UNIVERSITY
of
GLASGOW

Proteomic analysis of the mouse mammary gland

A thesis submitted for the degree of
Doctor of Philosophy at the University of Glasgow

By

Claire Rebecca Davies

Division of Cancer Sciences and Molecular Pathology
Faculty of Medicine
University of Glasgow
Glasgow
G11 6NT

August 2004

The research reported within this thesis is my own work
except where otherwise stated, and has not been
submitted for any other degree

Claire Rebecca Davies

Abstract

The mouse mammary gland develops mainly postnatally by highly regulated phases of invasive growth, ductal branching, differentiation, apoptosis and tissue remodelling during each pregnancy cycle. Detailed analyses of these stages of development were performed using proteomics in order to identify proteins that are regulated in these stages of development. Sixteen time points spanning the key stages of development were analysed and data mined by 2-D gel electrophoresis. Digital copies of the proteome images revealed the expression profiles of all the protein features detected across the developmental time points. These expression profiles were interrogated by creating two databases, and from these databases 125 protein features were selected for mass spectrometry analysis. Selection was based on the following criteria; features that were exclusively present in particular phases of development, and features that were altered during the switch from lactation to involution. Fifty nine protein identifications were determined by mass spectrometry.

In addition to analysing these stages of development, specialised structures which drive ductal morphogenesis, the terminal end buds, were isolated and analysed by 2-D gel electrophoresis. Forty four protein features which showed differential expression between the terminal end buds and the adult virgin gland were chosen for mass spectrometry analysis. Of these protein features, 24 produced protein identifications.

Detailed analyses were performed on selected protein identifications using western blot, immunohistochemistry, real-time RT-PCR and data available from a microarray study. Investigations on one particular protein identification, annexin A2, were extended to breast cancer tissue samples.

Not all the protein features selected for mass spectrometry analysis revealed protein identifications. The expression profiles of these features were assessed again and 28 were chosen for reanalysis by mass spectrometry. The data generated for one particular protein identification, BTF3, indicates a possible regulatory role in mouse mammary gland development.

This study has used proteomics as a tool for identifying proteins which regulate mouse mammary gland development. Many proteins were identified, and the follow-up studies performed highlighted a few as being important in this area of research.

Abbreviations

°C	degrees Celsius
µg	microgram
µl	microlitre
1-D	1-dimensional
2-D	2-dimensional
A_{260}	absorbance at 260 nm
A_{280}	absorbance at 280 nm
AMP	adenosine monophosphate
APES	3-aminopropyltriethoxysilane
APS	ammonium persulphate
ARF	ADP ribosylation factor
ASPP	apoptosis stimulating protein of p53
ATP	adenosine triphosphate
Av	adult virgin
BCA	bicinchoninic acid
BLAST	basic local alignment search tool
bp	base pairs
BRCA1	Breast cancer gene 1
BSA	bovine serum albumin
BTF3	basic transcription factor 3
C-	carboxy-
CapZ beta	F-actin capping protein beta subunit
cDNA	complementary DNA
CHAPS	3[(cholamidopropyl)dimethylammonio]-1-propane sulphonate
CLARP	caspase-like apoptosis regulatory protein
cm	centimetre
cpm	counts per minute
cRNA	complementary RNA
C_T	threshold cycle
Da	Dalton
DAB	3,3-diaminobenzamidine
dATP	2' deoxyadenosine triphosphate
DCIS	ductal carcinoma *in situ*
dCTP	2' deoxycytosine triphosphate

DEPC	diethyl pyrocarbonate
dGTP	2' deoxyguanosine triphosphate
dH_2O	distilled H_2O
DIGE	differential gel electrophoresis
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2' deoxy (nucleotide) triphosphate
DOC2	disabled homolog 2
DTT	dithiothreitol
dTTP	2' deoxythymidine triphosphate
E.coli	*Escherichia coli*
ECM	extracellular matrix
EDTA	ethylenediamine tetra acetic acid
EEF2	elongation factor 2
EH	eps15 homology
ER	endoplasmic reticulum
ESI	electrospray ionisation
EtBr	ethidium bromide
FAM	6-carboxy-fluorescein
FRET	fluorescence resonance energy transfer
g	centrifugal force equal to gravitational acceleration
g	gram
GAP	GTP-ase activating proteins
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GDP	guanosine diphosphate
GH	growth hormone
GTP	guanosine triphosphate
h	hour
hnRNP	heterogeneous nuclear ribonucleoprotein
HRP	horseradish peroxidase
HSL	hormone sensitive lipase
ICAT	isotope-coded affinity tagging
IEF	isoelectric focusing
Ig	immunoglobulin
IHC	immunohistochemistry

Inv	involution
IPG	immobilised pH gradient
kb	kilobases
kDa	kiloDaltons
L	lactation
L-15	Liebovitz L-15 medium
LC	liquid chromatography
LCIS	lobular carcinoma *in situ*
LHS	left-hand side
M	molar
mA	milliamp
MALDI	matrix-assisted laser desorption ionisation
MBSU	Molecular Biology Support Unit
MCI	molecular cluster index
MCM	mini-chromosome maintenance
MCM2/3	DNA replication licensing factor MCM2/3
ME/MP	metabolic enzymes/mitochondrial proteins
MELANIE	Medical ELectrophoresis Analysis Interactive Expert group
mg	milligram
min	minutes
ml	millilitre
mm	millimetre
mM	millimolar
MMPs	matrix metalloproteinases
MOPS	3-(N-morpholino)propane-sulphonic acid
Mr	molecular weight
mRNA	messenger RNA
Mu-adaptin 3A	adapter-related protein complex 3 mu 1 subunit
MudPIT	multidimensional protein identification technology
N-	amino-
NaOAc	sodium acetate
NCB	sodium bicarbonate cotransporter
ng	nanograms
NL	non linear
nm	nanometre
OGS	Oxford GlycoSciences

ORC	origin of recognition complex
P	pregnancy
PAGE	polyacrylamide gel electrophoresis
PAP	PYK2-associated protein
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDA	piperazine diacrylyl
PEM	protein expression map
pI	isoelectric focusing point
PKA	protein kinase A
pmol	picomole
Pol I	RNA polymerase I
pre-RC	pre-replicative complex
PTFE	polytetrafluoroethylene
PTM	post-translational modification
PTRF	polymerase I and transcript release factor
PYK	Protein tyrosine kinase
Q-TOF	quadrupole time-of-flight
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
RT	room temperature
RT-PCR	reverse transcriptase-PCR
s	second
SDS	sodium dodecyl sulphate
SELDI	surface enhanced laser desorption/ionisation
ser	serine
SHWFGF	Sir Henry Wellcome Functional Genomics Facility
SLAM	signalling lymphocytic activation molecule
SLRP	small leucine-rich proteoglycan
Stat3	signal transduction and activation of transcription 3
TAG	triacylglycerol
TAMRA	6-carboxy-tetramethyl-rhodamine
TBE	tris-borate EDTA
TBS	Tris buffered saline

TCP1	*t* complex polypeptide 1
TE	tris-EDTA
TEB	terminal end bud
TEMED	N,N,N',N'-tetramethylethylenediamine
T_m	melting temperature
TOF	time of flight
t-PA	tissue plasminogen activator
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer RNA
TTF	transcription termination factor
tyr	tyrosine
U	unit
UV	ultraviolet
V	volt
v/v	volume per volume
VAT-1	vesicle amine transport protein 1
Vh	voltage hour
w/v	weight per volume
WDR1	WD-repeat protein 1

Table of Contents

Index of Figures

Chapter 1 Introduction

Figure

Chapter 2 Materials and Methods

Figure

Chapter 3 Preparation of the mouse mammary gland

Figure

Chapter 4 Generation of the mouse mammary gland proteome by 2-D gel electrophoresis

Figure

Chapter 5 Candidate protein annotations

Figure

Chapter 6 Reanalysis of selected MCIs by mass spectrometry

Figure

Index of Tables

Acknowledgements

I would like to acknowledge the following people for their help and support during my PhD. I am grateful to my supervisors Prof. Barry Gusterson and Dr. Martin Page for their guidance and encouragement throughout this project.

I would like to thank everyone in Prof. Barry Gusterson's lab, including those who taught me IHC. I would particularly like to thank Joanna Morris for proof reading through my thesis and for all her encouragement, help and support. During my time at Oxford GlycoSciences Matthew Griffiths helped me enormously with the proteomics part of this study. He was extremely generous with his time and provided additional guidance to my PhD, thank you.

I am grateful to my advisor Nicol Keith who provided me with advice on several occasions.

I am grateful to the Pathological Society of Great Britain and Ireland for their studentship.

I would like to dedicate this thesis to my parents, and thank them for their lifetime of love and encouragement. I cannot express how grateful I am to them. Thanks to my sisters, Rachel and Sarah for being there for me. Thanks to Glenise and John Radford for reading my thesis. Finally I would like to thank my boyfriend Jonathan Radford for his unfailing love, support and confidence in me; you always make me feel happy.

Chapter 1
Introduction

This introduction provides an overview of the developing mammary gland both structurally and morphologically. A combination of proteomics and mammary gland development was the basis of this thesis and therefore various proteomic techniques are discussed. The intention was to dissect key events and critical time points in mammary gland development using proteomics. Each results Chapter has an introduction focussing on the areas relevant to the content of each Chapter.

1.1 The mouse mammary gland

1.1.1 Postnatal development of the mouse mammary gland

The unique aspect of the mammary gland is that unlike most organs it develops predominantly postnatally and therefore different stages of the developing gland can be defined. Its development advances in distinct stages, namely prepubertal, pubertal, resting virginal, pregnant, lactating and involuting (apoptosis of the gland). As the gland progresses through these phases of development, extensive tissue remodelling occurs which is achieved by the processes of proliferation, differentiation and apoptosis. Development is also dependent on the circulating levels of mammotrophic hormones, in particular oestrogen, progesterone and prolactin. There are other levels of control involved in mammary growth, such as epithelial-stromal interactions, i.e. interactions with the ductal epithelial components of the extracellular matrix (ECM).

The prepubertal gland is composed of a matrix of connective tissue and a special type of adipose tissue known as the fat pad. At puberty the gland develops by movement through the mammary gland of specialised club-shaped structures called terminal end buds (TEBs). During pregnancy another major change takes place with the development of lobuloalveolar structures along ductal branches. These structures are composed of a single layer of secretory alveolar epithelial cells, which are surrounded by an outer layer of myoepithelial cells. The final stage of development is involution, where extensive apoptosis of the epithelial cells occurs, resulting in extensive tissue remodelling (Figure 1.1). These stages are described in more detail below to provide the context of the project.

1.1.2 Composition of the gland

The mammary gland is a complex tissue composed of multiple cell types; epithelium, adipose tissue and connective tissue stroma. Dispersed within the gland are blood vessels, nerves, smooth muscle fibres, lymph nodes and a lymphatic system (Kaye *et al.*, 1995). There are two main compartments of the mammary gland, the epithelial (or parenchymal) compartment and the stromal (or mesenchymal) compartment. The parenchymal

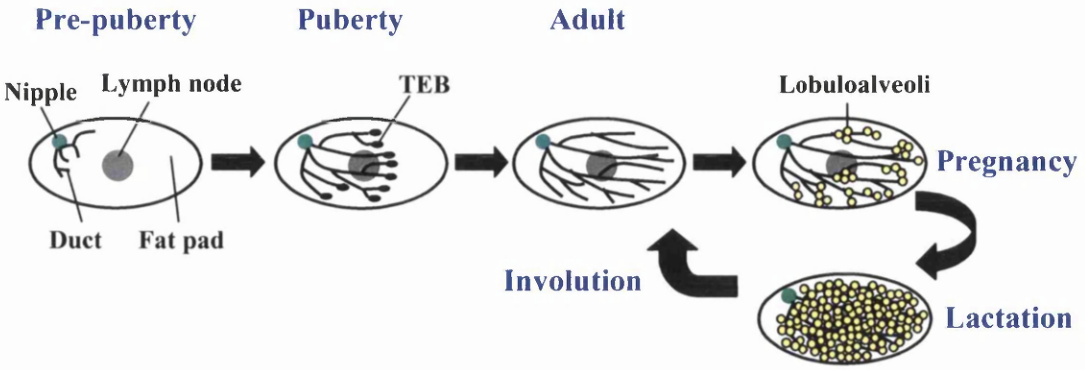


Figure 1.1 Postnatal stages of mouse mammary gland development. The prepubertal stage is a resting phase of development which continues until puberty. The release of ovarian hormones at puberty results in ductal branching and invasion by TEBs. At 10 weeks of age the fat pad is filled with ductal epithelium. Alveoli form along the ductal side branches and differentiate into lobular alveoli. During lactation the gland is filled with milk and the majority of the gland is composed of epithelial cells. Apoptosis of the epithelial cells occurs during involution resulting in much tissue reorganisation. The gland is remodelled back to its state prior to pregnancy.

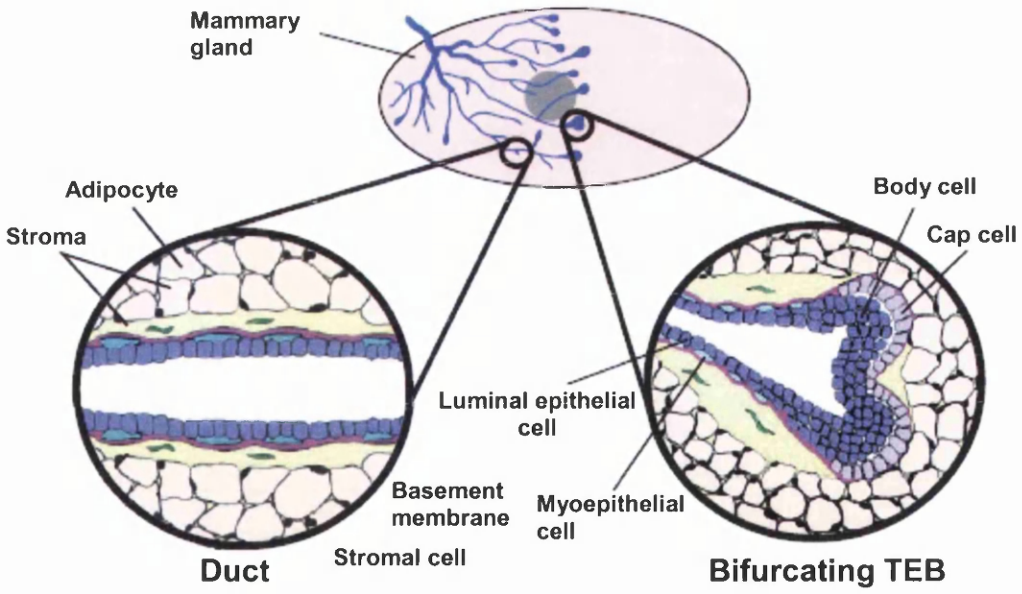


Figure 1.2 Cell types and structures of the mouse mammary gland. The epithelial cells are in contact with the lumen of the ducts and myoepithelial cells. Following these cells are layers of basement membrane, stroma and adipocytes. Most of the TEB is composed of body cells which become ductal epithelial cells and the tip is surrounded by a layer of cap cells which differentiate into myoepithelial cells (Wiseman and Werb, 2002).

compartment comprises of different epithelial structures, namely luminal epithelium of ducts, TEBs, alveoli and myoepithelial cells which vary morphologically and functionally (Figure 1.2).

The apical surface of the luminal epithelium is in contact with the lumen of the ducts and alveoli. Epithelial cell shape ranges from cuboidal or columnar in ducts, to pyramidal in non-secreting alveoli and flattened in secreting cells (Richardson, 1947). The myoepithelium surrounds the luminal epithelium of ducts and alveoli and its shape and thickness varies during development. The layer of myoepithelial cells is discontinuous around the alveoli during pregnancy and lactation, and forms a thick continuous layer during early involution. Although the alveoli regress during this phase of development, myoepithelial cells can reside around these sites (Emerman and Vogl, 1986). Evidence for the myoepithelial rather than the luminal cells secreting basement membrane, is the detection of the basement membrane proteins laminin intracellularly within the myoepithelium and not the epithelium layer (Warburton *et al.,* 1982). The myoepithelium synthesises and secretes basement membrane which separates the epithelium from connective stroma (Pitelka, 1980; Ferguson *et al.,* 1992; Gordon and Bernfield, 1980). The myoepithelial layer is between the luminal epithelium and basement membrane (Glukhova *et al.,* 1995; Koukoulis *et al.,* 1991; Gould *et al.,* 1977; Hollmann, 1974; Kim and Clifton, 1993; Russo and Russo, 1978; Moore *et al.,* 1987).

The connective tissue stroma differs in density, thickness and composition around the ducts and alveoli, being denser around the ducts (Van Zwieten, 1984). The stroma is rich with adipocytes which rapidly and reversibly change in size and number during lactation and involution. A major difficulty in analysing the effects of mammary stroma arises from the heterogeneity of the cell types that are present. Mammary stroma is composed of mesenchymal cells, adipocytes, fibroblasts, pericytes, endothelial cells and mast cells, all of which have an effect on the epithelium (Gouon-Evans *et al.,* 2000). The effect of each type of stromal cell in mammary gland development must be identified in order to determine its role in this organ.

1.1.3 The fat pad

The fat pad plays a crucial role in mammary development and is a mediator of endocrine action and growth regulation. Despite this the properties of the fat pad have received little attention. Understanding these properties may prove helpful in elucidating the importance of the fat pad in mammary gland development. Numerous studies have shown that it has an

essential role in morphogenesis of the mammary epithelium in the adult gland and in mammary pattern formation (DeOme *et al.,* 1959; Sakakura *et al.,* 1982; Daniel *et al.,* 1984). This property of the stroma has been used to develop the transgenic fat pad, a technique whereby the epithelium of the gland is removed to leave a "cleared" fat pad. Genetically manipulated epithelial cells expressing a gene of interest are introduced into the gland and the effect of this gene on mammary gland development is assessed (DeOme *et al.,* 1959; Edwards *et al.,* 1988). The cleared fat pad comprises of adipocytes, fibroblasts, endothelial and immune cells all of which could influence mammogenesis (Gouon-Evans *et al.,* 2000). Similar gland development is seen by transplantation of the epithelium to other sites of adipose tissue, which demonstrates that mammary epithelial growth is not specific to mammary adipose tissue (Ormerod and Rudland, 1986).

There are two kinds of adipose tissue, brown adipocytes which are energy dissipating cells and white adipocytes which are energy storing cells (Klaus, 1997). It had been assumed that the adipose in the mammary gland was white (Hovey *et al.,* 1999) but both types are present in the postnatal mammary gland. Brown and white adipocytes can be distinguished by their lipid inclusions. Brown adipocytes have multilocular lipid inclusions whereas white adipose tissue is unilocular (Nechad, 1986). As mammary development progresses there is a substantial change in the lipid content of the fad pad. The fat pad's weight increases from the adult virgin stage to gestation and decreases during lactation, as adipocytes lose most of their fat content. The brown adipocytes are a component of the mammary stroma during the first few weeks of life. Their presence decreases as mice age and they are undetectable prior to adulthood. Brown adipose tissue is located between the nipple and lymph node around the rudimentary ductal system. Investigations indicate that brown adipocytes negatively regulate mammary epithelial differentiation in early postnatal development (Gouon-Evans and Pollard, 2002).

1.2 Structural changes of the mouse mammary gland

1.2.1 Virginal gland

At birth the mouse mammary gland is composed of a rudimentary ductal system (Figure 1.3A). During puberty as a response to hormonal changes, these primitive ducts lengthen and undergo extensive branching into secondary and tertiary ducts. They slowly elongate from the nipple into the fat pad which at this stage consists mainly of adipose tissue (Figure 1.3B). The ductal system which develops eventually fills the mammary fat pad (Williams and Daniel, 1983; Daniel and Silberstein, 1987; Richert *et al.,* 2000). At the start of puberty bulbous TEBs appear at the ductal tips and invade mammary stromal tissue

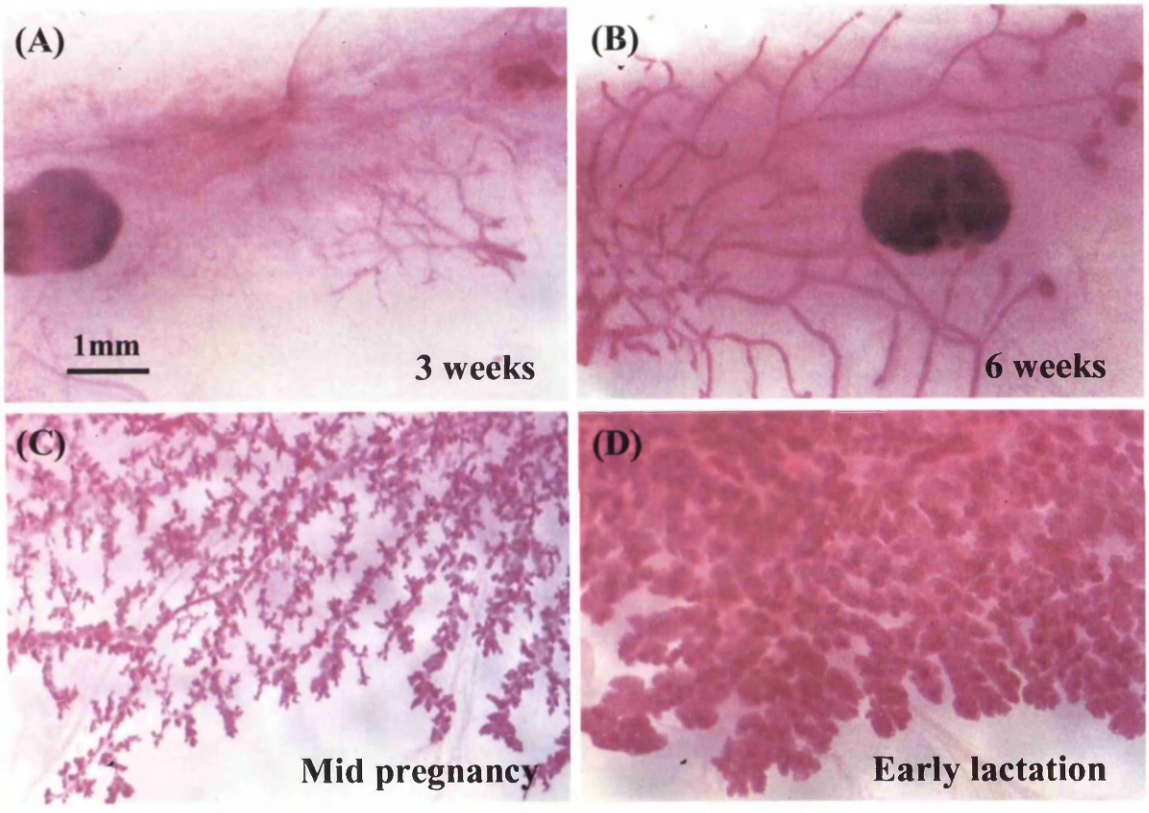


Figure 1.3 Wholemounts of postnatal mammary gland development. The morphology of the gland is visualised by carmine (A) Adult virgin 3 weeks (B) Adult virgin 6 weeks (C) Mid pregnancy (D) Early lactation. Magnification 16x.

(Figure 1.4). These are highly proliferative structures which propel ductal growth by generating differentiated ductal and myoepithelial cells. TEBs control directional growth, branching, turning and pattern formation of the mammary gland ductal system (Williams and Daniel, 1983; Daniel and Silberstein, 1987). The TEB consists of two histologically distinct cell types; cap cells and body cells. The cap cells appear in the outer layer of the TEB structure and differentiate to become myoepithelial cells surrounding the ducts. The majority of cells in the TEB are the body cells, which give rise to ductal epithelial cells. These are found as multicellular layers within the central cavity of the bud that surround the lumen (Dulbecco *et al.,* 1982; 1983; Pitelka and Hamamoto, 1977; Williams and Daniel, 1983; Humphreys *et al.,* 1996).

The TEBs act as control points and are influenced by systemic hormones and local growth factors. Systemic hormones supply stimulation for growth, and local factors influence the direction of growth and spacing of the ducts. These signals eventually lead to regression of the buds as they reach the limits of the fat pad and regress. This occurs at about 10 weeks of age when the mammary fat pad is filled by the branching network of ducts, which cease to elongate as they reach the margin of the fatty stroma (Daniel and Silberstein, 1987; Faulkin and DeOme, 1960). TEBs play a key role in mammary development as they are involved in proliferation, differentiation and apoptosis (Williams and Daniel, 1983; Sapino *et al.,* 1990; Humphreys, 1999; Humphreys *et al.,* 1996). Their importance is extended to epithelial mesenchymal interactions and ECM remodelling (Daniel and Silberstein, 1987; Cunha and Hom, 1996; Robinson *et al.,* 1999).

