the expression of reporter genes, our study of endogenous *Hmgcr* expression suggested that this problem with the promoter is intrinsic to the transgene construct itself. In fact, RT-PCR detection of *Hmgcr* transcripts in mammary tissue from pregnant animals (d12.5-13.5) confirmed that endogenous expression was approximately 5-fold higher in mammary tissue from pregnant animals than that found in the same tissue in their virgin littermates.

In contrast to the findings in mammary tissue, liver tissue from both pregnant and virgin animals showed high levels of expression. This supports evidence that *Hmgcr* is highly expressed in the liver, and is not subject to an estrogen-mediated regulation of expression. It is worth noting the slightly reduced expression measured in the virgin liver in the RT-PCR experiment (section 5.3.2). This was thought to have resulted from an unknown reaction variable. The main finding in this experiment, of high levels of *Hmgcr* expression in liver tissue relative to mammary tissue, agrees with previous reports that showed that the basal activity of the *Hmgcr* promoter in hepatic cells was 8-10 fold higher than that observed in other cell lines (Tam *et al.*, 1991) and that, *in vitro*, high levels of *Hmgcr* expression in hepatic cell lines does not require stimulation via activated estrogen receptor (ERα) (Di Croce *et al.*, 1999).

5.4.3 Published studies on the *Hmgcr* promoter

The results presented above suggest that the *Hmgcr* promoter is not "ideal" in several respects. In particular, as evidenced by the lack of expression in the virgin mammary gland, the true "housekeeping" properties of this gene are questionable. Tam's use of the *Hmgcr* promoter as a 'ubiquitously active promoter' was

justified, at the time, on the previous description of *Hmgcr* as a housekeeping gene (Tam and Tan, 1992). However, after a more thorough examination of the literature regarding this promoter, we found a report that described an ERE in the *Hmgcr* gene and the subsequent measurement of its activity in artificial constructs with reporter genes (Di Croce *et al.*, 1999). These experiments showed that, in mammary cell lines, expression from the *Hmgcr* promoter is upregulated by estrogen and this explained why, in some tissues, the promoter might lack constitutive activity.

The *Hmgcr* promoter has other sites (the SRE and cAMP binding sites) via which regulation may occur. The SRE has been tested with HMG-β-gal constructs and promoter activity has been found to be down-regulated in the presence of sterols and upregulated when sterols are at either present at low concentrations or are totally absent (Skalnik *et al.*, 1988; Smith *et al.*, 1990). While it was not investigated in the H253 mouse, cholesterol levels may also play a role in regulation of the transgene expression, possibly via a sterol level feedback loop. It is possible that maintaining the animals on a low cholesterol, low LDL diet, supplemented with either estrogen or drugs that stimulate estrogen production or *Hmgcr* promoter activity, may counteract the negative regulation of expression from the *Hmgcr* promoter's regulatory sites and effect transgene expression in a broader range of tissues.

The tissue-specific transgene expression patterns observed in our studies of the H253 mouse are not unique; other "housekeeping" gene constructs also give ambiguous expression. For example, a CAT reporter gene under the regulation of a cytomegalovirus (CMV) immediate early gene 1 enhancer/promoter, while expressed in all 28 tissues examined, had expression level variations of up to 100,000 fold between individual tissues (Furth *et al.*, 1991). Another study that tested the activity of the same promoter found expression in only 24 of 28 tissues examined (Schmidt *et al.*, 1990).

As well as the intrinsic properties of a promoter *per se*, another possible complication is that, at a high copy number, a transgenic promoter may be methylated and/or not in the correct chromatin conformation. In the case of the

Hmgcr promoter, it has been suggested that this could occur as the result of incomplete coverage of DNA binding proteins that normally act to protect the promoter from methylation (Mehtali *et al.*, 1990). If such methylation patterns varied from tissue-to-tissue due to this phenomenon, this effect could have serious implications for studies like ours that rely on constitutive expression of the transgene as a marker of X-inactivation.

5.4.4 Does HMG-*lacZ* transgene expression reflect X-inactivation choice in the mammary gland?

We were not able obtain a homozygous H253 animal for use as a control in this study as our collaborators were unable to generate such an animal due to in-house breeding problems. Because of this, we assessed expression of the HMG-*lacZ* construct in hemizygotes by measuring X-inactivation profiles in a number of 1.0 X 0.8 mm^2 frames of intestine and skin epithelia, as an alternative control. We believed that if a 1:1 ratio of blue-staining to unstained cells were observed in epithelial populations then the mammary epithelia would almost certainly display a similar profile, allowing the study to proceed despite the absence of the +ve control. Our observations showed a staining ratio that was close to the desired 1:1 ratio of stained to unstained cells (with percentage staining of 48% for epidermis and 55% for intestinal epithelial). In keeping with this assumption, other authors have suggested adult intestinal tissue from homozygous H253 animals displays ubiquitous transgene expression (see Figure 1 in Seymour *et al.*, 2004). This suggests that the transgene expression system is working reliably, both with respect to the tissues examined as controls in the present study, that exhibited staining close to the expected 50%, and to the efficacy of detecting *lacZ* expressing cells in all tissues in H253 mice. Assuming that entire glands are derived from a large pool of mammary stem cell progenitors, each of which displays a random choice of inactive-X chromosome, a similar 'normal' pattern of X-inactivation is very likely to exist in the mammary glands from H253 hemizygotes.

5.4.5 Future investigations

Confirmation of the importance of estrogen in the positive histochemical staining seen in mammary tissues from pregnant, but not virgin, H253 animals could be carried out using experiments that monitor changes in levels of expressed *Hmgcr* and *lacZ* in virgin animals after the administration of estrogen. If estrogen was found to be important for the constitutive expression of the transgene in mammary tissue, animals at different stages of development could be induced to express detectable levels of β-gal by administering estrogen. This would facilitate an extension to the observations of clonality in gestational animals to include examination of glands at important developmental time-points, such as in the embryo and pre-pubertal animal, and of changes in glands occurring throughout pregnancy, lactation and involution. However, there is a strong possibility that artificially increased estrogen levels may affect glandular growth. This may affect the interpretation of the staining data as it relates to anatomical structures normally found at the particular stages of development studied.

Studies of cellular clonality, via assessment of the inactivation of different X chromosomes as a marker of different cellular clones, require an inactivated X-linked transgenic reporter construct that is highly expressed in a broad range of tissues. Therefore, one must conclude that the H253 mouse line, which exhibits marked tissue-specific expression patterns, has too many innate experimental limitations for recommendation as a general model animal in which to observe cellular clonality in the mammary gland.

It is recommended, therefore, that to ensure maximum utility for studies of cellular clonality in the mammary gland, an alternative model system is found. We suggest that an interesting alternative is an animal with an X-linked transgene construct that includes a reporter gene, such as the bacterial *lacZ* gene, under the expressional regulation of the human *ubiquitin C* gene promoter (Schorpp *et al.*, 1996). Ubiquitin C has been found to be expressed in all tissues examined and is involved in non-lysosomal intracellular protein degradation. Particularly high expression in spleen, lung and salivary glands has been reported. Schorpp (1996) found that transgenes under the transcriptional regulation of this promoter are well expressed in mammary tissue and this would make it potentially useful as an

alternative to the HMG-*lacZ* construct used in this study. However, as the construction of an entirely new X-linked transgenic animal for studying clonality is a lengthy and difficult process, it would make better sense to test the utility of alternative existing transgenic animals for studying cellular clonality in the murine mammary gland.

Other transgenic mouse lines that express non-invasive cellular markers from X-linked transgenes exist and some of these may be suitable for the study of cellular clonality in mammary tissue. Hadjantonakis (1998) developed the D4/XEGFP mouse line that has an X-linked *EGFP* gene. Nakanishi (2002) produced over 100 GFP lines, including the X-linked transgenic line 152A in which the transgene is under transcriptional regulation of the chicken β-actin promoter and a cytomegalovirus enhancer. This transgene was found to be subject to X-inactivation and was ubiquitously expressed in embryonic and adult homozygous tissues. In future studies, a mouse line with both X-linked HMG-*lacZ* and XEGFP transgenes in *trans* could be bred. Detection of each transgenic product could be achieved using a combination of histochemistry and fluorescent detection strategies for the HMG-*lacZ* and XEGFP products, or immunohistochemistry could be used on serial thin sections of entire glands to detect each marker. Subsequently, images of these sections could then be reconstructed to produce a 3D model of mammary cellular clone and patch size and distribution.

Chapter Six

Mammary Cellular Clonality in H253 Mice

Summary

A complete analysis of cellular clonality in mammary glands from H253 transgenic mice was carried out by direct histochemical staining of whole mount glands for β-galactosidase, the expressed product of the HMG-*lacZ* reporter gene construct. In total, four glands were examined for their cellular clonal profiles by entire gland staining and this was followed by measurement of staining in intermediate major ductal structures and in individual terminal end buds/ducts (TEBs/TEDs). The total percentage of blue-staining (transgene-expressing) cells observed across 4 glands was 46%, with the remaining 54% of ductal tissue comprising red, counterstained, cells. It was found that the predominant clonal pattern in mouse mammary tissue was polyclonality (involving 57% of all ductal termini examined); this suggests that a large number of precursor cells contribute to the mouse gland ductal tissue. Terminal ductal structures, particularly those distal to the gland origin, appeared to be derived from a precursor pool of decreasing size, as was evidenced by the large fraction of homogeneously staining terminal structures (21% and 22% for blue and red termini respectively) in many peripheral lateral ducts and TEBs/TEDs. On the basis of these staining levels, it was estimated that two to three progenitor cells contribute to each TEB/TED (of around 500 cells), an estimate that was supported by statistical analysis. A model of murine mammary gland cellular clonality is proposed.

6.1 Introduction

6.1.1 Mouse mammary gland development revisited

The murine mammary gland arises from an embryonic ectodermal thickening that invades the underlying mesenchyme. At birth, the gland comprises of a collection

of branched epithelial cords. The most significant growth of the mammary ductal network occurs after puberty and requires intensive proliferative activity from tightly clustered "balls" of cells, the terminal end buds (TEBs), at the end of each growing duct (See Chapter One, also reviewed in Smith and Medina, 1988a).

The TEB is a specialized structure that exhibits a population of progenitor cap cells at the tip of the bud. Proliferation of these cells gives rise to both the mammary epithelial and myoepithelial lineages (reviewed in Hennighausen and Robinson, 1998; Humphreys *et al.*, 1996). It is thought that cap cells may represent the primary progeny of the 'light cells', which are the suspected mammary stem cells and are found throughout the mammary ducts in the adult. In culture, these cells give rise to myoepithelia and luminal and intermediate epithelia, through a series of increasingly differentiated progeny (Li *et al.*, 1998; Smith and Medina, 1988b). Accordingly, the cap cell is described as a bipotent stem cell (SC) with the capacity to give rise to both of the distinct primary cellular lineages in the mammary gland (Williams and Daniel, 1983). Further evidence of widely dispersed mammary stem cells has come from experiments where BrdU pulse-labelled mammary cells were transplanted into a cleared mammary fat pad. After 5-8 weeks, label-retaining cells (the presumed SCs and their immediate progeny) were observed in transitional units located about 250 μm apart along the ducts, and within larger TEBs (Kenney *et al.*, 2001).

SC anchoring to the basement membrane (via integrins) is likely to be important in directing the "...definition of form..." of the growing duct and probably acts by modulating responses to mitogenic factors (Alford *et al.*, 1998), such as the fibroblast derived growth factors and the matrix metalloproteinases from stromal tissues (reviewed in Simian *et al.*, 2001).

6.1.2 Estimating progenitor numbers

The clonal outgrowth of epithelia, by SC and transiently activated (TA) cell progeny, occurs from the pool of progenitors found within the embryonic mammary bud (Ellis, 1998). While the exact numbers of stem cells in individual mammary gland progenitor populations cannot be observed directly, progenitor numbers can be estimated by studying patterns of X-chromosome inactivation. Methods that use statistical interpretation of the relative numbers of cells

displaying X-linked markers have been developed for studying approximate progenitor pool sizes for erythrocytes (<10 cells) and for scalp hair follicles (4 to 6 cells) in humans (reviewed in Nesbitt, 1971).

To apply this method to the measurement of the number of precursor cells for the mammary gland requires the assumption that the gland is derived from a precursor pool that is large and heterogeneous. This ensures that, after clonal expansion, an approximately equal number of SCs with either one or the other parental X-chromosome active are found throughout the gland. For accurate analysis there is also an assumption that no preferential growth of one stem cell 'line' over another occurs during mammary development (McMahon *et al.*, 1983). In practice, this is difficult to test at the level of the entire gland and, therefore, in this study, conclusions about progenitor numbers were made only for individual terminal ducts and buds.

6.1.3 Studies of cellular clonality in the mammary gland

Thomas (1988) used G6PD isozymes as a cellular marker of X-inactivation choice and found heterogeneous expression patterns throughout the mouse mammary gland. This indicates that polyclonality is the predominant clonality pattern in the murine gland. Previously, *in vitro* approaches that involved detecting X-linked markers of both diseased and normal breast tissue from heterozygous human samples (as discussed in Chapter 1) have been used to measure patterns of cellular clonality (Diallo *et al.*, 2001; Going *et al.*, 2001; Noguchi *et al.*, 1992; Noguchi *et al.*, 1994a; Noguchi *et al.*, 1994b; Noguchi *et al.*, 1995; Thomas *et al.*, 1988; Tsai *et al.*, 1996), and the findings of these studies are summarised in the table below (Table 6.1).

Animal	Marker used	Approach	Authors	Normal	Tumour
Mouse	G6PD histochemistry	Whole mount	Thomas *et al.*, 1988	PC	n/a
Human	HUMARA	*in vitro*	Diallo *et al.*, 2001	MC	MC
	PGK	*in vitro*	Noguchi *et al.*, 1992	PC	MC
	PGK	*in vitro*	Noguchi *et al.*, 1994	PC	MC
	PGK	*in vitro*	Noguchi *et al.*, 1995	PC	MC
	HUMARA	*in vitro*	Going *et al.*, 2001	PC	70% MC
		in vitro	Tsai *et al.*, 1996	MC	N/A

Table 6.1. Summary of experimental findings from studies of clonality in normal and diseased human breast tissue. G6PD – glucose 6-phosphate dehydrogenase, HUMARA – human androgen receptor, PGK – 3-phosphoglycerate kinase, PC-polyclonal, MC-monoclonal.

The sometimes contradictory results obtained in these studies of cellular clonality in human mammary tissue show that there are inherent limitations in *in vitro* approaches. Possible errors in some of these experiments are thought to be caused by inadvertent sampling of non-target regions, for example, stromal regions within the sampled area (Noguchi *et al.*, 1992). Such sampling error may be avoided by direct *in situ* observation of markers of clonality, that is, within either ducts/lobules in section (see discussion in Chapter 4) or of entire intact ducts in whole mount mammary glands. In addition, Diallo *et al.* (2001) found that the healthy human gland is comprised of large homogenous patches of separate cellular clones. They warn that this may make it difficult to accurately measure cellular clonality in human breast tumours, as a polyclonal tumour is very likely to exhibit a homogeneous staining pattern, despite a potential origin from more than one stem cell.

6.1.4 Ductal clonality in mouse mammary tissue

Initially, patterns of X-inactivation in the H253 mouse were analysed to confirm that the particular breeding used (with H253 and C57 lines) would give the expected random X-chromosome inactivation profile (detailed in Chapter 5). Next, histochemical staining of the whole mount tissues was performed, and image analysis of both whole mount and serially-sectioned (10 μm) mammary tissue from H253 mice hemizygous for the HMG-*lacZ* transgene, was carried out to measure the size and distribution of cellular clones in the gland. The clonal profile of terminal ducts was measured in high magnification images by counting

heterogeneously and homogeneously staining ductal termini. From these data, estimates of stem cell numbers contributing to individual TEB were obtained.

To our knowledge, this study represents the first use of a transgenic animal for direct examination of cellular clonality in the mammary gland.

6.2 Methods

6.2.1 Detection of β-galactosidase in mammary tissue

Five adult mammary whole mount glands (four thoracic glands and one inguinal gland – see section 5.2.1) from pregnant animals (>12.5 <14.5 d.p.c) were removed and placed immediately in fixative, as detailed in Chapter 5. After 1-2 h of fixation, the glands were subjected to 3x 2 min washes in PBS and for a further 5 min in a buffer (*lacZ* wash) comprising 2 mM $MgCl_2$, 0.01% NaDOC, 0.02% NP-40 in PBS (pH 7.8) – as per Method 2 (detailed in Chapter 5, section 5.2.4). The *lacZ* wash buffer was then replaced with the same buffer supplemented with 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide and 1 mg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) dissolved in dimethylformamide (DMF). The staining reaction was followed visually for the first 3-4 h and usually allowed to continue for at least 12 h. If overnight staining was carried out, the solution was adjusted to 20 mM Tris-HCl and the pH of the solution adjusted to 7.8 before overnight incubation at 30°. After incubation, the tissues were washed as before (2x 5 min in PBS) and post-fixed for 6 h in 4% PFA.

The glands were then dehydrated in a graded ethanol series, cleared of fatty tissue by soaking for 30 min in acetone, hydrated and counter-stained in carmine alum as detailed in Chapter 2, section 2.4.6. Then the glands were dehydrated, and finally cleared in xylene for 1 h before mounting with D.P.X. mounting medium between glass microscope slides.

6.2.2 Thin sections

After X-gal staining, one of the five glands used in the study (Gland E, thoracic # 3 mammary gland) was paraffin-embedded (see Chapter 2, section 2.4.3) and 10 μm sections produced in the transverse plane using a microtome.

6.2.3 Whole mount mammary gland clonality measurement

Low resolution/magnification images, encompassing the entire gland, were analysed after separation of the different staining patterns with Photoshop™ image analysis software using techniques developed with advice obtained from various online digital imaging discussion forums (see section 6.2.3.1). Subsequently, measurement of histochemical staining patterns in high resolution/magnification images was carried out.

This high magnification analysis involved counting the number of homogeneously (exclusively blue- or red-staining) and heterogeneously staining (mixed blue- and red-staining) termini in selected frames from four whole mount mammary glands dissected from hemizygous H253 animals, as detailed below in section 6.2.3.2.

6.2.3.1 Entire gland examination: low magnification

Adjustment of hue can be used to reduce the staining to just three colours which can then be individually removed from an image, allowing the selective removal of blue-staining, mixed blue/red staining (seen as magenta) and red-staining portions, and the calculation of the percentage of each staining type. This process also requires a calculation of total ductal area by creation of a software-facilitated 'gradient map' that selects authentic ductal tissue away from lightly-stained stromal tissues in the image.

A limitation of the method used is that the low magnification images collected, while exhibiting superior depth of focus compared to images collected at higher magnifications, have a far lower resolution. Hence, for informative analysis of staining in these regions, one must consider the fact that such low magnification images tend to be blurred to a point where ducts that contain predominantly blue-staining cells may appear as solid blue-staining. Conversely, ducts that are

predominantly red-staining, exhibiting just a few *lacZ*-expressing (blue-staining) cells may appear solid red and be scored as homogeneously stained in these low resolution images.

Therefore, at these low resolutions, the imaging approaches used in the image analysis may affect the accuracy of the staining data. This was acceptable for the initial overview examination, as it allowed determination of the degree of staining within each gland, but was inappropriate for assessment of ratios of clonal structures. Hence, the low magnification results were supplemented with, and viewed along-side, higher magnification images.

6.2.3.2 High resolution examination at high magnification

High magnification measurement of the glands allowed counts of blue and red-staining TEBs/TEDs to be made. This data was used to determine the ratio of homogeneously- to heterogeneously-staining TEBs/TEDs, which is a prerequisite for estimating stem cell numbers contributing to these structures. Data collected "locally" could be extrapolated to allow gland-wide estimates of ratios of blue to red-staining TEBs/TEDs. High magnification assessment was, therefore, the preferred method for measurement of patterns of cellular clonality in the mammary gland.

6.2.6 Image data collection: flat images

Histochemically stained whole mount mammary glands from hemizygous females were examined using both high and low magnification with a Leica microscope. Either 5 or 13 megapixel TIFF images were collected using a Zeiss AxioCamHR digital camera.

6.2.7 Combining multiple low resolution frames to produce entire gland image:

Entire whole mount gland images were produced by digitally "stitching" multiple frames together. This ensured maximal resolution and gland coverage at 2x final magnification.

All images were opened into Photoshop Elements™ software and brightness/contrast adjusted as necessary to provide clear definition of staining patterns. The photo-stitching function was applied and the result saved as a Photoshop file (*.psd) for later analysis.

6.2.8 Calculating total viewable ductal area:

An imaging method that selected ductal regions from the surrounding matrix was developed. This simple method allowed for rapid determination of ductal area in whole mount mammary images:

A photo-stitched image (as above) was imported into Photoshop™.

Ducts were selectively visualized by manipulation as follows: IMAGE>ADJUSTMENTS>GRADIENT MAP (Grey)

Brightness/contrast was again adjusted and the 'Threshold' image (IMAGE>ADJUSTMENTS>THRESHOLD) command used to convert tones in the greyscale range <128 (between 1-128) to white and greyscale tones between 129-256 to black.

A histogram (IMAGE>HISTOGRAM) was read to determine the percentage of the frame that was black (corresponding to ductal tissue). This value was used to find the ductal area by multiplication of the percentage black by the total frame area. If an approximation of the actual entire duct area was required, then this number (the viewable ductal 'diameter') was multiplied by π, assuming that half of the duct is visible from the viewpoint. Total ductal area is provided for reference only and, for the purposes of the present calculations, perspective and other considerations were deemed unnecessary.

6.2.9 Calculating area of blue, red and mixed staining from low resolution images

Hue adjustment of the original stitched entire gland picture (using Photoshop 7™ software), allowed segregation of blue, red and magenta (purple-pink) shades: IMAGE>ADJUSTMENTS>HUE/SATURATION.

Subsequent to an initial separation of the three component colours, each was isolated by editing away the other colours by digitally adjusting individual colours

(red, blue or magenta) to grey, and then using the 'replace colour' command: (IMAGE>ADJUSTMENTS>REPLACE COLOUR) to remove the residual grey tones.

Finally, the area of each set of coloured ducts was measured by conversion to greyscale (IMAGE>MODE>GREYSCALE), adjustment to desirable brightness/contrast, and thresholding to convert to a black and white image.

The threshold image was assessed for the percentage of each individual staining colour as detailed above (in the total ductal area calculation).

Note: For the low magnification images below, blue and the 'blue/magenta' shades were grouped together, as were red and 'red/magenta' shades.

6.2.10 Calculation of cells per cm^2 of ductal tissue

Representative regions of the duct that exhibited good resolution of individual cells were chosen from a number of locations in the gland. Individual cells were counted within each of these regions and an estimate of average cells/cm^2 determined. For this study, a TEB or TED was defined as a budding structure comprising approximately 500 cells at the terminus of a lateral branch.

6.2.11 Cell counting by percentage staining

In order to measure the proportion of each staining type in mammary ducts, high resolution images (3600 X 3000 pixels; 10 MB TIFF images) were collected at either 100x or 200x magnification.

The images were enhanced by various software-based manipulations to maximize resolution of cells displaying blue-staining nuclei from 'non-staining' (counterstained only) cells. These manipulations included basic brightness/contrast adjustment, hue adjustment (sometimes with specific selection of either staining type, that is, of red-stained cells or of cells with histochemically-stained blue nuclei) and removal of areas where muscle had been co-dissected with the gland.

Ductal area calculation (as detailed above) was performed for the region under analysis and this figure was used to determine percentage of the region staining

either blue or red (counter-staining). The determined percentage staining was multiplied by the known cells/cm^2 to determine approximate cell numbers staining either blue or red.

The best glands for further analysis of cell number were those that exhibited a general blue to red cell ratio of close to 1:1. Blue nuclei within such regions were counted and this number doubled for an actual cell count to give an estimate of actual cell numbers in TEBs etc.

6.2.12 Analysis of cellular clonality

Whole mount analysis was carried out as detailed above. Images were "stitched" to allow whole gland assessment of blue, red and intermediate staining (magenta shades indicated a mixture of blue and red cells, and therefore polyclonality). The results of this analysis were collated and an estimate of percentage involvement of each staining type determined for each gland.

Finer detail of cellular clonality was analysed by sequential examination of higher magnification images from each whole mount gland. The magnification used was 100x for all of the images collected. Camera-based optical zooming gave further magnification (up to 3x).

6.2.13 Statistics

When approximate staining ratios for TED were known, the χ^2 test was applied to test the hypothesis that TEBs, found to include around 500 cells, were derived from either two or three stem cells. The hypotheses were that average findings from 10 frames in each gland examined did not deviate from a ratio of either 1:2:1 (for a two stem cell origin) or 1:6:1 (for a three stem cell origin) of blue:polyclonal:red cells. (p values greater than the $p_{(0.05)}$ indicated that the data did not deviate significantly from the null hypothesis).

6.3 Results

6.3.1 General entire gland analysis of clonality

6.3.1.1 GLAND A

Gland A whole mount: low magnification analysis

Gland A, an inguinal mammary gland from a hemizygous H253 mouse, gave staining patterns that showed clear resolution of blue- and red-stained portions (Figure 6.1). This was because a light counter-staining (in carmine alum) was used, so as not to saturate the histochemical (blue) stain.

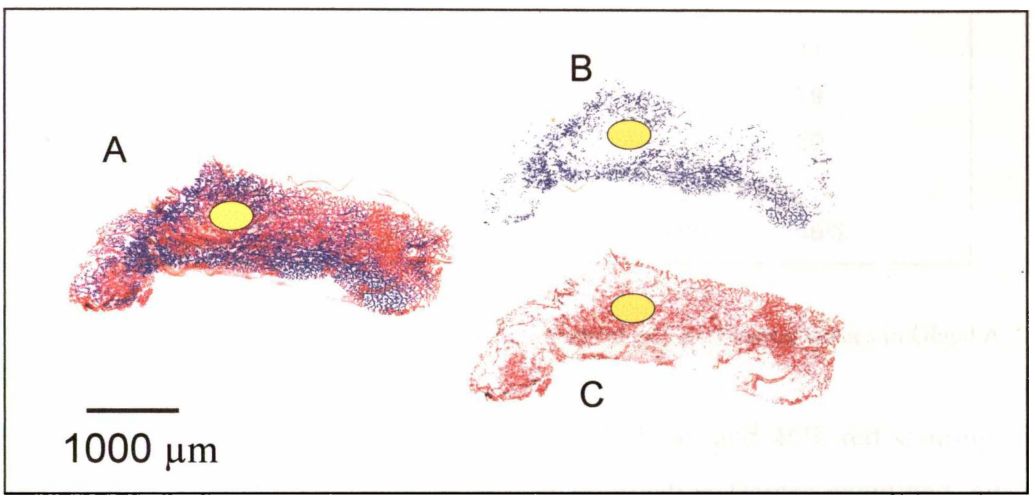


Figure 6.1. Gland A low magnification entire gland image. A) Showing hue-adjusted gland A (#4 mammary) and colour dissection for B) blue and C) red. The yellow circle represents the primary lymph node. Note the general distribution of blue- and red-staining duct. Peripheral ducts generally display a predominance of red-staining.

As detailed in the table below (Table 6.6), general low magnification analysis of staining in this gland showed relative percentages of blue- and red-staining at 30% and 44% respectively, with the remainder showing intermediate (magenta) staining 26%. The percentage of red-staining tissue may have been overestimated as the counterstain often stained stromal elements as well as the desired target ductal mammary epithelia.