1.2.2 Pregnancy

The duration of gestation in Balb/C mice is normally around 21 days. During pregnancy highly co-ordinated processes of proliferation, differentiation, apoptosis and tissue remodelling take place to allow for the expansion and formation of alveoli in the gland. At the onset of pregnancy the rate of proliferation of epithelial cells increases over the number of adipocytes formed. This results in rapid and intense proliferation of the ductal branches and the formation of alveolar buds, as is seen during postpubertal development (Nandi, 1958). These alveolar buds form along the walls of the mammary ducts and progressively cleave and differentiate into individual alveoli. At the start of the third week of gestation, lobuloalveolar structures, which are composed of many alveoli predominate in the gland (Figure 1.3C). These will eventually become the milk secreting lobules during lactation. During late gestation the alveolar epithelial cells produce milk proteins and lipids for the preparation of lactation. The individual cells of the lobuloalveoli expand due to the

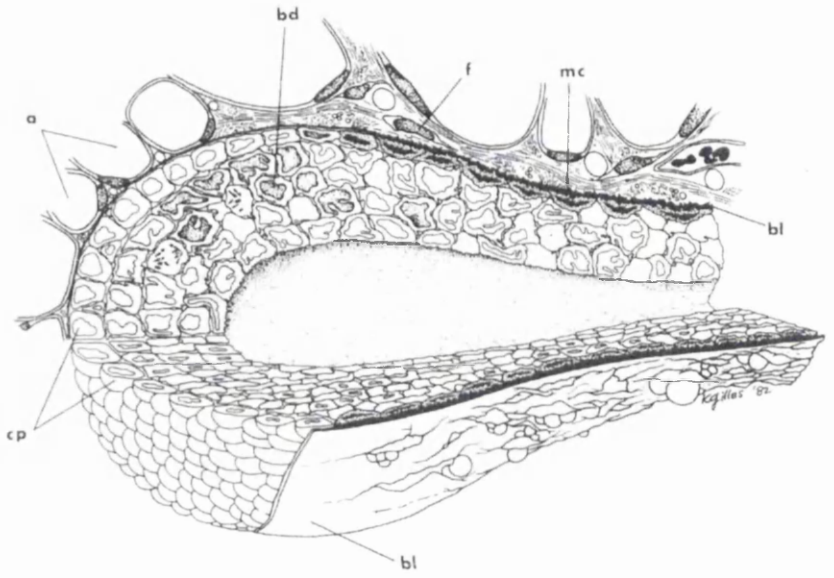


Figure 1.4 Terminal End Bud. Adipocytes (a) abut cap cells (cp) at the tip. Fibrous components and fibrocytes (f) comprise the connective tissue. The basal lamina (bl) represented as a cutaway to expose the underlying cap cells. Cap cells are cuboidal but become flattened toward the midregion of the end bud, then differentiate into myoepithelial cells (mc) in the neck region. The basal lamina overlying myoepithelial cells in the midregion is 14 times thicker than that at the tip. The majority of the TEB is composed of body cells (bd) which differentiate into ductal epithelial cells (Daniel and Silberstein, 1987).

accumulation of secretory products and cause the thickness of the gland to increase. By late pregnancy the alveoli fill the majority of the fat pad and are the most dominant structures within the gland. Myoepithelial cells still surround the alveoli but are not as continuous, so that the epithelial cells can be in contact with the basement membrane. This may be crucial for full development and milk secretion (Adams and Watt, 1993; Barcellos-Hoff *et al.,* 1989; Ferguson *et al.,* 1992; Lee *et al.,* 1985; Howlett and Bissell, 1993). The vast increase in the number of epithelial cells is accompanied by a decrease in surrounding stroma to allow contact between the epithelium and adipocytes (Neville *et al.,* 1998; Elias *et al.,* 1973). The predominance of the alveolar structures continues into lactation.

1.2.3 Lactation

Within three days of parturition, the alveoli appear very swollen as they are engorged with milk. Most of the cells in the mammary gland undergo a dramatic increase in size during pregnancy (Figure 1.3D). Extensive cell proliferation also takes place three to four days postpartum. The gland does not change its appearance for the duration of lactation, which lasts for approximately three weeks. Milk production increases each day and finally peaks in the second week of lactation (Knight and Peaker, 1982). After the onset of lactation, the luminal epithelium changes shape and cells become flattened. Milk fat globules can be seen within the epithelial cells which secrete large amounts of milk proteins from their apical membrane into the lumen of the alveolus that they surround. Myoepithelial cells form a discontinuous layer around each alveolus. The myoepithelial cells contract in response to oxytocin and in doing so force milk from the alveolus into the connecting ductal system (Richardson, 1947; Dulbecco *et al.,* 1986). At the start of lactation the gland is composed of approximately a third of adipocytes, but as lactation progresses the fat in the adipocytes is metabolised to enable the alveoli to expand and completely fill the gland (Elias *et al.,* 1973; Neville *et al.,* 1998). During lactation the adipocytes are present as long projections within the interstitial space of the alveoli (Elias *et al.,* 1973; Blanchette-Mackie *et al.,* 1995).

1.2.4 Involution

Lactation will continue until the pups are weaned or the litter is removed, and then the gland undergoes a process called involution. This is when extensive apoptosis and tissue remodelling occur within the gland (Quarrie *et al.,* 1996). Involution is initiated by the build up of milk once it is no longer removed from the gland. This process is reversible within two days if suckling is once again initiated within this time. This is because local

factors related to milk accumulation are sufficient to induce apoptosis. However if systemic lactogenic hormones persist, alveolar cell death is no longer induced.

The two phase process of involution is regulated by local and systemic factors. During the reversible phase of involution the gland remains histologically similar to the lactating gland and no major morphological changes have taken place. In this first stage of apoptosis the expression of cell cycle control genes is altered as well as survival genes such as bcl-2 (Yang and Korsmeyer, 1996). Local signals are sufficient to induce programmed cell death in this phase of involution, and it is independent of proteinase activity unlike the second phase of involution (Li *et al.,* 1997).

After two days, involution is irreversible (48 h after removal of litter). The secretory epithelial cells of the alveoli begin to apoptose and the basement membrane starts to remodel. At the start of the second stage of apoptosis the alveoli seem to be intact and the epithelial structures still appear to be in an organised state. As involution becomes established, the alveolar epithelial cells begin to apoptose and the alveoli collapse into clusters of epithelial cells.

The second phase of involution is dominated by proteinase activation such the matrix metalloproteinases (MMPs) (Lund *et al.,* 1996). The apoptotic epithelial cells are phagocytosed not only by macrophages (professional phagocytes), but also by semi-professional phagocytes (e.g. fibroblasts) and neighbouring epithelial cells (Fadok *et al.,* 1998; Monks *et al.,* 2002). Epithelial cells which are undergoing apoptosis are cleared prior to their lysis, thus preventing toxic and immunogenic substances from being released. Macrophages engulf the apoptotic epithelial cell and produce an anti-inflammatory reaction (Voll *et al.,* 1997; Fadok *et al.,* 1998; Gao *et al.,* 1998). It appears that semi-professional phagocytes and epithelial cells initially remove apoptotic cells until macrophages have matured and have the ability to phagocytose (Savill *et al.,* 1990).

The production of secretory products ceases and much cellular debris is found in the lumen of the mammary tree. At this point the volume of the fat pad occupied by adipocytes increases and the epithelium decreases. The removal of epithelial cells enables the stroma surrounding the collapsed alveoli to increase. The epithelium appears very disorganised, yet the stroma surrounding the ducts keeps its morphology. Approximately six days after the start of involution the majority of cell death has occurred, all the alveoli have collapsed into clusters, and the stroma and epithelium begin to rearrange (Strange *et al.,* 1992). The

gland eventually regresses until only a highly branched ductal system and a few alveoli remain. At this stage, a thin layer of epithelial cells is surrounded by myoepithelial cells, which rest on the basement membrane, and a dense layer of stroma envelops the ducts. The gland is very similar in appearance to the adult virgin mouse except that it is slightly more differentiated, as a few alveoli are present.

1.3 Lactogenesis

Lactogenesis is the term given to mammary differentiation and the processes involved to achieve lactation. This occurs in two stages. The first begins during mid pregnancy with the synthesis of milk components and can be distinguished morphologically by the presence of lipid droplets in the cytoplasm of the epithelial cells. The second stage of lactogenesis commences at parturition which is characterised by an increase in expression of milk proteins and movement of the lipid droplets and casein micelles from the cytoplasm to the lumen of the alveoli. Progesterone is no longer present during this stage and is replaced with prolactin and oxytocin which are stimulated by suckling.

The mammary gland has a highly organised secretory pathway during pregnancy and lactation. Most secreted milk proteins such as caseins are synthesised by the epithelial cells. They are synthesised on ribosomes of the rough endoplasmic reticulum and translocated to the Golgi by secretory vesicles. The contents of the secretory vesicles are released by their fusion to the apical plasma membrane (Franke *et al.,* 1976). The organisation of this pathway was defined many years ago (Saacke and Heald, 1974). Some milk proteins are derived from blood plasma such as immunoglobulins and the iron carrier transferrin. These enter the epithelium by endocytosis at the basolateral surface and are subsequently transported to the apical surface by transcytosis (Rothman, 1994). Much of the information about the secretory mechanisms and transport processes involved in the regulation of milk secretion in the mammary gland has been acquired from knowledge of these pathways in other cell types (Rothman, 1994).

1.4 Hormones and growth factors control mammary gland development

The development of the mammary gland is controlled by a variety of reproductive hormones including oestrogen, progesterone, prolactin and oxytocin, and growth factors such as epidermal growth factor and transforming growth factor-α (Topper and Freeman, 1980). Genetic manipulation of the mouse using transgenic techniques and knockouts has provided the opportunity to investigate the molecular and cellular mechanisms of normal mammary gland development.

Ovarian hormones are responsible for the ductal outgrowth which takes place postpuberty, and they also play a role in alveolar proliferation. However, many other complex hormonal growth factor and ECM interactions are involved. Deletion of the oestrogen receptor gene using knockout mice confirmed that oestrogen was required for ductal morphogenesis (Korach *et al.,* 1996). Oestrogen receptors are present on mammary stromal fibroblasts and adipocytes, and their expression is modulated during development (Fendrick *et al.,* 1998). Cunha and colleagues demonstrated that the mammary stroma was the mediator of oestrogen-stimulated growth in the mouse (Cunha *et al.,* 1997). This was achieved by using wildtype and oestrogen receptor negative stromal and epithelial tissue recombinants. It is thought that after its role in ductal morphogenesis oestrogen causes the induction of progesterone receptors in the luminal epithelial cells (Hovey *et al.,* 2002). Oestrogen is present throughout pregnancy with rising levels towards late pregnancy (Mizoguchi *et al.,* 1997).

Progesterone plays a definitive role in the formation of ductal side branches and levels begin to rise during early pregnancy to contribute to the development of lobuloalveolar structures (Atwood *et al.,* 2000; Ichinose and Nandi, 1966). The signalling of progesterone is mediated through stromal and epithelial progesterone receptors (Humphreys *et al.,* 1997).

Prolactin signalling is important for the proliferation and differentiation of the lobuloalveolar structures during pregnancy (Topper and Freeman, 1980). It induces the formation of alveoli that secrete milk products, including casein. Maintenance of the alveoli requires the hormones oestrogen, progesterone and growth hormone. Other circulating hormones such as cortisol and insulin, and local factors also contribute to the development of the mammary gland.

The development of transgenic and knockout mice has enabled scientists to explore the role of growth factors and determine how they affect the development of the mammary gland. Growth factors are able to regulate cell behaviour by ECM remodelling. Growth hormone (GH) has been shown to be as essential as oestrogen in mammary gland development. GH acts through the GH receptor in the stromal cells of the mammary gland (Feldman *et al.,* 1993; Walden *et al.,* 1998, Kleinberg *et al.,* 1990). The function of GH is mediated by insulin-like growth factor I to induce TEB formation and ductal morphogenesis during mammary gland development (Ruan and Kleinberg, 1999). GH

receptors have also been detected in the epithelium, and therefore GH may have a role in differentiation and/or function in the luminal epithelial cells (Ilkbahar *et al.,* 1999).

1.5 Functional changes of mouse mammary gland development

The mammary gland is a complex tissue consisting of epithelial cells which have adhesive interactions with the ECM. The ECM provides an essential physical support for most cells, but its direct interaction with them is essential for maintaining normal tissue homeostasis and function. During pregnancy, lactation and involution, interactions between the mammary epithelial cells and the surrounding ECM contribute to the signals needed for their proliferation, differentiation and survival. Therefore these mesenchymal-epithelial interactions play an important role in development and tissue remodelling. Stromal cells appear to influence epithelial cells by growth hormones and/or by altering the ECM. Regulation during pregnancy is dependent on the hormones oestrogen and progesterone and in turn these hormones are maintained by epithelial-stromal interactions (Haslam, 1986; Haslam and Counterman, 1991; McGrath, 1983).

The appropriate multicellular environment is needed for mammary epithelial cells to function normally. If these cells are isolated and not cultured under the correct conditions, the cells do not differentiate and will undergo apoptosis. Reconstituted basement membrane derived from Englebreth-Holm-Swarm tumour (Matrigel) has been used extensively to investigate the role of the ECM in the function of epithelial cells. The importance of the ECM for mammary differentiation was demonstrated using this system. Mammary epithelial cells rapidly lost their ability to secrete most milk proteins when cultured on plastic. However, when cultured on Matrigel, the mRNA levels of milk proteins were greatly enhanced (Li *et al.,* 1987). Further evidence for the importance of the basement membrane in mammary development was found when isolated mammary cells, in the presence of reconstituted basement membrane, were able to form alveoli. These cells expressed β-casein in the presence of basement membrane without collagen I (Streuli *et al.,* 1991). Epithelial cells require cell-cell adhesion systems for polarity and interaction with other cells. The basement membrane also provides polarity to the epithelial cells.

The ECM is essential for the control of mammary epithelial survival (Streuli and Gilmore, 1999). Stromal cells produce matrix proteins such as collagen I, laminin and fibronectin. These matrix proteins are able to interact with epithelial cells by cell surface receptors including integrins and initiate signal transduction pathways. Cells primarily adhere to the underlying ECM through integrins, which are transmembrane glycoproteins (Hynes, 1992).

Matrix proteins, their receptors and integrins are important for alveolar development during pregnancy, and differentiation of epithelial cells during lactation (Streuli *et al.*, 1995). Although the influence of the ECM on proliferating epithelial cells is not well understood, epithelial-stromal interactions play a major role in branching morphogenesis by enabling TEBs to invade stromal tissue (Daniel *et al.*, 1984).

Detachment of the epithelial cells from the ECM leads to apoptosis, whereas integrin mediated attachment prevents apoptosis from occurring. Cells have specific requirements for the type of ECM that promotes survival, since mammary epithelial cells undergo apoptosis on collagen 1 or fibronectin basement membranes (Boudreau *et al.*, 1995) but are able to survive on a laminin-rich basement membrane (Pullan *et al.*, 1996). These phenotypic effects are partly due to the integrin cell surface receptors, because blocking the β1 integrin subunit induces apoptosis (Farrelly *et al.*, 1999). Different integrin subunits have different abilities to suppress apoptosis and therefore adhesion to a particular extracellular matrix does not necessarily mean signalling survival. The role of integrins in cell survival is achieved by functioning as a linker between the extracellular matrix and the actin-containing cytoskeleton. They provide a mechanism whereby they recruit scaffolding proteins that organise cytoskeletal proteins and signalling molecules.

The structural link of integrins with the actin cytoskeleton regulates cell shape and provides a means for migration. There are many different responses to adhesion which are important for growth and survival. For example activation of the actin cytoskeleton, signalling through enzyme pathways such as protein tyrosine kinases, and cell shape promoting cell survival (Miyamoto *et al.*, 1996; Plopper *et al.*, 1995). Transmembrane proteins such as integrins serve as adhesion molecules, have cytoskeletal links and are important in signalling pathways. These are factors that are critical in determining the fate of a cell.

MMPs are a family of secreted proteases and their expression pattern suggests an important role in the morphological and functional changes of the mammary gland. The activity of MMPs is evident in branching morphogenesis and involution. Ongoing investigations examining MMPs in branching morphogenesis are determining which mediate branching elongation and which mediate lateral branching (Wiseman *et al.*, 2003). The expression of MMPs increases dramatically during involution coinciding with the major tissue remodelling phase of this stage of development (Lefebvre *et al.*, 1992; Talhouk *et al.*, 1992; Li *et al.*, 1994; Lund *et al.*, 1996). MMPs can degrade various

proteinace components of the ECM and are responsible for remodelling of the basement membrane and stroma to allow the gland to return to its resting adult state (Werb, 1997).

The ECM may be focal to understanding how different cell types communicate with each other. Mammary epithelial cells are not only regulated by hormones and growth factors but also by integrin-ECM connections. Cell adhesion is essential for tissue organisation and maintaining normal homeostasis of mammary epithelial cells.

1.6 Proteomics and two-dimensional (2-D) gel electrophoresis

As proteomics was used as a technique to determine proteins involved in mammary gland development, an overview of this technology is given below.

1.6.1 Historical overview of 2-D gel electrophoretic proteomics

The term "proteome" was coined by Wilkins (Wilkins *et al.,* 1996b) to describe global patterns of gene expression at the protein level. Proteomics is the experimental tool used to study the proteome, which uses a number of complex techniques for resolving, quantitating and characterising proteins. The concept of proteomics was introduced scientifically more than 15 years earlier during the evolution of 2-D gel electrophoresis (O'Farrell, 1975; Klose, 1975).

The combination of the separation of proteins by charge (isoelectric point (pI)) and molecular weight (Mr) created the 2-D gel profile of the protein sample. The analysis of a proteome by 2-D gel electrophoresis resolves thousands of proteins. The proteomes produced from the different samples analysed are compared by the protein feature (spot) patterns produced. The expression levels of each protein in each proteome can be determined and used as a method of assessing the importance of the protein to the biological system being manipulated. Candidate protein features that are of importance can be excised from the gel, enzymatically digested and subsequently identified by mass spectrometry. The number of proteins that are analysed by mass spectrometry are always limited due to cost and time constraints.

Before modern DNA techniques were introduced, 2-D gels seemed to be the only approach suitable for determining the cellular function. The importance of proteomics is its ability to study post-translational modifications (PTMs) within a cell. Phosphorylation, methylation, acetylation and processing of proteins as well as many other modifications are extremely

important for protein function as they can determine activity, stability, localisation and turnover. Many signalling pathways are also known to mediate their effects through PTMs.

In the late 1970s, researchers started to build protein databases and catalogue the protein expression data produced from 2-D gel electrophoresis. However, the constant drawback of this technique was at the level of identification of the proteins due to the lack of sensitive and rapid analytical methods present at the current time. 2-D gel protein databases such as that created by Celis and colleagues (Celis *et al.,* 1990) have since offered global approaches to study the function of groups of proteins.

The most significant breakthrough for proteomics in the 1990s was the identification of proteins separated by 2-D gel electrophoresis using mass spectrometry. Mass spectrometry has essentially replaced the classical technique of Edman sequencing (Edman and Begg, 1967) which chemically degraded a protein or peptide from its amino (N)-terminus and then identified the released amino acids (Hewick *et al.,* 1981). The powerful technique of highly sensitive mass spectrometry has enabled the rapid identification, characterisation and sequencing of femtomolar levels of proteins extracted from 2-D gels. Algorithms are now used to achieve the rapid identification of a protein by correlating the mass spectrometry data to sequence databases. The success of earlier studies (prior to the release of the human and mouse genomes) in identifying a protein via mass spectrometry was heavily dependent on the accessibility to sequence databases, both public and private. Public access to all genome sequencing projects has enabled proteomics to advance in many research disciplines.

1.6.2 Proteomics versus genomics

The advances of molecular genetic techniques have led to the high-throughput sequence analysis of many free-living organisms including the mouse genome which have made a crucial contribution to understanding the biology of whole organisms (Waterston *et al.,* 2002). Researchers have since realised that the sequencing of genomes is not sufficient for determining the biological function of a cell. It is proteins that are most commonly the functional molecules of a cell, and where most regulatory processes take place. Therefore, they are most likely to reveal a gene's true clinical and scientific significance. For this reason, it is proteins that are directly targeted for drug development. Proteins which are specific to a particular disease can be identified by quantifying the abundance of proteins present in diseased and normal samples. Proteomics has therefore been of great benefit in the field of research and drug discovery, as it can quantitatively identify specific proteins

which are related to a particular disease. For of these reasons there is now much more focus placed on proteomics.

The assignment of a protein function will require analysis of expression, localisation and structure of the proteins encoded by genomes. Information about post-translational control and modifications of a protein can rarely be determined from the detail of its gene sequence, as not all structural modifications and quantitative changes are controlled by the same gene. It is apparent that a large number of proteins undergo PTMs. Whilst the genome's content does not change, the proteome is continually being modified. Proteins are constantly changing their levels of expression according to the physiological and developmental state of the cell. Proteomics addresses areas that DNA analysis cannot approach such as synthesis rates, expressions levels, protein-protein interactions and PTMs. Therefore proteomics is needed to supplement the genomic data available by studying the patterns of protein expression (Dove, 1999; Pennington *et al.,* 1997; Wilkins *et al.,* 1995c).

The application of DNA chips and microarrays have had a profound effect on the investigations of gene function. They provide high throughput data of mRNA expression in a cell and tissue at a given moment. The advances of microarray technology have enabled thousands of differentially expressed genes to be analysed in parallel. After identifying the genes which have interesting expression changes, search tools and electronic information are readily available to decipher their implications and function with regards to the system being investigated. Microarray profiling is also less time consuming and labour intensive compared to proteomics and the past few years have produced a flurry of studies using this technology in a variety of biological systems. However, it is well known that gene expression at the mRNA level does not always correlate with protein expression levels which is also true for the reverse, as regulation occurs at the transcriptional and translational level (Gygi *et al.,* 1999b; Anderson and Seilhamer, 1997). mRNA expression data must be subsequently validated with protein expression profiles. Despite this, proteomics and microarray analyses are complementary as they both investigate the molecular organisation of a cell, one at a protein level and the other at the gene level. The data from each technique can enhance the effectiveness of the other. In combination, the data generated may lead to a better understanding of the regulatory processes involved in normal and diseased tissues.

One of the main challenges facing these approaches is their analyses, interpretation and ability to deal with the large amounts of data produced. Thus there is a need to develop simple and rapid methods of validating this information.

1.6.3 Sample preparation

A reproducible sample preparation is crucial in proteomic analyses. Reproducibility can also be maximised during 2-D gel electrophoresis by running sample sets simultaneously. Until recently 2-D gel electrophoresis technology relied on the use of carrier ampholytes to generate a pH gradient. However, this method failed to generate reproducible high resolution protein gels. This can be attributed to the poor reproducibility in generating the carrier ampholytes and their drift during focusing towards the cathode. The introduction of immobilised pH gradients (IPGs) overcame the problems of reproducibility and introduced the ability to select the pH range of interest. Wide range gradients are useful for initial analysis of a sample but may only represent a small proportion of the proteome sample due to inadequate spatial resolution. Selecting a smaller pH range can improve protein separation and resolution and prevent protein features from overlapping. The gels which cover a narrow pH range are known as 'zoom gels' (Gorg *et al.,* 2000). IPGs are pH gradients that are covalently grafted by vinyl bonds to a polyacrylamide-supporting matrix (Bjellqvist *et al.,* 1982; Gorg *et al.,* 1988; 2000). A general guide for focusing these IPG strips is to commence with a lower voltage to enable the sample to be absorbed into the gel and then the voltage is progressively increased to a maximum. Focusing times vary; however, shorter focusing times are favoured for better resolution and prevention of basic end streaking. If focusing times are too short they will result in both horizontal and vertical streaking. Prior to 2-D electrophoresis the IPG strips are equilibrated to allow the focused proteins to interact with sodium dodecyl sulphate (SDS) for migration purposes. In most 2-D gel proteomic studies, IPG strips are the method of choice for first dimensional separation of the protein sample.

The mammary gland is a particularly heterogeneous tissue and this contributes to the complexity of the sample preparation. In order to reduce the sample complexity for 2-D gel analysis and thus prevent loss of particular types of proteins, sample preparation techniques can be refined. One approach which is particularly favoured, is prefractionation preparations of samples, as they are able to enrich low abundant proteins (Pasquali *et al.,* 1999). Subcellular fractionation separates organelles based on their physical properties and is a method which firstly disrupts the cellular organisation by homogenisation. Secondly, the homogenate is fractionated into the different populations of organelles by

centrifugation (e.g. nuclei, heavy mitochondria, cytoskeletal components and plasma membrane; light mitochondria, lysosomes, and peroxisomes; and cytosol) (Pasquali *et al.*, 1999). However, there are disadvantages to this approach as fractionation disrupts important cell to cell interactions in the mammary gland and the results may not give an accurate representation of the processes involved in mammary gland development.

Efforts have been made to remove proteins from sample preparations which are known to obscure many areas of 2-D gels such as serum albumin and immunoglobulins. However there is the consideration that this may also remove other proteins of potential interest (Dunn, 1987; Lollo *et al.*, 1999).

Protein solubility is of major importance in proteomics in terms of reproducibility and in obtaining a representative sample preparation of the proteome investigated. One particular challenge is the solubility of membrane proteins. The degree of recovery of membrane proteins is variable due to solubility (Santoni *et al.*, 2000). A combination of CHAPS (3[(cholamidopropyl)dimethylammonio]-1-propane sulphonate), thiourea and urea has the greatest effect for solubilising these particular types of proteins (Perdew *et al.*, 1983; Molley *et al.*, 1998; Rabilloud *et al.*, 1997). SDS is very effective at solubilising proteins, although one drawback to using it is its anionic effect. This effect can be neutralised by the addition of zwitterionic detergents (such as CHAPS) to the lysis buffer used.

1.6.4 Visualisation and detection of protein features

The absence of a technique equivalent to PCR to amplify small amounts of proteins dictates that proteomics is limited to proteins that are detectable above a certain level of expression. The identification of low abundance proteins is an important area to research, as these proteins may consist of receptors, signal transduction and regulatory proteins.

Improvements have been made in detecting low abundant proteins by the introduction of fluorescent staining methods such as the OGT stain (Oxford GlycoSciences) (molecule on the basis of Hassner *et al.*, 1984) which can detect proteins to the femtomolar level. This particular fluorescent stain has successfully been used in several breast studies (Page *et al.*, 1999). It was used in an extensive human breast cancer cell line study to detect changes between breast cancer cell lines and cell lines derived from normal or benign disease tissue (Harris *et al.*, 2002). Fluorescent stains provide a greater sensitivity, and a larger dynamic range compared to organic dyes such as Coomassie blue which is only capable of detecting between 30-100 ng of protein (Brush, 1998). Silver staining can also be used and is more

sensitive than Coomassie blue. Despite the numerous methods used for silver staining, none combines convenience, sensitivity and speed (Rabilloud, 1992; Rabilloud, 1999).

1.6.5 Image processing

At the start of the 1980s software was created for the analysis of 2-D gel imaging (Anderson *et al.,* 1981). Evaluating the reproducibility of protein features in a data set and analysing the results cannot be performed without computer image analysis software. Firstly to use the computer based processing programs the 2-D gel images are converted into digital copies. Early image processing packages such as Elsie, Tycho, Melanie and Quest were originally created in academia and these have now been dramatically improved to the commercially available program Melanie III (GeneBio, Geneva, Switzerland). Melanie III is able to determine noise and artefacts so that they are not included during subsequent analyses. These programs perform operations such as detecting and quantifying a protein feature from a gel image, pairing features between duplicate gels and creating synthetic master gels. Computer imaging programs are able to determine the positions of the proteins in terms of pI and Mr and the shape of the protein for measuring abundance. Pairing protein features correctly is a crucial process in proteomics for successful analyses of the proteomes. Once this is completed the database can be interrogated to identify proteins which appear to have biologically interesting expression profiles with respect to the system being investigated. Once again the success of the image processing algorithms is reflected in the similarity of the gel replicates. Consistency is needed at the protein separation stage and during data formatting. The next step is to give the selected protein an identification using mass spectrometry.

1.6.6 Mass spectrometry

Mass spectrometry has increasingly been the most popular choice of analysis for the identification and characterisation of proteins and has led to incremental advances in standard proteome technology. Its ability to provide information on PTMs by determining the mass shifts of peptide fragments has also drawn attention to its use. Mass spectrometers consist of an ion source, a mass analyser for measuring the mass to charge ratio (m/z) of the ionised peptides and a detector for determining the number of ions at each m/z. Tandem mass spectrometry and matrix-assisted laser desorption ionisation (MALDI) are the most commonly used techniques to ionise proteins or peptides for mass spectrometry analysis. MALDI-time of flight (MALDI-TOF) has been used in numerous high-throughput applications (Karas and Hillenkamp, 1988; Yates, 2000; Patterson and Aebersold, 1995). MALDI measures the mass of a peptide and is normally the chosen mass

spectrometry technique used to analyse relatively simple peptide samples. MALDI ionises the sample on a dry crystalline matrix by applying short laser pulses, and mass separation of ions is based on their time of flight (Figure 1.5). Tandem mass spectrometry such as the quadrupole time-of-flight (Q-TOF) is a more sensitive application, detecting proteins at the femtomole level and it is preferred for partial peptide sequencing (Fenn *et al.,* 1989; Loboda *et al.,* 2000). Peptides are initially separated by liquid chromatography (LC) which is coupled to the mass spectrometer. The peptides are in solution and are subsequently ionised by electrospray ionisation; the resolved peptides are dissociated into either being a carboxy (C)- or N- terminal fragment. The peptide fragmentation ion spectra generated can be used to propagate sequence databases for a protein identification (Figure 1.6). The advantage of this technique is that peptide sequence information is much more specific for determining a protein identification than the peptide masses that are produced by MALDI analysis. It is also the preferred method for analysing complex protein samples. The advantages of MALDI-TOF are its simplicity, mass accuracy, high resolution and sensitivity.

Microfluidic technology has received much attention over the past years which is partly due to genomics and proteomics. The combination of this system to mass spectrometry has enabled direct, high throughput mass spectrometric analysis of picomole (pmol) amounts of peptides by electrospray ionisation mass spectrometry procedures (Lion *et al.,* 2003).

Despite the advances made with mass spectrometry there are still problems that are associated with its use in proteomics. A large amount of tandem mass spectrometry spectra data is due to noise or contaminants rather than peptides from the sample. A considerable amount of time and expense will already have been spent to get from protein separation to protein selection. The main delays of these stages are at the bioinformatic level. Even the final stage of identifying the protein by mass spectrometry is fraught with spending an enormous amount of time on interpreting the data which may yield no protein identification. The use of algorithms which are able to match spectra data to sequence have been an important development to proteomics, and these are continually being refined and improved. However, yet again this has its drawbacks because the results produced are based on significance scores which can introduce false positives. Manual analysis of the spectra data is able to increase confidence in a peptide sequence match.

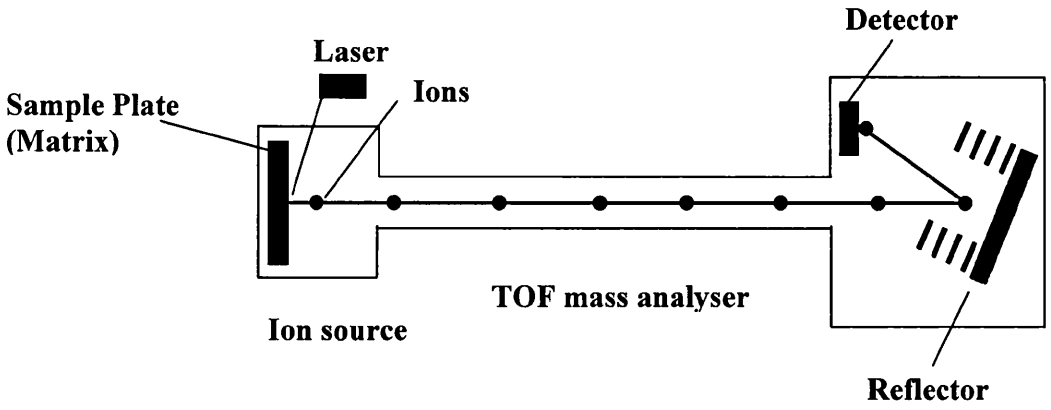


Figure 1.5 MALDI time-of-flight mass spectrometer. Samples are ionised out of a crystalline matrix by a laser beam. The ions are accelerated to high kinetic energy and are separated down a flight tube, this is known as time-of-flight (TOF). Ions are turned around by a reflector to compensate for slight differences in kinetic energy. The mass-to-charge (m/z) ratio is determined according to the time taken for an ion to reach the detector. Lighter ions will arrive first.

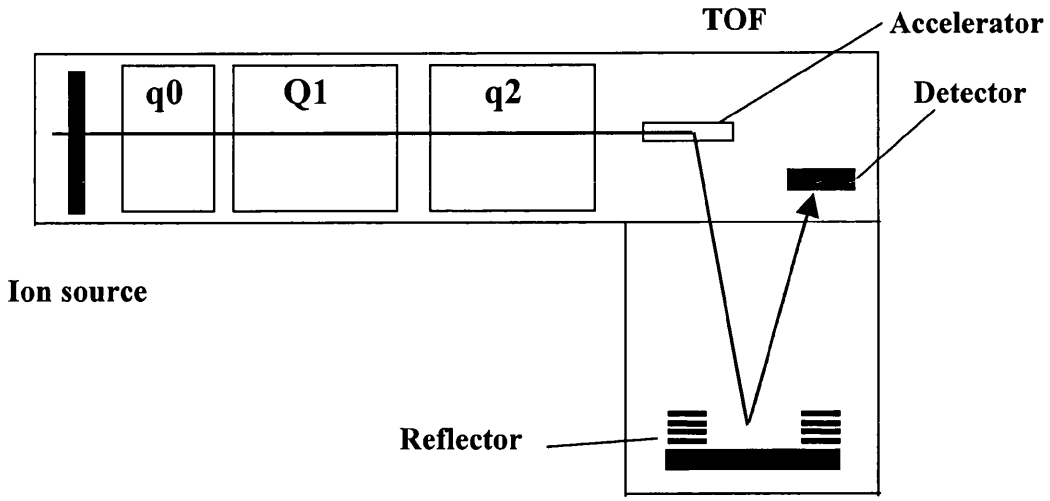


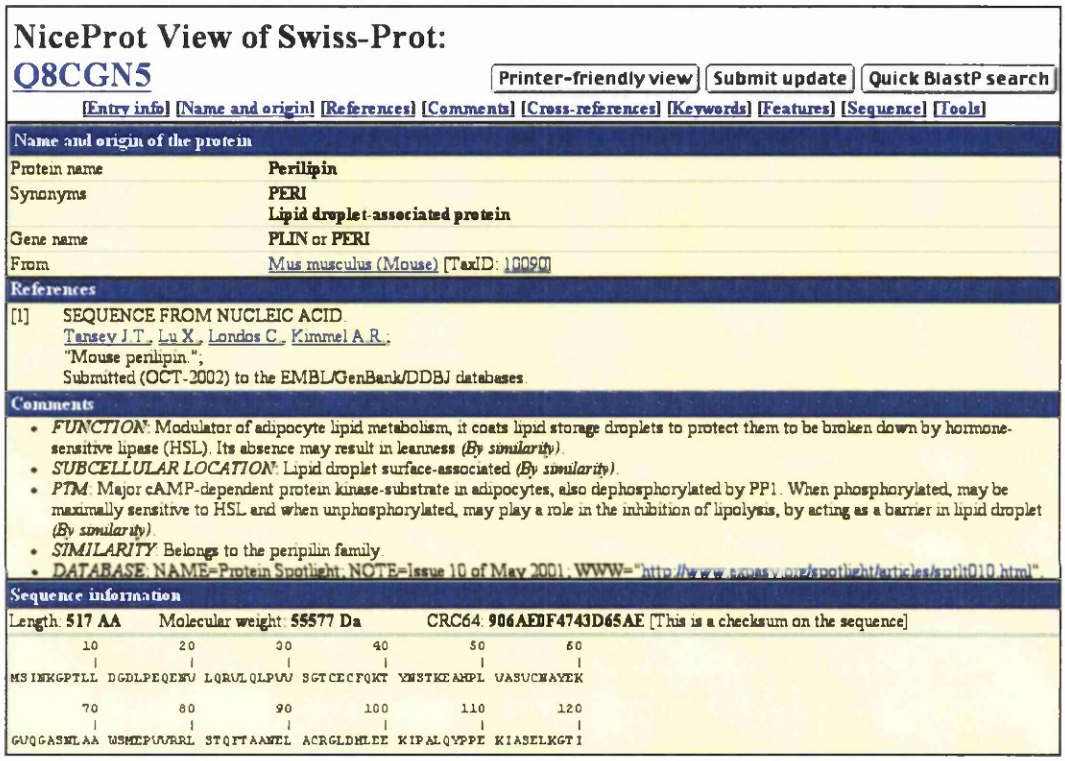
Figure 1.6 Quadrupole time-of-flight mass spectrometer. Samples are most commonly ionised by electrospray ionisation for QTOF mass spectrometry. The quadrupole is a mass filter which selects by time-varying electric fields between four rods. The ions enter a vaccum and are focused in the first quadrupole section (q0). Ions are mass separated in Q1 and then fragmented in a collision cell (q2). The mass of the ions are measure by a time-of-flight analyser.

As the field of proteomics is continuing to expand these problems are more likely to be addressed, algorithms will be refined and the number of hits by mass spectrometry increased due to the availability of most species' genome sequences.

1.6.7 Bioinformatics

Bioinformatics is required at many levels of 2-D gel proteomics, firstly at the 2-D image analysis stage and secondly at mass spectrometry stage. The advances in hardware and software development have made it possible to correlate mass spectrometric data with sequence databases. More sophisticated and accessible database searches with integrated algorithms have led to rapid protein identifications. Sequest (Eng *et al.,* 1994) and Mascot from Matrix Science (Perkins *et al.,* 1999) are algorithm programs that are used to correlate fragment mass spectral data of peptides with amino acid sequences in a protein database. The protein sequence databases used are *in silico* digested according to the specificity of the enzyme used in mass spectrometry (for example trypsin) to generate theoretical peptides. Sequest performs a cross-correlation analysis which scores these peptides to determine the best match to the experimental data obtained. Mascot uses probability based scoring which determines the probability of a match between the experimental data and theoretical fragmentation information being a random event. Sequest does not allow searching on the internet whereas the newer program Mascot is Web accessible.

To date the most comprehensive database used in proteomics is ExPASy (Appel *et al.,* 1994; Bairoch and Apweiler, 1997; http://www.expasy.ch/). ExPASy hosts databases such as SWISS-PROT (Figure 1.7) and TrEMBLE, SWISS-2DPAGE and PROSITE and access to these databases is free to academic institutions. SWISS-PROT is maintained at the Swiss Institute for Bioinformatics in Geneva, Switzerland and the European Bioinformatics Institute. This internet accessible proteome database contains protein sequences from different species and includes information such as protein expression, function, structural domains, PTMs and citation information. The database also cross-links to other well known on-line sources of information such as European Molecular Biology Laboratories (EMBL) (www.ebi.ac.uk/embl) and Ensembl (www.ensembl.org). Information is provided on theoretical Mr and pI values, but these must be used with caution because if the protein has been modified post-translationally it can alter these values. ExPASy provides access to a number of databases and analytic tools and is continually being refined to support the advances made in proteomics.



NiceProt View of Swiss-Prot:

Q8CGN5

Printer-friendly view | Submit update | Quick BlastP search

[Entry info] [Name and origin] [References] [Comments] [Cross-references] [Keywords] [Features] [Sequence] [Tools]

Name and origin of the protein

Protein name	**Perilipin**
Synonyms	**PERI** **Lipid droplet-associated protein**
Gene name	**PLIN** or **PERI**
From	Mus musculus (Mouse) [TaxID: 10090]

References

[1] SEQUENCE FROM NUCLEIC ACID.
Tansey J.T., Lu X., Londos C., Kimmel A.R.;
"Mouse perilipin.";
Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.

Comments

- *FUNCTION*: Modulator of adipocyte lipid metabolism, it coats lipid storage droplets to protect them to be broken down by hormone-sensitive lipase (HSL). Its absence may result in leanness *(By similarity)*.
- *SUBCELLULAR LOCATION*: Lipid droplet surface-associated *(By similarity)*.
- *PTM*: Major cAMP-dependent protein kinase-substrate in adipocytes, also dephosphorylated by PP1. When phosphorylated, may be maximally sensitive to HSL and when unphosphorylated, may play a role in the inhibition of lipolysis, by acting as a barrier in lipid droplet *(By similarity)*.
- *SIMILARITY*: Belongs to the perilipin family.
- *DATABASE*: NAME=Protein Spotlight; NOTE=Issue 10 of May 2001; WWW="http://www.expasy.org/spotlight/articles/spotlt010.html".

Sequence information

Length: **517 AA** Molecular weight: **55577 Da** CRC64: **906AE0F4743D65AE** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
         |          |          |          |          |          |
MSINKGPTLL DGDLPEQEVV LQRVLQLPVV SGTCECFQKT YNSTKEAMPL VASVCNAYEK