High magnification analysis of cellular clonality within Gland A

Ten high-resolution, high-magnification images, representing 676 ductal terminal structures, were analysed for percentage blue, polyclonal, and red-staining of termini, and for total percentage of red and blue epithelial tissue in the frames (Table 6.2).

Gland A		Termini			Total epithelia	Total epithelia
Frame	Number of termini	Blue	Red	Mixed	% Blue	% Red
1	42	29	0	13	83	17
2	81	16	38	27	26	74
3	55	27	0	28	85	15
4	76	0	0	76	41	59
5	116	28	62	26	39	61
6	78	56	22	0	65	34
7	76	31	0	45	78	22
8	61	24	9	28	59	41
9	38	4	14	20	21	79
10	53	20	15	18	40	60
Total	676	**235**	**160**	**281**		
		34%	**24%**	**42%**	**54%**	**46%**

Table 6.2. Table showing staining patterns of both termini and total ductal tissues in Gland A.

The total percentage blue and red staining (54% blue- and 46% red-staining) for all epithelial tissue across all of the high magnification frames examined, agrees with the 1:1 staining ratio expected if X-inactivation is occurring normally in the mammary gland. The majority of termini in Gland A exhibited heterogeneous staining indicating polyclonality (42% of termini). Sometimes this was a biased heterogeneity, such as in Figure 6.5 (B), where staining of regions within the same termini (or even duct) was heterogeneous but skewed towards overall red- or blue-staining. This effect was particularly common within peripheral regions of the gland, where general heterogeneity (polyclonality) predominates and homogeneity (just red- and or blue-staining cells) was found within smaller terminal ducts and termini, such as in Figure 6.5 (C)-(M). Figure 6.6 (A) shows two adjacent polyclonal ducts each displaying this "biased" staining; the duct on the left is composed predominantly of transgene non-expressing cells while the nearby duct has a predominance of transgene-expressing cells against a background of ducts exhibiting a polyclonal pattern. Homogeneity was observed in 58% of termini examined (34% blue and 24% red-staining termini); these termini were often found at the ends of lateral branches from larger polyclonal ducts.

6.3.1.2 GLAND B

Gland B whole mount: low magnification analysis

Low magnification imaging of Gland B, a thoracic mammary gland from a hemizygous H253 mouse, revealed a predominance of red-staining cells (Figure 6.2). However, closer inspection, at 40x and higher magnification, showed blue-staining regions throughout the gland, even within those regions that appeared solid red after hue adjustment of low magnification images. The blue stain was far less intense than that observed for either gland A or C and may indicate an artefact from staining reaction variations or, possibly, the effect of storage times. Another possibility is that cells within this tissue were expressing less of the β-gal transgene product than the other glands.

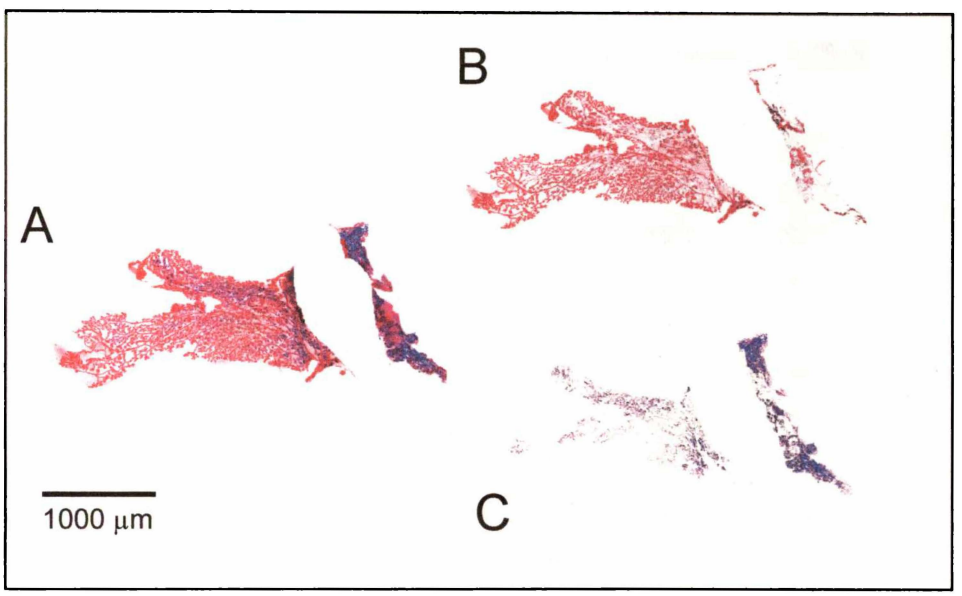


Figure 6.2. Gland B low-magnification entire gland image. In A) the entire analysed area of the hue-adjusted gland is shown. B) Shows just the red-staining regions and C) just the blue-staining regions. (A portion of the gland was not analysed because overlying muscle obscured the field).

After elimination of much of the red background staining in the image by manually erasing background portions, blue-staining cells were found to comprise 39% of the total stained ductal network. The remainder of the tissue (61%) was red-staining. No magenta staining tissue was detected in this gland.

High magnification analysis of cellular clonality within Gland B

Analysis of high magnification images of Gland B revealed staining similar to that observed for Gland A. A number of representative images collected to show general staining patterns throughout the gland are shown in Colour Plate Figure

6.7. Polyclonal regions, that showed a predominance of cells with one staining type, were frequently observed and are shown in Figure 6.8 and, in fine detail, in Figure 6.9.

Ten random frames, with 1213 termini, were analysed for staining patterns and the entire frame examined for percentage of blue- and red-staining cells (Table 6.3).

Gland B		Termini			Total epithelia	Total epithelia
Frame	Number of termini	Blue	Red	Mixed	% Blue	% Red
1	126	18	36	72	27	73
2	38	5	18	15	13	87
3	102	33	33	36	44.6	56.4
4	87	12	45	30	16.4	83.6
5	57	7	20	30	33	67
6	164	73	55	36	49	51
7	179	10	36	133	47	53
8	175	22	38	115	53	47
9	148	9	64	75	25	75
10	137	12	39	86	71.3	28.7
Total	**1213**	**201**	**384**	**628**		
		17%	**31%**	**52%**	**38%**	**62%**

Table 6.3. Table showing staining patterns of both termini and total ductal tissues in Gland B.

Across the entire mammary gland, red-staining epithelia predominated (62%). This high proportion of red-staining was also reflected in analysis of individual red, homogeneously staining, termini that were observed in 31% of cases, whereas blue termini were only seen in 17% of all termini.

Both high- and low-magnification estimates of percentage blue-staining in Gland B indicated that total blue-staining was in the range of 38-39% of the gland. Magenta staining, indicating polyclonality, was frequently observed (in 52% of termini). While Figure 6.7 (A), (E) and (H) show general mixed clonality, a

number of the representative frames, for example, (B)-(D), (F), (G) and (I), show a predominance of one staining type. Sometimes, distinctly homogeneously staining ductal structures were observed, such as in Figure 6.7 (C), in areas with adjacent blue and red-staining ducts. This same effect is seen in images (A)-(E) from Gland B (Figure 6.8), that show adjacent termini displaying different predominant staining patterns.

6.3.1.3 GLAND C

Gland C whole mount: low magnification analysis

Low magnification, gland-wide, analysis of Gland C, a thoracic mammary gland from a hemizygous H253 mouse, revealed a high percentage of blue-staining (42%) relative to intermediate magenta (37%) and red-staining (21%) (Figure 6.3).

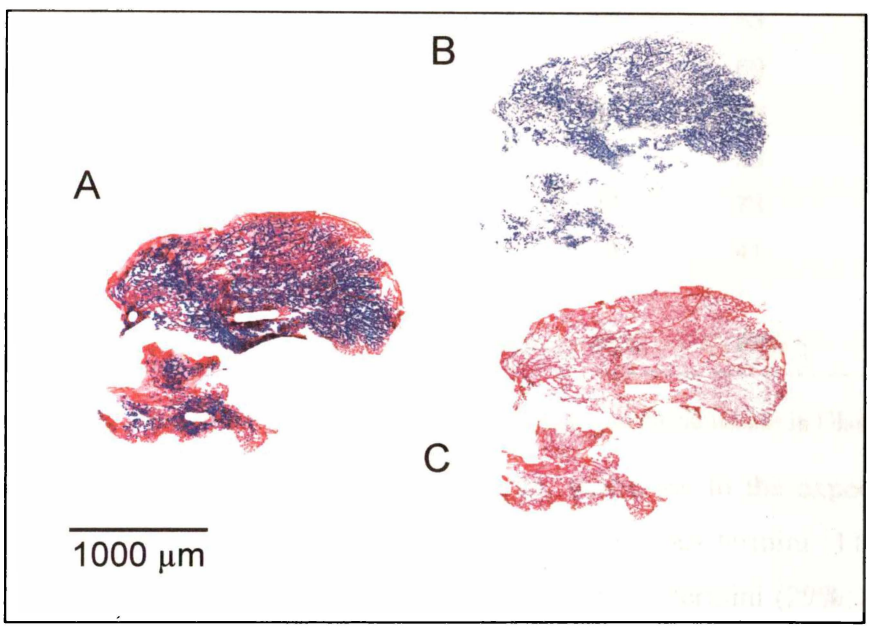


Figure 6.3. Gland C: low magnification image of the entire gland. In 'wide-field' at low-magnification, blue and red regions seemed equally well distributed throughout the gland but blue (at 42%) predominated the total stained regions with the remainder staining magenta (37%) and red (21%).

It is possible that red stain represents more of the actual area as the calculation of percentage involvement sometimes relies on arbitrary assignment of colour level cutoffs, that is, some of the staining scored as 'magenta' may have in fact been 'red'.

High magnification analysis of cellular clonality within gland C

Ten high resolution/high magnification images were collected from Gland C (examples are shown in Figures 6.10 and 6.11) and scored for relative numbers of red and blue-staining cells, and for total red and blue-staining within all the structures in the frame (Table 6.4).

Gland C		Termini			Total epithelia	Total epithelia
Frame	Number of termini	Blue	Red	Mixed	% Blue	% Red
1	48	29	0	19	72	28
2	43	20	0	23	81	19
3	61	38	0	23	90	10
4	40	2	0	38	44	66
5	69	0	5	64	37	63
6	45	6	0	39	31	69
7	69	16	0	53	24	76
8	117	47	0	70	67	33
9	147	23	74	50	27	73
10	88	31	0	57	59	41
Total	**727**	**212**	**79**	**436**		
		29%	**11%**	**60%**	**53%**	**47%**

Table 6.4. Table showing staining patterns of both termini and total ductal tissues in Gland C.

Total ductal staining, of 53% blue and 47% red, was close to the expected 1:1 ratio. The number of exclusively red-staining homogeneous termini (11%) was less than half the number of homogeneously blue-staining termini (29%). Image analysis of high magnification images of Gland C also showed the presence of high numbers of heterogeneously stained termini, totalling 60% of all termini (as seen in Figure 6.10). Figure 6.10 (G) and (J) show polyclonal ducts with different predominant cell staining types (blue- and red-staining, respectively) and Figure 6.10 (A), (B) and (E) show adjacent ducts with "skewed" or biased heterogeneity, with predominance of one staining type over the other. As in the other glands analysed, it is possible that some homogeneously staining structures were genuinely monoclonal, but it was not possible to test this in the present study (see discussion).

6.3.1.4 GLAND D

Gland D whole mount: low magnification analysis

Low magnification examination of the entire ductal tissues from Gland D, a "number 3" thoracic mammary gland from a hemizygous H253 mouse, showed very limited blue-staining in a predominantly red-staining gland (Figure 6.4).

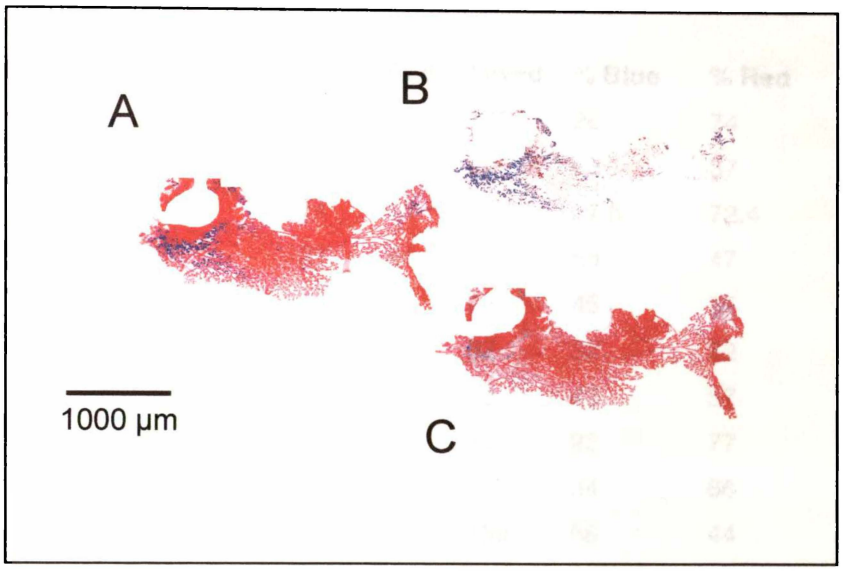


Figure 6.4. Gland D low magnification entire gland image. A) Gland D showed B) a low percentage of blue-staining (18%) with C) a predominance of red-staining (62%). Intermediate stained regions were indicated by the 20% of ducts that gave a magenta staining pattern (magenta shades are merged with red and blue in this figure).

After Photoshop-based image analysis, including colour discrimination, this gland showed a low percentage of blue-staining, with only 18% of the gland exhibiting a strong blue stain, 62% showing red-staining and 20% showing intermediate, magenta, staining. The preponderance of red-staining seen in this analysis may have been due, in part, to weak transgene expression resulting in weak blue-staining within a gland that had strongly stained red ducts and stromal tissue. It is more likely, however, that these data reflect the limited resolution obtained using analysis of entire glands from low magnification images.

High magnification analysis of cellular clonality within gland D

A number of high magnification images were collected from Gland D. These images show a range of staining profiles, including all of the staining types found in the other glands examined (Colour Plates, Figures 6.12 and 6.13). Ten randomly selected images were analysed for relative numbers of red-, mixed- and

blue-staining termini, and for total percentage blue- and red-staining cells within the frame (Table 6.5).

Gland D		Termini			Total epithelia	Total epithelia
Frame	Number of termini	Blue	Red	Mixed	% Blue	% Red
1	261	14	72	175	26	74
2	96	73	4	19	63	37
3	141	14	48	79	27.6	72.4
4	103	21	24	58	53	47
5	96	7	6	83	45	55
6	93	18	23	52	38	62
7	97	3	24	70	43	57
8	92	0	21	71	23	77
9	116	3	6	107	34	66
10	165	0	6	159	56	44
Total	**1260**	**153**	**234**	**873**		
		12%	**19%**	**69%**	**41%**	**59%**

Table 6.5. Table showing staining patterns of both termini and total ductal tissues in Gland D.

The predominant staining pattern across all epithelia tissue in the frames examined was red-staining (59%) while heterogeneity was clearly the major staining pattern in termini (with 69% heterogeneous staining relative to 12% and 19% for homogeneously staining blue and red termini, respectively). Higher magnification (40x) inspection of all regions in the examined frames that appeared to be totally 'red' staining (see Figure 6.4) revealed a predominantly heterogeneous staining pattern in both ducts and termini. When the higher magnification images are examined, one must conclude that the low magnification analysis used, in particular the hue-adjustment step that enhances blue and magenta by image saturation, did not give an accurate measurement of staining in this gland.

Colour Plates

Gland A

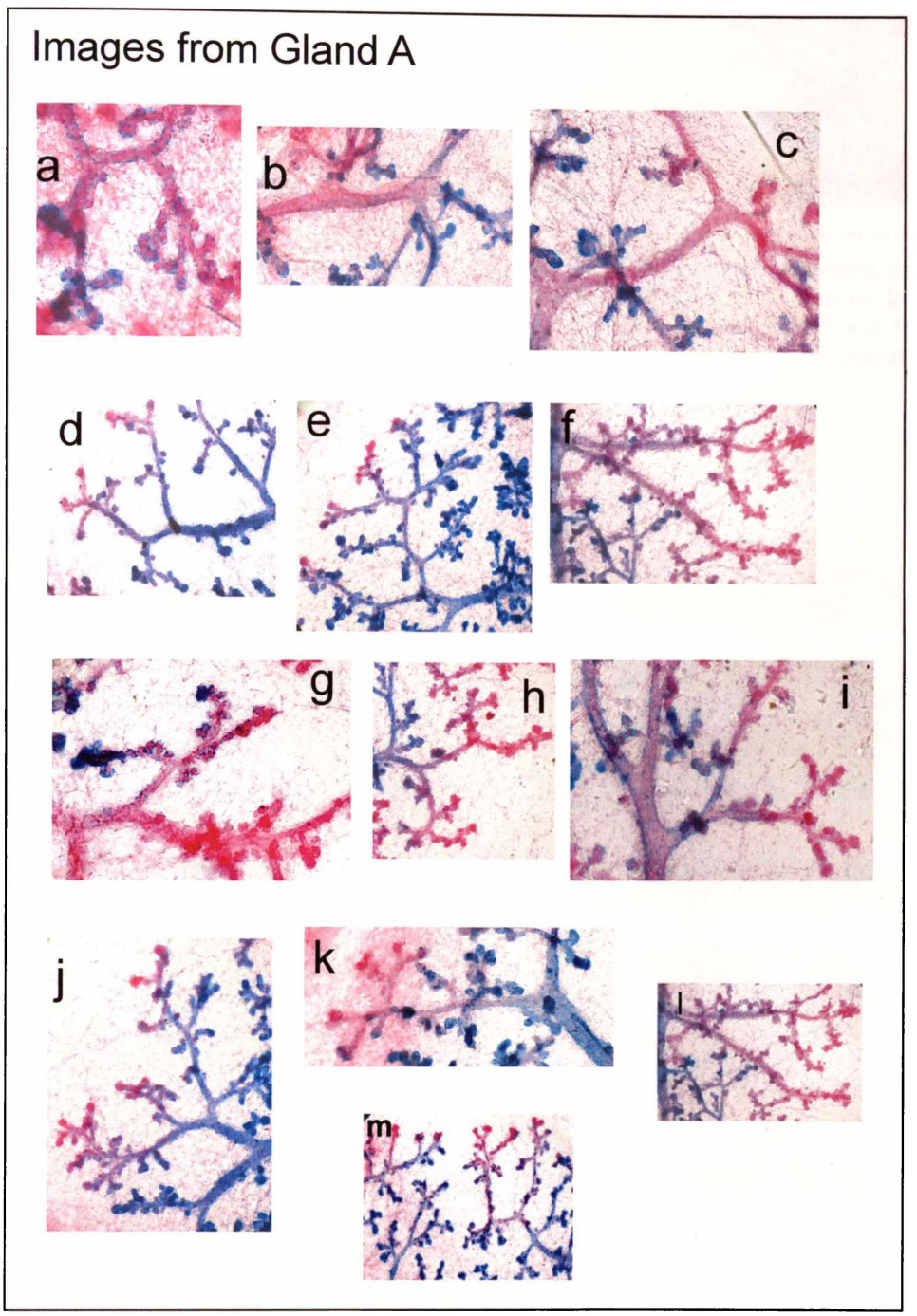


Figure 6.5. Images from gland A: number 4 mouse mammary gland showing ducts that stained heterogeneously and TEBs/TEDs with both homogenous and heterogeneous staining. In some cases the staining in ducts and termini appeared to be biased, with a predominance of either blue- or red-staining cells. (A) Polyclonal ducts, (B) and (C) Mainly heterogeneously staining polyclonal ducts with some homogeneously staining structures, (D)-(M) Terminal structures that originate as polyclonal ducts but terminate in homogeneously red-staining regions (see note in the discussion). The base of image (M) is 500 μm. Images (A)-(C), (G) and (I) were optically enlarged, up to 3x using the Nikon camera.

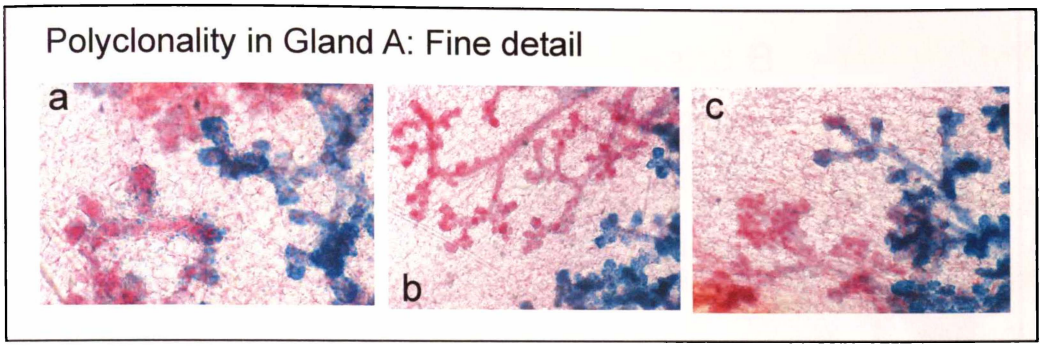


Figure 6.6. Polyclonality (biased) suggested in ducts from Gland A. A common staining pattern found in Gland A was of ducts that stained with a predominance of either red or blue nuclei, but always including some cells of the other staining type. The base of images (B) and (C) is approximately 1000 μm, while image (A) was optically zoomed 3x and has a horizontal axis length of around 300 μm.

Gland B

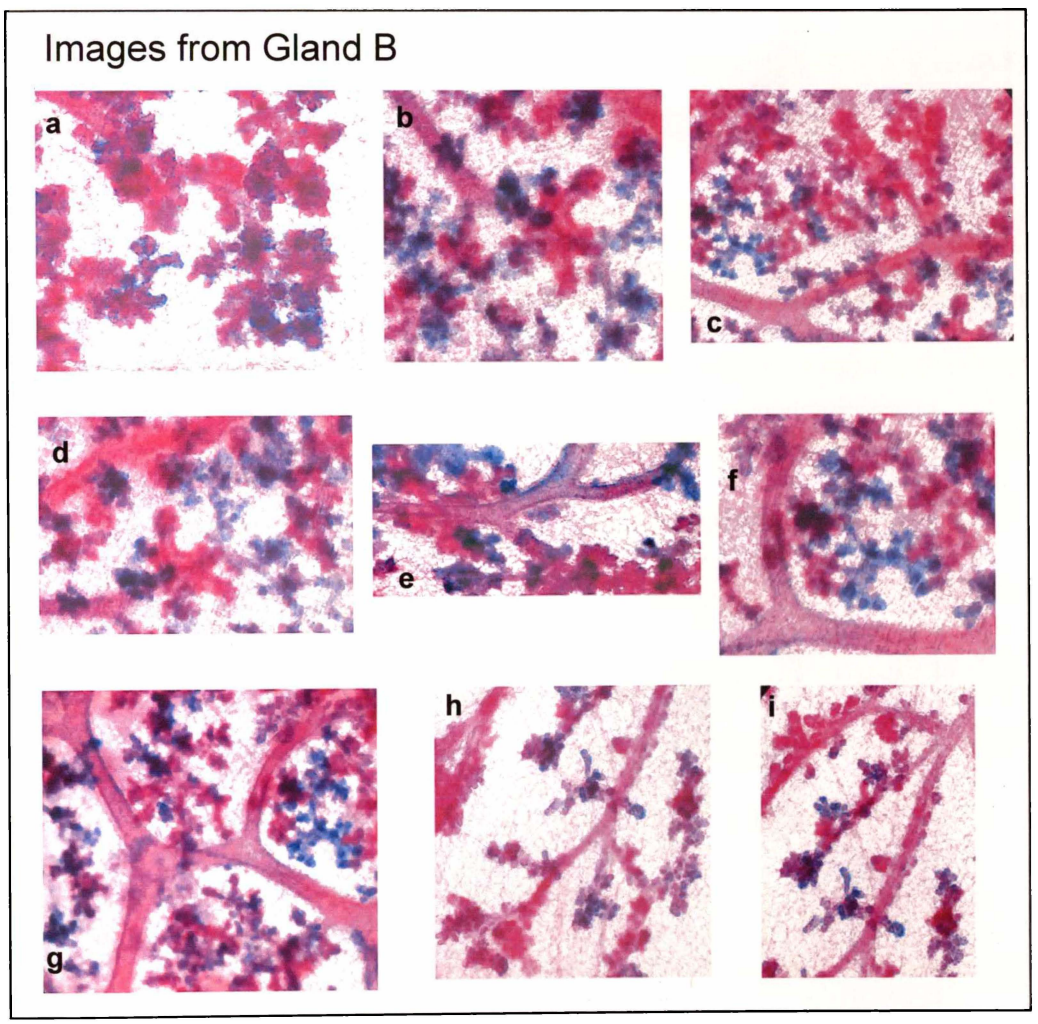


Figure 6.7. A selection of images from Gland B. Heterogeneous staining, suggesting polyclonality, was the predominant staining pattern observed in this gland. Homogeneous staining was seen much less frequently than heterogeneous staining. This is shown in the figure below and in image (F) above.

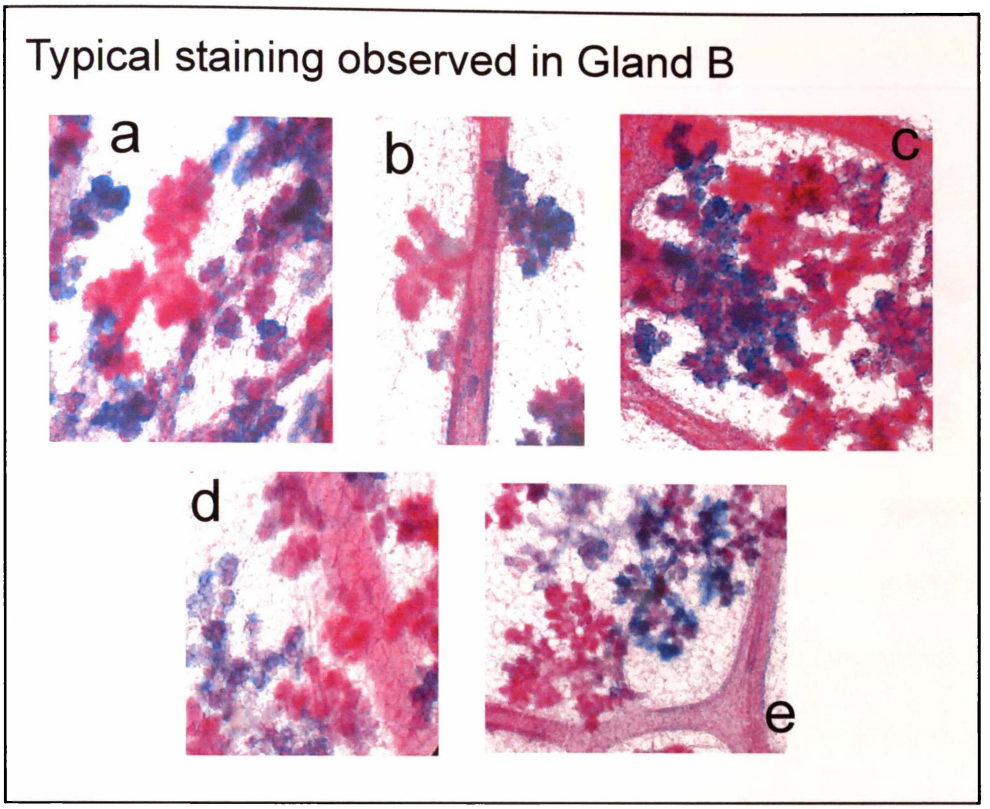


Figure 6.8. Heterogeneously staining ducts from Gland B showing bias for one staining pattern. (A) Homogeneously staining terminal ductal regions (both blue and red) within a polyclonal background. (B) Adjacent lateral buds showing different homogeneous staining after origin from the same polyclonal primary ductal structure. Images (C)-(E) show fine detail of both heterogeneously and homogeneously staining ducts and termini.