        70         80         90        100        110        120
         |          |          |          |          |          |
GVQGASMLAA WSMEPVVRRL STQFTAAMEL ACRGLDMLEE KIPALQYPPE KIASELKGTI
```

Figure 1.7 SWISS-PROT entry for Perilipin. This image represents a proportion of data that is provided by this proteome database. Information on the function, location, PTM and sequence of the protein is shown including appropriate references.

Although proteomics has increased in popularity there are areas in this field which are less popular, such as protein isoform changes and modifications as they are difficult to characterise. The use of databases such as SWISS-PROT can help in this area as they provide enormous amounts of information on protein processing and modifications. Programs have been created that can search for mass differences that correspond to a particular modification. The peptide sequence and mass data are entered into the system and sites where potential modifications could occur are determined (Wilkins *et al.,* 1999). The number of different proteins expressed by a genome is probably vastly underestimated as genomic information does not predict all PTMs.

Even though there has been much progress made with proteomics there is still the need to develop software tools that can identify the most likely identification based on protein and nucleic acid sequence of an unidentified peptide.

1.6.8 Alternative methods to 2-D gel electrophoresis

Alternative methods used to separate proteins for proteomic analysis include protein chip assays, direct analysis by mass spectrometry (Link *et al.,* 1999) and affinity tag techniques (Gygi *et al.,* 1999a). It is possible with advancing technology to apply microarray methods to proteomics in the form of protein biochips or protein arrays. Protein arrays enable thousands of proteins to be interrogated simultaneously with minimum sample consumption. These are more powerful than microarrays as they can analyse protein function, modification and regulation on a large scale (Zhu and Snyder, 2001; Zhu *et al.,* 2001; MacBeath and Schreiber 2000). This relatively new technology will be advantageous to the pharmaceutical industry due to its high throughput analyses. Depending on the configuration of the arrays they can measure protein expression levels and protein interactions. Antibody arrays were a starting point for proteomic microarray technology. However, a drawback with this technique was the potential lack of specificity (Haab *et al.,* 2001). One option which scientists are using to reduce this problem is antibody fragments (Borrebaeck *et al.,* 2001). A number of companies are developing antibody protein chip strategies. Another approach to protein array technology has been to attach proteins onto a glass slide which can be used to test for protein interactions (MacBeath and Schreiber 2000).

An alternative method to protein chip technology uses mass spectrometry. This technique has been termed SELDI (surface enhanced laser desorption/ionisation) and was developed by Ciphergen Biosystems. A MALDI target surface is modified with affinity ligands which

capture the proteins of interest and these are subsequently identified by MALDI-TOF mass spectrometry (Merchant and Weinberger, 2000). The arrays are available with different chromatographic properties which include hydrophobic and hydrophilic surfaces. These chips have also been designed to have pre-activated surfaces which are used for experiments such as antibody-antigen binding studies. Although this technology can provide much information about the proteins within a sample, it is not a quantitative method like 2-D gel electrophoresis.

Numerous alternative protein arrays will undoubtedly be created in the future and will accelerate the use of microarray technology to proteomics. Improving current methods of protein array fabrication, processing and analysis will also be of benefit to protein function microarrays.

Isotope-coded affinity tagging (ICAT) is a quantitative proteomic method which is being used as an alternative to 2-D gel separation (Figure 1.8) (Gygi *et al.*, 1999a). It has been used to study differential protein expression in biological systems. The ICAT reagent consists of an affinity tag (biotin) which isolates ICAT labelled peptides, a linker region for incorporating stable isotopes and finally a thiol specific reactive group (i.e. cysteines). The stable isotopes are generated using linkers composed of either eight deuterium atoms (heavy reagent) or eight hydrogen atoms (light reagent). One sample is derivatised with the heavy reagent and the other sample with the light reagent. The samples are combined, digested and the peptides produced are selectively isolated by avidin affinity chromatography to retrieve only the cysteine containing peptides. This reduces the complexity of the sample for subsequent analysis by mass spectrometry. Mass spectrometry provides information on quantitative protein levels by measuring the peak ratios of isotopes which differ by 8 Da, and a tandem mass spectrometry provides the protein identification. Despite its quantitative abilities and automated processes this method has drawbacks. For example proteins must have a cysteine residue due to the specificity of the reactive group.

Proteomics is often used for comparison of different protein profiles. However, a few of the drawbacks of 2-D gel proteomics are gel-to-gel variation and time spent on creating databases for comparing protein expression changes between different samples. Difference gel electrophoresis (DIGE) is a modification of the 2-D gel electrophoresis technique, as it enables samples to be run simultaneously on the same 2-D gel (Figure 1.9)

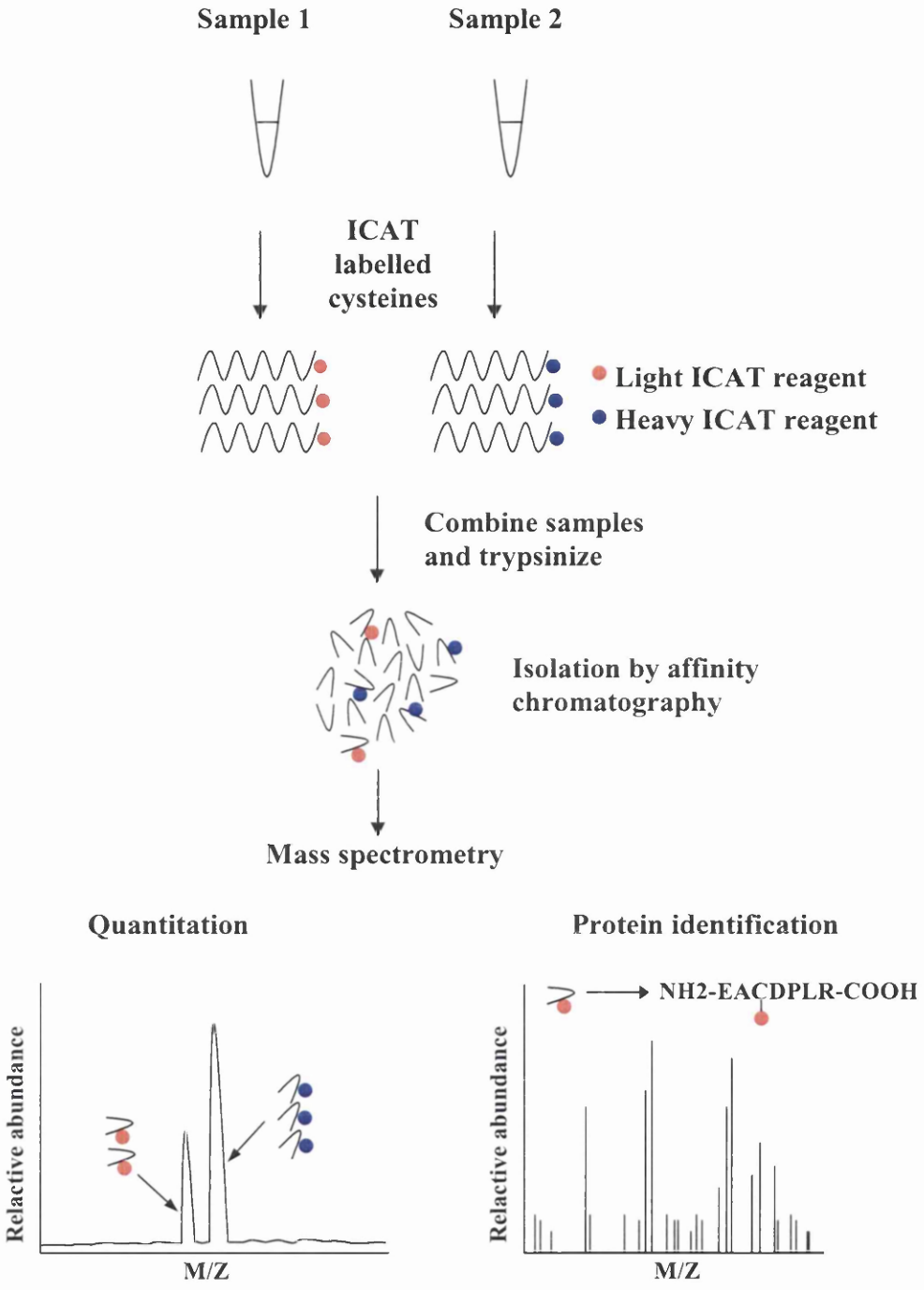


Figure 1.8 Isotope-coded affinity tagging. The analysis of two protein samples which are labelled with either ICAT heavy or light reagents. The samples are mixed and digested with trypsin and ICAT labelled peptides are isolated by affinity chromatography and analysed by a tandem mass spectrometer. M/Z (mass : charge ratio).

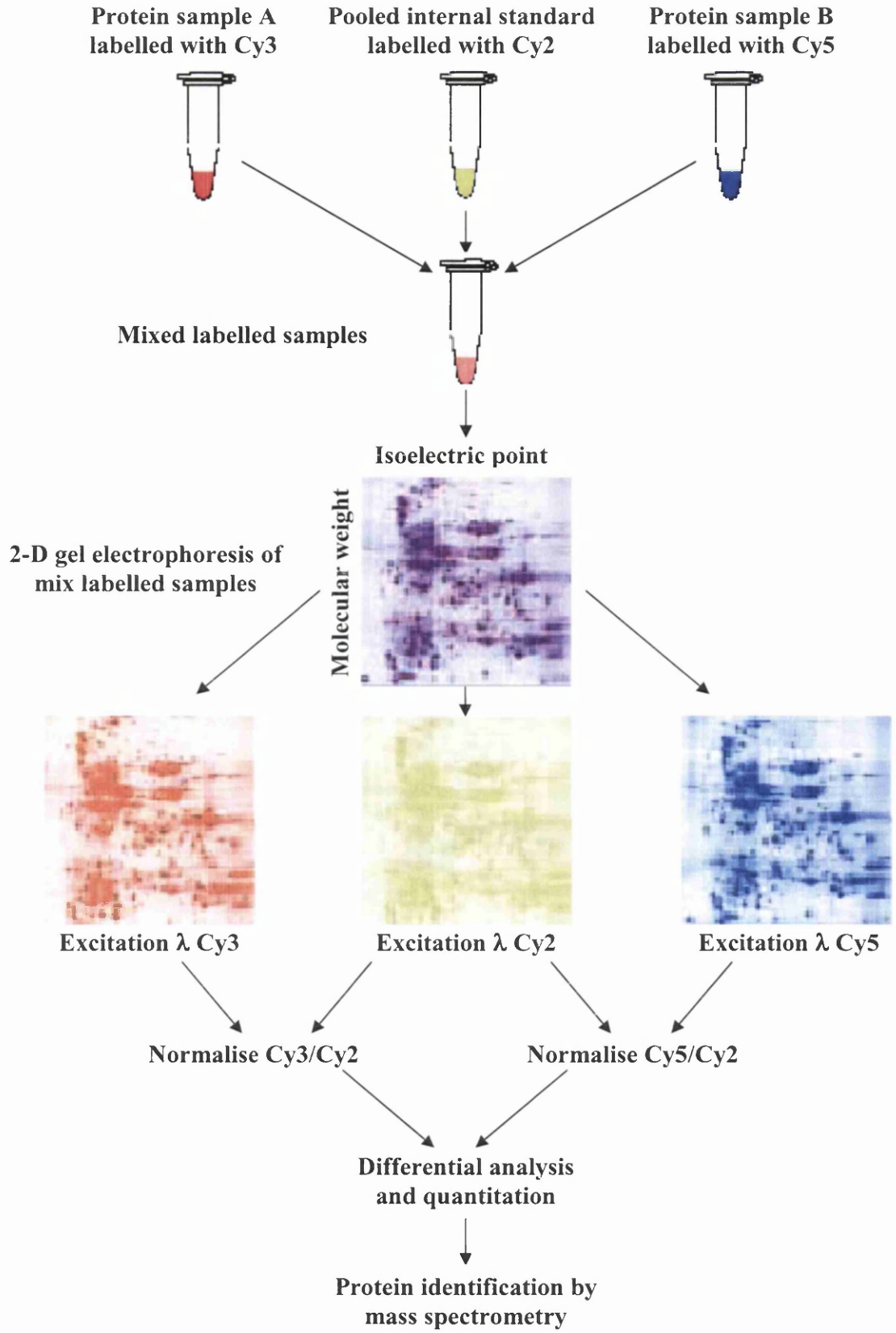


Figure 1.9 Outline of the 2-D DIGE technique. Samples A and B are labelled with a fluorescent dye (Cy3 or Cy5) and a pooled sample of A and B is labelled with Cy2 (internal standard). The samples labelled with Cy2, Cy3 and Cy5 are mix and separated on the same 2-D gel. The protein patterns of each sample can be visualised by illuminating the gel with specific excitation wavelengths. Samples A and B are subsequently normalised by the pooled internal standard sample. Differential expression analyses can be performed and the protein features of interest determined by mass spectrometry.

(Unlu *et al.*, 1997; Tonge *et al.*, 2001). This is made possible by fluorescently labelling protein samples with different cyanine dyes, Cy2, Cy3 and Cy5. A disadvantage of this technique is the limited number of samples that can be analysed. The samples are labelled with Cy3 or Cy5, and a pooled sample is made up of equal amounts of all the samples in the experiment and acts as an internal control (Cy2). The samples are applied to the same gel, separated simultaneously according to their isoelectric focusing point and subsequently by Mr using 2-D gel electrophoresis. The technique avoids gel-to-gel variation and enables protein expression values between different samples to be compared directly. Again there are problems with this methodology, such as sensitivity i.e. detecting low abundance proteins, and protein migration changes. As a result of labelling the proteins with fluorophores, they migrate at a slightly higher mass compared to those proteins that are not labelled. This leads to complications when protein features are excised for mass spectrometry (Unlu *et al.*, 1997). Despite the advantages that this technique offers, projects which have large sample comparisons render this technique unsuitable.

In the last few years progress has been made in developing alternative methods to separate proteins for proteomics such as the techniques mentioned. 2-D gel electrophoresis remains the leading choice for separating complex protein samples in most proteomic based projects, despite the limitations mentioned such as visualising low abundance proteins, hydrophobic and very basic proteins, and the reproducibility of gel replicate. The advantage of this methodology is its sensitivity, high resolution and ability to separate thousands of proteins from complex protein samples.

1.7 Identifying developmentally regulated proteins during mouse mammary gland development - Aims

At the beginning of this study it was recognised that proteomics and microarray technology could be used to identify genes that were associated with mammary gland development, specifically the major changes during pregnancy and at the lactation involution switch. The primary objective was to develop an appropriate extraction procedure to isolate high quality protein and RNA which was suitable for these techniques. Subsequently the second objective was to produce high quality data from triplicate samples using stages of mammary gland development. This project was to focus on the proteomics aspects, and other members of the research group would take forward the RNA profiling data. The correlation of these data would provide important information on comparisons of the two techniques and also assist with validation of the proteomics information.

A collaboration was formed which combined Oxford GlycoSciences expertise in proteomics and Professor Gusterson's in breast studies. After establishing techniques for producing reproducible 2-D gels from mammary tissue and TEBs, proteins would be identified which were associated to specific stages of mammary development. Having identified known and novel proteins, the next objective would be to validate the results. As the processes involved in breast development (invasion of connective tissue by epithelium, high proliferation rate and tissue remodelling) are similar to those that occur in breast cancer, it was predicted that genes regulating these processes would also be common to both. The creation of a mammary gland proteomic database could be used as a basis for breast cancer proteome studies. It was the intention that candidate proteins selected because of their interest to normal mammary gland development would also be examined in normal and cancerous breast tissue.

Chapter 2
Materials and Methods

2.1 Mouse mammary gland

2.1.1 Mouse husbandry

The mice used in this study were female Balb/C, and male Balb/C mice were used for mating (Charles River U.K.). Mice were cared for in accordance with the guidelines established by the Beatson Institute. They were kept at a constant temperature of 21 (+/- 2) °C with a 12 h light/dark cycles, housed in conventional cages bedded with aspen chip, and provided with food (SDS CRME breading diet) and tap water *ad libitum*.

2.1.2 Tissue collection

All mammary tissue was collected from the fourth inguinal gland of female Balb/C mice. The mice were culled by cervical dislocation, placed on a cork board ventral side up and restrained using metal pins. 70% (v/v) ethanol was applied liberally to the body. The mammary glands were then exposed by a ventral midline inverted Y incision midway between the fourth and fifth nipples and laterally down each hind limb. The abdominal skins, with the mammary glands attached were secured to the corkboard. The gland was removed by gripping the fat pad behind the nipple and peeling it from the skin using blunt ended scissors. The mammary gland was collected from the following developmental stages: i) 10 week old virgin, ii) pregnancy days 4.5, 8.5, 12.5, 14.5, 17.5 post coitum, iii) lactation days 1, 3, and 7 post partum and iv) involution, where the pups were removed after 7 days of lactation and the tissue collected on day 1, 2, 3, 4, 5, 10 and 20 post weaning. In the text and figure legends these stages are referred to as: Av10, P4.5, P8.5, P12.5, P14.5, P17.5, L1, L3, L7, Inv1, Inv2, Inv3, Inv4, Inv5, Inv10 and Inv20 respectively. Mice were mated in the evening, and the morning of the next day was counted as 0.5 days of pregnancy. Females were checked for the presence of a copulation plug in the vagina (vaginal plug). This consists of seminal coagulated proteins from the male seminal fluid and in most strains it can be identified easily.

2.1.3 Terminal end bud extraction

TEBs from the fourth inguinal mammary gland of Balb/C mice were isolated from animals aged 5-6 weeks and weighing between 16-18 g. The glands were removed from the mice and chilled with Liebovitz L-15 medium (Invitrogen). 12 glands were coarsely chopped with scalpel blades and digested for 30 min at 37°C in 10 ml of 1 mg/ml (w/v) collagenase type II (Sigma) in L-15 medium. To prevent further digestion after the specified time, chilled L-15 medium was added to the samples. The samples were centrifuged at 500 x *g* for 5 min at 4°C. The pelleted material was rinsed and resuspended in 1-2 ml chilled L-15 medium before transferring to a gridded contact dish (Nunc). TEBs were removed from the

other surrounding matter using a 10 µl micropipette (Gilson) and visualising the TEBs under a dissecting microscope. The samples were frozen at -80°C until a sufficient number had been collected for either protein or RNA extraction. 1110-1200 TEBs were collected for RNA extraction and 450-500 pooled TEBs for protein extraction.

2.1.4 Preparation of mouse mammary gland wholemounts

Mammary glands were prepared for wholemount analysis using a technique described by Edwards *et al.* (Edwards *et al.*, 1992). The excised glands were spread out onto clean, dry microscope slides and teased out flat, until they were in the orientation which is seen *in vivo*. The glands were fixed in Carnoys [25% (v/v) glacial acetic acid and 75% (v/v) absolute alcohol] for 2 h. The fixed glands were then washed for 1 h in 70% (v/v) ethanol and further rinsed for 5 min in dH_2O. The wholemounts were stained overnight with Carmine dye [2 mg/ml Carmine (Sigma), 5 mg/ml aluminium potassium sulphate (Sigma) and 0.1 mg/ml Thymol (Sigma)]. Following this the tissues were dehydrated in a series of increasing amounts of ethanol. 70% (v/v) ethanol for 45 min, 95% (v/v) ethanol for 30 min, and 100% (v/v) ethanol for 30 min (x2). The wholemounts were placed in Methyl Salicylate (Sigma) to remove fat from the gland and it was also used for storage purposes. To examine the wholemounts they were temporarily transferred to 100% (v/v) ethanol.

2.2 Protein extraction and quantification

2.2.1 Extraction of protein for 2-D gel electrophoresis

Protein was extracted from the fourth inguinal mammary gland. A section of tissue was taken rather than the whole gland, as a uniform sample was desired. A transverse section of the gland was cut approximately 3 mm in width, past the lymph node and distal from the nipple (Figure 2.1).

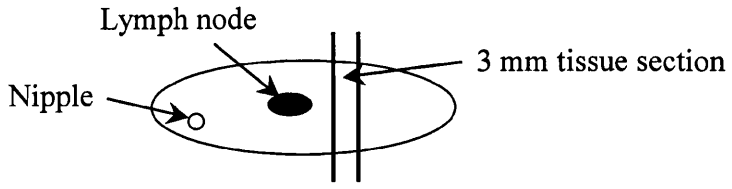


Figure 2.1 Schematic of a fourth right mouse mammary gland. The diagram indicates the area where a 3 mm section of tissue was taken for protein extraction.

The section of tissue was first washed in Dulbecco's Modified Eagle Medium (DMEM) for 5 min at room temperature (RT). Secondly the gland was 'milked of blood'. This term describes the technique used to expel blood contained within the mammary gland

vasculature. Digital pressure was applied to the gland to expel the blood followed by a 5 min wash in DMEM. To extract protein, the tissue was roughly cut with scissors in 2-D lysis buffer (see Appendix 1) and then homogenised with an Eppendorf pestle (Sigma) in a 1.5 ml tube. The resulting homogenate was incubated at 10°C for 10 min and then centrifuged at 20,000 x *g* for 5 min at 10°C to pellet any insoluble material. The supernatant (protein sample) was separated from the insoluble material, snap frozen and stored at –80°C. The concentrations of the protein samples were later adjusted to 0.5 µg/µl for running 2-D gels.

2.2.2 Quantitation of total protein in 2-D lysis buffer

To estimate the protein concentration of a protein sample in 2-D lysis buffer, the Pierce Coomassie Plus protein assay was used. It is based on the Coomassie dye changing colour on binding to protein. The assay is based on an Absorbance shift from 465 nm to 595 nm on protein binding. The Coomassie Plus reagent was equilibrated to RT before use. This assay was prepared in a 96-well plate. A range of standards was prepared prior to the assay from a stock of 2 mg/ml bovine serum albumin (BSA) provided. The standards had been diluted in 2-D lysis buffer to give final concentrations of 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 and 1.5 µg/µl of BSA in a total volume of 400 µl. 2-D lysis buffer was used as a blank. Various dilutions of the samples for estimation were prepared using 2-D lysis buffer as a diluent. 10 µl of each sample, standard and blank, were used in triplicate. To each well that was used 300 µl of Coomassie Plus reagent was added and then a cover was applied to the plate. The plate was shaken for 1 min at RT and then read at an absorbance of 600 nm. A standard curve was plotted in Excel using the BSA standard absorbance readings. A linear line of best fit calculated the protein concentration of the sample (Bradford, 1976).

2.2.3 Extraction of protein from 2-D to 1-D lysis buffer

Protein that had been extracted in 2-D lysis buffer was transferred into 1-D lysis buffer (2% (w/v) SDS and 63 mM Tris-HCl, pH 7.4) for Western blotting by acetone extaction. 6 volumes of cold acetone (-20°C) was added to the sample to precipitate the protein. This was stored overnight at-20°C and then centrifuged at 14,000 x *g* for 5 min at 4°C to pellet the protein. The sample was washed in 70% ethanol, centrifuged at 14,000 x *g* for 5 min at 4°C, air dried and resuspended in 1-D lysis buffer.

2.2.4 Quantitation of total protein in 1-D lysis buffer

To estimate the protein concentration of a protein sample in 1-D lysis buffer, the Pierce bicinchoninic acid (BCA) protein assay reagent kit was used (Smith *et al.,* 1985). In this

reaction Cu^{+2} is reduced to Cu^{+1} by protein in an alkaline medium (Biuret reaction). Detection is based on a purple complex formed between the cuprous ion and BCA with the absorbance maximised at 562 nm. This assay was prepared in a 96-well plate. A range of standards were prepared prior to the assay from a stock of 2 mg/ml BSA provided. The standards had been diluted in 1-D lysis buffer to give final concentrations of 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 and 1.5 µg/µl of BSA in a total volume of 400 µl. 1-D lysis buffer was used as a blank. Reagents A and B (provided in the kit) were mixed to 50:1 reagent A to B (BCA reagent) in a sufficient volume for 200 µl per sample and standard. Dilutions of the samples for estimation were prepared using 1-D lysis buffer as a diluent. 10 µl of each sample, standard and blank, were used in triplicate. To each well 200 µl of BCA reagent was added and then a cover was applied to the plate. The plate was shaken for 1 min at RT and subsequently incubated at 37°C for 30-60 min. The plate was allowed to cool to RT before reading the absorbance at 560 nm (Thermo Spectronic Biomate™ 3 Spectrophotometer). A standard curve was plotted in Excel using the BSA standard absorbance readings. A linear line of best fit calculated the protein concentration of the sample.

2.3 Oxford GlycoSciences 2-D gels

2.3.1 2-D gel preparation

Precautions were taken to minimise contamination of the mouse mammary protein samples to foreign material such as human keratin which if detected by mass spectrometry would give misleading results. Dissection and protein extraction of the mammary gland was performed while gowned and gloved. All 2-D preparations were performed in an area of controlled entry to minimise exposure to these contaminants. All operators in this area were gowned, gloved, masked and hair netted. The gels are cast, focused, electrophoretically separated, scanned, stored and cut for mass spectrometry under these conditions and in the controlled area dedicated to this technology.

2.3.2 Plate preparation

The 2-D gels were cast between a step plate and a plain plate. The step plate was wiped with absolute ethanol and then coated with 2% (w/v) dimethyldichloro silane in octamethylcyclotetrasiloxane (Rapel; BDH) to prevent the gel from adhering. The plate was air-dried, heated and rinsed with dH_2O and absolute ethanol. The plain plate was coated 3x with 0.4% (v/v) bind-silane (Amersham Pharmacia) in absolute ethanol to adhere the gel to the plate. One ml of bind-silane was added with each covering per plate, the plates were then air-dried for 1 h.

2.3.3 Casting 2-D gel

9-16% gradient acrylamide 2-D gels were made using:

Heavy acrylamide: 1.067 (w/v) Piperazine diacrylyl (PDA) and 40% acrylamide

Light acrylamide: heavy acrylamide 25% (v/v)

Light ammonium persulphate (APS): 1.2% (w/v) APS and 0.5% $Na_2(SO_3)_2$

Heavy APS: 3.2% (w/v) APS and 0.5% $Na_2(SO_3)_2$

Light TEMED: 0.6% (v/v) TEMED (N,N,N',N'-tetramethylethylenediamine)

Heavy TEMED: 1.4% (v/v) TEMED

After casting the gel a layer of butanol was added to the top of the gel to create a barrier between the air and the gel. After 5 h slab buffer (94 mM Tris-HCl pH 8.8) was added to the gel tank overnight. The gels were then stored in a cassette at 4°C until electrophoresis. The gels were 1 mm thick 20 x 20 cm width/height. Gels were stored for a maximum of one month before use.

2.3.4 2-D gel electrophoresis

Triplicate samples of the 16 stages of development investigated and duplicate samples of the TEBs were separated by 2-D gel electrophoresis. In the first dimension, IPG gels 18 cm long (Immobiline DryStrip 3-10 non linear (NL), Amersham Pharmacia Biotech) were brought to RT and rehydrated with 120 μg of solubilised sample and bromophenol blue. For gel rehydration the IPG gels were placed gel side up in solution, and to minimise evaporation of sample a layer of mineral oil was added. A 16 h overnight incubation at 20°C was required to rehydrate the IPG gels (Sanchez *et al.*, 1997). The IPG gels were placed in focusing trays in which damp electrode wicks had been placed over the anode and cathode. Electrode wicks must be blotted with tissue paper to remove excess liquid as it may cause streaking. The IPG gels were handled with forceps touching only the outside edges. The outer edges of the IPG strips were placed over the wicks gel side down, and covered with mineral oil to improve contact. The gels were focused in four phases: a ramping time of 10 min to reach 300 V; 300 V for 2 h; a gradient increase from 300 V-3500 V over 2 h, and then 3500 V for 18 h. Immediately after focusing, the IPG gels were equilibrated in 6 M urea, 2% (w/v) SDS, 2% (w/v) DTT, 50 mM Tris pH 6.8, and 30% (v/v) glycerol for 15 min at 20°C, in preparation for running the second dimension. A 9-16% acrylamide gel (2-D gel) was added to an electrophoresis tank to run the samples in the second dimension. To this a top buffer of 25 mM Tris base, 0.192 M glycine and 0.1% (w/v) SDS was added to improve the current. A 10x concentration of this solution was added to the rest of the tank. A pre-run of the 2-D gel was performed at low voltage

(10 mA) for 30 min to remove any unwanted acrylamide. A 2% dithiothreitol (DTT), 0.5% agarose and trace bromophenol blue solution was added on top of the 9-16% acrylamide gel to act as a stacking gel and focus the proteins. The IPG gels were placed in their allocated position on the top of the 2-D gels, and each was run at 10 mA for 10 min, followed by a further 6 h at 60 mA. Keeping the 2-D gels within the casting glass plates they were immediately fixed in 40% (v/v) ethanol:10% (v/v) acetic acid overnight for 16 h. The gels were washed for 30 min in 0.5% (w/v) SDS:7.5% (v/v) acetic acid to replace SDS removed during fixing. After washing, the gels were stained with a fluorescent stain OGT 1238 (0.06% OGT and 7.5% acetic acid) (Oxford GlycoSciences) (Hassner *et al.*, 1984). Sixteen-bit monochrome fluorescent images at 200 μm resolution were obtained by scanning the gels with an Apollo III linear fluorescence scanner (Oxford GlycoSciences). The intensity of the fluorescence signals was directly proportional to the abundance of each protein feature (spot) present.

2.4 2-D gel analysis

2.4.1 Analysis of gel images

The scanned images were processed with a custom version of Medical ELectrophoresis Analysis Interactive Expert group (MELANIE) III (GeneBio). The pI and Mr of each feature had been calculated by bilinear interpolation between landmark features on each image previously calibrated with respect to *Esherichia coli* (Swiss 2-D service ECO). Two different preparations were made using 1:1 and 2:1 ratio of sample: *E.coli* and standard loading conditions were used. Eleven landmarks proteins of *E.coli* were used and have been listed in table 2.1.

Individually resolved protein features were enumerated and quantified based on their fluorescence signal intensity. Intensity was measured by summing pixels within each feature boundary and recorded as a percentage of the total feature intensity of the image. The index which represented the pI, Mr and sample, and the percentage intensity values representing abundance (mean of replicate) for each gel feature were entered into a database (Rosetta^{TM}; Oxford GlycoSciences) to form a protein expression map (PEM). In order to compare PEMs from different samples, each protein was assigned a unique identifier number, the molecular cluster index (MCI). Thus a protein present in one PEM could then be compared to the other PEMs in the set. An 'image alignment' algorithm was used in order to create the electronic database that integrated all the MCIs present in the developmental stages collected and TEB proteomes (gel images). The PEMs produced were warped and aligned onto a single common geometry in order to compare the

individual features within the data set using a single database. Initially MELANIE III matched the gels to each other with checks performed by in-house programs and human operators. The level of error incurred such as mismatching features is ~0.1% (false positives and false negatives) when this alignment algorithm is applied (Wilkins *et al.*, 1996a). This is sufficiently low for the construction of this type of database. Differential analysis of the data generated from the proteomes produced was undertaken using Rosetta™.

Table 2.1 *E.coli* landmark proteins

***E.coli* landmark**	**pI**	**Mr**
THP8	4.39	13035
THP7	4.64	27664
THP14	4.47	56134
THP10	4.45	105904
THP4	5.00	51686
THP11	5.39	30847
THP3	5.82	113700
THP12	6.36	47178
THP13	7.23	22298
THP5	8.98	11638
THP9	9.27	35639

E.coli was used to landmark proteins from the proteome images produced in this study. Eleven landmarks were used and their pI and Mr (Da) have been given. *E.coli* reference sample Swiss 2-D service ECO.

For this project two databases were created, reasons for which have been explained in greater detail in Chapter 4. The first database included Av10, P4.5, P8.5, P12.5, P14.5 and P17.5, Inv5, Inv10 and Inv20. This database was called the developmental database and did not include any of the lactation and early involution samples. MCIs (features) which were selected for mass spectrometry from this database were identified by Rosetta™. Only MCIs that were detected in 2/3 triplicates were retained. An MCI was not selected from this list if it appeared to be present in both pregnancy and involution. The feature had to present exclusively in pregnancy and absent in involution or vice versa for mass spectrometry selection. The duplicate proteome images of the TEBs were compared to the Av10 proteomes and only MCIs unique to TEBs were selected for mass spectrometry.

Certain stages of development were excluded from the development database as the samples contained large amounts of milk proteins. As these excluded samples were still of interest, a second database was created which included all 16 stages of development and was named the Lactation/Involution database. Only MCIs that were detected in triplicate were retained which was achieved using Rosetta™. More stringent selection criteria were used for this database to compensate for the poorer quality lactation 2-D gels. Unlike the developmental database each MCI selected for mass spectrometry was manually checked using MELANIE III. Firstly, this was to ensure that the identified feature was not within the smeared areas of the gel. Secondly, this was done to check that the patterning of the surrounding MCIs to the one of interest could confirm that it was the same feature in all of the gels that it was detected in. Comparisons were made between lactation and involution.

2.4.2 Protein identification by mass spectrometry

Selected proteins features of interest were excised from the 2-D gels using a software driven robotic cutter. The gel pieces were transferred to a 96-well plate and digested with trypsin to produce a tryptic peptide pool. The tryptic peptide pool was divided using 10% for MALDI-TOF and 90% for analysis by Q-TOF. A mass list of peptides from each selected feature was determined using a MALDI-TOF spectrophotometer (Voyager-DE STR; Applied Biosystems) with a mass range of 800-2000. Fragmentation spectra from 1 Da mass windows (obtained using the MALDI mass list) were recorded using a nanospray ionisation source on a Q-TOF instrument (Micromass). The fragmentation spectra were converted to centred spectra and used to search the private Incyte database and a non-redundant NCBI sequence database (Released ID 244) with the Sequest (Eng *et al.*, 1994) and Oxford GlycoSciences internal computer programs.

2.5 Bio-Rad 2-D gels

2.5.1 Immobilised pH gradient gel strip rehydration

To 80 µg protein suspended in 2-D lysis buffer 0.2% (w/v) Resolytes (containing trace amounts of Bromophenol blue) (BDH Laboratory Supplies) were added resulting in a total volume of 185 µl. Resolytes enhance protein solubility and improve separation during 1-D focusing. The 185 µl sample was pipetted into a rehydration tray to which an 11 cm pH3-10 ReadyStrip immobilised pH gradient (IPG) gel strip (Bio-Rad) was added gel side up to avoid the introduction of bubbles. Once 2 ml mineral oil (Sigma) was overlaid on top of the sample the IPG strips were rehydrated for 15 h at RT.

2.5.2 1-D Isoelectric Focusing

Damp paper wicks were placed over the cathode and anode of the focusing tray. IPG strips were focused using a PROTEAN IEF Cell (Bio-Rad). The PROTEAN isoelectric focusing (IEF) Cell system is conducted at high voltage and low current due to the low ionic strength within the IPG strips. During IEF the current decreases whilst the voltage increases as the proteins migrate to their pI positions. The IPG strips were positioned in the correct orientation and covered with a sufficient amount of mineral oil. A low voltage (250 V) conditioning step was applied to the strip for 15 min to remove salt ions and charged contaminants. When this was completed, a linear voltage ramping step to increase the voltage to 8000 V proceeded. The current of each strip did not exceed 50 μA. Once maximum voltage was achieved the strips were focused for 35,000 Vh. A hold step of 500 V was incorporated into the run to prevent diffusion of focused proteins.

2.5.3 Equilibration of IPG strip

After completion of focusing, the strips were drained of oil to reduce horizontal streaking caused by unabsorbed protein. The strips were equilibrated for 30 min at RT with equilibration buffer prior to running the second dimension (see Appendix 1).

2.5.4 Criterion^{TM} Cell 2-D gel electrophoresis

IPG strips were loaded onto an 8-16% Tris-HCl Criterion gel (Bio-Rad). An additional well on the gel was loaded with an unstained marker MARK12 for sizing proteins on the membrane after transfer. The marker was visualised using Ponceau S. A 1x Tris-Glycine running buffer was used and the gels were run at 180 V for approximately 1.5 h.

2.5.5 Semi-Dry transfer of 2-D gel

Proteins were transferred from 2-D gels onto Nitrocellulose membrane (0.2 μm) using a TE 70 Semi-Phor Semi-Dry transfer unit (Amersham Pharmacia Biotech). The stack for transfer was assembled in the following order. A mylar mask was placed over the anode with a rectangular section removed from the centre. The shape removed from the mask was slightly smaller than the size of the 2-D gel being transferred. Three sheets of Whatmann 3MM paper cut to gel size and soaked in transfer buffer (see Appendix 1) were applied on top followed by membrane, gel and a further 3 sheets of wet Whatmann paper. The maximum current achieved during transfer was 0.8 mA/cm^2, transfer proceeded for 1 h. Blots were briefly stained with Ponceau S to visualise and label the marker ladder.

2.5.6 2-D western labelling

The same protocol was used for 2-D western labelling as for 1-D western blots.

2.6 1-D electrophoresis, western blotting

2.6.1 Protein electrophoresis

Protein electrophoresis was performed using the Novex NuPAGE™ Electrophoresis system. The Xcell II™ kit was used and 17-well 4-12% Bis-Tris-HCl (Bis (2-hydroxyethyl) amino-tris(hydroxymethyl) methane-HCl) buffered (pH 6.4) polyacrylamide gels. The running buffer used was 1x NuPAGE® MES SDS running buffer, diluted from a 20x MES SDS Running Buffer stock solution (Invitrogen) (see Appendix 2). For reduced protein samples one sixth of a volume of NuPage® Antioxidant was added to the upper chamber running buffer. The gels were run at a constant 200 V with an expected starting current of 100-115 mA/gel and a final current of 60-70 mA/gel, for approximately 50 min.

Samples were prepared under denaturing and reducing conditions. To the protein samples (20-40 μg) 4x NuPAGE® sample buffer (Invitrogen) was added with 40 mM DTT. The final sample solution was heated to 70°C for 10 min, and loaded onto the gel. An unstained marker MARK12 from Invitrogen was used for sizing the proteins on the membrane after transfer. The marker was visualised using Ponceau S.

2.6.2 Western blotting

Proteins were transferred from SDS-Page gels onto Nitrocellulose membrane (0.2 μm) (Schleicher and Schuell), using a Novex XCell II™ Blot module. The membrane, gel, Whatmann 3MM paper and four blotting pads were soaked in transfer buffer. Transfer buffer was used at 1x, diluted from a 20x stock solution using dH_2O and 10% (v/v) methanol (see Appendix 2 for transfer buffer). After soaking, these were assembled in the blot module in the following order. Two blotting pads were placed into the cathode core of the module, and then added to this was Whatmann paper, the gel, membrane, Whatmann paper and finally 2 blotting pads. Once the module was assembled it was filled with transfer buffer and the outer chamber was filled with deionised dH_2O. Transfer was performed at 30 V constant for 1 h.

2.6.3 Western labelling

After transfer the membrane blots were removed from the module and briefly stained with Ponceau S to determine the efficiency of the transfer. In order to visualise the transfer of

proteins on the blots, the stain was partially removed with dH_2O and then washed further with phosphate buffered saline (PBS) (pH 7.3) to remove the stain completely. The blots were blocked in blocking solution (PBS, 0.1% (v/v) Tween-20, 5% (w/v) Marvel milk or 1% BSA (depending on antibody)) for 1 h at RT to remove none specific binding of the antibody. The primary antibody (concentration depending on antibody) was diluted in the blocking buffer and incubated at RT for 1 h or overnight at 4°C. 3x 15 min washes were performed using PBS, 0.1% (v/v) Tween-20 on a horizontal shaker at room temperature. The secondary antibody (concentration depending on antibody) was diluted in blocking buffer and incubated at room temperature for 1 h. 3x 15 min washes were performed using PBS, 0.1% (v/v) Tween-20 at room temperature.

2.6.4 Western signal detection

Horseradish peroxidase (HRP) conjugated secondary antibodies were used to detect signals on western blots. The HRP activity can be detected using either chemiluminescence or 3,3-diaminobenzamidine (DAB) substrates.

Chemiluminescence detection using the ECL™ western blotting analysis system (Amersham Pharmacia) was performed by adding equal volumes of reagent 1 and reagent 2 to the filter, incubating at RT for 1 min and then exposing the blot, covered in Saran wrap, to ECL film (Amersham Pharmacia) for different lengths of time before developing using an X-OMAT film processor.

2.6.5 Coomassie blue staining of SDS-PAGE gels

After electrophoresis when required the gels were stained with Coomassie Brilliant Blue (Sigma) on an orbital shaker. After a sufficient time of staining (~ 40 min) the gels were destained with 30% (v/v) methanol and 10% (v/v) acetic acid on a shaker. The destaining solution was frequently changed until sharp protein bands on a clear background appeared on the gel.

2.6.6 Primary and secondary antibodies

Table 2.2 shows the primary and secondary antibodies used for western labelling and immunohistochemistry, including the dilutions used.

Table 2.2 Antibodies used for western labelling and immunohistochemistry

Antibody and Source	Concentration and Use
Anti-Perilipin (guinea pig polyclonal, Research diagnostics)	1:200 (immunohistochemistry) 1:2500 (western)
Anti-MCM2 (mouse polyclonal, donated by G. Williams)	1:1000 (immunohistochemistry)
Anti-MCM3 (goat polyclonal, Santa Cruz Biotechnology, Inc)	1:200 (immunohistochemistry) 1:1000 (western)
Anti-Rab11 (mouse monoclonal, BD Biosciences)	1:1000 (western)
Anti-Annexin II (mouse monoclonal, BD Biosciences)	1:100 (immunohistochemistry) 1:5000 (western)
Anti-PTRF (chicken polyclonal, donated by P. Grummt)	1: 200 (western)
Anti-WDR1 (rabbit polyclonal, donated by M. Lomax)	1:1000 (western)
Anti-Lumican (rabbit polyclonal, donated by T. Ishiwata)	1:1000 (western)
Anti-Ki67 (rat monoclonal, Dako)	1:75 (immunohistochemistry)
Anti-actin (mouse monoclonal, Abcam)	1:10000 (western)

2.7 Immunohistochemistry (IHC)

2.7.1 Human breast tissue collection

Human breast tissues were used from archives at the North Glasgow Pathology department, and had local research ethics committee approval.

2.7.2 Paraffin embedded tissue sections

Tissues were fixed in 10% (v/v) buffered formalin for 24 h before being processed through graded alcohols (70% (v/v), 90% (v/v) and 100%) to remove residual H_2O, to xylene, to remove the alcohol. They were finally embedded with molten paraffin wax under vacuum conditions. This automated process was performed by a Miles Scientific VIP tissue processor, which lasts 18 h.

Tissue sections (3 μm thick) were cut using a Leitz microtome and mounted on pre-treated microscope slides coated with 2% (v/v) 3-aminopropyltriethoxysilane (APES). The slides were air-dried and later oven-dried at 60°C overnight. The sections were next de-paraffinized in xylene (2x 5 min) followed by rehydration through 100% alcohol (2x 5 min), 70% (v/v) alcohol (5 min) to tap water (at least 2 min). Prior to antigen retrieval sections were pretreated with 3% (v/v) hydrogen peroxide for 10 min to remove endogenous peroxidases and then rinsed in tap water.

2.7.3 Citrate buffer/EDTA antigen retrieval on paraffin embedded sections

Three methods of antigen retrieval were used on paraffin embedded tissue. The first two methods involved adding the sections to either boiling 10 mM citrate buffer pH 6.0 or 1 mM EDTA and subsequently cooking in a microwave pressure cooker at full pressure for 2 to 4 min (times varied according to the antibody used). Sections were cooled in running tap water and later rinsed in either PBS pH 7.4 or Tris buffered saline (TBS) (100 mM Tris-HCl, 138 mM NaCl, and 27 mM KCl) pH 7.4 washing buffers (depending on the primary antibody). All incubations from this point were performed at RT. Non-specific binding was blocked by using 10% (v/v) serum (DAKO) diluted in PBS/TBS. The species used depended on the primary and secondary antibodies. After 20 min blocking, the sections were incubated with the appropriate antibody for 30-60 min depending on the primary antibody, using an optimised dilution prepared in blocking solution. Sections were washed 3x in 10 min with washing buffer and a biotinylated secondary antibody was applied for 30 min (type used depended on the primary antibody). The sections were washed again 3x in 10 min. The signal was amplified using avidin-biotin-peroxidase complex system (DAKO) which was added for 30 min. This step was not required if the

secondary antibody was HRP-conjugated. Sections were washed for 5 min and the peroxidase activity was visualised using DAB to produce a permanent end product. Slides were counterstained with haematoxylin (Harris's), dehydrated and mounted in mounting medium using 22 mm x 40 mm coverslips. Appropriate tissue sections as positive and negative controls were included. Microscopic images were acquired by a Zeiss Axiophot microscope and processed using Axiovision 3.1 software.

2.7.4 SDS antigen retrieval of paraffin embedded sections

The third antigen retrieval method used on paraffin embedded sections was SDS. Sections were incubated at RT with 1% (w/v) SDS in TBS for 5 min and then washed 3x 5 min with TBS. After washing the sections were blocked with 10% (v/v) serum diluted with TBS for 1 h at 37°C (serum from host species of secondary antibody). The appropriate dilution of the primary antibody diluted in blocking buffer was applied to the slides and incubated at 37°C for 2 h. The slides were washed 3x 5 min with TBS and covered with an HRP-conjugated or biotinylated secondary antibody for 1 h at 37°C. When an HRP-conjugated secondary antibody was used the avidin-biotin amplification step was not necessary. Sections were washed 3x 5 min in TBS and the remainder of the protocol used for the citrate buffer/EDTA antigen retrieval method starting from DAB staining was followed.

2.7.5 Frozen fixed tissue sections

Dissected tissues were frozen immediately in OCT (Tissue Tek) with liquid nitrogen and stored at -80°C until use. Frozen sections were serially sectioned at 5 μm thickness on a Bright Instrument microtome and mounted on glass slides coated with APES and stored at -80°C. For IHC, sections were air-dried, then fixed in acetone for 10 min. All incubations from this point were performed at RT. The sections were washed with PBS 2x 5 min, blocked in 10% (v/v) serum for 30 min, incubated for 30 min with the appropriate primary antibody diluted to the desired concentration in blocking solution, and washed again with PBS 3x 3.5 min (this has been described in more detail under section 2.7.2). Sections were incubated with a FITC-conjugated antibody diluted to the desired concentration in PBS for 30 min, washed with PBS 3x 3.5 min and then counterstained for 1 min with 1:100 dilution of propidium iodide (stock 500 μg/ml) which is a nuclear stain. Slides were mounted in Vectashield mounting medium (Vector) using 22 mm x 40 mm coverslips, and stored at 4°C away from light.

Samples were viewed by fluorescence microscopy, carried out using a Wang Biomedical microscope and a fluorescein filter. Images were captured by a Camedia digital camera C-2020 Z and processed using Adobe Photoshop 7.0.

2.7.6 Counterstaining with Haematoxylin

Slides were stained with haematoxylin (a nuclear stain) for 30 s and transferred to running tap water. To remove excess stain, slides were dipped in 1% (v/v) acid alcohol 3x and again transferred to running tap water. Slides were placed in Scott's tap water (81.14 mM anhydrous $MgSO_4$ and 41.66 mM $NaHCO_3$) for 30 s which blues the haematoxylin and rinsed in running tap water. Five minute incubations were used for dehydrating the slides from 70% (v/v) (1x) to 100% alcohol (2x) and finally to xylene (3x). Sections were mounted with permount using 22 mm x 40 mm coverslips.