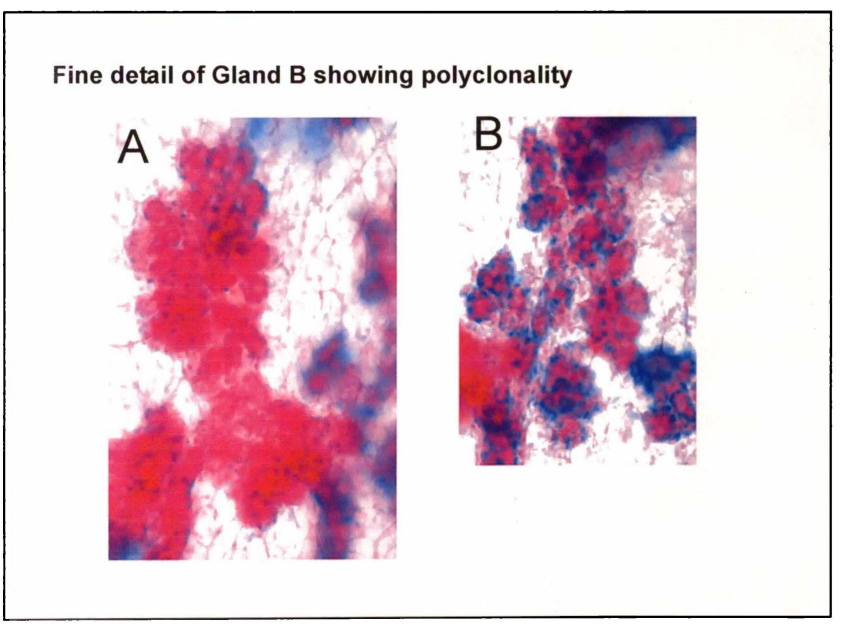


Figure 6.9. Fine detail images from Gland B showing heterogeneity (polyclonality). Images (100x magnification) showing polyclonal terminal ductal structures exhibiting strong bias for (A) red-staining over blue-staining or (B) blue-staining over red-staining.

Gland C

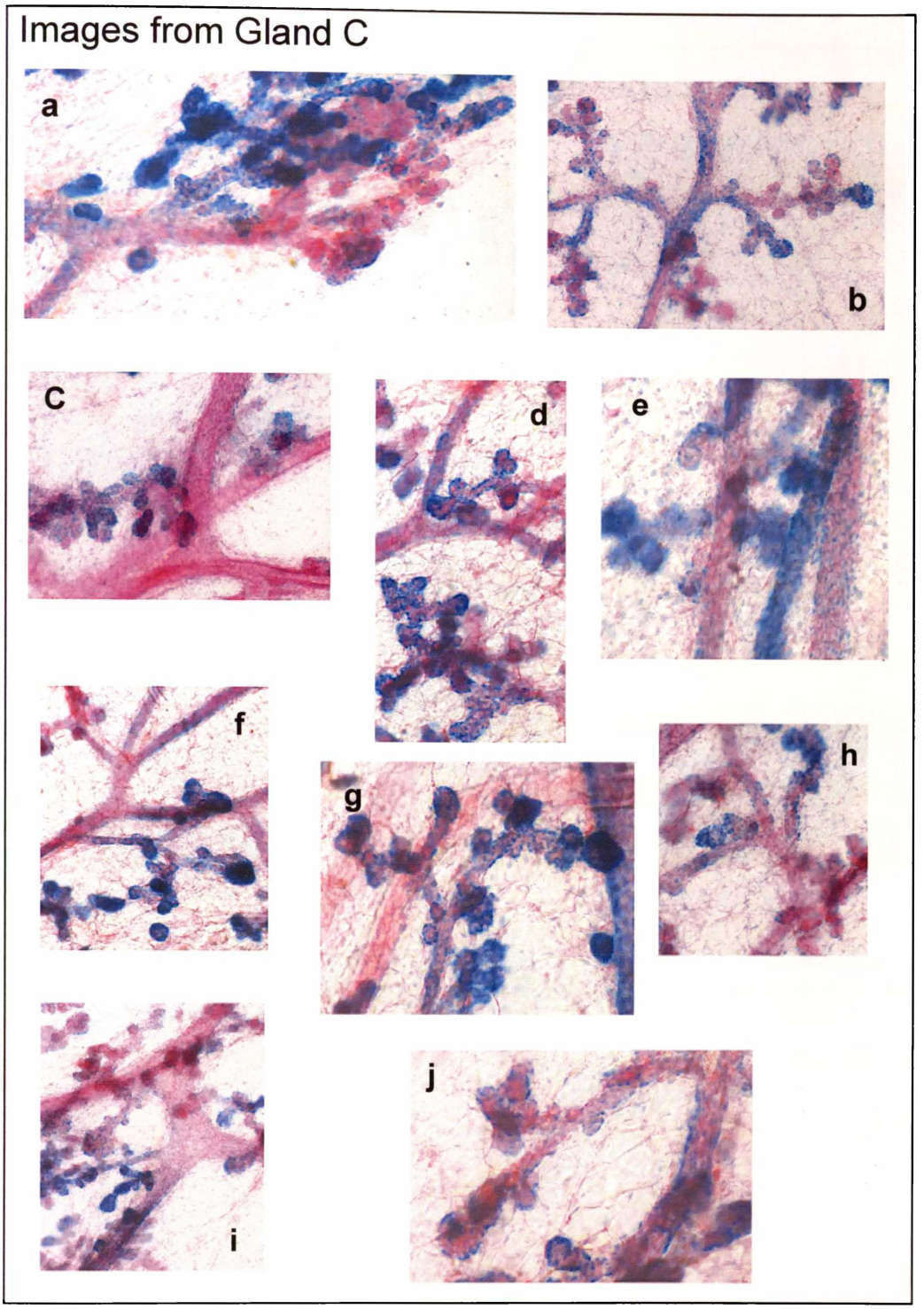


Figure 6.10. Representative images from Gland C at 100x magnification. Showing (D), (G), (H) the predominant polyclonal pattern with (A)-(C), (E), (J), (I) frequent observation of skewed polyclonality or (F) ducts exhibiting homogeneous staining of terminal structures at the peripheral regions of the gland. The scale is the base of (B) which is 1000 μm across, although optical zooming via the digital camera used allowed up to 3x magnification, e.g. this produced a horizontal scale of 300 μm for image (A).

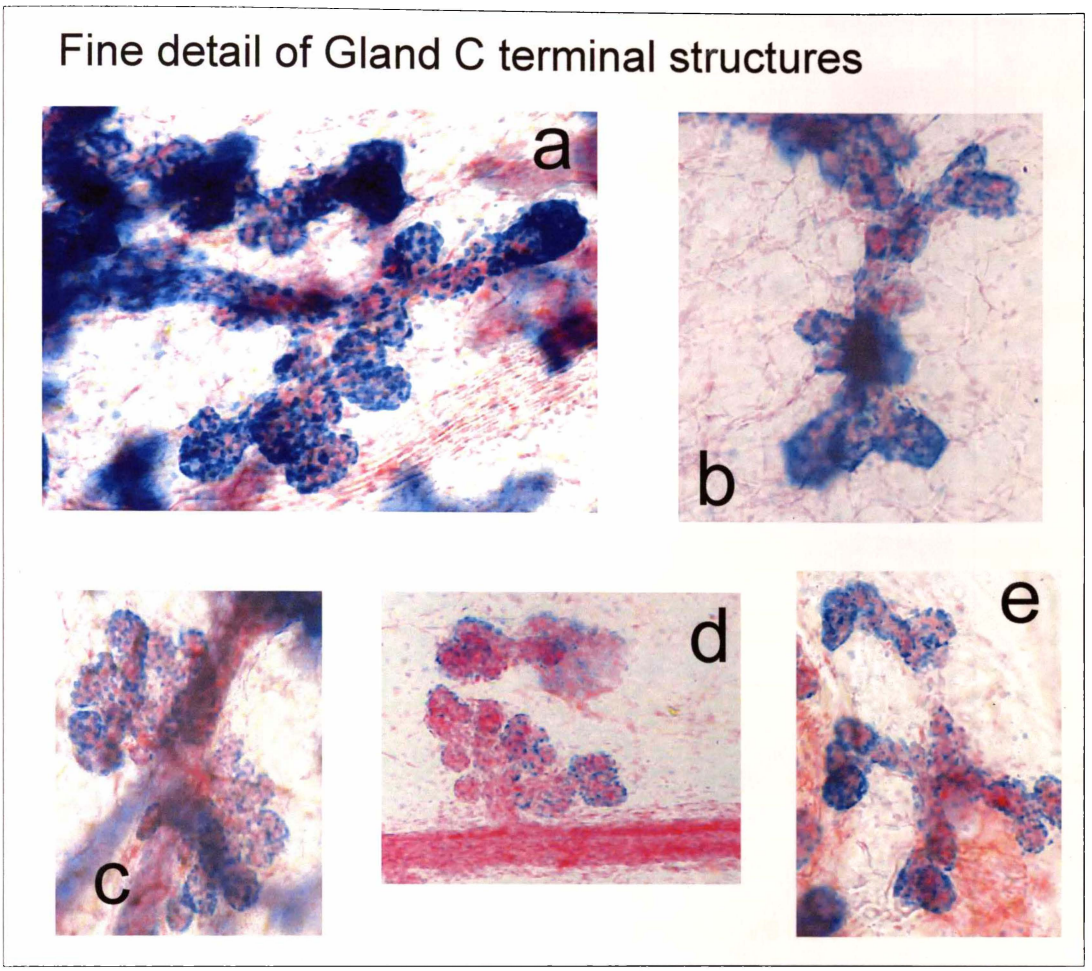


Figure 6.11. Showing high-magnification images of structures within Gland C. In these frames, heterogeneous (Polyclonal) patterns predominate. Although, image (D) shows a duct with markedly skewed staining that branches to produce more evenly polyclonal lateral structures.

Gland D

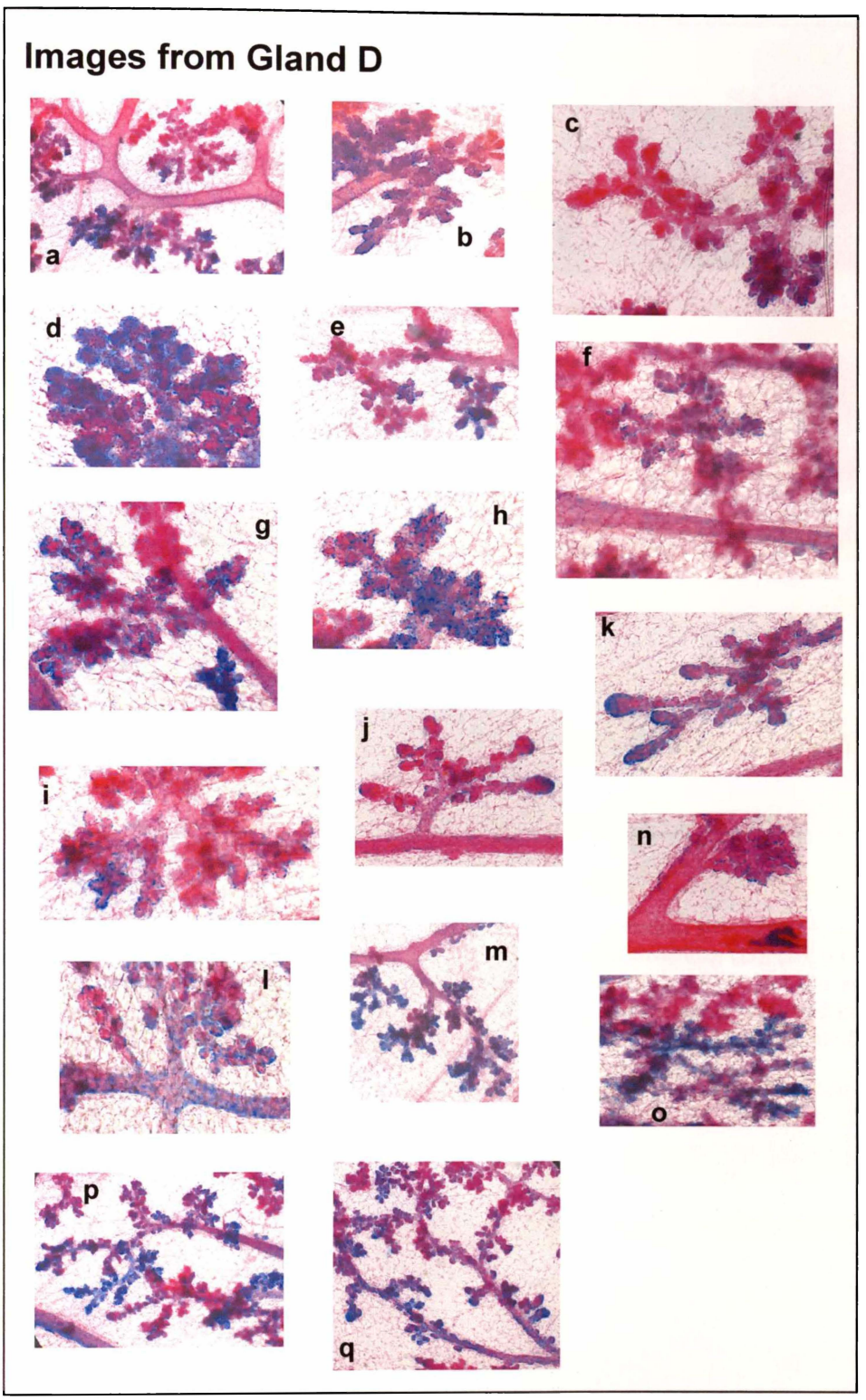


Figure 6.12. Images collected from gland D. Both homogenously and heterogeneously staining regions were observed in the gland, as were TEBs/TEDs showing biased heterogeneity. Image (A) (magnified below) shows all three staining types and has a baseline scale of approximately 1000 μm.

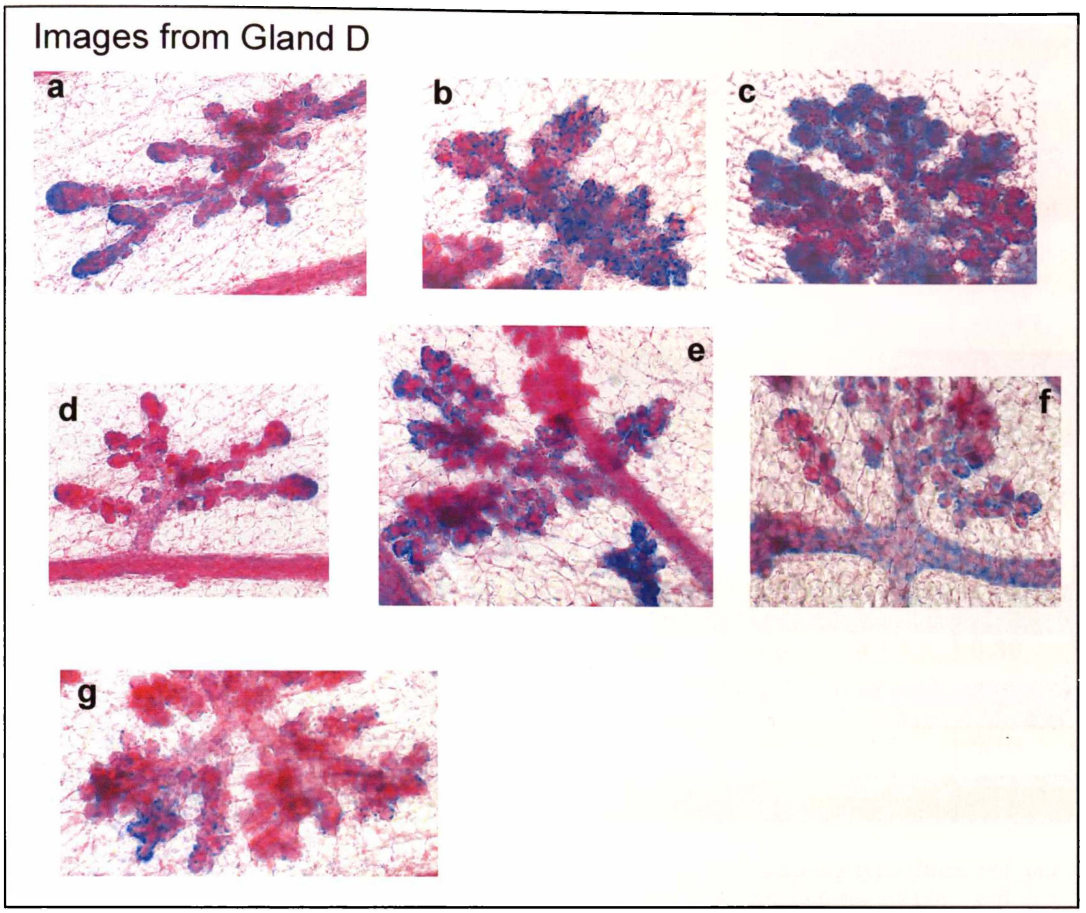


Figure 6.13. Fine detail of termini in images from Gland D. Most terminal structures (A)-(C), (F) show a mixed population of both staining types. A minority of such structures (D), (E), (G) show homogeneous staining within some ducts. The horizontal axis of image (A) is 300 μm and the magnification at the microscope is 100x. Some camera based optical zooming was used in these images to give a horizontal axis of ≈300 μm to the base of image (A) and a similar scale for the other images shown.

6.3.2 Summary of observed staining patterns

6.3.2.1 Whole mount low-magnification staining analysis

Analysis of staining patterns across entire mammary glands revealed relative staining percentages, as shown in the table below (Table 6.6), and in graph format (Figure 6.14).

	Gland A	Gland B	Gland C	Gland D
Total viewable ductal area cm^2	1.16	0.82	0.86	0.64
Total actual ductal area (cm^2)	3.64	2.56	2.69	2.01
Area of predominantly blue-staining ducts (cm^2)	0.45	0.30	0.34	0.11
Blue-staining (%)	30	39	42	18
Area of predominantly red-staining ducts (cm^2)	0.65	0.48	0.17	0.39
Red-staining (%)	44	61	21	62
Area of predominantly magenta staining ducts (cm^2)	0.39	0.00	0.29	0.13
Magenta (polyclonal) staining (%)	26	0	37	20

Table 6.6. Showing ductal area (viewable and total) and areas of each staining type (blue, red and magenta) expressed in cm^2 and as a percentage of total ducts. The percentages of blue, red and magenta staining are calculated relative to the viewable area only.

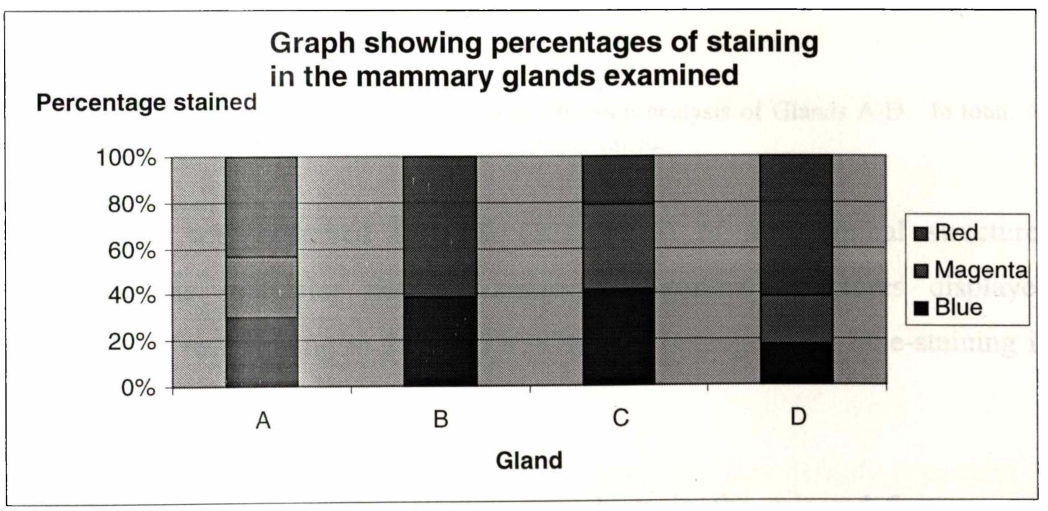


Figure 6.14. Graph showing relative contributions of each staining type to the mammary glands analyzed. Gland B lacked any discernable magenta staining. Furthermore, glands A and D were observed to have a relatively low percentage of blue-staining, with 30% and 18% blue, respectively.

As noted above, the percentage of red-staining shown in this graph (Figure 6.14) may have been overestimated due to the fact that red-staining of ducts was

analysed against a background of lightly stained stromal elements that were stained red.

6.3.2.2 High magnification analysis of staining in Glands A-D

High magnification image analysis of staining patterns in all of the glands examined revealed three staining patterns (Table 6.7). Terminal ductal structures and total ducts were analysed separately for clonality patterns in all observed frames.

	Number examined	% Blue termini	% Polyclonal termini	Red Termini	Entire frames % Blue	Red
Gland A	676	34	42	24	54	46
Gland B	1213	17	52	31	38	62
Gland C	727	29	60	11	53	47
Gland D	1260	12	70	18	41	59
	Combined percentage	21%	57 %	22%	46.5%	53.5%

Total termini observed	3876
Total blue termini	801
Total red termini	857
Total polyclonal termini	2218

Table 6.7. Showing combined data for high magnification analysis of Glands A-D. In total, 40 frames, exhibiting a total of 3876 duct termini, were examined.

Polyclonality was observed in 57% (2218/3876) of all terminal structures throughout the glandular tissue examined. Terminal structures displayed homogeneous red-staining in 22% (857/3876) and homogeneous blue-staining in 21% (801/3876) of all observations.

Total ratios of red to blue-staining within ducts in the selected frames were measured for each set of frames from each gland. Glands A and C, with percentage blue-staining of 54% and 53% respectively, showed staining consistent with the notion that the gland originates from a large progenitor population, with an equal representation of both transgene expressing and transgene non-expressing stem cells. In contrast, Glands B and D, with only 38% and 41% total

blue-staining ducts respectively, were somewhat low in transgene expressing cells (and this was also reflected by the analysis of clonality in terminal structures, where blue to red structure ratios of approximately 1:2 were found). The overall average staining for Glands A-D was 46.5% blue-staining and 53.5% red-staining. This combined average is in close agreement with the 1:1 ratio that is expected if the gland is sourced from a large pool of SC with about half of the SCs expressing the transgene from an active X chromosome.

6.3.2.3 Concordance between low and high magnification analysis of Glands A-D

The relative staining measured by low magnification analysis of all glands showed a slight overestimation of the area of red-staining total mammary tissue compared with the higher resolution analysis, with total red-staining of 57.4% and 53.75% from low and high resolution analysis, respectively. Individual gland variation detected by the two methods was most apparent in Gland A (57% vs. 47% red), Gland C (39% vs. 47% red) and Gland D (72% to 59% red), while Gland B showed almost perfect concordance between the results obtained using these different methods (61% vs. 62% red).

Not surprisingly, gland wide analysis of low magnification images by image manipulation and 'colour dissection' was found to lack the sensitivity of the high magnification technique. This was particularly apparent in the underestimation of blue-staining areas detected using this method and is most likely due to weak histochemical staining resulting in dominance of the counter-staining ducts.

6.3.2.4 Gland E in serial section

General clonal analysis

Selected portions of Gland E, mammary number 4, were serially-sectioned to produce a high magnification 'walk-through' analysis of the numbers of blue- and red-staining cells within ducts and their arrangement within these ducts (Figure 6.15, below).

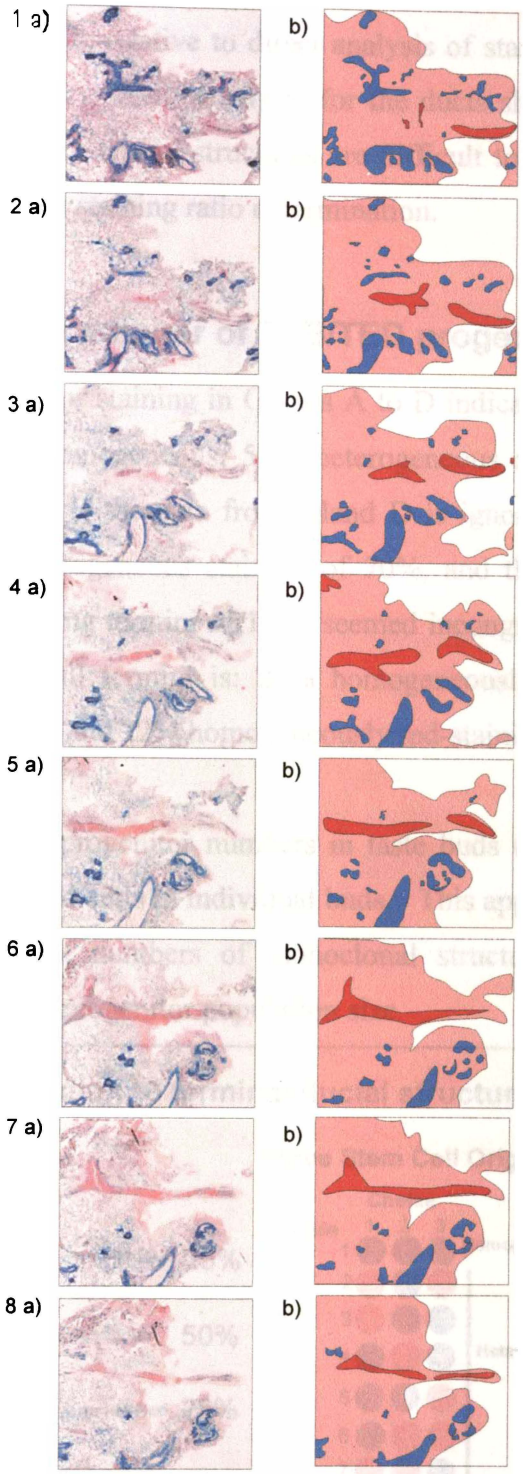


Figure 6.15. Heterogeneously staining (Polyclonal) ducts, with a bias for one staining type, were commonly observed. Sections from Gland E showing what seems to be a heterogeneously staining polyclonal (red stain predominant) duct penetrating a region predominated by other polyclonal (blue stain predominant) ducts. No ducts in these frames were found to exhibit shows complete homogeneity.

In the portion of this gland (~50 mm^2) analysed for clonality, the majority of ducts showed a heterogeneous (polyclonal) staining pattern. In the series shown in Figure 6.15, two polyclonal ducts, one with a bias for blue-staining cells, are

shown to exist adjacent to one another within the same field. A weakness of this serial sectioning approach, relative to direct analysis of staining in whole mount glands, is that the primary 'duct of origin' for the ducts shown is impossible to determine and that actual termini structures are difficult to identify for clonality analysis and blue/red cell staining ratio determination.

6.3.3 Estimating the number of TEB/TED progenitors

The combined average for staining in Glands A to D indicated a 1:3:1 ratio (that is, 21% blue-staining homogeneous, 57% heterogeneous and 22% red-staining homogeneous termini). If the data from Gland D is ignored as 'outlying' (the high percentage of heterogeneous staining of 70%, and the low percentage of homogeneous blue-staining termini of 12%, seemed incongruous), then corrected percentage staining in all termini is: 27% homogeneously blue-staining, 51% heterogeneously staining and 22% homogeneously red-staining.

Stone (2002) estimated progenitor numbers in taste buds using counts of blue-stained versus non-stained cells in individual buds. This approach, represented in Figure 6.16, shows that numbers of monoclonal structures are expected to decrease with increasing progenitor population size.

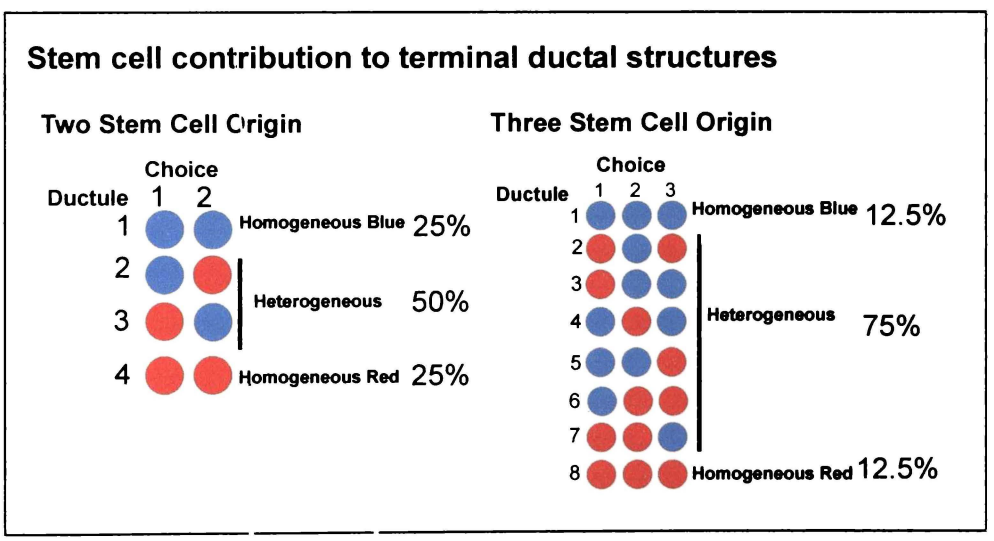


Figure 6.16. Schematic diagram showing cellular clonality ratios expected if either two or three stem cells contribute equally to a terminal duct. The experimental data supports both these patterns and, in practice, both may exist at different termini within the same mouse mammary gland. (Adapted from Stone *et al.* 2002 with modification).

Accordingly, if each structure was derived from two cells, then around 25% of buds would be blue-staining and, for three progenitors, authentic monoclonal blue-staining structures would be seen in only 12.5% of observations.

6.3.3.1 χ^2 test of staining data

The staining data obtained from the high magnification measurement of TED staining in four glands, A to D (see Tables 6.2-6.5), was subjected to χ^2 analysis. The results presented in Table 6.8 show that the observed staining patterns support the hypothesis that two or three distinct progenitor cells are involved in the formation of individual termini.

	2 progenitors ratio 1:2:1	3 progenitors ratio 1:6:1
Gland A	p=0.135	p< 0.01
Gland B	p=0.049	p< 0.01
Gland C	p=0.021	p< 0.01
Gland D	p< 0.01	p=0.087

Table 6.8. Table showing χ^2 test results (p-values) for glands A to D. The three values that were either greater than or very close to p(0.05) are shaded.

6.4 Discussion

6.4.1 Controls and sample suitability

Due to breeding problems in the colony that supplied the H253 mice for this study, we were unable to obtain tissues from homozygous H253 animals for use as a staining control in our experiments. The problems encountered in producing homozygous mice may, in part, reflect an inherent limitation in the H253 line, as Tam (1992) found that homozygotes were generally underrepresented in litters.

Despite the absence of the ideal control for this study, that is, histochemically-stained mammary tissue from a mid-gestational (d12.5-13.5) homozygous H253 mouse, a number of lines of evidence show that H253 expresses the *lacZ* transgene reliably. Tam (1992) showed that *lacZ* expression was ubiquitous from 2.5 dpc to at least 13.5 dpc in H253 homozygous embryos, and others have found that the transgene is ubiquitously expressed in intestinal and epidermal tissues of adult homozygous females or hemizygous males (Bennett *et al.*, 2003; Seymour *et al.*, 2004). In the present study, we found that the ratio of *lacZ* stained to unstained cells in both skin and small intestine from H253 hemizygotes was close to 1:1 (48% and 55% for skin and intestine, respectively; see section 5.3.6 in Chapter 5). We believe that this evidence is sufficient to justify using the H253 mice in studies of clonality in the mammary gland.

A number of preliminary conclusions about cellular clonality in the mammary gland can be made, particularly for those glands that show a ratio of about 1:1 blue- to red-stained cells. In this respect, all of the glands, A to D, seem to fit this requirement.

6.4.2 The majority of ducts display a polyclonal staining pattern

A heterogeneous staining was the most frequently observed pattern in ducts and this indicated polyclonality. We found no large continuous patches of cells of either staining type within these polyclonal larger ducts. This suggests that progeny of SC become "mixed" during the formation of larger ducts or that outgrowth from SC is minimal and produces very small cellular clones.

However, smaller ducts and termini were found to sometimes show an uneven heterogeneous staining pattern, with a predominance of one staining type (examples in Gland C; Figure 6.12 a-b,e). This observation led us to conclude that, while the majority of ducts in the gland may originate from a large population of SC and give rise to a largely heterogeneously staining ductal network, smaller structures, in particular those some distance from the gland origin, sometimes show staining that indicates derivation from a very small number of stem cells.

6.4.3 Terminal portions of the ductal network stain either heterogeneously or homogeneously for the transgenic marker

In general, our findings were in keeping with previous observations in the murine mammary gland of 1) predominant heterogeneity in larger ducts, indicating polyclonality and, 2) frequent homogeneous staining in termini, indicating the presence of a single clone with the same inactive-X choice or of a "patch" with similarly staining clones adjacent to one another (Thomas *et al.*, 1988). In our observations of 3876 terminal structures each comprising around 500 cells, homogeneous staining was detected in 43% (1658/3876) of termini (22% red-staining, 21% blue-staining), with the remaining 57% (2218/3876) of termini exhibiting a heterogeneous staining pattern, indicating polyclonality.

Outgrowth from a single stem cell is almost certainly capable of forming genuinely monoclonal TEBs. However, an alternative explanation, that better fits the data collected in this study, is that two or more stem cells with an identical inactive-X choice, and therefore identical transgene expression patterns, produce homogeneously staining TEBs. This is in agreement with a report by Kissenberth and Sandgren (2004), who found both homogeneous and heterogeneous staining in individual alveoli in sections taken from chimeric mammary glands. This possibility is discussed below in relation to the estimation of stem cell numbers contributing to individual termini (section 6.4.5).

Homogeneous staining for transgene non-expressing (red-staining) cells was observed at a high frequency in some peripheral regions of the gland. Examples

of the effect are shown, in the figures showing staining in Gland A; Figure 6.7 (D), (E), (F), (J), (L) and M). Moreover, this homogeneous staining was more common in the extremities of the gland space, particularly in the periphery of Gland A, an inguinal mammary gland that, in contrast to thoracic glands, is elevated along its longitudinal axis. Whether this anatomical feature is important remains to be determined.

There is a possibility that the mammary epithelia underlying the minimal fat pad found in peripheral regions of the gland is comprised of cells exhibiting active X-chromosomes bearing the HMG-*lacZ* transgene construct. However, if this was the case, then one would expect that the blue histochemical stain would be detected in these areas. A possible explanation for the lack of detectable staining is that these cells fail to express the transgene product in detectable quantities because of the absence of strong, stromally-mediated, upregulation of transgene expression (see Chapter 5, section 5.4.3, for an explanation of the hypothesized link between systemic estrogen levels and transgene expression in the H253 mouse mammary gland). This may be due to reduced expression of estrogen receptor (ER) in and around TEBs located in these regions, which are close to the borders of the fat pad, although there are no reports suggesting such an expression pattern. Mitogen-responsiveness would be lost by a reduction in expression of ER resulting in reduced expression of the transgene. In future studies, this effect could be tested by analysing histochemical staining patterns in H253 homozygotes.

6.4.4 Direct examination of whole mount tissues

Examination of staining patterns in *lacZ* histochemically stained and fat-cleared mammary glands under low magnification showed clearly demarcated regions of transgene-expressing blue-staining and 'counterstained only' red-staining ductal tissue. In some glands, hue-adjustment using image analysis software allowed separation of red-staining from blue-staining regions by replacement of the non-target colour with white, and measurement of the remaining area. However, this technique was found to be possible only in those glands that were lightly counterstained. The higher magnification images yielded far more information on clone size and distribution than obtained using the low magnification images.

Accordingly, in future studies, these high magnification images would be used in preference to the low magnification images.

6.4.5 How many stem cells contribute to single ductal termini?

Clonality studies that used coat colour as a marker of epidermal cell origin in mice exhibiting Cattanach's translocation (In7;X) revealed a low frequency of extreme skewedness of predominance of either active translocation-bearing or normal chromosomes (Nesbitt, 1971). This observation suggests that murine epidermis is derived from a large population of precursor cells. Alternatively, where a tissue is derived from a very small number of progenitors, a considerable stochastic skewing in cells is likely to be observed and this would be reflected in a departure from the 1:1 ratio of X_p to X_m inactive X chromosome choice. Extending these predictions further, if a single stem cell contributed to each terminal end bud or duct, then completely homogeneous staining should be observed in all termini. Furthermore, if each TEB was derived from two stem cells then a overall 'homogeneously staining red:heterogeneous:homogeneously staining blue' ratio of 1:2:1 would be observed in all termini (see schematic in Figure 6.16).

Statistical analysis of the 'observed' staining ratios and the ratios predicted from experimental observations ('expected') by the χ^2 test, showed that patterns in Glands A and B, with p-values $\geq p_{(0.05)}$ did not deviate from the expected ratio 1:2:1 ratio for blue:blue/red:red termini if TEB structures in these glands were derived from two stem cells. The results from the Gland C data did not support an origin of termini from either two or three stem cells. Gland D, with a p-value of 0.087 did not deviate from an expected 1:6:1 ratio, that is, suggesting that TEB structures in this gland may be derived from three stem cells.

While the observed overall ratios of blue-, red- and mixed staining indicates an origin of termini from two mammary progenitor SCs, the observation in all glands of terminal structures that were polyclonal but exhibited biased staining patterns, presumably as the result of stochastic effects, supports an origin of many termini from three division-competent SCs.

6.4.6 Do numbers of available mammary stem cells reduce with gland outgrowth?

We believe that one possible explanation for the observation of increased numbers of homogeneously staining terminal structures in peripheral regions of the glands could be that SC 'units' (that is, SCs and their proliferation-capable daughter cells) are reduced in number as gland growth proceeds, due to these units becoming committed to the formation of lateral branches. This reduction may increase the likelihood that a single terminal structure may be sourced from a low number of SCs (or even a single SC) and hence display a homogeneous staining pattern. In this situation, the gland would, in effect, begin to "run out" of stem cells as growth progresses towards the periphery of the gland, thereby placing a natural limitation on further growth.

If a reduction in available SC numbers does occur during murine mammary gland development then it is possible that a similar pattern of reduction exists in the human breast, in which entire TDLUs have been found to exhibit homogeneous expression of X-linked markers (see Diallo *et al.*, 2001), with a single patch encompassing a number of TDLU (Novelli *et al.*, 2003).

6.4.7 Models of normal murine mammary clonality

As discussed above, we found that large primary ductal structures were comprised of a heterogeneous mixture of cells of each staining type, that is, such structures had a polyclonal origin. In contrast, smaller ducts and lateral branches were sometimes observed to stain in a homogeneous pattern with 21% showing blue staining and 22% negative for the *lacZ* marker, while the remaining 57% of termini showed a polyclonal staining pattern.

These data suggest that, in normal mammary development, a heterogeneous starting population of mammary progenitors in the mammary bud expands to produce a primary ductal structure comprising many individual cellular clones (Figure 6.17). It has been reported that within such primary structures, transitional units occur on average every 250 μm and also possibly within larger TEBs (Kenney *et al.*, 2001). These transitional units contain SCs and their

immediate progeny (the transiently activated cells) that respond to particular systemic and stromally-derived signals by initiating lateral budding. As is shown schematically in the figure, homogeneous staining is observed when a lateral bud forms either from a single SC clone or from a small number of SCs that give the same staining pattern, that is, a stochastic effect. A similar observation would be made if total SC number in the mammary precursor pool was very low, hence resulting in an increased chance of a stochastic effect leading to a skewed (non 1:1 stained to unstained cells) profile. However, as widespread staining heterogeneity was seen in the glands examined in this study, we assume that the glands were indeed derived from large populations of progenitors.

Observations of glands from pregnant animals suggested that a reasonably large population of SCs must be involved in the initial population of the fat pad by mammary epithelia. Alternatively, mammary progenitors may be deposited along the primary ducts, where they remain until they are subjected to signals prompting their growth and the formation of lateral buds. In this model, some progenitor cells in the adult gland may be mitotically senescent because they are not within a stem cell niche (Schofield, 1983) that is subject to signals for growth. This could be due to the fact some such progenitors may not be correctly positioned to allow for even spacing between individual lateral buds, which is a characteristic of mouse mammary gland architecture. After the cessation of lactation, involution would reduce the mammary network back to a "resting state", with SC clones remaining as transitional populations for later repopulation of the fat pad, which is required in subsequent pregnancies.

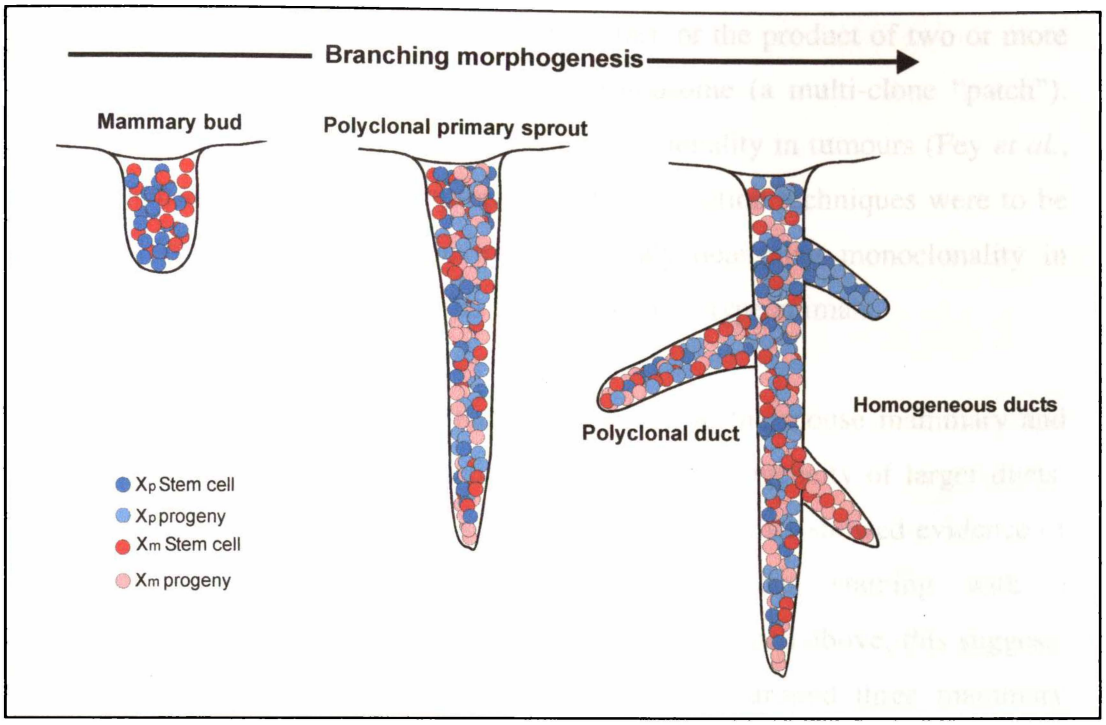


Figure 6.17. Schematic a showing a general model of clonality for the mouse mammary. Red and pink cells represent, respectively, SCs and their daughter cells that have an inactive maternally derived X chromosome. Dark blue and pale blue cells represent, respectively, stem cells and their daughter cells that are derived from cells with an inactive paternally-derived X-chromosome. As shown in this diagram, outgrowth from precursors in the mammary bud produces predominantly polyclonal ductal epithelia. Homogeneity in ducts results from outgrowth of a small number of SCs that both express the same marker.

Further investigation is required to test whether this alternative model more accurately describes clonality in the murine mammary gland. Such investigation would include evaluation of expression patterns of both spatial and temporal SC-specific markers so that growth from these SC reservoirs deposited during initial gland development could be measured.

6.4.8 General aspects of the H253 mouse as a mammary gland clonality model

It is not possible to confirm monoclonality by whole mount examination of histochemical staining patterns. While the observation of homogeneity in a terminal glandular structure (such as a lateral bud/branch) indicates that cells within the structure are derived from progenitors with the same inactive X chromosome choice (that produce identical staining), it does not prove a clonal origin from a single stem cell. This is because there is currently no method available to determine whether such a structure is the product of a single SC

(producing an authentically monoclonal structure), or the product of two or more SCs with the same choice of inactive X-chromosome (a multi-clone "patch"). DNA fingerprinting has been used to distinguish clonality in tumours (Fey *et al.*, 1988; Ng *et al.*, 2003) and, if appropriate microdissection techniques were to be applied, it may also be possible to conclusively determine monoclonality in individual TEB or TED using similar methods in wild-type animals.

True polyclonality is the predominant pattern seen in the mouse mammary and this is evidenced by direct observation of staining heterogeneity of larger ducts. Of these heterogeneously staining structures, a large number showed evidence of "stochastic skewing" effects, exhibiting heterogeneous staining with a predominance of one staining type over the other. As noted above, this suggests that these termini are derived from a small pool of around three mammary precursor cells, and this was supported by statistical analysis of patterns found in Gland D at higher magnification and by observations of various images from all of the glands in this study.

6.4.9 Cellular clonality in murine and bovine mammary tissue

As outlined at the beginning of Chapter 5, the mouse gland study was carried out in order to measure individual cellular clones in the murine mammary gland and to provide an experimental base upon which to base future bovine experiments; ultimately, both types of studies are envisaged as useful models for the study of human breast disease. It is, therefore, valuable to compare the two experimental approaches used in studying the murine and bovine glands, so that conclusions about general similarities in patterns of cellular clonality can be drawn.

The first and most obvious comment to be made is that it would be totally impossible to detect clonality in murine mammary blocks of the same size as those sampled in the bovine experiments. Recall that these were $30mm^3$ in volume; a block of equal size would comprise some 4% of the murine gland, which weighs around 700 mg. This sample would contain 100 TEBs and would include both ductal tissue (57% of which display a heterogeneous pattern) and terminal end ducts/buds (of which only about 43% exhibited homogeneous staining); clearly, homogeneity would never be detected in murine samples of this

size. Given that, in the mouse gland, homogeneity is only found in individual TED/TEB, one would presumably have to take much smaller samples (a fraction of a mg) in order to detect homogeneous *lacZ* expression patterns using the *in vitro* RT-PCR technique.

If the bovine mammary gland had a similar clonal architecture to the mouse gland, equally small samples would have to analysed and one would have to deal with the added complication that, because of the technical reasons one allele of the X-linked marker (SNP) could only be detected when that allele was present at a ratio equal to or greater than 3.5:1 compared to the alternate allele (Chapter 4, section 4.2.4.2). However, the preliminary results presented in Chapter 4 do indicate that regions of clonality in bovine tissue may be somewhat larger than in mouse, of the order of several mm^3 – if this proves to be the case, it would represent an important difference between the two mammary glands, and also make bovine *in vitro* RT-PCR analyses more feasible.

6.4.10 Future work:

- To confirm that expression of the transgene occurs in all cells in mammary tissue, a homozygous H253 animal needs to be sourced as a control. Presently, we assume that staining would be found to be ubiquitous in such glands and in tissues from male hemizygous H253 animals.
- If the transgene was found to be expressed at reasonable levels, perhaps after induction of estrogen production, then glands could be analysed from animals at various developmental time points such as embryonic, new-born, pre-pubertal, lactating, nonpregnant multiparous animals and from animals undergoing mammary involution. This would help to determine the clonal characteristics of the precursor population in the embryo and would show chronological changes occurring in the clonal architecture of the gland throughout the life of the animal.
- Alternative nuclear-specific counter-stains, such as nuclear fast red or brazilin, could be tested on whole mounts or thin sections to better distinguish the transgene expressing cells from the counterstaining non-expressing cells.

- An alternative model, such as a transgenic mouse expressing an X-linked EGFP gene, is likely to provide clearer insights into the size and distribution of cellular clones in murine mammary tissue. Transgene expression profiles in such alternative model animals would have to be rigorously examined to confirm their utility in cellular clonality studies.
- DNA fingerprinting techniques may be able to be used in determining the extent of authentic monoclonality of termini in mammary samples from wild-type animals (the inbred nature of lab strains make them unsuitable for such analyses).

Chapter Seven

General discussion

A significant limitation that arises when using X-chromosome inactivation based assays to investigate clonality of both normal and diseased tissue is that, when the patch size is large, polyclonality cannot be determined with complete certainty. This is because polyclonal structures may arise from stem cells in the same "patch" and hence display the same X-linked marker, despite their polyclonal origin (Schmidt and Mead, 1990). Thus, if the patch size in mammary tissue was large, encompassing many hundreds of termini, then it would be impossible to conclude anything about the patterns of clonality in the individual termini themselves. Obviously, this limitation is negated if the patch size is small, and therefore useful data about normal patterns of clonality can be obtained from tissues that exhibit small patches.

In the present study, data on clone and patch size and distribution in the murine mammary gland was obtained. These data indicated small patch sizes in the murine mammary gland, which means that the murine gland could be a suitable model in which to test theories about the frequency of mammary tumours that have a polyclonal origin but a clonal evolution. Such studies in the murine gland have the potential to give a better understanding of tumour origin and biology. It is likely that patterns of clonality in murine mammary tissue would approximate those is human mammary tissue, as the murine TEB and the human TDLU - and the bovine TDLU - have anatomical, and probably developmental, similarities.

Patch and clone size estimates for Murine mammary

Patch size, defined as the average size of groups of contiguous cells that share a common (X chromosome) epigenotype (Nesbitt, 1974), is important as it determines the extent to which conclusions about clone size and distribution can be made.

The existing experimental evidence indicates that patch size differs within and between species. In thin sections of the human mammary gland, the ductal epithelial patch size has been estimated to be between 0.08 and 0.45 mm^2, and these relatively large patches have been found to include entire TDLU (Tsai *et al.*, 1996). Another study, in which patches were measured using heterozygous expression of a temperature-sensitive human G6PD variant as a marker of X-inactivation, showed that patch size in the human breast is much larger than that which was observed in the Tsai study, and that entire TDLU, of between 1 and 8 mm^2 in diameter, may lie within the same patch (Novelli *et al.*, 2003).

In the present study, our experimental data showed that the ductal termini, with an estimated area of around 0.03 mm^2 when viewed in 2D, are comprised of two or three cellular clones, thereby making the average clone size between 0.015 and 0.01 mm^2. Around 40% of these termini stained homogenously, thereby suggesting that patches could be as large as 0.03 mm^2. However, as about 60% of termini showed heterogeneous staining, the average patch size is likely to be smaller than 0.03 mm^2 and that these patches may comprise of progeny from a single stem cell. Clone size can also be estimated in a quite independent manner, using the total ductal area for 'Gland A' (36 mm^2) coupled with the estimate of 900-1000 stem cells in the inguinal murine mammary gland (Kordon and Smith, 1998). Considered together, this would indicate an average clone size of around 0.04 mm^2 (36 mm^2 divided by 950 SCs). If clones in the thoracic glands are approximately the same size, then we can estimate that the thoracic mammary glands analysed in this study (Glands B to D), with an average total ductal area of 24 mm^2, have around 600 stem cells (24 mm^2 divided by 0.04 mm^2). Using this method it is possible to predict that patch size could be as large as 0.08 mm^2 (two clone diameters); although, using the present data, it is not possible to accurately determine the average patch size. Taken together, these data indicate that murine mammary patches are much smaller than those observed in humans, with the *maximum* observed patch sizes only just approaching the *minimum* patch sizes observed in the human gland. These relatively small patches could render the murine mammary gland a potentially useful model of clonal processes, because, if more than one clone exists in an examined region then this polyclonal pattern is likely be detected.

The general observation of "fine-grained patchiness" in patterns of cellular clones throughout the murine gland suggests that mammary tissue is similar, in this respect, to murine liver, which is comprised of $>10^6$ small clones and is thought most likely to result from considerable "mixing" of lineages during development (West, 1976). In contrast, we observed that skin from H253 hemizygotes is comprised of large clonal patches, with "mixing" only at the patch boundaries (see Figure A7 in Appendix 4), which agrees with observations of large transverse stripes in chimeric mouse epidermis (Schmidt *et al.*, 1988). It is likely that glands originating from this epithelium would fall within a homogenous patch, so if cells for the initial mammary gland progenitor population arose from a portion of the epidermal cell population, to form the mammary bud, then it would be expected that many of the glands formed would be completely monoclonal. However, monoclonality was not observed in any of the five glands analysed in this study, suggesting that the mammary committed stem cell population becomes determined before the formation of the epidermis and that this cellular population is maintained during formation of the epidermis.

As noted in the introduction (Chapter 1.14), it has been proposed that all epithelial organs are likely to all be formed by similar processes to those responsible for the formation of intestinal crypt-villi (Zajicek, 1995). In the intestine, polyclonal crypts in the neonatal mouse undergo a post-natal "purification" to monoclonality (Schmidt *et al.*, 1988). However, we found no evidence for such "purification" in the mammary gland, which would have been observed as mainly homogeneous individual termini and perhaps even larger structures. It is worth noting, however, that there are structural similarities between the large intestinal crypt, in which stem cells are thought to reside at the base of the crypt (reviewed in Potten *et al.*, 1997), and the TEB, where stem cells are thought to reside at the tip of the bud (Ormerod and Rudland, 1984).

Predominant polyclonality in mammary ducts and termini

Termini and ducts of the adult murine mammary gland from the H253 hemizygous animals used in our study stained in a heterogeneous manner, indicating polyclonality. This polyclonality, which is supported by the findings of others (Kisseberth and Sandgren, 2004; Thomas *et al.*, 1988), is perhaps most

interesting when viewed alongside the additional observations made by Thomas *et al.* (1988), who found large mammary epithelial patches during early development. The presence of these large patches was thought to be due to coherent clonal growth. From the present evidence, it is difficult to reconcile Thomas' observation of large patches in the embryonic gland, with the extensively polyclonal tissue seen in the adult gland. Although, according to the model proposed in the previous chapter (Chapter 6, section 6.4.7), stem cells are deposited in niches at the base of the luminal epithelia during the formation of simple primary mammary ducts, and these cells may be committed to the formation of individual lateral ducts. The high frequency of polyclonal termini detected in the glands indicates that, despite the relatively large size of the early mammary epithelial patch, stem cell clones from both X_p- and X_m- inactivated 'lineages' are represented in most lateral ducts. Perhaps the mammary gland is derived from a small number of progenitors and symmetrical division of stem cells occurs during the early stages of gland development, thereby explaining the number of individual stem cell derived clones seen in the adult gland. Alternatively, stem-cell-like transiently activated cells may be produced by asymmetrical division at an early stage of gland development and these may function in the formation of small polyclonal structures in the adult gland.