2.8 Nucleic acid isolation and quantification

2.8.1 Preparation of total RNA

Total RNA was prepared from the whole fourth inguinal mammary gland using TRIZOL ® Reagent (Invitrogen). Prior to dissection of the gland the lymph node was removed, then it was immediately snap frozen in liquid nitrogen. Disposable plastics, RNase-free were used for all the procedures and RNaseZap (Ambion) used to clean all work surfaces, equipment and instruments. The gland was crushed in liquid nitrogen using a pestle and mortar, which had been cooled on dry ice. To create a fine powder-like consistency, the gland was homogenised further using Mikro-Dismembrator (B. Braun Biotech International) contained in a polytetrafluoroethylene (PTFE) capsule cooled in liquid Nitrogen. Approximately 50-100 mg of tissue per 1 ml TRIZOL reagent was used for isolation. The homogenate was incubated at RT for 5 min to permit the complete dissociation of nucleoprotein complexes. 0.2 ml of chloroform was added per 1 ml TRIZOL reagent, the sample was shaken vigorously by hand for 15 s and incubated at RT for 2 min. The homogenate was then centrifuged at 12,000 x *g* for 15 min at 4°C. This separates the mixture into 3 phases: a lower red, phenol-chloroform phase (containing protein), an interphase (containing DNA) and a colourless upper aqueous phase (containing RNA). The aqueous phase was transferred to a fresh tube and 0.5 ml of isopropanol per 1 ml of TRIZOL reagent was used to precipitate the RNA. The sample was allowed to stand for 10 min at RT and then centrifuged at 12,000 x *g* for 10 min at 4°C. The supernatant was removed and the RNA pellet was washed once by adding 1 ml of 75% (v/v) ethanol (made with diethyl pyrocarbonate (DEPC) dH_2O, Sigma) per 1 ml TRIZOL reagent. The sample was vortexed and then centrifuged at 7,500 x *g* for 5 min at 4°C. The

RNA pellet was briefly air-dried, and resuspended in an appropriate volume of DEPC dH_2O. A suitable volume of the sample was used for RNA quantification, the remaining RNA was aliquoted and stored at –80°C.

2.8.2 Quantification of nucleic acids

Nucleic acid concentrations were estimated by spectrophotometry at A_{260}/A_{280} (Beckman spectrophotometer DU-600), where an OD of 1 at 260 nm corresponds to 50 µg/ml of double-stranded DNA and 40 µg/ml of single-stranded DNA and RNA. Readings were zeroed with the solution in which the samples had been diluted. However the nucleic acid purity of RNA was determined by diluting the RNA in 10 mM Tris buffer pH 8.1 because dH_2O can vary the ratio as it is not buffered. The ratio of A_{260}/A_{280} provided an estimate of nucleic acid purity. Values between 1.8 and 2.0 indicated pure preparations.

2.8.3 RNA purification

Qiagen RNeasy columns were used on RNA for Microarray chips and TaqMan analysis and this was applied according to the manufacturers' instructions to clean all RNA samples. The column isolates all RNA molecules that are longer than 200 nucleotides, and excludes most RNAs <200 nucleotides (such as 5.8S rRNA, 5S rRNA and tRNAs). RNA was eluted in a suitable volume of DEPC dH_2O.

2.8.4 DNase treatment of RNA

All RNA samples used for TaqMan analysis were DNase treated using Ambion's DNA free™ DNase treatment and removal agents according to the manufacturers' instructions. This is designed to remove contaminating DNA from RNA samples.

2.8.5 Assessment of RNA quality

For microarray analysis all RNA samples were assessed for RNA quality by using Agilent Technologies RNA 6000 Nano LabChip® Kit according to the manufacturers' instructions.

2.9 Northern blotting

2.9.1 Probe preparation for Northern blotting

A radiolabelled cDNA probe was used to generate a signal to determine the size and abundance of the target RNA across specific stages of mouse mammary gland development. Radiolabelling the cDNA probe was achieved using RadPrime DNA labelling system (Invitrogen). Thirty ng (21 µl) cDNA was denatured at 90°C for 5 min and immediately cooled on ice for at least 1 min. Leaving the cDNA on ice the following

components were added: 1 µl dATP, 1µl dTTP, 1µl dGTP, 20 µl 2.5x random primer solution (all provided in kit) and 5 µl α-^{32}P-dCTP (50 µCi). Lastly 1 µl Klenow fragment was added to the probe mixture and incubated at 37°C for 10 min. The probe was placed on ice for 1 min and applied to a Sephadex G50 NICK™ column (Amersham Biosciences) with 380 µl TE buffer pH 8 (1 M Tris-HCl and 0.5 M EDTA). This wash removed the unincorporated radioactivity, and a further 400 µl TE was added to collect the radiolabelled probe. The final concentration of probe for Northern blotting was 1.0×10^3 cpm per ml.

2.9.2 Northern electrophoresis

All equipment and surfaces used for gel preparation were cleaned with RNaseZap™ (Ambion) and rinsed with DEPC dH_2O. A 1% formaldehyde-agarose gel (see Appendix 3) was prepared in a fume hood and 1x MOPS was used as running buffer diluted from a 10x stock (see Appendix 3). A volume of 200 µl RNA loading buffer (see Appendix 3) was added to aliquots of 20 µg (19.8 µl) total RNA. To 2 µg (2 µl) RNA Millennium™ marker (Ambion) 5 µl RNA loading buffer was added. Samples and marker were placed at 65°C for 15 min and transferred to ice for 1 min. Each tube was centrifuged briefly to collect the contents and 5 µl loading dye (see Appendix 3) was added. Once the samples were loaded the gel was run at a constant 100 V for 3 h.

The gel was washed several times over 30 min with ultra pure dH_2O on a horizontal shaker to remove formaldehyde. The gel was stained in 1x running buffer containing 0.5 µg/µl EtBr for 10-20 min and destained with 2x washes of dH_2O for 20 min. The gel was photographed using a Digi Doc-It imaging system under UV light in order to visualise and measure the migration of the markers from the loading wells. This also enabled visualisation of loading by the intensity of the rRNA (18S and 28S) bands. If the intensity of the markers was too bright, the lane was cut from the gel and destained overnight.

2.9.3 Northern blotting

RNA was transferred from the agarose gel to a positively charged nylon membrane (Hybond-N+; Amersham). The loading wells were removed from the gel with a scalpel blade and then the gel was washed in 10x SSC (see Appendix 3) for 20-30 min. The following materials were assembled in the order stated to achieve transfer. Three dry pieces of Whatmann 3MM paper (cut slightly larger than gel) were placed on top of a 3 cm stack of paper towels (cut 1-2 cm wider than gel). Two pieces of wet Whatmann paper soaked in 10x SSC were added to the top of the stack followed by membrane and gel

(upright). Once the membrane and gel were centred and air bubbles removed, 3 pieces of wet Whatmann paper were stacked on top followed by a 1 cm thick sponge the same size as the Whatmann paper soaked with 10x SSC. Transfer was left for 1.5-2 h.

After disassembling the transfer apparatus the membrane was marked for orientation purposes. The RNA was crosslinked immediately after transfer using a UV Stratalinker (Stratagene) for 1200 μjoules x100. The blot was stored between Whatmann paper.

2.9.4 Northern hybridisation

The blot was rinsed with 2x SSC and then incubated with 20 ml Rapid-Hyb buffer (Amersham Biosciences) for at least 15 min at 65°C. At all times the blot was kept hydrated. Half the volume of prepared probe was heated at 90°C for 5 min, cooled on ice for 1 min and then added to the prehybridised blot. Hybridisation was allowed to proceed for 1.5 h at 65°C. The blot was rinsed with 2x SSC and washed for 20 min with 2x SSC, 0.1% (w/v) SDS at 60°C. A further 2x 20 min washes were applied using a higher stringency wash of 0.2x SSC, 0.1% (w/v) SDS at 60°C. After a final rinse with 2x SSC the blot was wrapped in plastic wrap and exposed to an imaging plate (Fujifilm). The signal was captured using a FUJIFILM FLA-5000 phosphoimager and a 635 nm laser. The images were processed using Advanced Image Data Analyser (AIDA) software (Fujifilm).

2.9.5 Stripping Northern blots

Blots were stripped with boiling 0.1% (w/v) SDS until solution was cold in order to hybridise blot with different probes.

2.10 Affymetrix microarrays

2.10.1 Oligonucleotide microarray hybridisation

Total RNA (10 μg) previously extracted from Av10, Av12, P1, P2, P3, P8.5, P12.5, P14.5, P17.5, L1, L3, L7, Inv1, Inv2, Inv3, Inv4 and Inv20 was used to produce biotinylated cRNA. The protocol followed has been described in the Affymetrix manual. All stages of development were repeated in triplicate from three different mice. Five μg of the labelled cRNA was checked for quality purposes by hybridising to an Affymetrix test chip (Test3). Following confirmation of the quality of labelled cRNA 15 μg of each sample was hybridised to the MG-U74Av2 chip (Affymetrix) for 16 h at 45°C in a rotating hybridisation oven. Chips were washed and stained with Streptavidin-phycoerythrin by using an automated fluidics system (Affymetrix). The chips were scanned with an Agilent scanner and the raw data were analysed by Microarray Suite 5.0 software. The results were

scaled to a target signal of 100 (arbitrary value) and the scaling factors of the individual experiments were no greater than three fold between each other. A total of 12,488 probe sets were interrogated.

2.10.2 Microarray data analysis

Details of the algorithms used are available at http//www.Affymetrix.com. The database used for analysis was constructed with Affymetrix MicroDB® software and was analysed with the Affymetrix Data Mining Tool® (DTM).

2.11 Polymerase chain reaction

2.11.1 Reverse transcriptase (RT)-PCR

Total RNA was reverse transcribed using Superscript™ II RNase H⁻ Reverse Transcriptase (Invitrogen). One to 5 µg of total RNA was added into a reaction containing 500 ng Oligo $(dT)_{15}$ primer (Roche) and 0.5 mM of each dNTP (Invitrogen) to a final volume of 12 µl dH_2O. The reaction was heated to 65°C for 5 min, and then placed on ice to denature the secondary structure of the RNA. To each reaction the following was added: 1x first strand buffer, 5 mM DTT and 40 U RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen). After a 2 min incubation at 42°C, 200 U Superscript™ II RNase H⁻ Reverse Transcriptase was added and the reaction incubated at 42°C for 50 min. Final reaction volume was 20 µl. The reaction was terminated by 15 min incubation at 70°C, and the RNA/DNA duplex degraded by the addition of 2 U of RNase H (Invitrogen) for 20 min at 37°C. A suitable amount of synthesised cDNA (0.5-1 µl) was used for each PCR reaction.

2.11.2 Standard PCR

Standard PCR protocols were used for routine amplification of cDNAs. 0.5 µl of template cDNA was used per reaction. Each reaction contained 1x PCR buffer (minus $MgCl_2$), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 µM forward and reverse primers and 1.25 U Taq DNA polymerase (Invitrogen), in a final volume of 25 µl. Cycling was performed in thin-walled dome-topped 0.2 ml PCR tubes (ABgene) in an Eppendorf Gradient thermocycler or a Biometra™ T3 thermocycler.

Cycling conditions were typically:

94°C for 3 min to ensure template denaturation.

25-35 cycles; denaturing at 94°C, 15 s/ annealing at 50-60°C, 30 s/ extension at 72°C, 1-3 min.

72°C, 5 min.

Annealing temperatures were dependent on the primers used. When multiple PCR reactions were run at the same time or same samples run using different annealing temperatures, the Eppendorf Gradient PCR machine was used and a gradient imposed across the heating block to account for the required annealing temperatures.

2.12 Real-time TaqMan RT-PCR

2.12.1 Primer and Probe design

The primers and probes were designed to produce an amplicon spanning an intron-exon boundary and such that either primer or probe overlapped two exons to prevent genomic amplification. The following guidelines were used for designing the primers and probes.

Designing probes:
- G-C content in the range of 30-80%
- No G on 5' end
- Sequence containing more Cs than Gs
- A T_m value of 10°C greater than the primers
- Avoiding runs of more than 4 of the same base
- Length of probe no longer than 30 nucleotides

Designing primers:
- G-C content in the range of 30-80%
- Avoiding runs of more than 4 of the same base
- A T_m value between 58-60°C
- Placing both primers close to the probe without overlapping

The five nucleotides at the 3' end should have no more than 2G and/or C bases.

2.12.2 Real-time TaqMan® PCR of gene expression

Oligonucloetide primers and Taqman probes were designed to meet specific criteria using Primer Express, version 2.0 (Applied Biosystems) from the GenBank database. The TaqMan probe consists of an oligonucleotide with a 5' fluorescent reporter dye, 6-carboxy-fluorescein (FAM) that is covalently linked to the 5'-end of the oligonucleotide. The reporter dye is quenched by 6-carboxy-tetramethyl-rhodamine (TAMRA) at the 3'-end. Separation of the quencher dye from the reporter results in increased fluorescence. The samples were placed in 96-well plates using a final 25 µl reaction volume. The reaction mix contained a final concentration of 300 or 900 nM (depending on the primer set) forward and reverse primers, 200 nM TaqMan probe, 0.1 µg cDNA sample and 1x PCR mastermix (TaqMan® Universal PCR Master Mix, No AmpErase ® UNG; Applied Biosystems). The samples were amplified in an automated fluorometer (ABI PRISM 7700 sequence Detection System; Applied Biosystems). Amplification conditions were 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 s at 95°C and 1 min at 60°C. Each sample was

repeated in duplicate for consistency and control reactions without reverse transcriptase were run for each sample. In addition, negative control reactions containing only dH_2O and no RNA were used to establish baseline levels of fluorescence. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an endogenous control for normalisation.

2.12.3 TaqMan analysis

The relative quantitation of gene expression was obtained by the comparative C_T method which uses the arithmetic formula $2^{-\Delta\Delta CT}$. Values are expressed relative to a reference sample, called the calibrator. First, the threshold cycle (C_T) of the target gene and the C_T of the endogenous control were determined for each sample. The C_T value is the cycle number at which exponential growth of the PCR product commences. As each sample was duplicated the values were averaged. The difference between the target C_T and endogenous control, called ΔC_T, were calculated in order to normalise for the differences in RNA extractions and the efficiency of the reverse transcription step. The ΔC_T values for each normalised experimental sample was subtracted from the ΔC_T of the calibrator, called the $\Delta\Delta C_T$ value. Lastly the amount of target, normalised to the endogenous reference and relative to the calibrator was calculated by $2^{-\Delta\Delta CT}$. Therefore these values demonstrate the *n*-fold difference of the experimental values relative to the calibrator.

To validate the $\Delta\Delta C_T$ calculation, the efficiencies of the target and endogenous control must be approximately equal. This can be assessed by analysing the ΔC_T in a dilution series of the template. Six dilutions of cDNA were prepared in duplicate and were amplified to obtain standard curves. The average C_T values of the target and endogenous control were determined and the ΔC_T was calculated. The log input of the total amount of RNA was plotted against ΔC_T. If the efficiencies of the two amplicons were approximately equal, the slope of the log input versus ΔC_T was < 0.1.

ΔC_T (sample) = C_T (target gene) - C_T (GAPDH)

$\Delta\Delta C_T = \Delta C_T$ (sample)- ΔC_T (calibrator)

relative expression = $2^{-\Delta\Delta CT}$

(ABI Prism Sequence Detection System User Bulletin #2)

2.13 DNA electrophoresis and purification

2.13.1 Agarose gel electrophoresis of DNA

DNAs were separated in 1% agarose in 0.5x TBE (90 mM Tris, 90 mM boric acid, pH 8.3, 2 mM EDTA) containing 0.1 μg/ml EtBr as described in Sambrook and Russell (2001),

using 0.5x TBE as the electrophoresis buffer. DNA was visualised under UV light, and sizes of fragments were compared to a 1 kb or 123 bp ladders (Invitrogen). Prior to loading, 6x loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol in dH_2O) was added to the samples to a final 1x concentration.

2.13.2 Purification of DNA from gels

DNA fragments were excised from the gel using a clean scalpel blade and the DNA extracted using the Qiagen Gel Extraction kit according to the manufacturers' instructions. DNA was typically eluted in 30 µl of dH_2O.

2.14 Oligonucleotide synthesis

Oligonucleotides were synthesised by the Sigma-Genosys custom primer service on a 0.025 µmol scale, purified by a minimum of desalting, and their quality assessed by MALDI-TOF analysis. Oligonucleotides were received as a lyophilised pellet, resuspended in dH_2O to a stock concentration of 100 µM and further diluted with dH_2O to a working concentration of 6.6 µM. All primers were stored at -20°C.

2.15 Automated DNA sequencing

Automated sequencing was performed at the Glasgow University Molecular Biology Support unit (MBSU). A single-stranded reaction was used with template and primer supplied at 10 ng/100 bp and 3.2 pmol respectively, with a PCR mix containing fluorescently labelled dideoxynucleotides. Samples were run on an agarose gel with the nucleotides being detected on an ABI automated DNA sequencer. Analysis was carried out by the Applied Biosystems automated sequence analysis programme, and the sequence data viewed and further analysed using Editview (version 1.0, free DNA sequencing software from Perkin Elmer).

Chapter 3

Preparation of the mouse mammary gland

3.1 Summary

This Chapter describes the collection of mammary glands at various stages of development and optimisation of the extraction procedure to obtain good quality protein for 2-D gel analysis. Sixteen stages of the developing mouse mammary gland were selected for detailed morphological and biological analysis. These were key time points of development which demonstrate the processes of proliferation, differentiation and apoptosis, which are iterated in pregnancy, lactation and involution. In order to observe the morphological changes, which take place during development, wholemounts were taken of both the left and right fourth inguinal mammary gland for each of the 16 stages of interest. Tissues were collected for RNA and protein extraction. The protein extracted from the mammary glands was of suitable quality for proteomic analysis. This technique was used as a tool to identify proteins that may be regulating proliferation, differentiation and apoptosis during mammary gland development. In addition, proteomic analysis was used to identify proteins specifically associated with invasion and migration by isolating specialised structures called TEBs from the pubertal gland and extracting protein in the same way as for whole glands.

3.2 Introduction

The first six months of this project were dedicated to standardising the conditions for dissection and collection of intact mammary glands and TEBs for proteomic analysis. This work was in collaboration with Oxford GlycoSciences. All gland collection and extraction procedures were carried out by myself, and the 2-D gels were run at Oxford GlycoSciences. One year was spent at Oxford GlycoSciences in order to understand each aspect of proteomics and, at the practical level, to run 2-D gels.

3.2.1 Investigation of the mouse mammary gland

An inbred mouse strain was chosen to investigate mammary gland development as it provided a uniform and reproducible system. The long term aim of the project was to identify proteins which regulated human breast development. However, due to the lack of availability of human tissue and the similarities that are shared between human and mouse breast development the mouse mammary gland was the most logical tissue to study. After identifying proteins regulating mouse mammary gland development these could then be related to human breast development.

Female mice have five pairs of mammary glands, three thoracic and two inguinal (Figure 3.1). The fourth inguinal gland was chosen for investigation in this study due to the ease of complete removal of the gland from the animal and rapid dissection. Less muscle is removed with the fourth gland which makes the tissue preparation cleaner.

The female mouse, when maintained in good condition, ovulates at intervals of four to five days. Ovulation occurs during each oestrous cycle and ductal cells give rise to alveolar buds. During the adult virgin stage of development some strains of mice do not have a functional ovarian luteal phase, and so the mammary gland consists entirely of ducts. The strain of mouse chosen for this study was Balb/C because females from this strain do not develop alveoli prior to pregnancy. This was desirable to identify clearly the morphological change from adult virgin to pregnancy (Fekete, 1938).

3.2.2 The developing mouse mammary gland

Sixteen stages of mammary development were focused on in this study, ranging from the adult virgin gland to pregnancy, lactation and involution. In newborn female mice, the mammary gland is composed of ducts, which are able to migrate from the nipple through the fat pad, which is driven by structures at the duct tips, known as TEBs. These

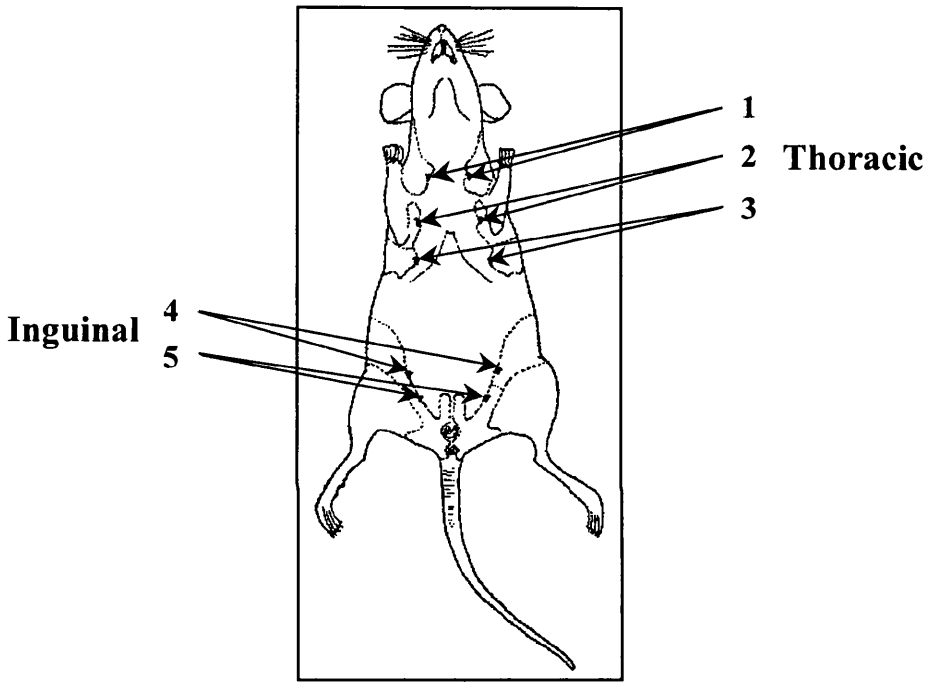


Figure 3.1 Five pairs of mouse mammary glands. Taken from Cole, 1933. The mouse has three pairs of thoracic and two pairs of inguinal mammary glands.

specialised structures were collected for analysis at both the RNA and protein level as were glands from the 16 stages of mammary gland development.

The TEBs regress before adulthood leaving an extensively branched ductal system to fill the whole fat pad. This stage of development is reached at about 10 weeks of age. Tissue samples from 10 week old adult virgin mice were collected since at this stage it was regarded as a normal resting mammary gland. Comparisons of the expression levels in pregnant, lactating and involuting glands could then be made with the resting virgin gland.