From the results of serial cell transplantation experiments, it has been theorized that 25-27 doublings of a single mammary stem cell in a surgically deepithelialized fat pad could produce an entire murine mammary gland (Kordon and Smith, 1998). However, normal mammary tissue is clearly polyclonal, and the normal colonization of the mammary fat pad is likely to occur via developmentally committed lobular and ductal progenitor cells as this provides a safeguard against aberrant stem cell behaviour by limiting the number of cell divisions required by any one stem cell (Kordon and Smith, 1998). These transplantation experiments also indicated that the entire murine inguinal mammary gland is comprised of approximately 2.5 x 10^6 cells. Taken together with the estimate of 950 stem cells in this mammary gland, one would calculate that outgrowth from each stem cell is responsible for around 2700 daughter cells. This number is difficult to reconcile with the observation of polyclonal staining patterns in termini (each of ~500 cells), indicating that a much lower number of

cells, perhaps $1/10^{th}$ of the theoretical daughter cell number (~250 cells, half of total cell number in the termini) were produced from each stem cell. Possible explanations for this difference are that either (1) in termini, clones are much smaller than average or, (2) transiently activated cells function as secondary stem cells during gland development. Our observations, of small clones of <500 cells distributed throughout the glands analysed, support the second suggestion.

Termini and Hyperplasic Alveolar Nodules (HANs)

Until quite recently, a basic tenet of tumour biology was that cancer is a clonal disease, with entire tumours being derived from individual mutant cells. However, experimental reports that disprove this general tenet are now plentiful (reviewed in Garcia *et al.*, 2000). In the case of mammary carcinoma, evidence for polyclonality has come from both PCR-based studies, in which polyclonality was demonstrated in 4 of 12 (33%) invasive breast carcinoma (Going *et al.*, 2001), and from studies of chromosomal banding patterns, where 27 of 79 (34%) primary breast carcinomas were demonstrated to contain cells having distinguishable cytogenetic differences (Pandis *et al.*, 1995), therefore representing different clones. In addition, other studies using human breast carcinoma culture have revealed that nearly 50% of benign and malignant tumours contain cytogenetically distinguishable (different) clones (reviewed in Heim *et al.*, 2001). Although a number of X inactivation based studies have stated that mammary tumours are generally monoclonal (Diallo *et al.*, 2001; Noguchi *et al.*, 1992), it is possible that the observed clonal patterns were simply a "pseudomonoclonality" resulting from the evolution of a single dominant clone from a tumour with a polyclonal origin (Teixeira *et al.*, 1994).

A particularly powerful technique for studying patterns of clonality in normal and precancerous murine mammary gland tissues involves the creation of chimeric glands. Of particular note, Kissenberth and Sandgren (2004) created a transplantation chimeric model of mammary gland hyperplasic alveolar nodules (HANs), by introducing into a cleared fat pad mammary epithelial cells from two different lines that both expressed the transforming growth factor-alpha (*TGF-α*) gene and one of two cellular markers, either human cytokeratin 18 (hCK18) or human placental alkaline phosphatase (R26-hPAP). In alveoli of chimeric glands,

they observed relatively large patches of brown (hCK18) staining with, for example, a section from one gland showing individual alveoli that displayed: 1) all brown staining apart from a single non-brown cell, 2) half brown and half non-brown staining, and 3) a mixture of both brown and non-brown staining cells These observations agree very closely with those obtained in our H253 mouse study. One notable exception was that we did not observe any (*lacZ*+ve) blue-staining alveoli in which just a single non-blue-staining cell was present. This difference is almost certainly due to our use of whole mounts which gave use the ability to view alveoli in their entirety whereas Kissenberth and Sandgren (2004) used thin sections of the mammary gland, which inevitability, as they pointed out, led to an underestimation of numbers of polyclonal alveoli. It was also of interest that we found many termini in which cells were either totally non-staining or totally blue-staining whereas Kissenberth and Sandgren did not observed such homogenous staining. This difference may simply reflect the fact that we were observing patch sizes originating from X-inactivation events, whereas the "patch" sizes observed in chimeric glands reflected a quite different phenomenon, namely the seeding of the fat pad with two types of mammary epithelial cells.

Kisseberth and Sandgren (2004) also examined the clonal profile of precancerous mammary lesions in the chimeric gland. Patterns of staining in these chimeric glands indicated that 38% (52 of 134) of hyperplasic alveolar nodules (HANs), were mixed chimeric (that is, definitely polyclonal). This led to the suggestion that heterogeneous HANs may represent a "target" population within which a clonal tumour might later arise by "clonal evolution", which is the selective outgrowth of a single lineage from a larger initial tumour clone population. The high number of polyclonal HANs also hinted at the possibility that cellular interaction, via direct cell-to-cell contact or paracrine factors, may be important in neoplasia formation. Our experimental findings clearly demonstrated that individual TEB have a polyclonal origin, specifically, from two to three stem cells, and this strongly suggests that intercellular interactions, like those suggested to occur in some HANs, could also operate in directing the formation of alveoli.

Future investigations

Although it would be possible to marginally improve the quality of the data we have obtained on clonality in the mouse mammary gland by using more robust detection systems, a basic limitation of all " X-inactivation two-marker" detection systems is the relatively large patch sizes. All of the data on the mouse mammary gland, including ours, is in broad agreement in that it indicates that all mammary ducts and at least 60% of TEBs are polyclonal. One would predict that a large amount of additional experimentation might only reap small amounts of new information.

However, one approach that could bring considerably more power to the analyses would be to extend the method of Kissenberth and Sandgren (2004) by using a large number of transgenic cell types to create a chimeric mammary gland. In principle, it would be relatively simple to set up such experiments if one could source a number of congenic transgenic animals that each expressed a distinct marker that was detectable immunohistochemically. If four such distinguishable cell lines were used in the construction of a chimeric transplantation mammary gland, then there is only a 6.25% (0.25^2x100) chance that a polyclonal structure with a 2-cell origin will stain homogeneously for one of the four cell line specific markers. Extending this suggestion further, if a 10 cell chimeric gland could be produced then the percentage chance of homogenous staining in termini with an origin from 2-cells would be reduced to just 1% (0.1^2x100); although constructing a gland from so many cell lines would be technically difficult. Accordingly, within a transplantation chimeric gland comprised of many different and distinguishable lineages, the actual number of cell lines responsible for an individual terminus, alveolus or HAN, could be measured with some degree of certainty.

While the power of the mouse mammary gland analyses might be quite considerably improved by employing a larger number of marked stem cells, it would be somewhat impractical to do such analyses in larger mammals such as farm animals, or in humans. In these animals, one is essentially limited to the "two-marker" approach based on inactivation of one or the other X-linked allele. So, in the immediate future, the main improvements that could be made to the

detection and measurement of stem cell clones revolve around better methods of detecting and distinguishing the expression of alleles of X-linked genes. Some methods of doing this have already been discussed in Chapter 4, and we believe that improvements in *XIST* RNA detection still has considerable potential, given that the expression of this gene is a canonical marker of X-inactivation, and that 32% of cattle are heterozygous for the SNP that we have identified in the RNA. Despite the fact that patches may still prove a confounding issue, there are a number of issues concerning cellular clonality in the bovine mammary gland that could be addressed:

1) What are the patterns of monoclonality in bovine TDLU?
2) Are the patterns observed in first lactations similar to those seen in subsequent lactations?
3) Are the patterns similar to those observed in murine studies, and are sizes of monoclonal regions, patches etc also similar?
4) Are there any correlations between regions that are clonal and those that do, and do not, express milk protein genes, especially in pregnant and/or beef animals? (Molenaar *et al.*, 1992)
5) Is the bovine gland a more suitable model of human disease than the mouse gland?

Beyond the relatively straightforward questions concerning the clonal origins of cells that constitute the mammary gland, there are a number of more complex biological issues that concern the interactions between these cells. Some of these revolve around the basic biology of the normal gland, such as the nature of the cell-to-cell or paracrine communication that is thought to exist between stem cell clones within individual termini/alveoli or across the entire gland, and how manipulation of such pathways affects mammary growth. Then there are a whole series of questions relating to precancerous changes in mammary tissue and the actual growth of mammary tumours. That the hyperplasia occurring in HANs appears to be polyclonal in origin is intriguing, and may be highly relevant to the observation that an appreciable fraction of (human) tumours are not monoclonal. And the question of just why a tumour is polyclonal is far from resolved. Could it reflect some basic biology of the mammary gland in which adjacent cells communicate and interact in some mutually stimulatory fashion, or are there some

quite bizarre phenomena taking place in neoplastic tissue, such as cells taking up DNA from adjacent apoptosing cells? (Holmgren *et al.*, 1999). Moreover, could the vast majority of breast tumours originate from polyclonal growth and it is only some type of later clonal evolution within the tumour that lead to the predominance of the clone that results in molecular analysis classifying the tumour as "monoclonal", whereas a much earlier analysis might have given a "polyclonal" result. And is this predominant clonal outgrowth closely related to the epithelial stem cells and should it be the main target of anti-cancer therapies?

Clearly, such future investigations extend well beyond the subject of this thesis, but the techniques and results presented here would provide a valuable framework for many of these studies.

References

Adami, J.G. (1901) The causation of cancerous and other new growths. *Br Med. J.*: 621-628.

Alford, D., Baeckstrom, D., Geyp, M., Pitha, P., and Taylor-Papadimitriou, J. (1998) Integrin-matrix interactions affect the form of the structures developing from human mammary epithelial cells in collagen or fibrin gels. *J Cell Sci* **111 (Pt 4)**: 521-532.

Alvi, A.J., Clayton, H., Joshi, C., Enver, T., Ashworth, A., Vivanco Md, M.M., Dale, T.C., and Smalley, M.J. (2002) Functional and molecular characterisation of mammary side population cells. *Breast Cancer Res* **5**: R1-8.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1997) *Current Protocols in Molecular Biology.* New York: John Wiley and Sons.

Avner, R., Wahrman, J., Richler, C., Ayoub, N., Friedmann, A., Laufer, N., and Mitrani-Rosenbaum, S. (2000) X inactivation-specific transcript expression in mouse oocytes and zygotes. *Mol Hum Reprod* **6**: 591-594.

Balasubramanian, S., Harrison, P., Hegyi, H., Bertone, P., Luscombe, N., Echols, N., McGarvey, P., Zhang, Z., and Gerstein, M. (2002) SNPs on human chromosomes 21 and 22 - analysis in terms of protein features and pseudogenes. *Pharmacogenomics* **3**: 393.

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

Barr, M.L., and Bertram, E.G. (1949) A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**: 676-677.

Barraclough, R., Fernig, D.G., Rudland, P.S., and Smith, J.A. (1990) Synthesis of basic fibroblast growth factor upon differentiation of rat mammary epithelial to myoepithelial-like cells in culture. *J Cell Physiol* **144**: 333-344.

Barres, B.A., Hart, I.K., Coles, H.S., Burne, J.F., Voyvodic, J.T., Richardson, W.D., and Raff, M.C. (1992) Cell death and control of cell survival in the oligodendrocyte lineage. *Cell* **70**: 31-46.

Baxevanis, A.D., and Ouellette, B.F.F. (2001) *Bioinformatics, 2nd ed.* New York: John Wiley & Sons, Inc.

Beletskii, A., Hong, Y.K., Pehrson, J., Egholm, M., and Strauss, W.M. (2001) PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proc Natl Acad Sci U S A* **98**: 9215-9220.

Bennett, W.R., Crew, T.E., Slack, J.M.W., and Ward, A. (2003) Structural-proliferative units and organ growth: effects of insulin-like growth factor 2 on the growth of colon and skin. *Development* **130**: 1079-1088.

Berdichevsky, F., Alford, D., D'Souza, B., and Taylor-Papadimitriou, J. (1994) Branching morphogenesis of human mammary epithelial cells in collagen gels. *J Cell Sci* **107 (Pt 12)**: 3557-3568.

Beutler, E., Collins, Z., and Irwin, L.E. (1967) Value of genetic variants of glucose-6-phosphate dehydrogenase in tracing the origin of malignant tumors. *N Engl J Med* **276**: 389-391.

Beutler, E., and Kuhl, W. (1990) Linkage between a *PvuII* restriction fragment length polymorphism and G6PD A-202A/376G: evidence for a single origin of the common G6PD A- mutation. *Hum Genet* **85**: 9-11.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., and Swift, S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **711**: 515-526.

Browett, P.J., Hoffbrand, A.V., and Norton, J.D. (1988) Use of the X-chromosome linked hypoxanthine phosphoribosyl transferase gene as a marker of cell monoclonality in hemopoietic malignancies. *Leuk Res* **12**: 321-326.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H. (1992) The human XIST gene: analysis of a 17kb inactive X-specific RNA that contains conserved repeats and is highly localised within the nucleus. *Cell* **71**: 527-542.

Busque, L., and Gilliland, D.G. (1996) The PGK-PCR clonality assay (PPCA). *Mol Biotechnol* **5**: 275-280.

Busque, L., Mio, R., Mattioli, J., Brais, E., Blais, N., Lalonde, Y., Maragh, M., and Gilliland, D.G. (1996) Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood* **88**: 59-65.

Buzin, C.H., Mann, J.R., and Singer-Sam, J. (1994) Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. *Development* **120**: 3529-3536.

Chepko, G., and Smith, G.H. (1997) Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tiss. Cell* **29**: 239-253.

Chepko, G., and Smith, G.H. (1998) Stem cells in mammary glands. Washington: www.mammary.nih.gov/reviews/development/chepko001/slides/main.htm

Christov, K., Swanson, S.M., Guzman, R.C., Thordarson, G., Jin, E., Talamantes, F., and Nandi, S. (1993) Kinetics of mammary epithelial cell proliferation in pituitary isografted BALB/c mice. *Carcinogenesis* **14**: 2019-2025.

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., and Duret, L. (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* **12**: 894-908.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996) XIST RNA paints the inactive X chromosome at interphase: evidence for a Novel RNA involved in nuclear/chromosome structure. *J Cell Biol.* **132**: 259-275.

Coles, H.S., Burne, J.F., and Raff, M.C. (1993) Large-scale normal cell death in the developing rat kidney and its reduction by epidermal growth factor. *Development* **118**: 777-784.

Collick, A., Bois, P., Grant, G., and Buard, J. (1998) Current and future contributions of transgenic mice to the analysis of germline toxicology. *Adv Exp Med Biol* **444**: 119-125; discussion 125-117.

Coucouvanis, E., and Martin, G.R. (1995) Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* **83**: 279-287.

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

Cunha, G.R. (1994) Role of mesenchymal-epithelial interactions in normal and abnormal development of the mammary gland and prostate. *Cancer* **74**: 1030-1044.

Cunha, G.R., and Hom, Y.K. (1996) Role of mesenchymal-epithelial interactions in mammary gland development. *J Mammary Gland Biol Neoplasia* **1**: 21-35.

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

Dandolo, L., Stewart, C.L., Mattei, M.G., and Avner, P.R. (1993) Inactivation of an X-linked transgene in murine extraembryonic and adult tissues. *Development* **118**: 641-649.

Di Croce, L., Vicent, G.P., Pecci, A., Bruscalupi, G., Trentalance, A., and Beato, M. (1999) The promoter of the rat 3-hydroxy-3-methylglutaryl coenzyme A reductase gene contains a tissue-specific estrogen-responsive region. *Mol Endocrinol* **13**: 1225-1236.

Diallo, R., Schaefer, K.L., Poremba, C., Shivazi, N., Willmann, V., Buerger, H., Dockhorn-Dworniczak, B., and Boecker, W. (2001) Monoclonality in normal epithelium and in hyperplastic and neoplastic lesions of the breast. *J Pathol* **193**: 27-32.

Dulbecco, R., Henahan, M., and Armstrong, B. (1982) Cell types and morphogenesis in the mammary gland. *Proc. Natl. Acad. Sci. USA* **79**: 7346-7350.

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999) Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum Mol Genet* **8**: 195-204.

Ellis, S., and Capuco, A.V. (2002) Cell proliferation in bovine mammary epithelium: identification of the primary proliferative cell population. *Tissue Cell* **34**: 155-163.

Ellis, S.E., Edwards, G., and Akers, R.M. (1995) Morphological and histological analysis of the prepuberal ovine mammary gland. *J. Dairy Sci.* **(Suppl. 1)**:: 157.

Ellis, S.E. (1998) Mechanisms controlling ductal morphogenesis in the ruminant mammary gland. In *Faculty of the Virginia Polytechnic Institute and State University* Blacksburg, VA: Virginia Polytechnic Institute and State University, pp. 101.

Engelhardt, J.F., Schlossberg, H., Yankaskas, J.R., and Dudus, L. (1995) Progenitor cells of the human airway involved in submucosal gland development. *Development* **121**: 2031-2046.

Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**: 759-767.

Ferguson, D.J., and Anderson, T.J. (1981) Morphological evaluation of cell turnover in relation to the menstrual cycle in the "resting" human breast. *Br J Cancer* **44**: 177-181.

Ferguson, D.J. (1985) Ultrastructural characterisation of the proliferative (stem?) cells within the parenchyma of the normal "resting" breast. *Virchows Arch A Pathol Anat Histopathol* **407**: 379-385.

Fey, M.F., Wells, R.A., Wainscoat, J.S., and Thein, S.L. (1988) Assessment of clonality in gastrointestinal cancer by DNA fingerprinting. *J Clin Invest* **82**: 1532-1537.

Fialkow, P.J. (1973) Primordial cell pool size and lineage relationships of five human cell types. *Ann Hum Genet* **37**: 39-48.

Fialkow, P.J. (1976) Clonal origin of human tumors. *Biochim Biophys Acta* **458**: 283-321.

Frank, T.S., Svoboda-Newman, S.M., and Hsi, E.D. (1996) Comparison of methods for extracting DNA from formalin-fixed paraffin sections for non-isotopic PCR. *Diagnostic Molecular Pathology* **5**: 220-224.

Furth, P.A., Hennighausen, L., Baker, C., Beatty, B., and Woychick, R. (1991) The variability in activity of the universally expressed human cytomegalovirus immediate early gene 1 enhancer/promoter in transgenic mice. *Nucleic Acids Res* **19**: 6205-6208.

Gale, R.E., Mein, C.A., and Linch, D.C. (1996) Quantification of X-chromosome inactivation patterns in haematological samples using the DNA PCR-based HUMARA assay. *Leukemia* **10**: 362-367.

Gale, R.E. (1999) Evaluation of clonality in myeloid stem-cell disorders. *Semin Hematol* **36**: 361-372.

Garcia, S.B., Novelli, M., and Wright, N.A. (2000) The clonal origin and clonal evolution of epithelial tumours. *Int. J. Exp. Path.* **81**: 89-116.

Gardner, R.L., and Lyon, M.F. (1971) X chromosome inactivation studied by injection of a single cell into the mouse blastocyst. *Nature* **231**: 385-386.

Gilbert, S.L., Pehrson, J.R., and Sharp, P.A. (2000) XIST RNA associates with specific regions of the inactive X chromatin. *J Biol Chem* **275**: 36491-36494.

Gilliland, D.G., Blanchard, K.L., Levy, J., Perrin, S., and Bunn, H.F. (1991) Clonality in myeloproliferative disorders: analysis by means of the polymerase chain reaction. *Proc Natl Acad Sci U S A* **88**: 6848-6852.

Going, J.J., Abd El-Monem, H.M., and Craft, J.A. (2001) Clonal origins of human breast cancer. *J Pathol* **194**: 405-411.

Going, J.J. (2003) Epithelial carcinogenesis: challenging monoclonality. *J Pathol* **200**: 1-3.

Goldman, M.A., Stokes, K.R., Idzerda, R.L., McKnight, G.S., Hammer, R.E., Brinster, R.L., and Gartler, S.M. (1987) A chicken transferrin gene in transgenic mice escapes X-chromosome inactivation. *Science* **236**: 593-595.

Guarini, G., and Onofri, E. (1993) [New horizons in medicine. Fractals]. *Recenti Prog Med* **84**: 438-442.

Gumbiner, B.M. (1992) Epithelial morphogenesis. *Cell* **69**: 385-387.

Hadjantonakis, A.K., Gertsenstein, M., Ikawa, M., Okabe, M., and Nagy, A. (1998) Non-invasive sexing of preimplantation stage mammalian embryos. *Nat Genet* **19**: 220-222.

Hadjantonakis, A.K., Cox, L.L., Tam, P.P., and Nagy, A. (2001) An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**: 133-140.

Haslam, S.Z. (1988) Local versus systemically mediated effects of estrogen on normal mammary epithelial cell deoxyribonucleic acid synthesis. *Endocrinology* **122**: 860-867.

Heim, S., Teixeira, M.A., and Pandis, N. (2001) Are some breast carcinomas polyclonal in origin? *J Pathol* **194**: 395-397.

Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993) Evolutionary conservation of possible functional domains of the human and murine XIST genes. *Hum Mol Genet.* **2**: 663-672.

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

Hermiston, M., and Gordon, J.L. (1995) Organization of the crypt-villus axis and evolution of its stem cell hierarchy during intestinal development. *Am J Physiol.* **268**: G813-822.

Hogg, N.A., Harrison, C.J., and Tickle, C. (1983) Lumen formation in the developing mouse mammary gland. *J Embryol Exp Morphol* **73**: 39-57.

Holmgren, L., Szeles, A., Rajnavolgyi, E., Folkman, J., Klein, G., Ernberg, I., and Falk, K.I. (1999) Horizontal transfer of DNA by the uptake of apoptotic bodies. *Blood* **93**: 3956-3963.

Hong, Y., Ontiveros, S.D., Chen, C., and Strauss, W.M. (1999) A new structure for the murine Xist gene and its relationship to chromosome choice/counting during X-chromosome inactivation. *Proc. Natl. Acad. Sci. USA* **96**: 6829-6834.

Hong, Y.K., Ontiveros, S.D., and Strauss, W.M. (2000) A revision of the human XIST gene organization and structural comparison with mouse Xist. *Mamm Genome* **11**: 220-224.

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

Hovey, R.C., McFadden, T.B., and Akers, R.M. (1999) Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. *J Mammary Gland Biol Neoplasia.* **4**: 53-68.

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

Joshi, K., Smith, J.A., Perusinghe, N., and Monoghan, P. (1986) Cell proliferation in the human mammary epithelium - differential contribution by the epithelial and myoepithelial cells. *Am. J. Pathol.* **124**: 199-206.

Jovanovic, L., Delahunt, B., McIver, B., Eberhardt, N.L., and Grebe, S.K. (2003) Thyroid gland clonality revisited: the embryonal patch size of the normal human thyroid gland is very large, suggesting X-chromosome inactivation tumor clonality studies of thyroid tumors have to be interpreted with caution. *J Clin Endocrinol Metab* **88**: 3284-3291.

Kenney, N.J., Smith, G.H., Lawrence, E., Barrett, J.C., and Salomon, D.S. (2001) Identification of stem cell units in the terminal end bud and duct of the mouse mammary gland. *J Biomed. Biotechology* **1**: 133-143.

Kisseberth, W.C., and Sandgren, E.P. (2004) Polyclonal development of mouse mammary preneoplastic nodules. *Cancer Res* **64**: 857-863.

Kleinburg, D.L. (1998) Role of IGF-I in normal mammary development. *Breast Cancer Res. Treat.* **47**: 201-211.

Kordon, E.C., and Smith, G.H. (1998) An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**: 1921-1930.

Kratzer, P.G., and Gartler, S.M. (1978) HGPRT activity changes in preimplantation mouse embryos. *Nature* **274**: 503-504.

Krumlauf, R., Chapman, V.M., Hammer, R.E., Brinster, R., and Tilghman, S.M. (1986) Differential expression of alpha-fetoprotein genes on the inactive X

chromosome in extraembryonic and somatic tissues of a transgenic mouse line. *Nature* **319**: 224-226.

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

Li, P., Barraclough, R., Fernig, D.G., Smith, J.A., and Rudland, P.S. (1998) Stem cells in breast epithelia. *Int. J. Exp. Path.* **79**: 193-206.

Lipschutz, J.H., Fukami, H., Yamamoto, M., Tatematsu, M., Sugimura, Y., Kusakabe, M., and Cunha, G. (1999) Clonality of urogenital organs as determined by analysis of chimeric mice. *Cells Tissues Organs* **165**: 57-66.

Lobe, C.G., Koop, K.E., Kreppner, W., Lomeli, H., Gertsenstein, M., and Nagy, A. (1999) Z/AP, a double reporter for cre-mediated recombination. *Dev Biol* **208**: 281-292.

Lochter, A. (1998) Plasticity of mammary epithelia during normal development and neoplastic progression. *Biochem. Cell. Biol.* **76**: 997-1008.

Luhovy, M., Liu, Y., Belickova, M., Prchal, J., and Prchal, J. (1995a) A novel clonality assay based on transcriptional polymorphism of X-chromosome gene p55. *Biology of blood and bone marrow transplantation* **1**: 81-87.

Luhovy, M., Liu, Y., Belickova, M., Prchal, J.F., and Prchal, J.T. (1995b) A novel clonality assay based on transcriptional polymorphism of X chromosome gene p55. *Biol Blood Marrow Transplant* **1**: 81-87.

Lyon, M.F. (1961) Gene action in the X-chromosome of the mouse. *Nature* **190**: 372-373.

Lyon, M.F. (1992) Some milestones in the history of X-chromosome inactivation. *Annu Rev Genet* **26**: 16-28.

Manni, A., Badger, B., Wei, L., Zaenglein, A., Grove, R., Khin, S., Heitjan, D., Shimasaki, S., and Ling, N. (1994) Hormonal regulation of insulin-like growth factor II and insulin-like growth factor binding protein expression by breast cancer cells in vivo: evidence for stromal epithelial interactions. *Cancer Res* **54**: 2934-2942.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* **11**: 156-166.

McMahon, A., Fosten, M., and Monk, M. (1983) X-chromosome inactivation mosaicism in three germ layers and the germ line of the mouse embryo. *J. Embryol. exp. Morph* **74**: 207-220.

Mead, R., Schmidt, G.H., and Ponder, B.A. (1987) Calculating numbers of tissue progenitor cells using chimaeric animals. *Dev Biol* **121**: 273-276.

Mehtali, M., LeMeur, M., and Lathe, R. (1990) The methylation-free status of a housekeeping transgene is lost at high copy number. *Gene* **91**: 179-184.

Migeon, B.R. (1998) Non-random X chromosome inactivation in mammalian cells. *Cytogenet Cell Genet.* **80**: 142-148.

Mighell, A.J., Smith, N.R., Robinson, P.A., and Markham, A.F. (2000) Vertebrate pseudogenes. *FEBS Lett* **468**: 109-114.

Mills, J.C., and Gordon, J.I. (2001) The intestinal stem cell niche: there grows the neighborhood. *Proc Natl Acad Sci U S A* **98**: 12334-12336.