A number of stages within the different phases of gland development were selected in order to have a representative sample of each of the main changes that occur. In pregnancy early (P4.5 and P8.5), mid (P12.5, P14.5) and late (P17.5) pregnancy samples were selected. An increase in ductal branching and the formation of alveoli are seen within the first days of gestation, hence the collection of P4.5. Lobuloalveoli have formed by P8.5 and by P12.5 to P14.5 they are mature structures. By late pregnancy (P17.5) the gland has increased in thickness and is predominantly alveolar structures.

As minimal changes take place within the first week of milk production, lactation samples were collected only on day one (first day of parturition) (L1), three (L3) and seven (L7).

During involution a cascade of apoptotic events take place. In order to focus on the switch from lactation to involution and also on the factors which steer the gland through the irreversible stage of involution, samples were collected from the initial stages of involution Inv1 (the first day after removal of the pups), Inv2, Inv3, Inv4, Inv5. The latter stages of involution, Inv10 and Inv20, were also collected, since at these time points the gland resembles that of an adult virgin. All 16 stages of development are shown in Table 3.1.

Table 3.1 Stages of mammary gland development studied

	Days		
Adult Virgin	**Pregnancy**	**Lactation**	**Involution**
10 weeks	4.5	1	1
	8.5	3	2
	12.5	7	3
	14.5		4
	17.5		5
			10
			20

3.3 Results

3.3.1 Comparison of fresh and frozen tissue for protein extraction

Trial tissue samples were collected from adult virgin female mouse mammary glands. The samples were tested on 2-D gels in order to optimise the protein extraction protocol prior to using the mammary developmental sample set. Triplicate samples of fresh and frozen tissue were used for the first trial samples run by 2-D gels, but fewer than 1000 protein features resolved on the gels compared to the 2000 expected. Figure 3.2A is the proteome image produced from a frozen sample and Figure 3.2B from a fresh sample. Table 3.2 shows the number of protein features detected in the first and second tissue preparation trials; the procedure used for trial 2 is described below. In the first trial the majority of the frozen samples had a greater number of protein features detected compared to the fresh samples. This was because they were more concentrated than the fresh samples and consequently a higher amount was loaded onto the gels.

The protein extraction protocol used on the fresh and frozen tissue was based on the wet weight of the tissue per volume of lysis buffer. This vastly underestimated the amount of tissue required to obtain a suitable concentration of protein for proteomics. A refined technique was applied to increase the yield of protein extracted from the gland. To increase the amount of tissue per volume of lysis buffer the dounce homogeniser was replaced with an Eppendorf homogeniser, which was able to contain the sample in a smaller volume and resulted in less loss of tissue.

The second set of trial samples that were extracted in this way had a concentration above 0.5 μg/μl. A concentration of ≥0.5 μg/μl enables the 120 μg protein loading target to be achieved. Again, fresh and frozen samples were collected in triplicate to assess reproducibility. The scanned images of these 2-D gels revealed that five out of the six samples in trial 2 had over 2000 protein features detected (Table 3.2). The 2-D gels revealed that the fresh samples produced a better proteome (gel image) (Figure 3.2D) than the frozen samples (Figure 3.2C) because they had resolved a greater number of higher molecular weight proteins. Only extreme changes are visualised on these gels and it must be remembered that the scanning software used to create these digital gel images can detect to levels that can barely be visualised.

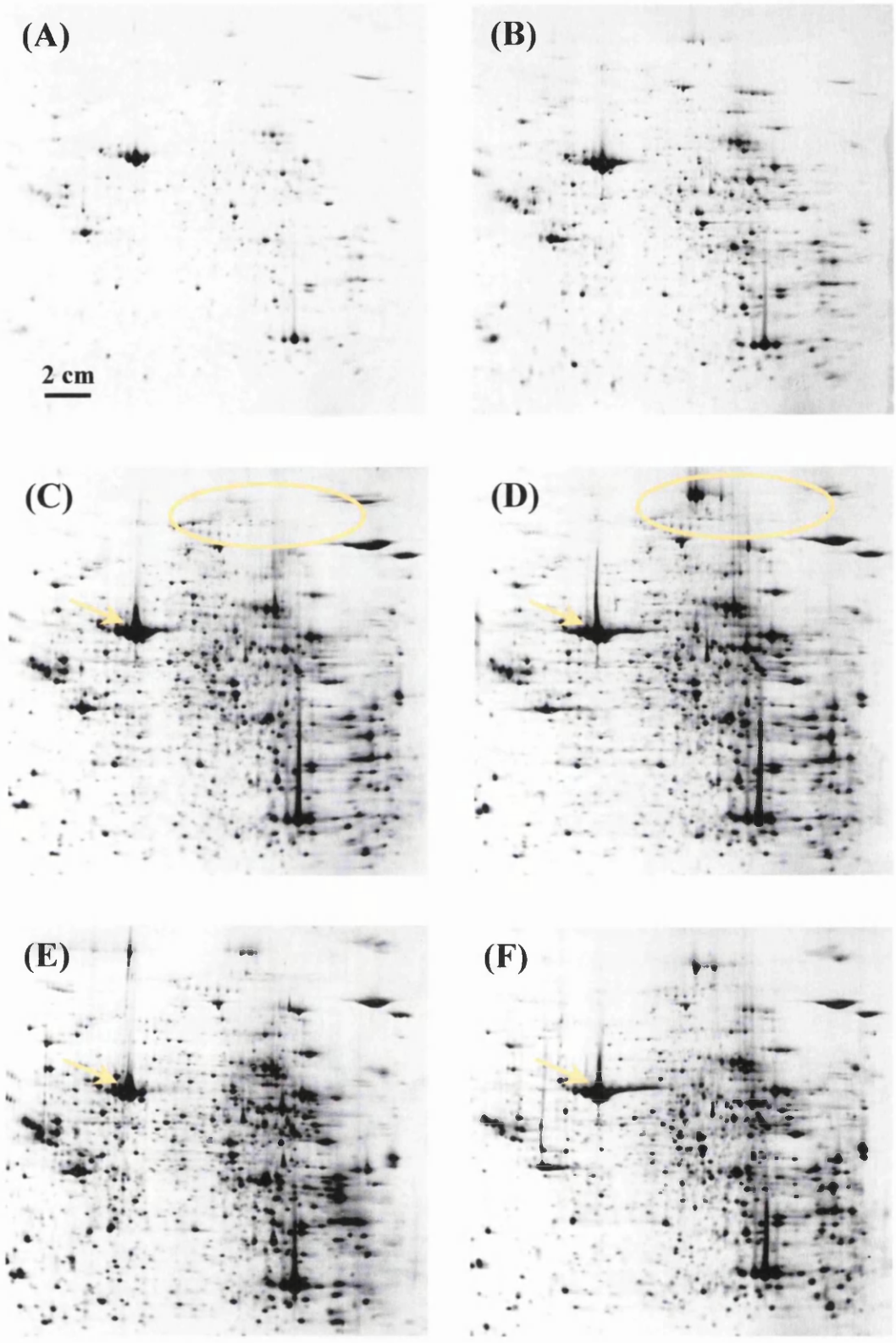


Figure 3.2 Fluorescent 2-D gel images of optimised tissue extraction protocol. 2-D gel images stained with OGT 1238 and scanned with an Apollo III linear fluorescent scanner. For scale, images are 40% of the original size. (A) Frozen (B) Fresh tissue extracted with a dounce homogeniser (<1000 features). (C) and (D) Frozen and fresh tissue (respectively) extracted with an eppendorf homogeniser (>2000 features). (E) Mammary gland washed in DMEM prior to extraction to reduce serum albumin contamination (indicated with yellow arrows). (F) Mammary gland tissue section 'milked' of blood. Yellow oval highlights the greater number of high molecular weight proteins detected in (D) compared to (C).

Table 3.2 Features detected in 2-D gel trial sample preparations

Trial 1	No. of features	Trial 2	No. of features
Frozen 1	659	Frozen 1	2231
Frozen 2	696	Frozen 2	2106
Frozen 3	632	Frozen 3	1972
Fresh 1	516	Fresh 1	2034
Fresh 2	568	Fresh 2	2265
Fresh 3	770	Fresh 3	2324

The primary images of the 2-D gels were processed using MELANIE III software which detected the fluorescently stained features resolved on the gels.

3.3.2 Washing the tissue to remove serum albumin contamination

Serum protein contamination is a major concern in proteomics. A volume of 2 µl of serum can generate a proteome image of 1000 protein features on 2-D gels run by Oxford GlycoSciences. This amount of serum contamination in a tissue sample will result in the obscuring of proteins surrounding the affected areas. Serum proteins are used as a marker of the amount of blood contamination present in a sample. Since this was a problem in the trial samples, an attempt was made to minimise the amount of serum albumin (i.e. serum contamination) in the samples. The first approach to this problem was to wash the glands for various lengths of time with DMEM. Visual comparisons were made of the 2-D gels which showed no obvious benefit in extending the washes with regards to reducing the levels of serum. Figure 3.2E shows the proteome image produced after washing for 5 min in DMEM. The yellow arrows on images C, D, E and F are pointing to the serum albumin feature. There appeared to be some benefit in washing the tissue in medium when comparing the image to Figure 3.2D which is of an unwashed fresh sample. The image produced from the unwashed protein sample was used as a reference gel as it demonstrated a sample which had not been subjected to further optimisation protocols such as washing with DMEM, 'milking' or digestion of the gland. These techniques are described later in this Chapter. In conclusion, 5 min washes were applied to subsequent protein extractions to reduce the level of serum albumin contamination. Previous experience at Oxford GlycoSciences had also found DMEM washes beneficial to serum removal.

3.3.3 'Milking' the tissue to remove serum albumin contamination

A second technique was developed to remove endogenous serum from the gland without any damage to the tissue. After dissecting a section of gland digital pressure was applied along the length of the gland thus blanching the tissue as blood was expelled. Duplicate

samples were prepared by this method known as 'milking the gland' and analysed on 2-D gels. The 'milked' gel image (Figure 3.2F) was compared to the unwashed unmilked protein sample (Figure 3.2D), and it was concluded that there was some advantage in using this technique. Therefore it was applied to subsequent protein extractions.

3.3.4 Proteome generation of the mouse mammary gland development set

These revised protein extraction methods were used on all 16 stages of development, which were collected in triplicate and run on 2-D gels. Visual analysis of the triplicate gel images revealed no gross differences and showed feature patterns complimentary to each other indicating their reproducibility. Similar levels of high abundant proteins were also visualised from the triplicate images (data not shown). The reproducibility of the sample preparation was shown by the 2-D images produced, as over 2000 features were resolved on each gel. The proteome images of these samples are at the back of this thesis in CD format.

3.3.5 Morphological analysis of the mouse mammary gland

To visualise the morphological changes taking place during mammary gland development wholemounts were collected from each of the 16 stages of interest. Representative stages were captured on a digital camera and are shown in Figure 3.3A-J. Each stage was collected in duplicate to verify whether there was any individual variation. Both the left and right gland of each mouse was wholemounted to establish whether developmental progression was identical on each side. Indeed, the left and right glands were identical to each other which enabled both sides to be manipulated.

3.3.6 Generation of the TEB proteome

These structures were captured on wholemount sections and Figure 3.4 highlights the bulbous club-shaped TEBs. A protocol described by Richards *et al.*, 1982 was modified to isolate these structures from the gland. The 2-D gel image produced from the TEB sample resulted in a large area of the gel being occupied by serum protein, thus preventing many proteins from being identified (Figure 3.5A). The TEB extraction technique was therefore revised by using medium containing no serum. The proteome image produced from the TEBs without serum (Figure 3.5B) compared to the TEBs with serum was completely different. Only 1689 features were detected from the TEB serum positive preparation compared to 2172 features resolved from the TEB serum free sample. This demonstrates the importance of reducing the levels of serum in a sample, since features may be masked by serum proteins when resolved on a 2-D gel.

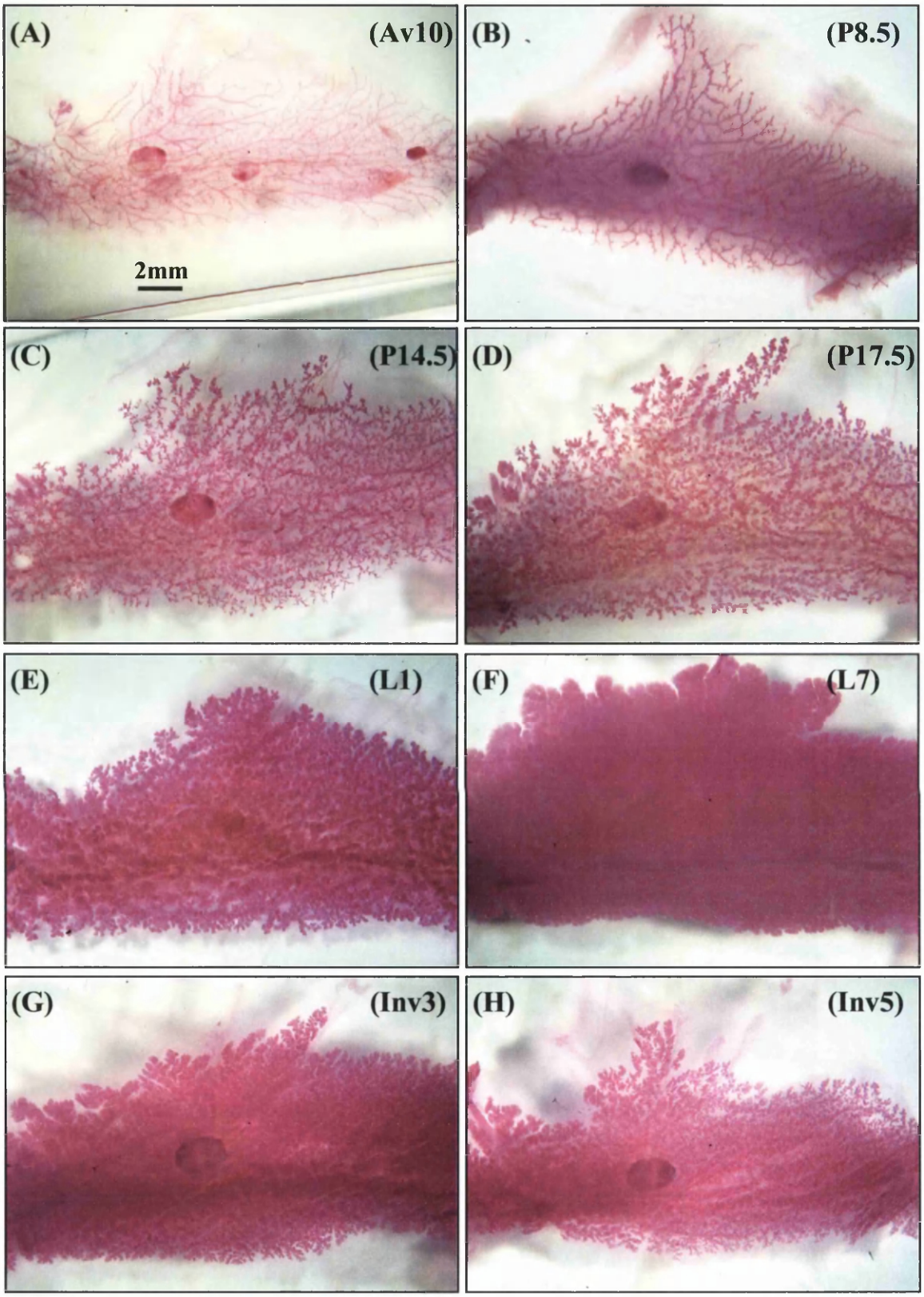


Figure 3.3 Wholemounts of representative stages of mammary gland development. These images are of wholemount mammary glands that have been stained with carmine which identifies the epithelium of the gland. The dark pink oval shaped structure lying almost central to the gland is the lymph node. For some of the glands more than one lymph node can be seen. (A) to (H) range from Av10 to Inv5. Photographs were taken at 6x magnification using a Zeiss SV11 microscope and a Zeiss KL1500 LCD light source. Images were captured on the Fuji fine pix digital camera.

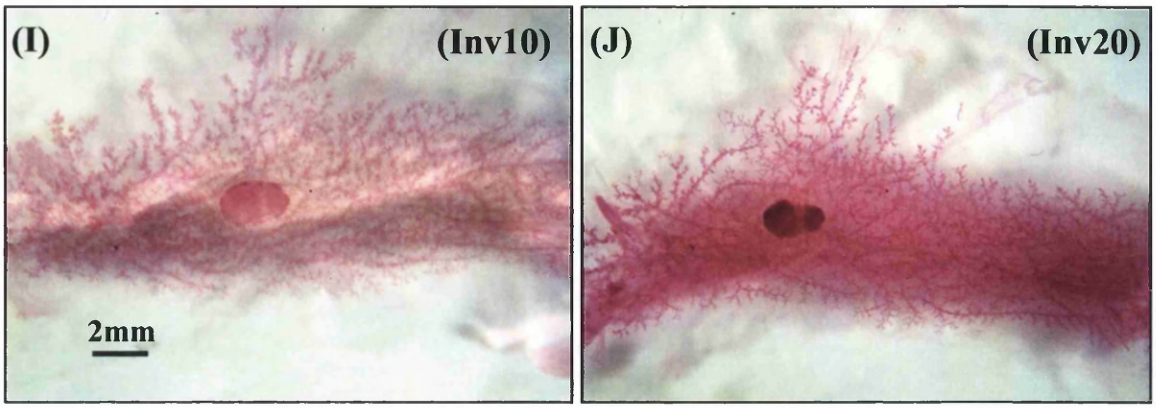


Figure 3.3 continued Wholemounts of representative stages of mammary gland development. (I)-(J) Inv10 and Inv20 the final stage of mammary gland development investigated.

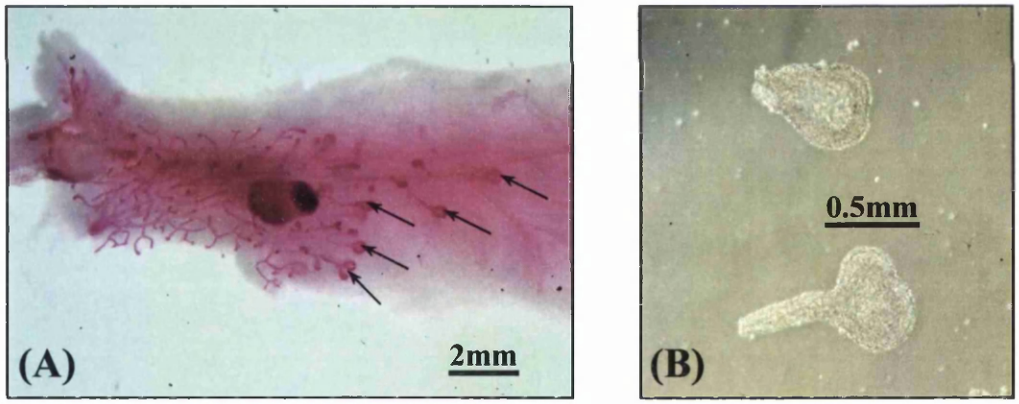


Figure 3.4 Structure of the TEB. (A) A wholemount of a gland taken from a six week old female mouse. The TEBs have just past the lymph node and have been indicated by black arrows. The wholemount has captured the *in vivo* morphological organisation of the mammary gland. Photograph taken at 6x magnification using a Zeiss SV11 microscope and a Zeiss KL1500 LCD light source. Images were captured on a Fuji fine pix digital camera. For scale, the length of the lymph node is approximately 2mm. (B) An image of isolated TEBs suspended in L-15 medium post digestion, photograph taken using a phase contrast microscope at 50x magnification.

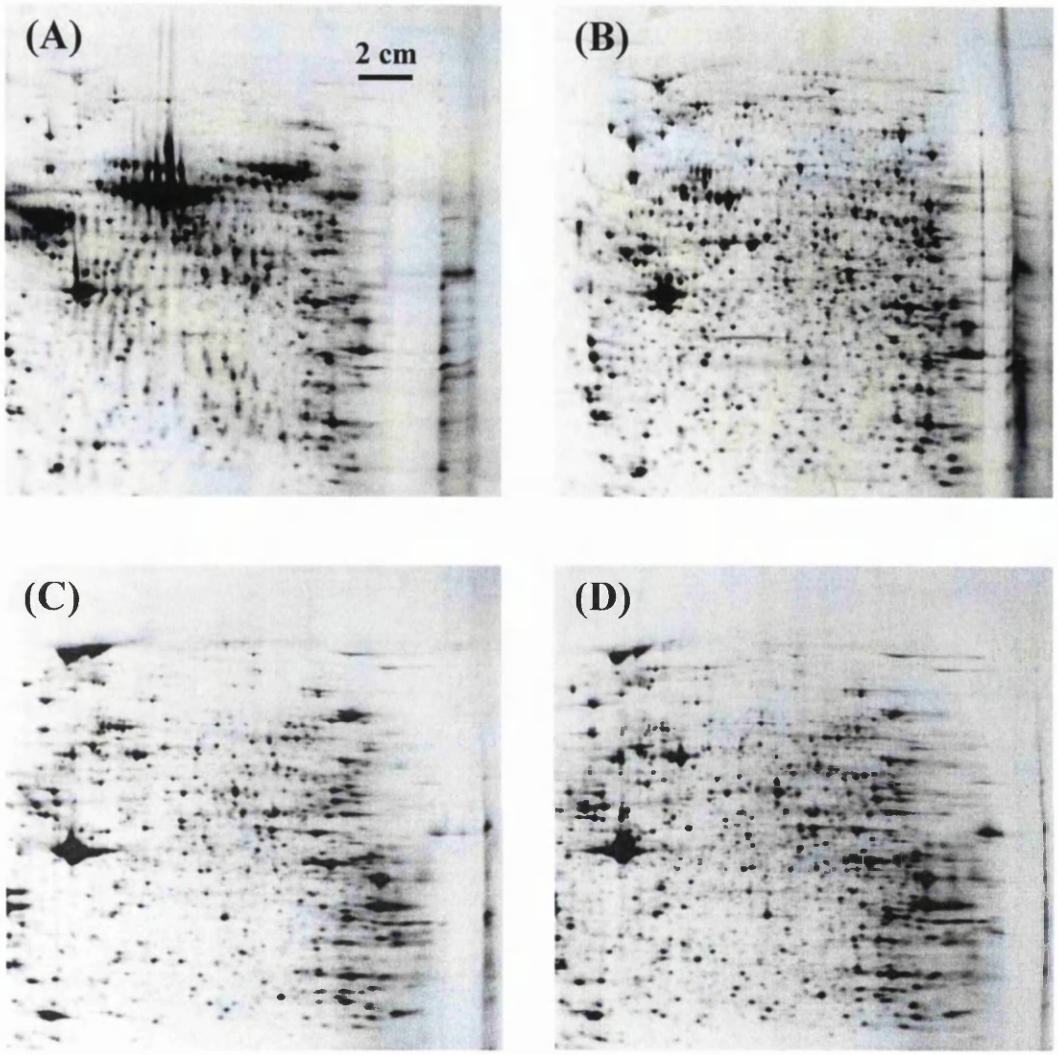


Figure 3.5 Fluorescent 2-D gel images of optimised TEB extraction protocol. 2-DE images were stained with OGT 1238 and scanned with an Apollo III linear fluorescent scanner. For scale, these images are 40% of the original size. (A) TEBs with serum, (B) TEBs without serum, (C) Digested glands washed 3x with L-15, (D) Digested glands not washed.

An additional complication associated with the TEB preparation was observed from the 2-D gel image. The image showed evidence of streaking at the basic end of the gel which is usually caused by salt. Streaking of gels to the degree seen in the TEB sample had not been detected in the tissue section samples. The culture medium used for digestion was not implicated as the cause of streaking for the following reason. L-15 medium and DMEM have similar salt concentrations and the gel image of the sample that had not been washed or 'milked' (Figure 3.2D) showed no difference in the degree of streaking at the basic end when compared to a sample washed in DMEM (Figure 3.2E). Again, as with serum contamination, the levels of salt needed to be reduced; otherwise protein features within the streaked area would not be incorporated into the proteome database which would be used later for expression analysis.

It was thought that the salt causing the basic end streaking was possibly from the collagenase used to digest the glands. This may have been due to residual amounts of collagenase being present in the suspended TEBs. To try and reduce the collagenase levels, additional washes with L-15 medium were made to the pelleted material before isolating the TEBs on contact dishes. However, the length of time taken to collect a sufficient number of TEBs for a 2-D gel limited further optimisation experiments performed with the TEB sample. Therefore an attempt to reduce the salt levels was made by taking 16-18 g mice and digesting their whole mammary glands using the normal digestion procedure. The pelleted material was washed 3x with L-15 medium and protein was extracted from the pelleted material and not from the TEBs. This was performed on two different samples for reproducibility purposes. As a control to determine whether streaking had been reduced by this technique, two samples were prepared from digested mammary glands which when pelleted were not washed 3x in culture medium.

Visual analysis of the gel images from the washed Figure 3.5C and unwashed digest samples Figure 3.5D showed that both had a higher degree of basic end streaking when compared to the gel images produced from tissue sections which had not been digested (Figures 3.2D, E, and F). This provided evidence that the digestion procedure was intensifying streaking at the basic end. When the duplicate washed and unwashed pellet images were compared, they showed that washing the pellet did not improve the basic end streaking on the gels. In conclusion as the removal of basic end streaking was not achieved, the previous optimised methodology for isolating TEBs was used. The TEB 2-D

gels passed the quality control checks performed by Oxford GlycoSciences. However, areas that were affected by basic end streaking were avoided for expression analysis.

3.4 Discussion

3.4.1 Mouse mammary gland tissue preparation

At the beginning of this project there were no publications linking proteomics and mouse mammary gland development. Hence, there was no technique devised to extract protein from this type of tissue which was of suitable quality for proteomic analyses. This Chapter has described the improvements which were made to the preparation of protein samples from mouse mammary glands.

The mammary gland is a fatty tissue particularly prior to pregnancy before the invasion of alveolar epithelial cells. This made complete homogenisation of the mammary gland difficult to achieve, but an alteration of the equipment used resulted in a smoother suspension than was previously obtained. This improvement increased the number of features detected on a 2-D gel to achieve the target of over 2000 protein features.

The extraction of protein from fresh tissue resolved a larger number of higher molecular weight proteins compared to frozen tissue. Frozen tissue would have been preferable to use since the protein could have been extracted when convenient. However, as the quality of the protein was of principal importance, protein was always extracted from fresh tissue.

A large area of the gel was occupied by serum albumin which may have masked proteins that had migrated in the same area as this protein. These masked proteins could have had potentially interesting expression profiles and therefore in order to detect them the levels of serum contamination needed to be reduced. This was achieved by a combination of washing and 'milking' the gland of blood. It is believed that only a minimal change was detected in the attempt to reduce the serum protein levels because the majority of it was endogenous. Many blood capillaries run through the mammary gland which become more vascularised during pregnancy and have most likely contributed to most of the serum proteins present on the 2-D gels.

3.4.2 Reproducibility of the mammary gland and its importance in proteomics

Wholemounting the left and right glands across development revealed that they were morphologically the same, and microarray data (not shown) also showed this at the RNA level. Therefore the left gland was taken for protein extraction and the right was taken for

RNA purposes. Reproducibility is crucial to the success of proteomics and sample preparation is a major contributor to reproducibility. The triplicate tissue section samples collected produced reproducible proteome images.

Quality control and curation of 2-D images also contribute to reproducibility in proteomics. Many 2-D gels fail image quality control checks because of visual imperfections such as those caused by gel overloading and poor resolution. All of the developmental mammary samples passed the quality control criteria which can be demonstrated at the bioinformatic level during curation. Once artefact features were discarded by an image processing algorithm, the triplicate gels were grouped into a single virtual feature group. The virtual image produced only contained features which were present in two or more of the individual gels. It was only these replicated features that were integrated into the master group for differential analysis. This was achieved by image processing software MELANIE III and differential software called Rosetta^{TM}. Therefore the replicate samples were of the highest quality.

3.4.3 A modified approach to TEB isolation

The TEB extraction procedure was both intricate and long, and there had been concern with regard to the quality of the protein and RNA extracted. However, good quality 2-D images were produced from the protein extracts and RNA quality control checks demonstrated good quality RNA. The TEB preparation (without serum) revealed almost 500 extra protein features compared to the TEB preparation (with serum). Many of these extra features had previously been masked by serum proteins.

A greater degree of streaking was found at the basic end of the TEB 2-D gel (without serum) when comparing it to the tissue section image. The attempts made to reduce the level of streaking were not effective and therefore the areas affected were not incorporated into the proteome database generated later on.

In conclusion of this Chapter, good quality protein was extracted from all stages of mammary gland development including the TEBs. As a result good quality reproducible proteomes were produced which were then used to identify features which changed either in their presence or absence across development, or demonstrated a change in expression levels across development.

Chapter 4

Generation of the mouse mammary gland proteome by 2-D gel electrophoresis

4.1 Summary

Detailed analyses of 16 time points of mammary gland development and TEB preparations were carried out using the technique of proteomics 2-D gel electrophoresis. The proteomes revealed the expression profiles of all the protein features detected across these developmental time points. The information generated from the proteome images was interrogated by creating three databases. One database contained only gel images from the adult virgin, pregnancy and mid-late involution stages called the developmental database; a second database included all 16 stages of mammary gland development, known as the Lactation/Involution database; and finally the TEB database was created which consisted of TEBs and adult virgin proteomes. The developmental database analysed the expression profiles of MCIs across development, and 81 MCIs were selected from it for mass spectrometry analysis. The second database was used to analyse the switch from lactation to involution, and 44 MCIs were selected from it for mass spectrometry analysis. A further 44 MCIs were selected from the TEB database when comparing the TEBs to the adult virgin proteomes. A total of 169 MCIs were selected for mass spectrometry analysis from these databases. The total number of MCIs with protein identifications (annotations) was 83; 29 annotations were identified from the developmental database, 30 from the Lactation/Involution database and 24 from the TEB database. Some of these MCIs had more than one protein annotation within the selected feature and were called clusters. These annotations were categorised into areas of their biological function which included metabolic enzymes and mitochondrial, signalling, serum, cytoskeletal, transport, milk, RNA processes, adipose associated, protein turnover proteins, and annotations which shared no common function to any of these were classed as 'other'.

Eight annotations were chosen from the developmental and Lactation/Involution databases for further work based on their biological interest, novelty to mammary gland development and quality of the peptide sequence that linked to the annotation. A further two MCIs with three annotations were pursued as they were hypothetical and unique proteins.

Analysis of the TEB annotations concluded that the majority of cytoskeletal proteins detected have previously been reported to be present in axonal growth cones. The proteomic data and the similarities of growth observed between TEBs and axonal growth cones such as directional growth and spatial patterning suggest that other proteins which are related to growth cone function could be important to TEBs.

4.2 Introduction

4.2.1 Overview of Proteomics