Mintz, B., and Baker, W.W. (1967) Normal mammalian muscle differentiation and gene control of isocitrate dehydrogenase synthesis. *Proc Natl Acad Sci U S A* **58**: 592-598.

Molenaar, A.J., Davis, S.R., and Wilkins, R.J. (1992) Expression of alpha-lactalbumin, alpha-S1-casein, and lactoferrin genes is heterogeneous in sheep and cattle mammary tissue. *J Histochem Cytochem* **40**: 611-618.

Molenaar, A.J., Wilkins, R.J., and Davis, S.R. (1996) Measurement of cell death by in situ end-labeling of ruminant mammary gland tissue. *Proc. N. Zealand Anim. Prod. Soc.* **56**.

Nakanishi, T., Kuroiwa, A., Yamada, S., Isotani, A., Yamashita, A., Tairaka, A., Hayashi, T., Takagi, T., Ikawa, M., Matsuda, Y., and Okabe, M. (2002) FISH analysis of 142 EGFP transgene integration sites into the mouse genome. *Genomics* **80**: 564-574.

Nesbitt, M.N. (1971) X chromosome Inactivation Mosaicism in the Mouse. *Dev Biol* **26**: 252-263.

Nesbitt, M.N. (1974) Chimeras vs X inactivation mosaics: significance of differences in pigment distribution. *Dev Biol* **38**: 202-207.

Ng, I.O., Guan, X.Y., Poon, R.T., Fan, S.T., and Lee, J.M. (2003) Determination of the molecular relationship between multiple tumour nodules in hepatocellular carcinoma differentiates multicentric origin from intrahepatic metastasis. *J Pathol* **199**: 345-353.

Nickerson, S.C., Akers, R.M., and Weinland, B.T. (1982) Cytoplasmic organization and quantitation of microtubules in bovine mammary epithelial cells during lactation and involution. *Cell Tissue Res* **223**: 421-430.

Nilsson, M., Barbany, G., Antson, D.O., Gertow, K., and Landegren, U. (2000) Enhanced detection and distinction of RNA by enzymatic probe ligation. *Nat Biotechnol* **18**: 791-793.

Noguchi, S., Motomura, K., Inaji, H., Imaoka, S., and Koyama, H. (1992) Clonal analysis of human breast cancer by means of the polymerase chain reaction. *Cancer Res* **52**: 6594-6597.

Noguchi, S., Aihara, T., Koyama, H., Motomura, K., Inaji, H., and Imaoka, S. (1994a) Discrimination between multicentric and multifocal carcinomas of the breast through clonal analysis. *Cancer* **74**: 872-877.

Noguchi, S., Motomura, K., Inaji, H., Imaoka, S., and Koyama, H. (1994b) Clonal analysis of predominantly intraductal carcinoma and precancerous lesions of the breast by means of polymerase chain reaction. *Cancer Res* **54**: 1849-1853.

Noguchi, S., Aihara, T., Koyama, H., Motomura, K., Inaji, H., and Imaoka, S. (1995) Clonal analysis of benign and malignant human breast tumors by means of polymerase chain reaction. *Cancer Lett* **90**: 57-63.

Novelli, M., Cossu, A., Oukrif, D., Ouaglia, A., Lakhani, S., Poulsom, D., Sasien, P., Carta, P., Contini, M., Pasca, A., Palmier, W., Tanda, F., and Wright, N.A. (2003) X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. *Proc Natl Acad Sci U S A* **10.1073**: online www.pnas.org/cgi/doi/10.1073/pnas.0437825100.

Oakburg, E.F., and Huckins, C. (1976) *Stem cells of renewing populations*: NY Academic Press.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A* **86**: 2766-2770.

Ormerod, E.J., and Rudland, P.S. (1984) Cellular composition and organization of ductal buds in developing rat mammary glands: evidence for

morphological intermediates between epithelial and myoepithelial cells. *Am J Anat* **170**: 631-652.

Pandis, N., Jin, Y., Gorunova, L., Petersson, C., Bardi, G., Idvall, I., Johansson, B., Ingvar, C., Mandahl, N., Mitelman, F., and et al. (1995) Chromosome analysis of 97 primary breast carcinomas: identification of eight karyotypic subgroups. *Genes Chromosomes Cancer* **12**: 173-185.

Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M.J., and Petersen, O.W. (1999) Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. *Dev Biol* **206**: 88-99.

Peng, H., S, R.L., Lee, C., Zheng, Q., Chaggar, R.K., Wright, N.A., Pan, L., and Isaacson, P.G. (2000) Clonality analysis of defined cell populations in paraffin-embedded tissue sections by RT-PCR amplification of X-linked G6PD gene. *J Pathol* **191**: 313-317.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996) Requirement for Xist in X chromosome inactivation. *Nature* **379**: 131-137.

Philp, J.A., Burdon, T.G., and Watson, C.J. (1996) Differential activation of STATs 3 and 5 during mammary gland development. *FEBS Lett* **396**: 77-80.

Potten, C.S. (1986) Cell cycles in cell hierarchies. *Int J Radiat Biol Relat Stud Phys Chem Med* **49**: 257-278.

Potten, C.S., and Morris, R.J. (1988) Epithelial stem cells *in vivo*. *J. Cell Sci. Suppl.* **10**: 45-62.

Potten, C.S., and Loeffler, M. (1990) Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development* **110**: 1001-1020.

Potten, C.S., Booth, C., and Pritchard, D.M. (1997) The intestinal stem cell: the mucosal governor. *Int. J. Exp. Path.* **78**: 219-234.

Povey, R.C., and Osborne, A.D. (1969) Mammary gland neoplasia in the cow. A review of the literature and report of a fibrosarcoma. *Pathol Vet* **6**: 502-512.

Propper, A.Y. (1978) Wandering epithelial cells in the rabbit embryo milk line. A preliminary scanning electron microscope study. *Dev Biol* **67**: 225-231.

Quarrie, L.H., Addey, C.V., and Wilde, C.J. (1995) Apoptosis in lactating and involuting mouse mammary tissue demonstrated by nick-end DNA labelling. *Cell Tissue Res* **281**: 413-419.

Richards, A.H., and Hilf, R. (1972) Influence of pregnancy, lactation and involution on glucose-6-phosphate dehydrogenase and lactate dehydrogenase in the rat mammary gland. *Endocrinology* **91**: 287-295.

Rudland, P.S. (1987) Stem cells and the development of mammary cancers in experimental rats and in humans. *Cancer Met. Reveiws* **6**: 55-83.

Rudland, P.S. (1991) Histochemical organization and cellular composition of ductal buds in developing human breast: evidence of cytochemical intermediates between epithelial and myoepithelial cells. *J Histochem Cytochem* **39**: 1471-1484.

Rudland, P.S. (1993) Epithelial stem cells and their possible role in the development of the normal and diseased human breast. *Histol Histopath* **8**: 385-404.

Rupert, J.L., Brown, C.J., and Willard, H.F. (1995) Direct detection of non-random X chromosome inactivation by use of a transcribed polymorphism in the XIST gene. *Eur J Hum Genet* **3**: 333-343.

Russo, I.H., and Russo, J. (1978) Developmental stage of the rat mammary gland as determinant of its susceptibility to 7,12-dimethylbenz[a]anthracene. *J Natl Cancer Inst* **61**: 1439-1449.

Russo, J., and Russo, I.H. (1987) Biological and molecular basis of mammary carcinogenesis. *Lab Invest* **57**: 112-137.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., and Altshuler, D. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928-933.

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

Sakakura, T., Sakagami, Y., and Nishizuka, Y. (1982) Dual origin of mesenchymal tissues participating in mouse mammary gland embryogenesis. *Dev Biol* **91**: 202-207.

Sakakura, T., Kusano, I., Kusakabe, M., Inaguma, Y., and Nishizuka, Y. (1987) Biology of mammary fat pad in fetal mouse: capacity to support development of various fetal epithelia *in vivo*. *Development* **100**: 421-430.

Schmidt, E.V., Christoph, G., Zeller, R., and Leder, P. (1990) The cytomegalovirus enhancer: a pan-active control element in transgenic mice. *Mol Cell Biol* **10**: 4406-4411.

Schmidt, G.H., Garbutt, D.J., Wilkinson, M.M., and Ponder, B.A. (1985) Clonal analysis of intestinal crypt populations in mouse aggregation chimaeras. *J Embryol Exp Morphol* **85**: 121-130.

Schmidt, G.H., Winton, D.J., and Ponder, B.A. (1988) Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine. *Development* **103**: 785-790.

Schmidt, G.H., and Mead, R. (1990) On the clonal origin of tumours - lessons from studies of intestinal epithelium. *Bioessays* **12**: 37-40.

Schofield, R. (1983) The stem cell system. *Biomed. Pharm.* **37**: 375-380.

Schorpp, M., Jager, R., Schellander, K., Schenkel, J., Wagner, E.F., Weiher, H., and Angel, P. (1996) The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice. *Nucleic Acids Res* **24**: 1787-1788.

Sell, S., and Pierce, G.B. (1994) Maturation arrest of stem cell differentiation is a common pathway for the cellular origin of teratocarcinomas and epithelial cancers. *Lab Invest* **70**: 6-22.

Seymour, P.A., Bennett, W.R., and Slack, J.M.W. (2004) Fission of pancreatic islets during postnatal growth of the mouse. *J. Anat.* **204**: 103–116.

Shibata, A., Tsai, Y.C., Press, M.F., Henderson, B.E., Jones, P.A., and Ross, R.K. (1996) Clonal analysis of bilateral breast cancer. *Clin Cancer Res* **2**: 743-748.

Shimohama, S., Rosenberg, M.B., Fagan, A.M., Wolff, J.A., Short, M.P., Breakefield, X.O., Friedmann, T., and Gage, F.H. (1989) Grafting

genetically modified cells into the rat brain: characteristics of *E. coli* beta-galactosidase as a reporter gene. *Brain Res Mol Brain Res* **5**: 271-278.

Silberstein, G.B., and Daniel, C.W. (1982) Glycosaminoglycans in the basal lamina and extracellular matrix of the developing mouse mammary duct. *Dev Biol* **90**: 215-222.

Silberstein, G.B., and Daniel, C.W. (1987) Investigation of mouse mammary ductal growth regulation using slow-release plastic implants. *J Dairy Sci* **70**: 1981-1990.

Simian, M., Hirai, Y., Navre, M., Werb, Z., Lochter, A., and Bissell, M.J. (2001) The interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithelial cells. *Development* **128**: 3117-3131.

Sinha, Y.N., and Tucker, H.A. (1966) Mammary gland growth of rats between 10 and 100 days of age. *Am J Physiol* **210**: 601-605.

Sinha, Y.N., and Tucker, H.A. (1969) Mammary development and pituitary prolactin level of heifers from birth through puberty and during the estrous cycle. *J Dairy Sci* **52**: 507-512.

Skalnik, D.G., Narita, H., Kent, C., and Simoni, R.D. (1988) The membrane domain of 3-hydroxy-3-methylglutaryl-coenzyme A reductase confers endoplasmic reticulum localization and sterol-regulated degradation onto beta-galactosidase. *J Biol Chem* **263**: 6836-6841.

Smith, G. (1996) Experimental mammary epithelial morphogenesis in an *in vivo* model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res. Treat.* **39**: 21-31.

Smith, G.H., and Medina, D. (1988a) A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci* **89**: 173-183.

Smith, G.H., and Medina, D. (1988b) A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci* **90 (Pt 1)**: 173-183.

Smith, J.R., Osborne, T.F., Goldstein, J.L., and Brown, M.S. (1990) Identification of nucleotides responsible for enhancer activity of sterol regulatory element in low density lipoprotein receptor gene. *J Biol Chem* **265**: 2306-2310.

Smith, J.W., Townsend, D.E., and Sparkes, R.S. (1971) Glucose-6-phosphate dehydrogenase polymorphism: a valuable tool to study tumor origin. *Clin Genet* **2**: 160-162.

Srinivasan, K., Strickland, P., Valdes, A., Shin, G.C., and Hinck, L. (2003) Netrin-1/Neogenin interaction stabilizes multipotent progenitor cap cells during mammary gland morphogenesis. *Dev Cell* **4**: 371-382.

Stingl, J., Eaves, C.J., Zandieh, I., and Emerman, J.T. (2001) Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**: 93-109.

Stone, M. (1983) A general statistical model for clone/tissue studies using X-chromosome inactivation data. *Biometrics* **39**.

Swett, W.W., Matthews, C.A., and Graves, R.R. (1940) Extreme rarity of cancer in the cow's udder: a negative finding of vital interest to the dairy industry and to the consumer. . *J. Dairy Sci.* **23**: 437–446.

Takagi, N., and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**: 640-642.

Takagi, N. (2001) The role of X-chromosome inactivation in the manifestation of Rett syndrome. *Brain Dev* **23 Suppl 1**: S182-185.

Tam, P.P., and Tan, S. (1992) The somitogenic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. *Development* **115**: 703-715.

Tam, P.P., Williams, E.A., and Tan, S.S. (1994) Expression of an X-linked HMG-lacZ transgene in mouse embryos: implication of chromosomal imprinting and lineage-specific X-chromosome activity. *Dev Genet* **15**: 491-503.

Tam, S.P., Brissette, L., Ramharack, R., and Deeley, R.G. (1991) Differences between the regulation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase and low density lipoprotein receptor in human hepatoma cells and fibroblasts reside primarily at the translational and post-translational levels. *J Biol Chem* **266**: 16764-16773.

Tan, S.S., Williams, E.A., and Tam, P.P. (1993) X-chromosome inactivation occurs at different times in different tissues of the post-implantation mouse embryo. *Nat Genet* **3**: 170-174.

Teixeira, M.R., Pandis, N., Bardi, G., Andersen, J.A., Mandahl, N., Mitelman, F., and Heim, S. (1994) Cytogenetic analysis of multifocal breast carcinomas: detection of karyotypically unrelated clones as well as clonal similarities between tumour foci. *Br J Cancer* **70**: 922-927.

Thomas, G.A., Williams, D., and Williams, E.D. (1988) The demonstration of tissue clonality by X-linked enzyme histochemistry. *J Pathol* **155**: 101-108.

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

Tsai, Y.C., Lu, Y., Nichols, P.W., Zlotnikov, G., Jones, P.A., and Smith, H.S. (1996) Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. *Cancer Res* **56**: 402-404.

Tucker, H.A. (1987) Quantitative estimates of mammary growth during various physiological states: a review. *J Dairy Sci* **70**: 1958-1966.

Uejima, H., Lee, M.P., Cui, H., and Feinberg, A.P. (2000) Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. *Nat Genet* **25**: 375-376.

Vanin, E.F. (1985) Processed pseudogenes: characteristics and evolution. *Annu Rev Genet* **19**: 253-272.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., and Feinberg, A.P. (1985) Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science* **227**: 642-645.

Wagner, K.U., Boulanger, C.A., Henry, M.D., Sgagias, M., Hennighausen, L., and Smith, G.H. (2002) An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* **129**: 1377-1386.

Wainscoat, J.S., and Fey, M.F. (1990) Assessment of clonality in human tumors: a review. *Cancer Res* **50**: 1355-1360.

Weaver, R.F., and Hedrick, P.W. (1997) *Genetics*: WC Brown Publisher.

Weiss, D.J., Liggitt, D., and Clark, J.G. (1997) *In situ* histochemical detection of beta-galactosidase activity in lung: assessment of X-Gal reagent in distinguishing lacZ gene expression and endogenous beta-galactosidase activity. *Hum Gene Ther* **8**: 1545-1554.

West, J.D. (1976) Patches in the livers of chimaeric mice. *J Embryol Exp Morphol* **36**: 151-161.

Williams, J.M., and Daniel, C.W. (1983) Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol* **97**: 274-290.

Wilson, P.J., Meaney, C.A., Hopwood, J.J., and Morris, C.P. (1993) Sequence of the human iduronate 2-sulfatase (IDS) gene. *Genomics* **17**: 773-775.

Wolpert, L., and Gingell, D. (1970) Striping and the pattern of melanocyte cells in chimeric mice. *J. Theor. Biol.* **29**: 147-150.

Woodward, T.L., Dumont, N., O'Connor-McCourt, M., Turner, J.D., and Philip, A. (1995) Characterization of transforming growth factor-beta growth regulatory effects and receptors on bovine mammary cells. *J Cell Physiol* **165**: 339-348.

Wright, N.A., and Alison, M. (1984) *The biology of epithelial cell populations.* Oxford: Clarendon press.

Wu, H., Fassler, R., Schnieke, A., Barker, D., Lee, K.H., Chapman, V., Francke, U., and Jaenisch, R. (1992) An X-linked human collagen transgene escapes X inactivation in a subset of cells. *Development* **116**: 687-695.

Zajicek, G. (1995) Streaming organism. *Med Hypotheses* **45**: 403-407.

Zeps, N., Dawkins, H.J., Papadimitriou, J.M., Redmond, S.L., and Walters, M.I. (1996) Detection of a population of long-lived cells in mammary epithelium of the mouse. *Cell Tissue Res* **286**: 525-536.

Zeps, N., Bentel, J.M., Papadimitriou, J.M., D'Antuono, M.F., and Dawkins, H.J. (1998) Estrogen receptor-negative epithelial cells in mouse mammary gland development and growth. *Differentiation* **62**: 221-226.

Appendices

Appendix 1. Electropherogram showing relative position of SNP2 in the bovine *Xist* gene

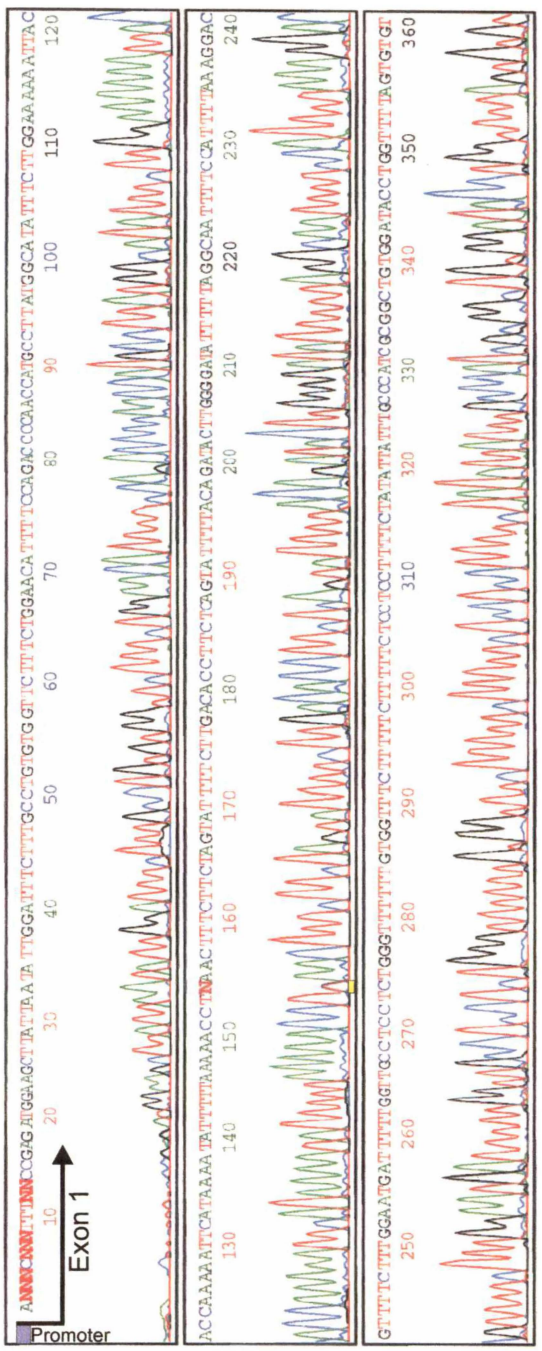


Figure A.1. Electropherogram showing the position of the identified polymorphism (SNP2) relative to *Xist* promoter in a heterozygous animal. SNP2 is indicated by a grey box highlighted by a yellow bar. The promoter begins upstream of the indicated sequence and terminates close to the beginning of the sequence in the electropherogram.

Appendix 2: DNA Sequences obtained during study

A.2.1 AF531754 pG6PD: Bos taurus glucose-6-phosphate 1-dehydrogenase (G6PD) pseudogene ('exons' 9-13)

The entire sequenced pG6PD clone insert was submitted for intron/exon boundary prediction by Genemachine (http://genemachine.nhgri.nih.gov/).

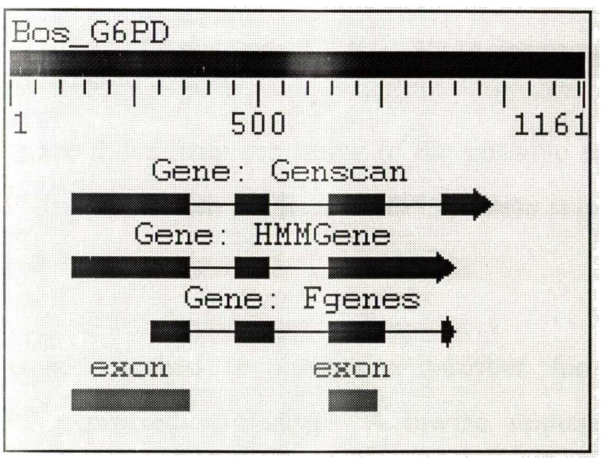


Figure A.2. 'Genemachine' exon/intron prediction for pG6PD.B analysed on Sequin sequence submission software. The exons predicted by GENSCANW approximate those predicted by the other software algorithms, such as MZEF, used. Note the predicted repeat regions predicted by 'REPEAT_masker'.

GENSCAN

1 130-365 3rd phase read, 5′ GG insertion
VPFILRCGKALNERKAEVRLQFRDVAGDIFRQQCKRNELVIRVQPNEAVYTKMMTKKPGMFFNP
EESELDLTYGNRYK
2 457-524 1st phase read, 3′ TC insertion
NVKLPDAYERLILDVFAGARCT
3 650-759 3rd phase read, 5′ GT, 3′ AG insertion
DELREAWRIFTPLLHHIEREKARPIPYVYGRYVYAE
4 876-980 3rd phase read, 5′ CC, 3′ T insertion
GACGGRRAVKRVGFQYRGHLQVGEPPQALSRPPS

HMMGene

1 130-365 As GENSCAN 1
2 457-524 As GENSCAN 2
3 650-905 3rd phase read, 5′ GT, 3′ GA insertions
MNCGRPGGSSLHCSTTSNARRPGPSPTCTAGTSMQKVRRTRRGVDRGVTGVEGGRRGTLWLRAP
SPLPTACPPASRGLWRQTSC

F-Genes

1 288-366 1st phase read
MMTKKPGMFFNPEESELDLTYGNRYK
2 457-532 1st phase read
NVKLPDAYERLILDVFAGARCTSCA
3 650-759 2nd phase read
DELREAWRIFTPLLHHIEREKARPIPYVYGRYVYAE
4 876-905 2nd phase read, 3′ A insertion
RGLWRQTSCE

MZEF exons

1 130-365 As above
2 650-742 2nd phase read, , 5′ T, 3′ GA insertions
DELREAWRIFTPLLHHIEREKARPIPYVYG

The peptides shown are those from the frame of the putative exons that showed similarity by simple alignment with published G6PD protein sequences.

ESTs

Published ESTs were searched to determine whether the clone sequence corresponded to any expressed sequences. A bovine sequence (gi:17887640) showed strong similarity to regions of the clone sequence as shown in the clustalX alignment below (Figure A.3).

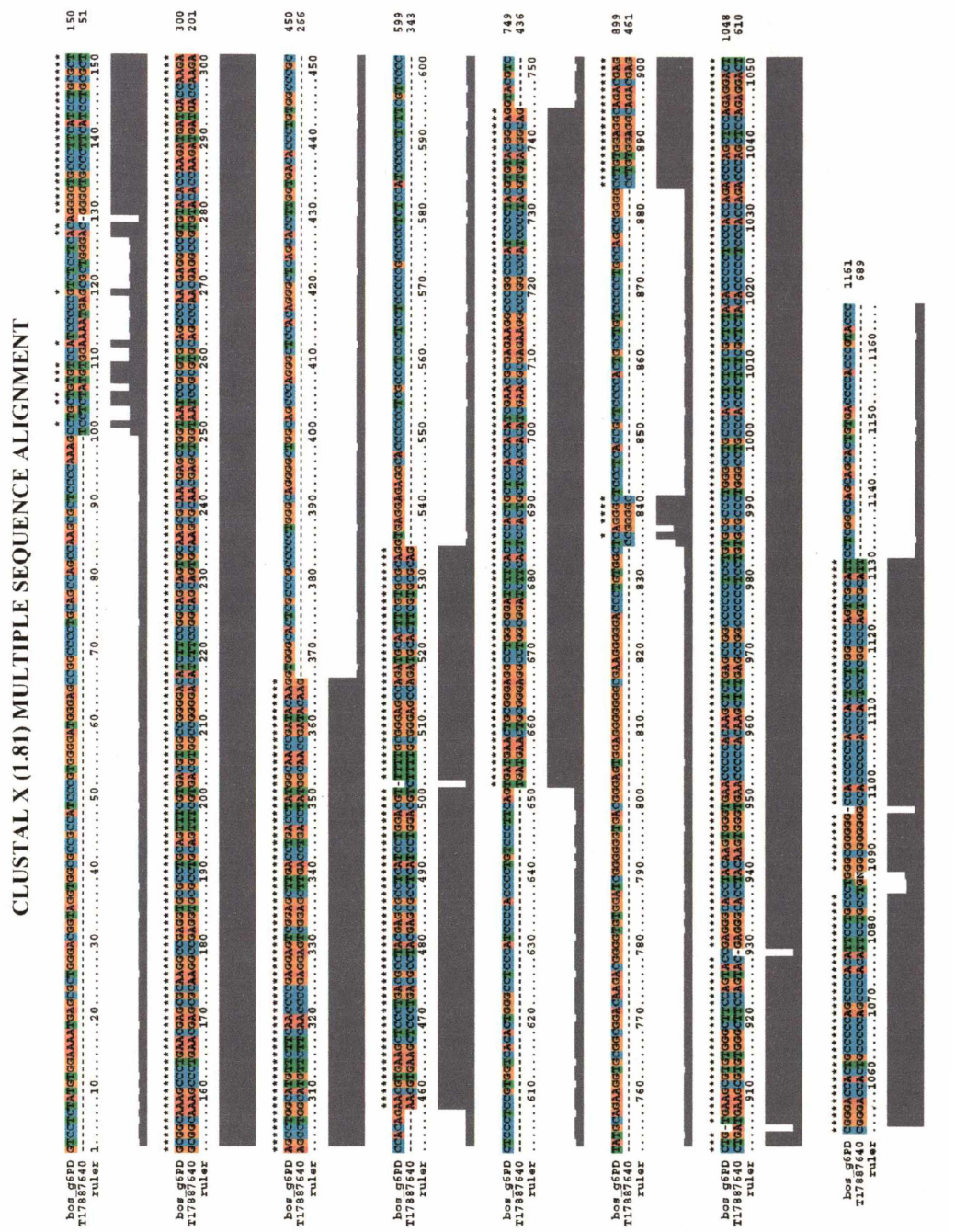


Figure A.3. A clustalX sequence alignment showing pG6PD clone and the bovine EST gi:17887640. (from a Normalized bovine leucocyte EST library project). Significant similarity is found in regions corresponding roughly to the putative coding regions predicted by the 'Genemachine' suite of exon/intron prediction algorithms. This alignment was considered in a description of the clone sequence.

The EST data supported the presence of frame-shifted exons between positions:

1. 99-365 (ignoring a less favourable start at 130) to encode:

PAVSIPRLLTGVPFILRCGKALNERKAEVRLQFRDVAGDIFRQQCKRNELVIRVQPNEAVYTKMMTKKPG MFFNPEESELDLTYGNRYK

2. 366-532 (ignore 3′ G insertion) to encode:

NVKLPDAYERLILDVFAGARCTSCA

3. 649-742 to encode:

DELREAWRIFTPLLHHIEREKARPIPYVYGRY

4. 875-1143 to encode (ignore 5′ CC and 3′ CT that result in short 'ORF'):
AGACGGRRAVKRVGFQYRGHLQVGEPPQALSRPPSCAPWACPPLSLYTPSHQTQLQRTRDHCPQ
PHIPALGGGHPPPTPRPSRIPRPAAL

The coding regions predicted using EST similarity as a guideline (reading real frame portions of the data only) were assembled into putative proteins and submitted to BLASTP searching.

gi|14579295|gb|AAK69185.1|AF326207_1 (AF326207) glucose-6-phosphate dehydrogenase [Mus musculus]
Length = 515

Score = 248 bits (632), Expect = 6e-65
Identities = 122/155 (78%), Positives = 127/155 (81%), Gaps = 1/155 (0%)

Query: 11 GVPFILRCGKALNERKAEVRLQFRDVAGDIFRQQCKRNELVIRVQPNEAVYTKMMTKKPG 70
GVPFILRCGKALNERKAEVRLQFRDVAGDIF QQCKRNELVIRVQPNEAVYTKMMTKKPG
Sbjct: 351 GVPFILRCGKALNERKAEVRLQFRDVAGDIFHQQCKRNELVIRVQPNEAVYTKMMTKKPG 410

Query: 71 MFFNPEESELDLTYGNRYKNVKLPDAYERLILDVFAGARCTSC-ADELREAWRIFTPLLH 129
MFFNPEESELDLTYGNRYKNVKLPDAYERLILDVF G++ +DELREAWRIFTPLLH
Sbjct: 411 MFFNPEESELDLTYGNRYKNVKLPDAYERLILDVFCGSQMHFVRSDELREAWRIFTPLLH 470

Query: 130 HIEREKARPIPYXXXXXXXXXXXXXXXXXXFQYRG 164
I+REK +PIPY FQY G
Sbjct: 471 KIDREKPQPIPYVYGSRGPTEADELMKRVGFQYEG 505

The search revealed that the regions corresponding to EST regions 1-3 gave strong similarity to published G6PD sequences (*Mus musculus* in the above example). However the region beyond these, corresponding to EST4 above, had no significant similarity to any published sequence. A BLASTX comparison of the EST itself shows strong similarity to published G6PD protein sequences only for the first 524 of its 684 bp sequence. This may indicate, as is common for EST data, that sequence quality is poor. Usually this is because the sequence is obtained with only a 1X sequencing coverage, and average 3% error, c.f. human genome project that boasts an error rate of only 0.01% (Baxevanis and Ouellette, 2001). Curiously, sequence alignment of the EST with the pG6PD.B clone gives reasonable similarity in the same region. This may be due to the coding bovine G6PD protein having a quite unique C-terminal sequence. Another possibility is that the EST is itself the expressed (and untranslated) product of a pseudogene (supported by the fact that the regions of the EST corresponding to exons of G6PD are themselves not represented as sequences of lengths that are multiples of three – as one would expect to find for coding exons).

BLASTX

If frameshift aberrations in the predicted exons are ignored, a protein with strong similarity to all but the 3'-most of the exons predicted by GENSCAN and F-genes (see above) gives significant expect values with a number of published G6PD protein sequences.

Repeatmasker

A single repeat element was identified in the clone sequence using an online repeat searching engine (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). The repeat was:

Simple hexanucleotide repeat: (CCCCCT)n, positions 547-604 forward strand

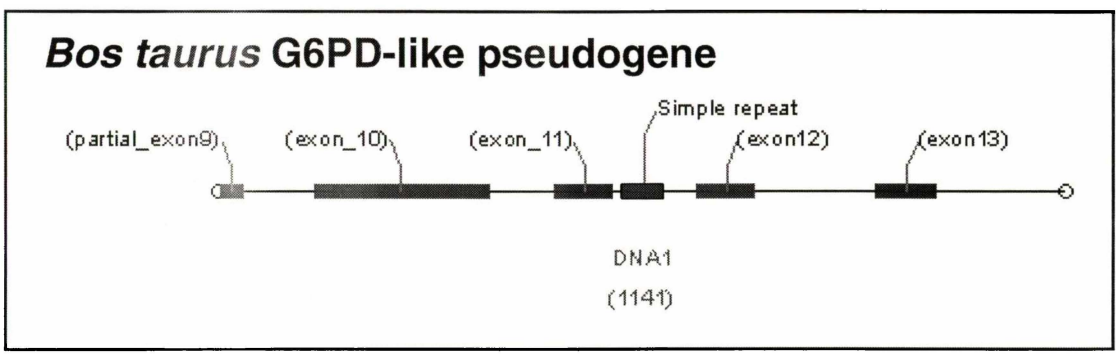


Figure A.4. Showing regions of similarity to published *G6PD* sequences (at the gene, EST and protein level) for the 1141 bp submitted sequence. The brackets around the 'exons' indicate that there is lack of evidence for first frame coding for the sequence indicated. This fact alone seems sufficient to allow a description of the sequence as a pseudogene (as detailed below).

Summary of *in silico* data for pG6PD.B description

Exon/intron boundary prediction software allowed the description of a number of putative exons (Figure A.2). Unfortunately most of these exons (9 from a total of 13) were not first frame reading exons. Furthermore, the single most relevant EST found (from a bovine EST library) did give not evidence for in-frame coding from the clone sequence.

In the absence of experimental confirmatory data (or of a high accuracy sequence) it is assumed that the clone pG6PD.B displays elements of an unprocessed (duplication-type) pseudogene that exhibits strong similarity to published glucose-6-phosphate dehydrogenase sequences at the gene and protein levels (Figure A.4).

The sequence was submitted to Genbank where it has the reference gi:AF531754

```
LOCUS       AF531754          1161 bp    DNA     linear   MAM 16-AUG-2002
DEFINITION  Bos taurus glucose-6-phosphate 1-dehydrogenase pseudogene, partial
            sequence.
ACCESSION   AF531754
VERSION     AF531754
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Jacobson,G.M. and Wilkins,R.J.
  TITLE     Bovine G6PD fragment from Stratagene genomic library #945701
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1161)
  AUTHORS   Jacobson,G.M. and Wilkins,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2002) Biology, The University of Waikato,
            Hillcrest Rd, Hamilton 2001, New Zealand
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     gene            <1..>1161
                     /gene="glucose-6-phosphate 1-dehydrogenase"
                     /pseudo
     repeat_region   547..604
                     /note="simple repeat"
                     /rpt_unit=CCCCCT
BASE COUNT     194 a   441 c   339 g   187 t
ORIGIN
        1 gtcctctatg tggaaaatga gcgctgggac ggtaggtggc gccgccatcc cgtggggatg
       61 ggagccggcc cctgcagcca gccaagcgct ccccaaagcc tgctgtgtcc atcccccgtc
      121 tcctcacagg ggtgcccttc atcctgcgct gcggcaaagc cctgaacgag cgcaaggccg
      181 aggtgcgcct gcagtttcgt gacgtggccg gggacatctt ccggcagcag tgcaagcgca
      241 acgagctggt aatccgcgtg cagcccaacg aggccgtgta caccaagatg atgaccaaga
      301 agcctggcat gttcttcaac cccgaggagt cggagcttga cctgacctat ggcaaccgat
      361 acaaggtggg cactcgcccg ccccctgggc aggggctggc agcccagggc tccacagggc
      421 tcagcacctt ggtgacaccc tgtggcccgc ccacagaacg tgaagctccc tgacgcctac
      481 gagcgcctca tcctggacgt ttttgcggga gccagatgca cttcgtgcgc aggtgaggag
      541 aggcaccccc ctcgccctcc ctcctccccc gcccccctct catccccctc ttcgtccccc
      601 tccctccgtg gtcacactgg gcctcccatc cccacccctg tcccttcagt gatgaactgc
      661 gggaggcctg gcggatcttc actccactgc tccaccacat cgaacgcgag aaggcccggc
      721 ccatccccta cgtgtacggc aggtacgtct atgcagaagg tgcggcggac aagacggggt
      781 gtggatcggg gggtgacggg agtggagggg gggcgaaggg ggaccctgtg gctcagggct
      841 ccctcaccgc tccccactgc ctgtcccccc gccagccggg gcctgtggag gcagacgagc
      901 tgtgaagcgt gtgggcttcc agtaccgagg gcacctacaa gtgggtgaac ccccacaagc
      961 tctgagccgg ccccctcct gtgcgccctg ggcctgccca cctctctcgc tctacacccc
     1021 ctcccaccag acccagctcc agaggactcg ggaccactgc ccccagcccc acattcctgc
     1081 cctgggcggg ggccacccca cacccactcc tcggcccagt cgcattcctc ggccagcagc
     1141 actgtgaccc cacccgtacc c
```

A.2.2 pBHAP: *Bos taurus* Huntingtin-Associated Protein-1 (HAP1) like pseudogene

A suite of exon/intron boundary prediction software, including GENSCANW (as above) HMM, F-genes and the exon likelihood assessment predictor algorithm MZEF, was applied to the sequence by submission to Genemachine (http://genemachine.nhgri.nih.gov/) Figure A.5.

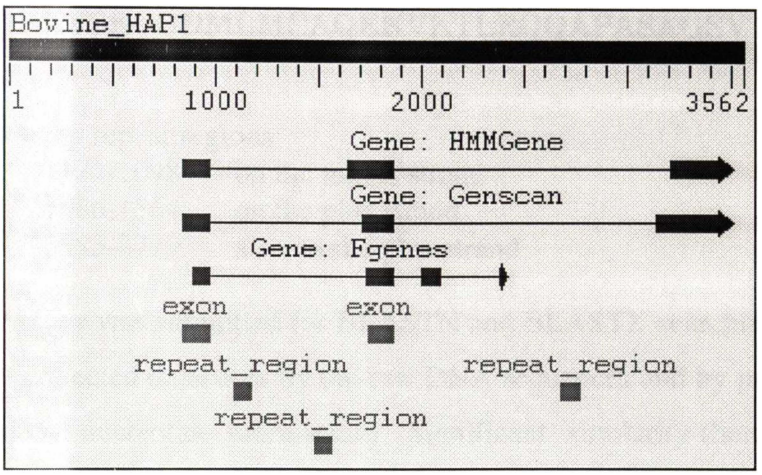


Figure A.5. 'Genemachine' exon/intron prediction analysed on Sequin sequence submission software. The exons predicted by GENSCANW approximate those predicted by the other software algorithms, such as MZEF, used. Note the repeat regions predicted by 'REPEAT_masker'.

The above prediction corresponds to potential exons occurring at:

GENSCAN:

1 849-979 Frameshift (5'C, 3'G)
2 1715-1870 Codes for 52 aa peptide SLWAGSQLQDLQEKYSECGDMLMEAQEKVKTLRQQAPA SAGSVTHYTYTVPL
3 3145-3514 Frameshift

HMMGene:

1 849-979 Frameshift (5'C, 3'G)
2 1715-1870 Codes for a 74 aa peptide ASRPLWERVGKASRLGVPGSHQSLWAGSQLQDLQEKYSE CGDMLMEAQEKVKTLRQQAPASAGSVTHYTYTVPL
3 3145-3514 Frameshift

Fgenes:

1 899-979 Codes for a 27 aa peptide MENYEQQQREVAQLQGQVTKLQRCCRS
2 1736-1870 Codes for a 45 aa peptide LQDLQEKYSECGDMLMEAQEKVKTLRQQAPASAGSVTH YTYTVPL
3 2003-2094 Frameshift
4 2386-2424 Frameshift

MZEF

1 849-979 Frameshift
2 1742-1870 Codes for a 43 aa peptide

DLQEKYSECGDMLMEAQEKVKTLRQQAPASAGSVTHYTYTV
PL

And the following repeat regions:

SINE/MIR	1172-1098	on the minus strand
LINE/L2	1486-1564	on the plus strand
SINE/MIR	2685-2779	also on the plus strand

The pBHAP clone was submitted for BLASTN and BLASTX searching (that is, a query of 'nr' collected databases by the raw DNA sequences and by protein query from a translated nucleotide submission). Significant similarity (here defined as expect values of less than 1) was required of the search. The submitted sequence gave significant hits against a number of published human and rodent Huntingtin Associated Protein (HAP1) sequences. These sequences were considered with special attention to published human HAP1 gene and protein (AJ012824 and P54256 respectively).

Rationale for annotation of the pBHAP clone

Weigh BLASTN (nr and EST) data

The first predicted exon (849-979) was supported by genscan, hmm, mzpz and partially by Fgenes predictions, and corresponds to exon 7 of the HAP gene. This region, while showing strong similarity to HAP1 exon 7 when read in the 3rd phase (that is, starting from position 3) was only 131 bp in length and seems to be frame-shifted (by the insertion of a 'C' at the 5′ and a 'G' at the 3′ end of the exon).

BLASTN shows some similarity between pBHAP position 1355-1453 and exon 8 (position 9325 to 9399) of the human HAP1 gene (AJ012824). This similarity was supported by back-translation of the exon 8 portion of the human HAP1 protein, which produced the DNA sequence:

ATGTACGGAGCYGARACYGARAAGYTBCARAAGCARYTBGCYTCY
GARAAGGARATYCARATGCARYTBCARGARGAR

This sequence was shown to be 74% similar to 1354-1432 bp of the submitted bovine sequence. Although corresponding to the approximate position of human exon 8, the pBHAP sequence has a TTA stop codon at the putative centre of the exon (whereas the human protein has a TTC codon for serine) and lacks a splice

donor (AG) site immediately 5′ to the putative exon. It is suspected that if the sequence has coding potential then it is likely that exon 8 is not transcribed. Moreover, it should be noted that this region was not predicted by exon boundary identification software.

Another putative coding exon was predicted by HMM (1649-1870) Fgenes (1736-1870) and Genscan (1715-1870). BLASTX similarity searching supported a protein coding region between 1715 and 1870 (by similarity with the protein from AJ012824). This region, the only exon with definite coding potential was very similar to exon 9 of the human sequence.

A region with similarity to human exon 10 was predicted by the Fgenes algorithm. This region was shown to correspond almost perfectly to exon 10 of the human HAP gene AJ012824. Unfortunately, as is characteristic of ancestral 'exons' within pseudogenes, this region was shown to be frame-shifted (with a total length of 92 bp) and therefore unlikely to be coding.

An 'exon 11-like' region, a 39 bp 'exon' residing between positions 2386 and 2424 of the submitted sequence, was predicted using F-genes but none of the other algorithms. This region, although at the gene level, exhibiting an adjacent consensus slice acceptor site (GT), would encode a protein that was truncated relative to published HAP1 protein sequences.

Finally, an exon corresponding to exon 12 in human HAP1 sequences, was predicted by both Genscan and HMM algorithms. This putative exon sequence was partial and truncated by insertion into the cloning vector (therefore no more sequence from this exon was obtainable from the clone under study). While the incompleteness of the sequence made it impossible to predict potential for valid coding from this exon, a 1st phase truncated read from position 3278 of the submitted sequence was detected.

The two putative coding exons (corresponding to exons 9 and 12) that seem consistent with known HAP1 genes would yield a protein product of 120 residues:

SLWAGSQLQDLQEKYSECGDMLMEAQEKVKTLRQQAPASAGSVTHYTYTVPLVPA EERVTEEEELVSEEAEAWEDLEPEVDETTQMNVVTSALEASGLGPLRLDIKYVLQQL ANWQDARFRQQLRQKMLQKGECSRGGLPLASLTSC (truncated)

This predicted protein fragment was submitted for BLASTP similarity searching and the following representative result obtained against one high scoring homologue:

gi|10241694|emb|CAC09418.1| (AJ012824) huntingtin-associated protein 1 [Homo sapiens]
Length = 671

Score = 100 bits (249), Expect = 3e-21
Identities = 52/68 (76%), Positives = 57/68 (83%)

Query: 53 VDETTQMNVVTSALEASGLGPLRLDIKYVLQQLANWQDARFRQQLRQKMLQKGECSRGGL 112
+DE T+MNVVTSALEASGLGP LD+ YVLQQLANWQDA +R+QLR KMLQKGEC G L
Sbjct: 599
LDEATRMNVVTSALEASGLGPSHLDMNYVLQQLANWQDAHYRRQLRWKMLQKGECPHGAL 658

Query: 113 PLASLTSC 120
P AS TSC
Sbjct: 659 PAASRTSC 666

Score = 77.0 bits (188), Expect = 4e-14
Identities = 37/52 (71%), Positives = 43/52 (82%)

Query: 1 SLWAGSQLQDLQEKYSECGDMLMEAQEKVKTLRQQAPASAGSVTHYTYTVPL 52
S+W GSQLQDL+EKY +CG ML+E QE+VKTLRQQ P S GS THY Y+VPL
Sbjct: 426 SVWVGSQLQDLREKYMDCGGMLIEMQEEVKTLRQQPPVSTGSATHYPYSVPL 477

The product of these exons showed good similarity to human HAP1. As noted above, exons 10 and 11 in the clone sequence are frame-shifted; if they were coding they would account for around 45 aa of the 122 residue gap in similarity with the human sequence under comparison.

As mentioned above, sequence coverage for most of the material submitted to Genbank was 1-times only. It is possible that this low coverage resulted in frameshift reads seen in exons 7, 8 and 10 (and possibly exon 11). However, it is more likely that the sequenced clone is a fragment of a pseudogene with strong similarity to Huntingtin-associated protein. Such pseudogenes are not unique as a human HAP1 pseudogene has been previously identified (gb: NG_000875; HAPP).

Repeat identification

A number of repeat elements were identified in the clone sequence by Repeatmasker. The repeats were:

MIR3 repeat SINE/MIR (class/family) positions 1089-1172	reverse strand
L2 repeat LINE/L2 (class/family) positions 1486-1564	forward strand

(a third monotreme type repeat was identified but is not shown).

The relative positions of these repeats are shown on the pseudogene diagram below (Figure A.6):

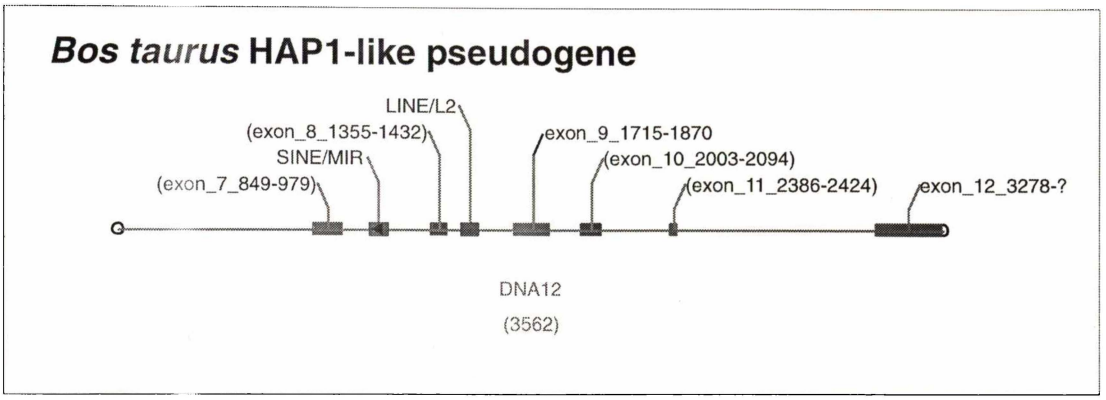


Figure A.6. Annotated diagram of the pHAP1-like (pseudogene) showing exons with predicted coding potential (9 and 12) and the mis-framed exons 7, 8, 10 and 11.

Summary

In the absence of either 1) experimental confirmation of expression from the putative coding region within the pBHAP clone, 2) of 3-fold or better sequence coverage, or 3) of complete '(pseudo)gene' sequence, it is difficult to conclusively characterize the clone sequence.

However, prima facia evidence suggests that the pBHAP clone is most likely to be an unprocessed pseudogene with strong similarity to published Huntingtin-associated protein-1 sequences at the DNA and protein levels (Figure A.6). That is, due to the frequency of reading frame shifts found within the exons (a few miscalled bases may be expected of sequencing misreads) it is thought very unlikely that the sequence represents elements of a coding gene. The clone sequence was hence described in the submission as a *Bos taurus* HAP1-like pseudogene.

The LINE and SINE repeat family elements predicted by repeatmasker as detailed above, were included and the annotated sequence submitted to Genbank where it has the accession number gi:AF531755.

LOCUS AF531755 3562 bp DNA linear MAM 16-AUG-2002
DEFINITION Bos taurus huntingtin-associated-like protein (HAP1) pseudogene, partial sequence.
ACCESSION AF531755
VERSION AF531755

KEYWORDS .
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3562)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Bos taurus HAP1-like partial gene sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3562)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2002) Biology, The University of Waikato,
Hillcrest Rd, Hamilton 2001, New Zealand
FEATURES Location/Qualifiers
source 1..3562
/organism="Bos taurus"
/db_xref="taxon:9913"
gene <1..>3562
/gene="HAP1"
/note="similar to huntingtin-associated protein"
/pseudo
repeat_region complement(1104..1178)
/note="MIR3"
/rpt_family="SINE/MIR"
/rpt_type=dispersed
repeat_region 1492..1570
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=dispersed
BASE COUNT 677 a 999 c 1043 g 843 t
ORIGIN
1 ttgaggtggc aacagtgaac aacagttgaa gtgagggtgc agttgccctc accccaaatt
61 tggccttcct tccctcaggt tccctgaatg acccccctggt cccaggcccc tggagtagga
121 gtcctcacat tctgttgctt ctgcctcgta ggaaaggctc aagtctcctt cctcttcccc
181 atcccagcta cctctagcgc tggccacctg ttacatggtc tctctgcctc cttccctgct
241 cttccaggcc cttctgcaca cagcccaaag gagttttcta aagccagact attggtcctc
301 tgcctaaacc gcccaagtaa ctccctattg gccaaaggat aaaccctacc ctccctaccc
361 cagccctagt catctggacg ctgcccactg ctccagcact ctctcctggt gtccggtggt
421 tccctctggg cccctccaca ctctgcctgg gctctgttgc cacttctgcc tggtgagacc
481 tccttgccct gtggaggtcc tagctccata gatcttgcca ggctggactg ggtgcggcgg
541 ctgggaccct ggatgcactc tttacactcg ctgcacctta gttcataatt tgcactgttg
601 tgatttgttt ttccctcagt ctccacactc gccttgagct cagtgtgggc tggggccggt
661 cttgctgact gcttgcccag caccctgctc tgtgtcagtc agtaaacgtt tattgaaaag
721 aagacccgag tgagggatgt taagccctgt ggaaccttgg gtacctggag gggtgcaggc
781 agacaggggc atctgggaga ctgagggccc tggggagcca gcgctgacct tgccccactc
841 ctccccagct gaggccagcc agcagatggc caagctgtca gaggtgctgg cgctcaggat
901 ggagaactat gaacagcagc agagggaggt cgcccagctg cagggccagg tcacgaagct
961 acagcggtgt tgccgatcgg tgagtcggga catatgcggg gcttcctcct cctgtcacag
1021 ccccctggaa tccatgcccc cttctctctg cccccacccc aggagactca gctcagaggc
1081 ctctctcttg tgtcagagtc agagttggag aagacatcag ggagcctctc atcctgtccc
1141 cttgtctcaa gtcgggggct gagcctaaag agagacacag ggtgttccca ggaaacaggc
1201 cacatgactc caccttcttt ggccttttcc catgtgactt ctgttttaaa gagctgctga
1261 cactcactga cgtgtctggg gagacttgct ggcagcaggg gctgcaggga tgtggggaga
1321 cctggggtgg gggttcactg gccttaggtc cttccagtat ggggctgaga ctgagaagtt
1381 gcagaagcag ctggcttagg agaaagaaat ccaggtgaag ctccaggatg aggtgaggct
1441 tcccctgagc tcccaggacc cttgggtggc tgcctgagtc taaccagctc cttgaacagg
1501 ggctgacgct gtcctttcca catactcagg ctctggcaca gggcctggca ggtagcagtt
1561 cagttggggc ctgaatggag tagagtccag gggtcacact gtgtcattat atgggtctag
1621 agtctgggcc cgttgcctct ttgcccaggc ttcccggccc ctttgggaga gagtgggcaa
1681 ggcgtccagg ctgggtgtgc ctgggtccca tcagagcctg tgggcgggtt cccagctgca
1741 ggacctgcag gagaagtact cagagtgtgg ggacatgctg atggaagccc aggagaaggt
1801 gaagaccctc cgccagcagg ccccggcatc tgctggctct gtcacccact acacatacac
1861 tgtacctctg gtgaggcccc cggctcctgc ctgttacctc tgtgcgccct gtgtctatct
1921 gcagtccctt cctgggtctg gaggtgtgtg tgtggggtct tctgctgaac cttctggtaa
1981 tgtcccccat ctctgtccat aggaggcact tcctgatttc caggagaccc tgactgagga
2041 gctcagaatg tctataagga gaattatctc agaccctgtg ttttttatgg aaaggtacac
2101 actgggcccc agtcccttgg ctgccccaca ctcactcttg cctacctggt cacagcagcc
2161 tttctggtaa tgggaggggt ttgctcgatc aatcttctcc cctgccgtgg tgtctttcct
2221 cctgcctcat cggtccccgg cccttccact caacaccagt cctgtgtcca cgttctgggt
2281 ctgtcacttc tccctgggat ttttctgcac cgtgggtccc cagggatttc acactttcag
2341 cttcccattg ccttgggaca ccaaaggccc ttttctcttt tgaaggaatt atgaagtgac
2401 tccagaggag acatcagacc tggggtaagc cgaccagggt ctccctcttc acttgctctc

```
2461 acttttaagc atccagacaa ccccctgcca ctggctttgt gacctgcagc cccggagaag
2521 ttctggctct cgggctgact ggctgagatg tactgggagg gatactctct aacccttgc
2581 cttcaaattt tgccccatgg gtggccctct gattagaggc ctctccaagc aggtactgaa
2641 gcctttgctt ctcccaacca gtgcctaaga ggcaaatggg tgctggccta ggagccagga
2701 atgctgggac taagtcccag cttgagccct ttactggctg tgtgtccttg gacaagtcac
2761 ttgacctgtc tgagccttac taaaatggat ttgttcctat ggcaactaga cctgggatgg
2821 gttatgaaag ttgaataatg aaggagtggg gcagaggaga tgttattgct agtctgttct
2881 atgaagcaga gtgaagttac tgggggaagg tggggcttgt gcagtgttgg tgaccgttct
2941 tcttttgcc agggccatgg gctgggtgtt tgaggccttg agacttgtac ccagctgcct
3001 cctccctgg ctgggacaga acttgggagt aatcatcagg ttgagtggga gccctggtgg
3061 ctggagcccc gggctgcaga tgaagggaga ggctctggct tagagcgggt aggaaaagac
3121 tttcctgtct gcctgcccag ataggagctt cgctacagag aggagcgagc acaggagcag
3181 gggctcgagg ctggggaggg gctgacgccg accgaggatt ttgtgcctgt ggaagagttg
3241 gtgcctgagg agcagctggg ggtcatagaa gaggcagtgc cagctgagga gagggtaaca
3301 gaagaggagg aactggtgtc tgaggaagct gaggcctggg aggacctaga gccggaggtg
3361 gatgaaacaa ctcagatgaa cgtggtgacc tcagccctgg aggccagcgg cttgggcccc
3421 ttgcgcctgg acataaagta tgtcctccag cagctggcca actggcagga tgcccgtttc
3481 aggcagcagc tgaggcagaa aatgctccag aaaggtgagt gttcccgtgg gggcctccct
3541 ctggccagcc tgacaagctg ca
//
```

Discussion of sequence annotation

If the sequences described above are indeed true pseudogenes (and not merely the result of sequencing error) then they seem most likely to have arisen by gene duplication (c.f. 'processed pseudogenes' that characteristically lack intron regions due to their origin from retrotransposition of mRNA molecules) (Vanin, 1985). It seems that both types of these pseudogenes are not uncommon. A recent survey of human chromosomes 21 and 22 has described some 189 pseudogenes with an approximate ratio of 1:1 for processed to unprocessed pseudogene, and a genome wide estimate of around 7,000 and 9000 for each type respectively (relative to total gene number of around 25,000) (Balasubramanian *et al.*, 2002).

After gene duplication, a new copy of a gene may: be maintained unchanged, acquire new functions, or it may become neutralised to give rise to a pseudogene (a gene-like region that is non-functional because of either: a dramatically altered 'frame-shifted' protein coding potential, or due to failure of transcription or translation) (Vanin, 1985). Most of the predicted exons (by splice consensus sequences) in the sequences under discussion have undergone frame-shift changes, presumably after insertion or deletion mutations. Furthermore as the sequenced region does not include potential promoter sites, or earlier exons, it is difficult to judge conclusively whether other portions of the gene have better potential for coding. Hence, these factors combine to make it difficult to conclusively determine the likelihood of the production of mature coding RNA from either sequence.

It is predicted that the described *g6pd* pseudogene is located on the bovine X chromosome as it is common for such products of gene duplication to reside close to the genes original loci (Mighell *et al.*, 2000).

Future classification of the region may include re-sequencing of the region and RT-PCR with primers on flanking 'exons' to determine whether the predicted exons are found in any expressed gene product (presumably a brain cDNA library would provide the best template DNA for such an experiment).

Appendix 3. Other sequences obtained

A.3.1 AF181665 *Bos taurus* SINE sequence [gi:6003528]

A 432 bp fragment that gave significant similarity to other SINE on Genbank. This sequence was produced from a spurious band that was gel eluted after it co-amplified with a smaller (desired) product.

LOCUS AF181665 432 bp DNA linear MAM 30-JUL-2002
DEFINITION Bos taurus Bov-tA2 sequence.
ACCESSION AF181665
VERSION AF181665.1 GI:6003528
KEYWORDS .
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 432)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-1999) Biology, University of Waikato, Hillcrest
Rd., Hamilton 2001, New Zealand
FEATURES Location/Qualifiers
source 1..432
/organism="Bos taurus"
/db_xref="taxon:9913"
repeat_region 102..306
BASE COUNT 122 a 91 c 111 g 108 t
ORIGIN
1 aggttcaaag taagggcttt tgtaaagtct ggttaccaga cccagagtcc aagctcagaa
61 catgtctgaa gtgtttttta gtcaaataat gtgctaaatt tgagcttccc tggcggcttt
121 catggtaaag agtctgcctg caactcagga gacccaggtt cgatccctgg gtcaggaaga
181 tctcctggag aaggaaatgg catccagctc cagtattctt gcctggaaaa tcccatggac
241 agaagagcct ggcgagtccc agtccatgca atcacaagga gtcggacacg actgagtgac
301 taacacttca atgtgctaaa tttagggtcc ttaaagtgtt ttctacttag gaagcaactg
361 gtcccaggat attgaataag ctgaagcttt gcttggagag tcagacagtc acagctgttg
421 gcagaagagg aa
//

A.3.2 AF104906 *Bos taurus* X-inactive specific transcript (*Xist*) gene, partial sequence. [gi:10181229] 24-AUG-1999

This sequence represents the product of amplification directly from a bovine genomic DNA template using universal primers.

LOCUS AF104906 2182 bp DNA linear MAM 19-SEP-2000
DEFINITION Bos taurus X-inactive specific transcript (XIST) gene, partial sequence.
ACCESSION AF104906
VERSION AF104906.5 GI:10181229
KEYWORDS .
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2182)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Bos taurus partial (exon 1) XIST sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1187 to 1740)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Dept. of Biology, University of Waikato, Hillcrest Rd, Hamilton 2001, New Zealand
REFERENCE 3 (bases 1 to 2182)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1999) Dept. of Biology, University of Waikato, Hillcrest Rd, Hamilton 2001, New Zealand
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 2182)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1999) Dept. of Biology, University of Waikato, Hillcrest Rd, Hamilton 2001, New Zealand
REMARK Sequence update by submitter
REFERENCE 5 (bases 1 to 2182)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Dept. of Biology, University of Waikato, Hillcrest Rd, Hamilton 2001, New Zealand
REMARK Sequence update by submitter
REFERENCE 6 (bases 1 to 2182)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2000) Dept. of Biology, University of Waikato, Hillcrest Rd, Hamilton 2001, New Zealand
REMARK Sequence update by submitter
COMMENT On Sep 18, 2000 this sequence version replaced gi:7322074.
FEATURES Location/Qualifiers
source 1..2182
/organism="Bos taurus"
/db_xref="taxon:9913"
/chromosome="X"
gene <1..>2182
/gene="XIST"
promoter <1..38
/gene="XIST"
mRNA 33..>2182
/gene="XIST"
exon 33..>2182

/gene="XIST"
BASE COUNT 457 a 473 c 533 g 719 t
ORIGIN
1 ttcttaaagc gctgcaattt gctgcgaccg ccatatttct tcttttcccg agatggaagc
61 ttattaatat tggatttctt tgcctgtgtg gttctttctg gaacattttc cagaccccaa
121 ccatgcctta tggcatattt cttggaaaaa attacaccaa aaattcataa aatattttta
181 aaaacctgaa ctttcttcta gtattttctt gacaccttct cagtatttta cagatacttg
241 gggatatttt taggcaattt tccattttaa aggacgtttt ctttggaatg atttttggtt
301 gcctcctctg ggtttttttgt ggtttctttt tctttttctc ctccttttct atattatttg
361 cccatcgcgg ctgtggatac ctggttttag tgtgtatata gataggtaga tagatatatt
421 attattgtat tatttttttac ccaacggggt catggatacc tgcctttat tttatttttt
481 tttaatttgc ccatcggggc cacggatacc tgcttttaat ttttttttc ccccttagcc
541 catcggggcc tcggatacct gctgtgtacc ccctctctcc ctaacctggc ccatcggggc
601 aatggatacc tgcctctttt ttaaatgtgt tgtttttttt ttttttcct tgcccatcgg
661 ggcctcggat acctgcttta attttttttt tcttgcccat cggggcctcg gatacctact
721 ttaatttttt ttccttgccc atcgggaccg cggacacctg cttagatttt tgttttacac
781 cacccatcgg ggctttatat ggttggaaaa gtgttgggtt ttgtggttcg ttgtactatc
841 tggaatgtct acaaaatttt gctgctaatc gtttgtgttg tgtgagtgga cctacggctt
901 tggctgggag atgactttgc agttaggcta gagggttggt taggctgggg aggaaagatg
961 gcggccactt cagatttgcc gcccagcttg gctgaggggt aactgtgcta agtatcacta
1021 ggaaggtaag atgaataatg cgacaggcaa aggaaatggt atgcattgca tgagttaaat
1081 gttttgaata agctgacttg gatcatgtta ggagcttggc atgaattgtg gtatcatgag
1141 gtgggaaaac ataggatcat cctatgtcat attacaaggg tctaatggaa aatgagcggg
1201 agaagaatta ggcgcagtgt tcaaaatggc gattttgacc ttgcagcatt gcttagcatg
1261 gctctctgct ttgttagagt gttcaaaatg gcggatccat tttgccgcag tgttccaatg
1321 gcgggaagcc acatcatggg tgtctttgtt ctagtgtgca gcatggcggt agaaatattc
1381 tgttacatag taaaagatgg ccgcttaagt acttgccgca gtctaaaaca tggcgggctt
1441 ttgtctctgc cgtgtgcatt tcctgatagg ttttgctgca gggacaatat ggctgacctt
1501 gtcatgtgga tatcatggca gtttgtcacg tggatatcgt ggcaggggtg tttgaccgtt
1561 acattcttgg cgggctttgc atcaggaggg cctgccgcat tgttaaagat ggcgtgcttt
1621 gccgcggaca aagtgaaagg agggattggc aatgttagat tgccgcgtgt cccacccaat
1681 cagaaagggt ggtagaatcg gtcacagcca gttagtggag gatggaatta gatgagttag
1741 catagcacct cgctaccgtc tctattcagc cagtcaggat tggccacatt tgtactaatc
1801 tcagtgggtg gtaccaaagt ctttccaagg acatttggcc tttccacctc ctccctcccc
1861 tctcactggc tccctcccct ccctcattgc cacttgcagt gctggatatt aggttgtgca
1921 ggcagaacct cacgccattc ctctgcattg gtgccgtggc cgaagccctt ccccacttcc
1981 cccttccccg ccccctgctc ttctgcgttg tggccagggg cagtgcgcca tgcctgctaa
2041 gtgtgaactt ggcggtgagt cgtggcaagg accagaatgg atcgcagatg atcgttggcc
2101 aactgttggc agaagaggaa tccctgcctt cctcaagagg aacacctacc ccgtggctaa
2161 tgctggggtc ggatttttga ta
//

A.3.3 AY136819

LOCUS AY136819 691 bp DNA linear MAM 20-AUG-2002
DEFINITION Bos taurus clone pBLS1 Bov-B LINE/BovB repeat sequence.
ACCESSION AY136819
VERSION AY136819
KEYWORDS .
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora;
Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 691)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Bos taurus LINE repeat sequence (Bov-B LINE/BovB)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 691)
AUTHORS Jacobson,G.M. and Wilkins,R.J.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-2002) Biology, The University of Waikato,
Hillcrest Rd., Hamilton 2001, New Zealand
FEATURES Location/Qualifiers

```
   source          1..691
                   /organism="Bos taurus"
                   /db_xref="taxon:9913"
                   /clone="pBLS1"
                   /clone_lib="Stratagene lambda library Cat number 945701"
   repeat_region   <1..>691
                   /rpt_family="Bov-B LINE/BovB"
                   /rpt_type=dispersed
BASE COUNT     239 a   138 c   152 g   162 t
ORIGIN
     1 gatcactcac ctagagccag acatccttga gtgtgaagtc aagtgggcct taggaagcat
    61 gacaatgaac aaagtgagtg gaggtgatgg aatcccagtt gagctgtttc aaatcctaaa
   121 agaagcagtg gccacaggac tgaaaaaggt cagttttcat tccaatcaca aagaaaggca
   181 atgccaaaga atgctcaaac taccgcacaa ttgctctcat ttgatatgct agcaaagtaa
   241 tgttcaaaat ccttcaagct aggtttcaat agtatgtgaa cttagaactt ccagatgtac
   301 aagctggatt tagaagaggc agaggaacca gagatcaaat tgccaacatc ttttggatca
   361 tagaaaaagc caattcagaa aacatctact tctgcttcat tgactatgct aaagcctctg
   421 cctgtatgga tcacaccaaa ctggaaaatt cttaaagaga tgggaatacc agaccacctg
   481 acctgcctct tgagaaacct gtatgcaggt caggaagcaa cagttagaac tggacatgga
   541 acaacagact ggttccaaat aggaaaagga gtatgtcaag gcagtatgtt gccaccctgc
   601 ttatttaact tatttgcaga gtacatcatg agaaatgcca ggctggatgg atcacaagct
   661 ggaatcaaga ctgctgggag aaatatcaat a
//
```

A.3.4 AY136820

```
LOCUS       AY136820          2082 bp    DNA     linear   MAM 20-AUG-2002
DEFINITION  Bos taurus clone pBLS2 Bov-B LINE/BovB and BOV-tA2 SINE/BovA repeat
            sequences.
ACCESSION   AY136820
VERSION     AY136820
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 2082)
  AUTHORS   Jacobson,G.M. and Wilkins,R.J.
  TITLE     Bos taurus Bov-B LINE/Bov-B and Bov-tA2 SINE/BovA containing
            repetitive region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2082)
  AUTHORS   Jacobson,G.M. and Wilkins,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2002) Biology, The University of Waikato,
            Hillcrest Rd., Hamilton 2001, New Zealand
FEATURES             Location/Qualifiers
     source          1..2082
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="pBLS2"
                     /clone_lib="Stratagene lambda library Cat number 945701"
     repeat_region   <1..1464
                     /rpt_family="Bov-B LINE/BovB"
                     /rpt_type=dispersed
     repeat_region   2018..2082
                     /rpt_family="BOV-tA2 SINE/BovA"
                     /rpt_type=dispersed
BASE COUNT     616 a   467 c   504 g   495 t
ORIGIN
     1 gatcactcac ctagagccag acatccttga gtgtgaagtc aagtgggcct taggaagcat
    61 gacaatgaac aaagtgagtg gaggtgatgg aatcccagtt gagctgtttc aaatcctaaa
```

```
 121 agaagcagtg gccacaggac tgaaaaaggt cagttttcat tccaatcaca aagaaaggca
 181 atgccaaaga atgctcaaac taccgcacaa ttgctctcat ttgatatgct agcaaagtaa
 241 tgttcaaaat ccttcaagct aggtttcaat agtatgtgaa cttagaactt ccagatgtac
 301 aagctggatt tagaagaggc agaggaacca gagatcaaat tgccaacatc ttttggatca
 361 tagaaaaagc caattcagaa aaacatctac ttctgcttca ttgactatgc taaagcctct
 421 gactgttgga tcacaacaaa ctggaaaatt cttaaagaga tgggaatacc aaccacctga
 481 cctgcctctt gagaaacctg tatgcaggtc aggaagcaac agttagaact ggacatggaa
 541 caacagactg gttccaaata ggaaaaggag tatgtcaagg cagtatgttg ccaccctgct
 601 tatttaactt atttgcagag tacatcatga gaaatgccag gctggatgga tcacaagctg
 661 gaatcaagac tgctgggaga aatatcaata acctagatat gcagatgaca ccactcttat
 721 ggcaggaaga gaagaggaac taaagagcct cttgatgagg ctggaagaga agagtgaaaa
 781 agctggctta aaacacattc aaaaaactaa gatcatggta tctggtccca tcacttcatg
 841 gcaaatggat ggggaaacaa tgcaaacagg gaccaacagt tttcttgggc tccacaatca
 901 ctacagatgg tgcctgcagc catgaagtta aaagacgctt gctctttgga agaaaagcta
 961 tgacaaaact aaacagagta ttaaaaagca gagacatcac tttgctgact aagttctatc
1021 tagtcaaagc tatggttttt gcagtaatca tgtacagatg tgagacttgg acaataacaa
1081 aggctgagtg ccaaagaatt gatgctttca aactgtggtg ctggagaaca ctcttgagaa
1141 tcccttggac agcaaggaga tcaaagcagt gcatcctaaa ggaaatcaac cctgaatact
1201 cattggaagg actcatgctg aagctgaagc tccaatactt tggccacctg atgcaaagag
1261 ccttttctca ttggaaaaga ccctgatgct aggatagact gaggacagga gaaggggacg
1321 acagaggatg agatgcttgg atggcatcac cgactcaatg gacacgagtt tgagtaaact
1381 ccgggagatg gtgaaggaca gggaagcctg gcgtgctgca gtccatgcat tctcaaggag
1441 ttgggcatgc ctgagtgaat gaacaacaac caggacctca tgcctctttt gcattgcaag
1501 ctcctagagc tgagattaca gctactgtcc ttctcaggtc cttcatgtct gtccttcatt
1561 ttcattgcac tgggcatccg gtgatagttg agcatcctca ccagggccct ctcatctctg
1621 tcacagggta gcaccaagat ctgagctgtg cactaacttc tctagcctgg gccctctgtc
1681 atggtgggga cactcagaga ggatcagctt gggggttccc accctgccct ccctgctccc
1741 acgtagtcat cacatgttgg gagtccagga agagcttagt ctctcctcca tttccggttc
1801 ctggccctac atggatcctc cacgctgtta agtgtcccac tcagtgatgc catgccagaa
1861 ggctttccaa ggggcttggg atgactctcc ccagaactca gctctactgg gtcccccaaa
1921 gtctgtaggt gtctgtggag ctgtgtgagg ggtccctgag cacaaagacc ccagcagggt
1981 caagaaaggc tcttttgttt ctttgctttt tctcctgagg gcttccctgg tggctcagat
2041 ggtaaagaat ccgcctgcaa tgccggagac cgaggttgga tc
//
```

Appendix 4.

lacZ staining in mouse intestine and skin

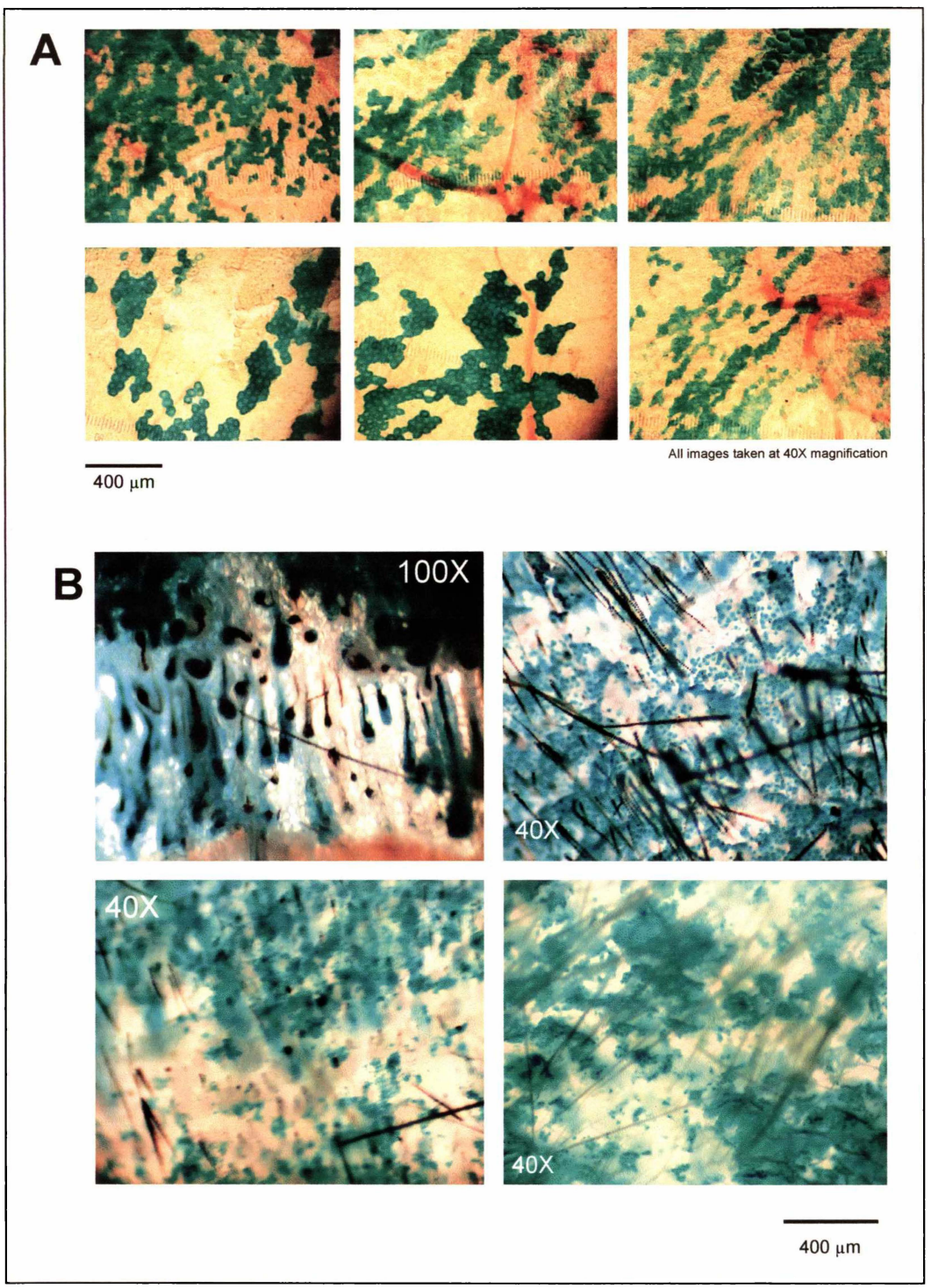


Figure A.7. Showing A) Samples of histochemically stained intestinal tissue and B) skin, from a homozygous Tam-*lacZ* animal. The images are of whole-mount tissues and were taken via a compound microscope using a Nikon 'Coolpix' camera. Magnification is indicated on the plates.

Appendix 5. Recipes

Neutralizing solution

0.5 M Tris-HCL (pH 7.4)

1.5 M NaCl

5X TBE buffer

108 g Tris base (final 890 mM)

55 g boric acid (final 890 mM)

40 ml 0.5 M EDTA, pH 8.0 (final 20 mM)

Adjust volume to 1 l

TE buffer

10 mM Tris-HCl (as recipe below)

1 mM EDTA, pH 8.0

Adjust to required pH (usually 7.4, 7.5, or 8.0)

1M Tris-HCl

Dissolve 121 g Tris base in 800 ml H_2O

Adjust pH (between pH 7.2-9) with 10 M HCl

Add H_2O to 1 l with mixing

50X TAE buffer

242 g Tris base

57.1 ml glacial acetic acid

37.2 g $Na_2EDTA.2H_2O$

H_2O to 1 l

Sodium acetate, 3 M (500 ml)

Dissolve 204 g sodium acetate.$3H_2O$ in 400 ml H_2O

Adjust pH with 3 M acetic acid (usually to pH 4.8 or 5.2)

Add H_2O to 0.5 l

10 M NaOH

Dissolve 400 g NaOH in 450 ml of H_2O

Add H_2O to 1 l

1M $MgCl_2$ (100 ml)

Dissolve 20.3 g MgCl2.6H_2O in 80 ml of H_2O

H_2O to 100 ml

Ethidium bromide (10 mg/ml)

Dissolve 0.1 g ethidium bromide in 10 ml of H_2O

Store at RT in an amber bottle.

0.5 M EDTA

Dissolve 186.1 g $Na_2EDTA.2H_2O$ in 700 ml H_2O

Adjust to pH 8.0 using 10 M NaOH

Add H_2O to 1 l

Lambda top agarose

10 g Tryptone

8 g NaCl

6 g Agarose

add H_2O to 1 l, mix well and autoclave

Luria-Bertani (LB) medium (broth/agar) (1 l)

10 g Bacto-tryptone

5 g LB base

10 g NaCl

(add 15 g/l agar before autoclaving if LB agar is required)

Bring to 1 l and autoclave for 20 min

Suspension medium (SM) (1 l)

5.8 g NaCl

2 g $MgSO_4.7H_2O$

50 ml 1 M Tris.HCl (pH 7.5)

0.01% gelatin

Bring to 1 l with ddH_2O and autoclave before use.

1X TSS

0.1 g LB

1 ml of 40% PEG 8000

100 µl 2M $MgSO_4$

Bring to 4 ml with deionised water

10 X MOPS buffer

0.2M MOPS (pH 7)

35 mM Sodium Acetate

10 mM EDTA (pH 8)

0.5 M EGTA (250 ml)

190.2 g EGTA

Bring to 250 ml with H_2O

Bring to pH 7.5 by addition of 1 M NaOH to dissolve

LacZ Fix #1

0.2% Glutaraldehyde

5 mM EDTA (pH 7.35)

0.1M $MgCl_2$

1 X PBS

LacZ wash buffer

2 mM $MgCl_2$

0.01% desoxycholic acid (Sodium salt)

0.02% Nonidet-P40

1 X PBS, adjust pH to 7.5

LacZ stain

1 mg/ml X-gal (initially dissolved in dimethylformamide)

5 mM Ferrocyanide

5 mM Ferricyanide

1 X PBS

Adjust to pH 7.8 before use

(may be supplemented with 20-100 mM Tris-HCl as a pH buffer)

LacZ fix #2

Dissolve 2 g of paraformaldehyde in ddH_20 at 60° with stirring by adding a few drops of 2 M NaOH, add a 1/10 V of 10 X PBS and adjust pH to 7.3. Add glutaraldehyde to a final concentration of 0.25%.

Bis-acrylamide/Acrylamide stock solution:

For 100 ml - 0.81 g bis-acrylamide
30 g Acrylamide
(bis-/acrylamide cross-linking ratio = 2.7).

Diethylpyrocarbonate (DEPC) treatment

Solutions used for RNA work were pre-treated by the addition of 1 ml of DEPC per litre of solution (0.1%). The solution was then mixed vigorously and autoclaved before use.

15% non-denaturing Acrylamide gel

For 20 ml total (that is 2 10 cm X 10 cm gels, 0.75 mm thick)

10 ml 30% bis-/Acrylamide.
4 ml 5X TBE buffer
6 ml High quality ultra-filtered water
160 µl 10 % ammonium persulphate
16µl TEMED stock solution.

10 X PBS (with or without DEPC)

80 g of NaCl, 2 g of KCl, 2 g KH_2PO_4 and 11.5 g of $Na_2HPO_4 7H_2O$ was dissolved in high quality water. Then solution was immediately autoclaved or, if the PBS was intended for RNA work, autoclaved 6 h after the addition of 1 ml/l of DEPC.

Proteinase K buffer

20 mM Tris pH 8.0

25 mM KCl

1.5 mM $MgCl_2$

0.5% Tween 20

Supplemented with 400 ng/µl of Proteinase K

Church and Gilbert hybridization buffer

1. Dissolve 44.5 g of $Na_2HPO_4.2H_2O$ in 500 ml of H_2O. Adjust pH to 7.2 with H_3PO_4 and autoclave.

2. Dissolve 50 g of SDS in 500 ml of H_2O and dissolve at 37°. Add 1 ml of 0.5 M EDTA.

Mix 1 and 2 and store at 37° before use

20X MOPS buffer

To make 400 ml, dissolve 16.75 g of MOPS, 2.7 g $NaOAc.3H_2O$ and 1.4 g of free acid EDTA in 390 ml of DEPC-H_2O. After adjusting pH to 7.4 with NaOH, filter through a 0.4 µm filter and store at 4°.

Formaldehyde gel loading buffer

1 mM EDTA pH 8

0.25% BPB

0.25% XC

50% Glycerol

Appendix 6.

Quantity of bovine DNA extracted from 1 l of retail milk product; How many cells does this equate to?

DNA from a 50 µl resuspension of total DNA from 240 ml of milk was quantified at of 1880 ng/µl. Hence, assuming no DNA loss during the extraction, there was 94 µg of DNA per 240 ml, or 390 ng/ml. A diploid bovine genome is equivalent to 6 pg of DNA, so, in 1 µl of milk there are 65 cells, in 1 ml of milk there are 390/0.006 (6.5×10^4) cells and in a litre of milk - from healthy animals - 6.5×10^7 mammary gland cells.