Proteomics can broadly be defined as large-scale systematic analysis of protein expression levels in biological systems. Proteins are firstly separated according to their isoelectric focusing point and then by molecular weight. The former is achieved using IPG strips and the latter by a polyacrylamide gel. The gels are subsequently stained with an appropriate dye that will highlight the distribution of the proteins. Sophisticated software scans the gel image to create a synthetic gel (Protein Expression Map (PEM)) and once all the images for a particular study have been scanned they are then curated. This is a process which landmarks the proteins according to their pI and Mr, and warps and aligns the PEM onto a single common geometry. Each protein feature detected is assigned a unique identifier number known as the molecular cluster index (MCI) number. This enables proteins in other PEMs to be compared. Once this process has been completed the images are interrogated to determine expression distributions of the features separated. The selected features of interest are subjected to tryptic proteolysis and the resulting peptide fragments are analysed by mass spectrometry in order to determine the protein identification of the feature of interest. Figure 4.1 overviews the steps taken during proteomics to achieve mass spectral data from a feature present on a 2-D gel image.

As previously mentioned in Chapter 3 this project was run in collaboration with Oxford GlycoSciences. Their proteomic unit ran the mammary gland samples on 2-D gels and curators then built a digital copy of the biological information held on each gel in order to create a database for differential analysis. To gain experience of these processes at Oxford GlycoSciences the manual selection of MCIs for mass spectrometry analysis, a method which was necessary for one of the databases created, was carried out by myself and a curator at Oxford GlycoSciences.

4.3 Results

4.3.1 Generation of the proteome databases

The objective of this study was to identify proteins that were differentially expressed during mammary gland development. This was achieved by analysing each individual MCI both quantitatively and qualitatively with Rosetta^{TM} software.

As a reference, a protein feature is a protein spot on a 2-D gel and each feature is assigned an MCI. Once an MCI has been selected for mass spectrometry, and a protein identification assigned to it, this is known as a protein annotation. In some cases there was

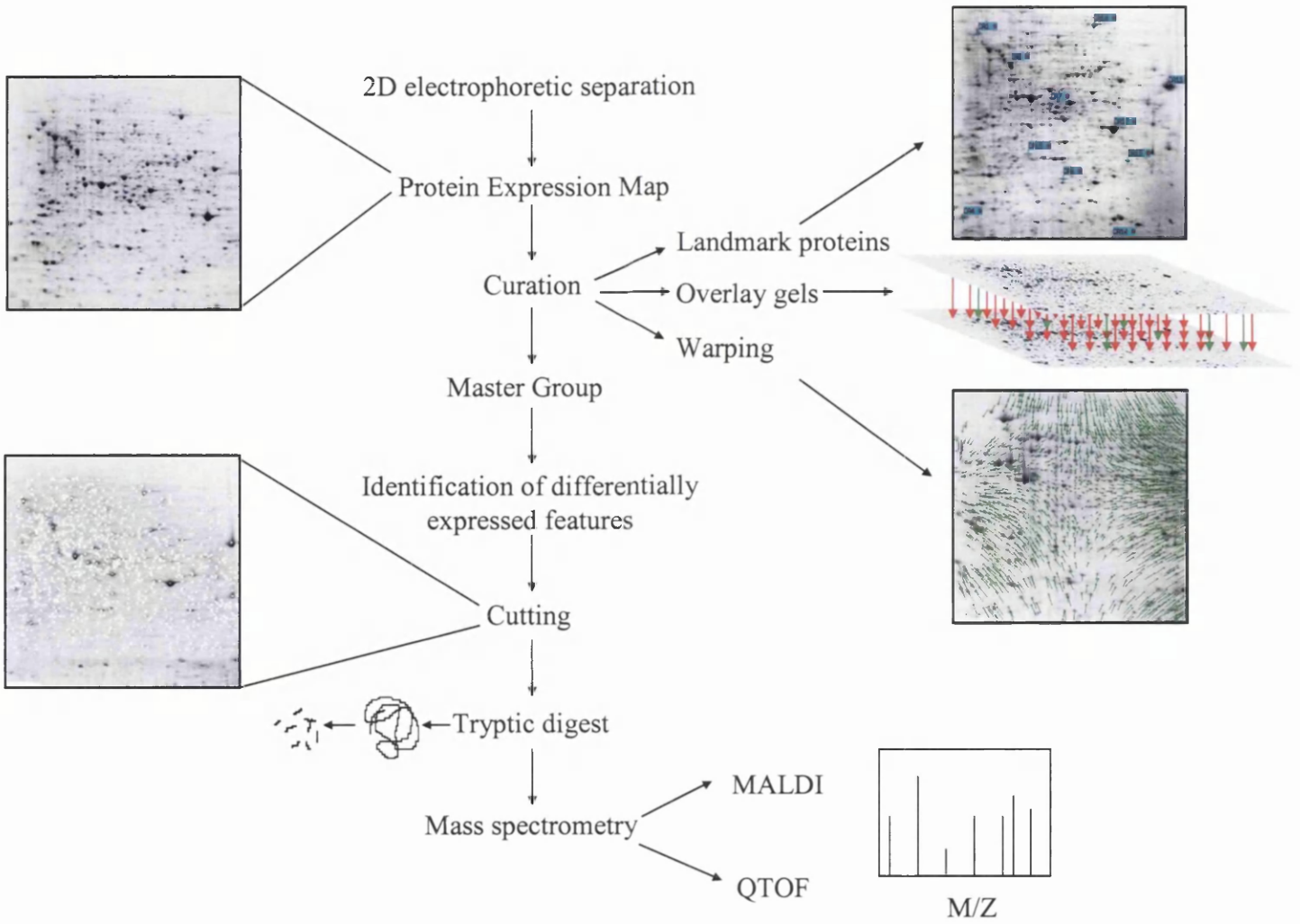


Figure 4.1 Overview of proteomics. This is flow diagram which highlights the key stages of proteomics starting with the separation of proteins by gel electrophoresis and finishing with mass spectrum data which reveals the mass:charge (M/Z) ratio of the peptide fragments detected by mass spectrometry.

more than one annotation for an MCI and these were called clustering annotations. These clusters have arisen from proteins that have co-migrated during separation on the 2-D gel. Therefore, in these instances the identification of the protein with the expression profile of interest cannot be unambiguously determined without further analysis.

Triplicate samples of each of the 16 stages of mammary gland development and duplicate samples of the TEBs were run on 2-D gels. The proteome images produced were scanned and processed using MELANIE III. Rosetta^{TM} software enabled a protein expression database to be created from these data (as previously described in materials and methods), which could then be interrogated for differential analysis.

Figure 4.2 shows images of an annotated MCI. These images demonstrate the pattern effect of the features surrounding the feature of interest which were then used as a guide to determine whether the same feature was present in the other gels within the study. Not all of the features on a gel can readily be seen by eye.

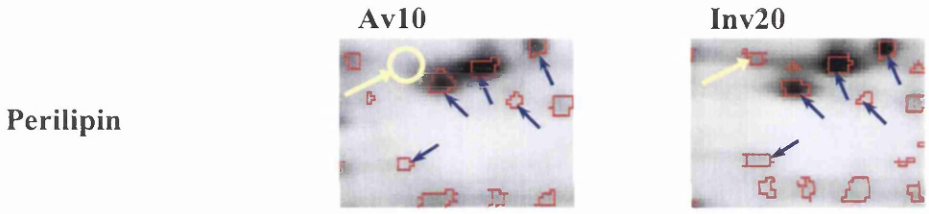


Figure 4.2 2-D gel images of perilipin. Computer imaging software has identified all detectable features present (highlighted in red) on the gels. A yellow arrow indicates the annotation of interest. A yellow circle indicates the area where the annotation would have appeared if the feature had been present in that stage. The features which were used as a guide are identified with blue arrows.

The first database created was composed of gel images from the Av10, pregnancy and early involution images. All of these stages passed Oxford GlycoSciences quality control criteria. This database was called the developmental database and was used to identify absolute changes of expression in pregnancy and involution. Differential analysis of the 16 mammary gland proteome images was accomplished by creating a second database. This was necessary because the lactation and early involution gels did not pass the Oxford GlycoSciences quality control criteria for automated analysis, as large areas of the gels were swamped by milk proteins (Figure 4.3A and B). The milk proteins were identified as smears across the 2-D gels; however, certain areas of these gels were not affected by the milk proteins and these areas were analysed manually. The database created to include these gels was called the Lactation/Involution database and contained all 16 stages of development. This was created primarily to determine changes of expression from lactation

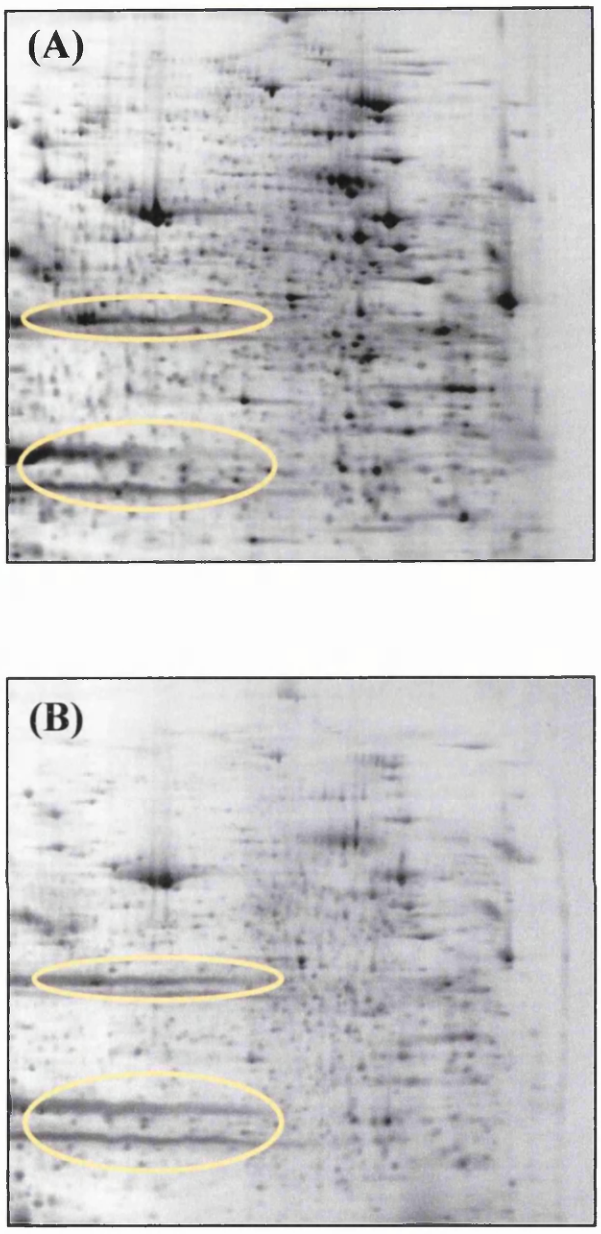


Figure 4.3 2-D gel images of L7 versus Inv1. These images are of representative samples containing high amounts of milk proteins which resulted in swamping certain areas of the gels. These areas have been highlighted by yellow ovals on the gel images (A) L7 (B) Inv1.

to involution which is an important process in mammary gland development, as it results in the commencement of apoptotic signalling to revert the gland back to its pre-pregnant state. All MCIs selected for mass spectrometry from this database were analysed manually to ensure that the changes were real. A third database was created called the TEB database in order to identify proteins which were specifically associated with the migration of TEBs. This was achieved by comparing the TEB proteome images to Av10.

The three databases analysed a total of 169 MCIs by mass spectrometry and identified 83 MCIs with a protein annotation, some of which were clustering proteins. Twenty nine MCIs were identified from the developmental database, thirty MCIs were determined in the Lactation/Involution database and 24 MCIs in the TEB database. Eight of the MCIs in the developmental database, 15 in the Lactation/Involution database and 10 in the TEB database had more than one protein annotation. Figure 4.4 summarises the steps taken to create these databases, the criteria used for selection for mass spectrometry analysis and archiving of the resultant annotations.

4.3.2 Identification and classification of annotations from the developmental database

The stages of development interrogated in the first database, known as the developmental database, were Av10, P4.5, P8.5, P12.5, P14.5, and P17.5, Inv5, Inv10 and Inv20. 7664 MCIs were identified in this database and the average number of MCIs detected per sample was approximately 2000 features.

An image processing algorithm was set to the specifications of Oxford GlycoSciences to determine gel artefacts. Artefacts such as gel imperfections, features created from residual OGT stain and areas of poor resolution were identified and remained in the databases but were not included in further analyses. As a result less than half the number of MCIs remained for protein expression analysis. A list of MCIs was created which selected all features present in at least two out of the three triplicate gels. A feature was present if it was identified in at least 2/3 triplicate gels; a feature was absent if it was identified in none or 1/3 triplicate gels. This list was subjected to further assessment and only MCIs which were present exclusively in pregnancy and absent in involution, or exclusively present in involution and absent in pregnancy were chosen for mass spectrometry analysis. This was called a presence/absence assessment. To determine the protein identifications of the selected MCIs, mass spectrometry analysis (MALDI and Q-TOF instruments) was used. The peptide sequences revealed by mass spectrometry were matched to sequence databases in order to identify the corresponding protein annotation.

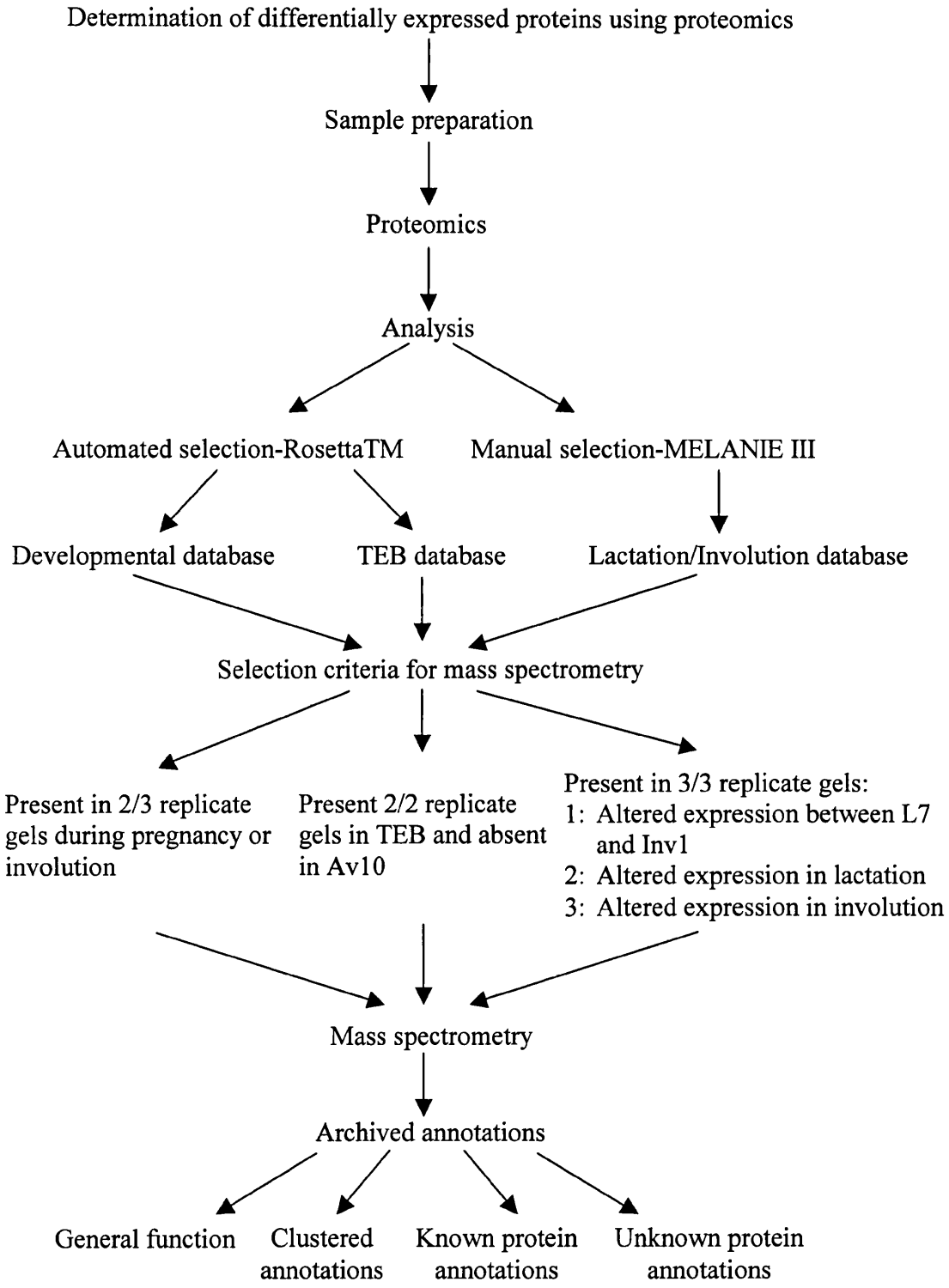


Figure 4.4 A flow diagram of the steps taken for determining differentially expressed protein annotations during mouse mammary gland development. This flow diagram can be used as a reference for the steps taken to determine a differential protein annotation from an extracted protein sample by proteomic analysis. Different criteria were set for three different databases interrogated. Once the annotations were identified and archived a small number were selected as candidate proteins which were subject to further investigation.

A total of 81 MCIs were selected for protein identification by mass spectrometry analysis from the developmental database. Their selection was based on their expression profiles across development. The gel containing the most abundant amount of protein for the MCI of interest was cut for mass spectrometry identification. The most abundant feature was used, because the greater the abundance of a feature used for mass spectrometry, then the greater the chance of obtaining a protein identification.

These 81 MCIs revealed 29 MCIs with annotated information. Eight of these MCIs were clustered annotations. Table 4.1 (see end of Chapter) lists each protein annotation according to the expression stage in which it was found to be present across mammary gland development. The data were then sorted by grouping the proteins according to their general function, and therefore Table 4.1 orders the annotations according to their functional group rather than the order that they were discovered. This was also the case for the Lactation/Involution and TEB databases. This enabled the data to be interrogated in a manner that could relate the function of the protein to the role it may play during development of the mammary gland. Protein annotations from the same cluster have been labelled with the same cluster letter.

In all three databases generated, the majority of the annotations were identified either as proteins classified as metabolic enzymes/mitochondrial proteins (ME/MP), or structural cytoskeletal proteins. Twelve ME/MP annotations were determined in the developmental database. Three were present during pregnancy and eight during involution. Carbonic anhydrase was classed as a ME/MP, but an error occurred with this selection as it was present both in the first three stages of pregnancy and in involution. Most of the annotations that were classified as cytoskeletal proteins were cytokeratins and actin associated proteins. This was the case for all three databases. A total of nine annotations were classed as cytoskeletal proteins in the developmental database.

Annotations were also grouped as adipose associated, milk, protein turnover, RNA processes, serum, signalling and transport proteins. Proteins that did not share a common function with the other annotations were classed under 'other'. Most annotations that have been classed as 'other' in all three databases are hypothetical or unique proteins with the exception of those proteins that cannot be classed under the other categories defined.

Predictably adipose associated and milk proteins were found to be expressed from these tissues. Perilipin was classed as an adipose associated protein and this was present in involution. A number of milk protein annotations were identified in the developmental and Lactation/Involution databases. During normal mammary development low levels of caseins are known to be present and there is also a huge surge in milk production during pregnancy and lactation. Four annotations were classified as milk proteins in the developmental database, all of which were casein proteins.

Eight protein annotations were classed as 'other' in the developmental database. One was a unique protein present in P17.5 and one was a hypothetical protein present in Inv5. Three annotations were identified under RNA processes and were present during involution. Unsurprisingly serum proteins were detected despite the efforts made to remove them from the samples and to avoid areas of known serum albumin identifications on 2-D gels.

Many signalling proteins will be required in the regulation of all phases of development during pregnancy, lactation and involution. Therefore careful consideration was taken with this group of annotations as follow-up proteins. Four signalling protein annotations were identified in the developmental database. Contrapsin expressed during P14.5 and annexin A2 expressed in Inv10 were of primary interest. The final annotation Ras-related protein Rab11 was classified as a transport protein, and was present during pregnancy.

The location of all of the annotations mentioned in this database are shown on a 2-D gel in Figure 4.5. This image highlights the wide area of coverage of the gel used for selection of MCIs for mass spectrometry.

Sequest and Oxford GlycoSciences internal programs were used to search for the identifications of the MCIs from the public (NCBI) and Incyte sequence databases. In particular nr, nr_prip, nr_rat-mouse sequences from the NCBI database and Lifeseq from the Incyte database were used to interrogate the peptide data from mass spectrometry.

The remaining 52 MCIs out of 81 selected for mass spectrometry analysis from the developmental database provided no protein identifications, which may have been due to a number of possibilities, for example insufficient amounts of protein for mass spectrometry, or loss of gel segment containing the feature for mass spectrometry analysis. The greatest impact on a protein identification was the information present in the sequence databases searched. At the time of this study there was restricted sequence information in the

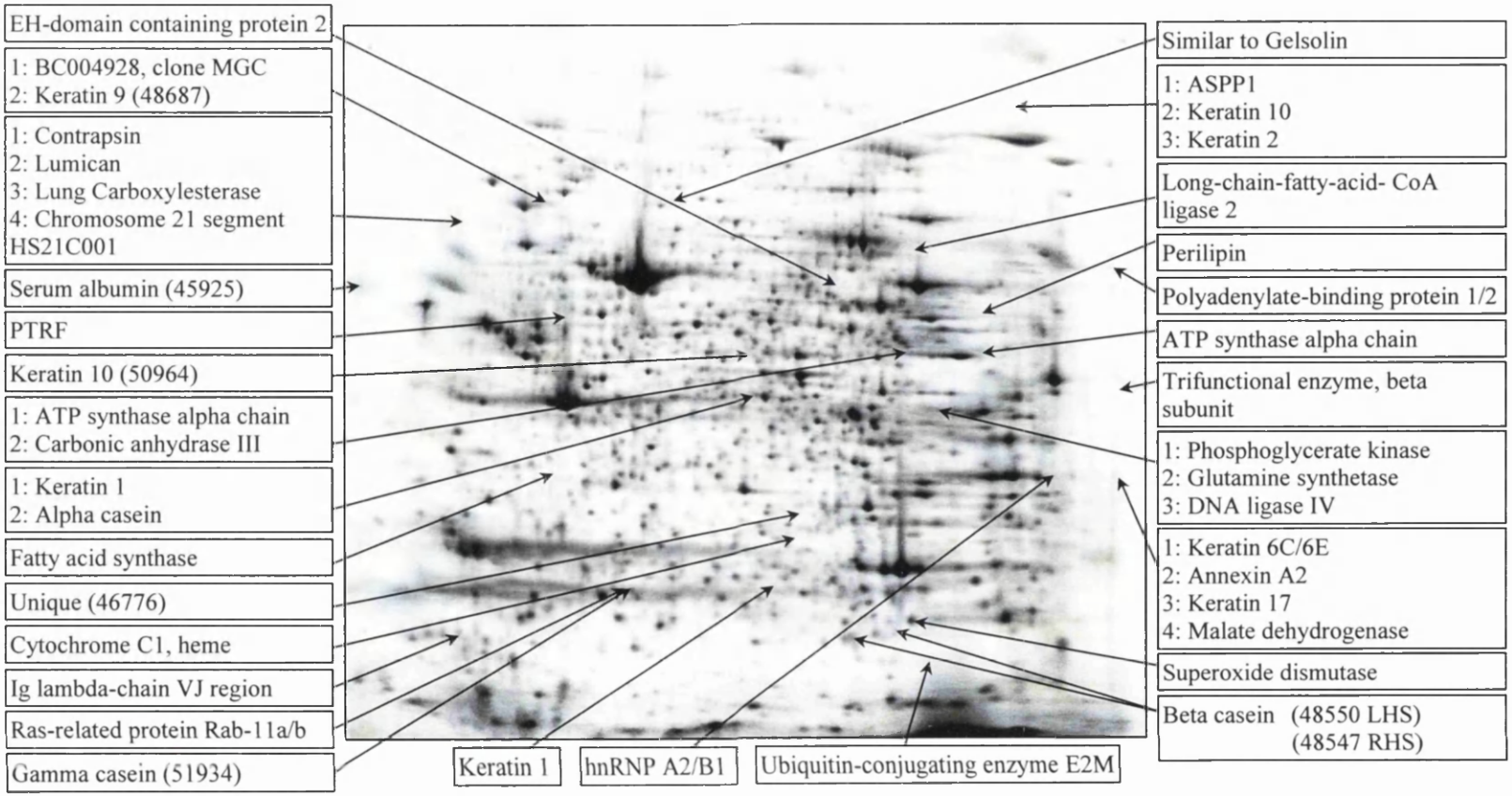


Figure 4.5 Location of annotations identified in the developmental database. Annotations are located according to their pI and Mr on a 2-D gel image. All annotations that clustered have been group together.

available databases searched for matching the peptide sequences generated by mass spectrometry to a protein.

These sequence databases were searched twice and the second time round more annotations were identified in all three databases created. The second round of sequence searches also removed annotations and this data has not been shown. Although the second sequence search was performed in October 2002, this data was not released for a further 10 months. Changes to the algorithms used by Sequest and Oxford GlycoSciences internal programs will have contributed to some of the altered annotations. Only annotations from the second round of searches have been listed in the tables provided for all three databases created (Tables 4.1, 4.2 and 4.3). The additional protein annotations identified from the second round of sequence database searches have been highlighted in these tables with a red asterisk, and those annotations that were altered by the second search have been highlighted by two asterisks. The Lactation/Involution and TEB databases were most affected by the second round of sequence database searches. Unfortunately at the time of the second sequence search the list of candidate proteins had been selected. Time limitations restricted the number of annotations selected for follow-up. Therefore, although some of the new protein annotations were potentially interesting, they were not selected as follow-ups from the developmental and Lactation/Involution databases.

4.3.3 Identification and classification of annotations from the Lactation/Involution database

The second database created concentrated on differential expression changes found during lactation and involution. This database was called the Lactation/Involution database. All 16 stages of mammary development were included, and only MCIs that were present in all three triplicate gels or absent in all three triplicate gels were considered. This was because the number of MCIs selected for mass spectrometry was limited to 44, and for quality assurance purposes as some gels were smeared with milk proteins.

The selection criteria used for mass spectrometry analysis were as follows. Where altered expression was used, this required a fold change limit of at least two or greater for significance: 14 MCIs selected were present in L7 and absent in Inv1; 7 MCIs were absent in L7 and present in Inv1; 8 MCIs were selected from an altered expression level from L7 to Inv1; 7 MCIs were present during all of lactation and finally 8 MCIs had an increase in expression from Inv1 to Inv20. A total of 30 out of the 44 MCIs selected from the Lactation/Involution database had protein annotations which had been identified using

mass spectrometry. Table 4.2 (see end of Chapter) lists these MCIs according to the general classification of the protein annotation. As with the developmental database, some of the MCIs were clusters (15 clusters) which had more than one protein identification. Again, as with the developmental database, not all selected MCIs provided an annotation.

Twenty of ME/MP annotations were identified in the Lactation/Involution database. The majority of these annotations were added after the second sequence search. Of the six cytoskeletal proteins identified in this database four were cytokeratin annotations. One annotation was classified as an adipose associated protein which increased in expression during involution. Four milk protein annotations were identified: two lactadherin annotations which increased in expression from L7 to Inv1, and two casein proteins. The casein expression profiles in this database were expanded to integrate the adult virgin, pregnancy and involution stages. It was possible to do this for the Lactation/Involution database, since each MCI selected for mass spectrometry was manually checked to determine their authenticity. The expanded profile of gamma casein is shown in Figure 4.6 and demonstrates the typical increase of expression during lactation and subsequent decrease during involution.

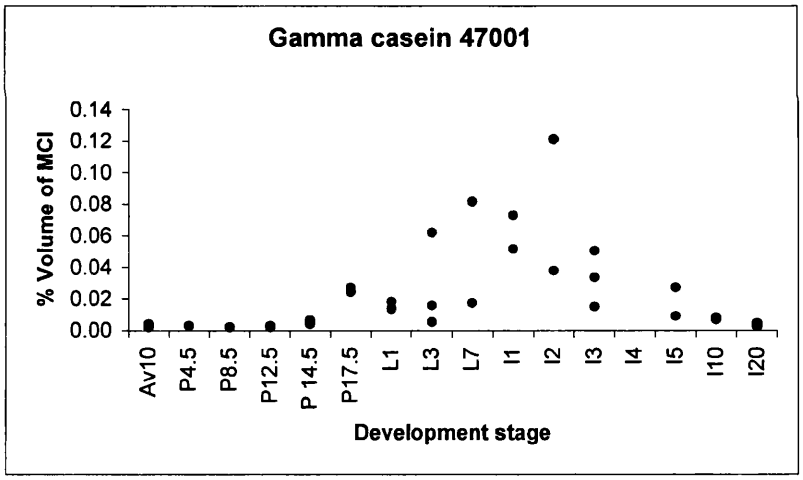


Figure 4.6 Expression plot of gamma casein across mammary gland development. The expression profile for gamma casein peaks during lactation. The values plotted are the percentage volume of the MCI on the 2-D gels; this is based on the breadth and depth of the feature in the gel. For selection purposes only MCIs which were present in 3/3 triplicate gels from the Lactation/Involution database were chosen as there was a limitation of 44 MCIs for mass spectrometry. Nevertheless the rule still applies that for a feature to be present it is found in at least 2/3 triplicate gels and for a feature to be absent it has to be found in one or none of the triplicate gels.

Sixteen annotations were classified as 'other' proteins. Four of the five unique proteins were present in L7 and absent in Inv 1. Three of the four hypothetical proteins were present throughout lactation. Of particular interest in this category was DNA replication

licensing factor mini-chromosome maintenance (MCM)3 (MCM3), a protein known to license DNA replication. This protein increased in expression from early to late involution.

Three protein turnover annotations were identified in the Lactation/Involution database. Two of these were proteasomes and both showed an increase in expression during involution. Only one annotation was classified under RNA processes (heterogeneous nuclear ribonucleoprotein D), and two MCIs were identified as serum proteins both of which were serum albumin. Five signalling protein annotations were identified; WD-repeat protein 1 (WDR1) was of particular interest showing an increase in expression from L7 to Inv1.

The final classification to mention in this database was the transport proteins. Four annotations were classed as transport proteins and none of these were considered for further follow-up work.

The location of all of the annotations identified in the Lactation/Involution database is shown in Figure 4.7. Again a wide area of coverage of the gel was used for selection of MCIs for mass spectrometry.

4.3.4 Identification and classification of annotations from the TEB database

The TEB proteomes were repeated in duplicate and not triplicate as for the other samples because of the difficulty in achieving sufficient numbers of TEBs. The database created included the TEB and the Av10 images. The comparison of the TEBs to the Av10 images was chosen, as the Av10 samples represent that of a resting virgin gland whereas the TEBs are involved in proliferation, differentiation and apoptotic processes. Since only duplicate samples of the TEBs were run, both gels had to contain the MCI for it to be counted as being present. A total of 108 MCIs were identified as being present in TEBs and absent in Av10. Of these, 44 were selected for mass spectrometry analysis. Twenty four MCIs provided a protein identification. Ten of these MCIs were clusters and the remaining 14 were single protein annotations. Table 4.3 (see end of Chapter) lists all the proteins identified in the order of their general function.

The positions of these annotations are shown on a 2-D gel image (see Figure 4.8). Based on the function of the annotations 17 were identified as cytoskeletal proteins, some of which were duplicated, for example keratin 2a (48442 and 49394). This group of proteins were mainly cytokeratin proteins (1, 2a, 8, 9, 10 and 19) and actin associated proteins

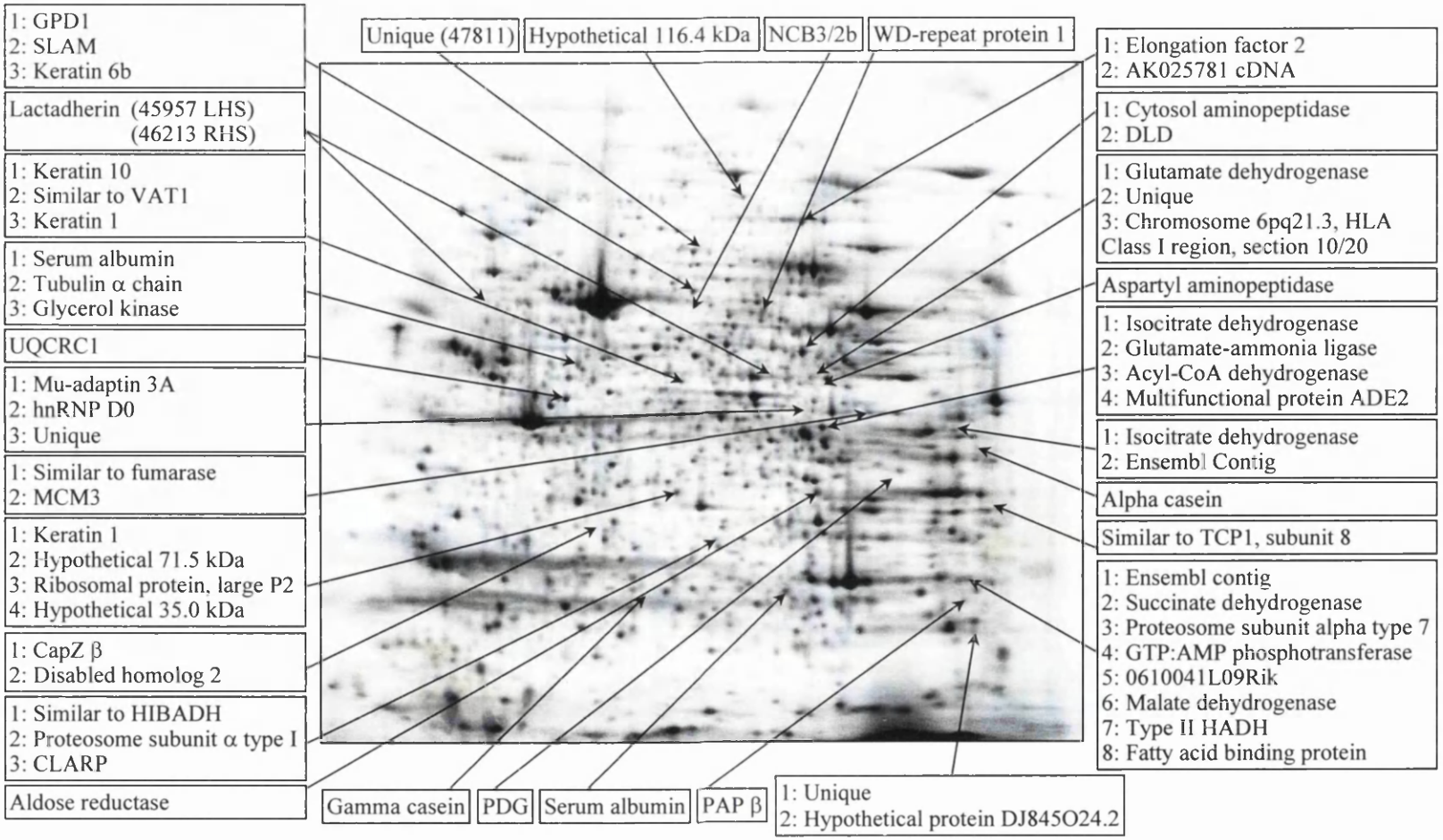


Figure 4.7 Location of annotations identified in the Lactation/Involution database. Annotations are located according to their pI and Mr on a 2-D gel image. All annotations that clustered have been group together.

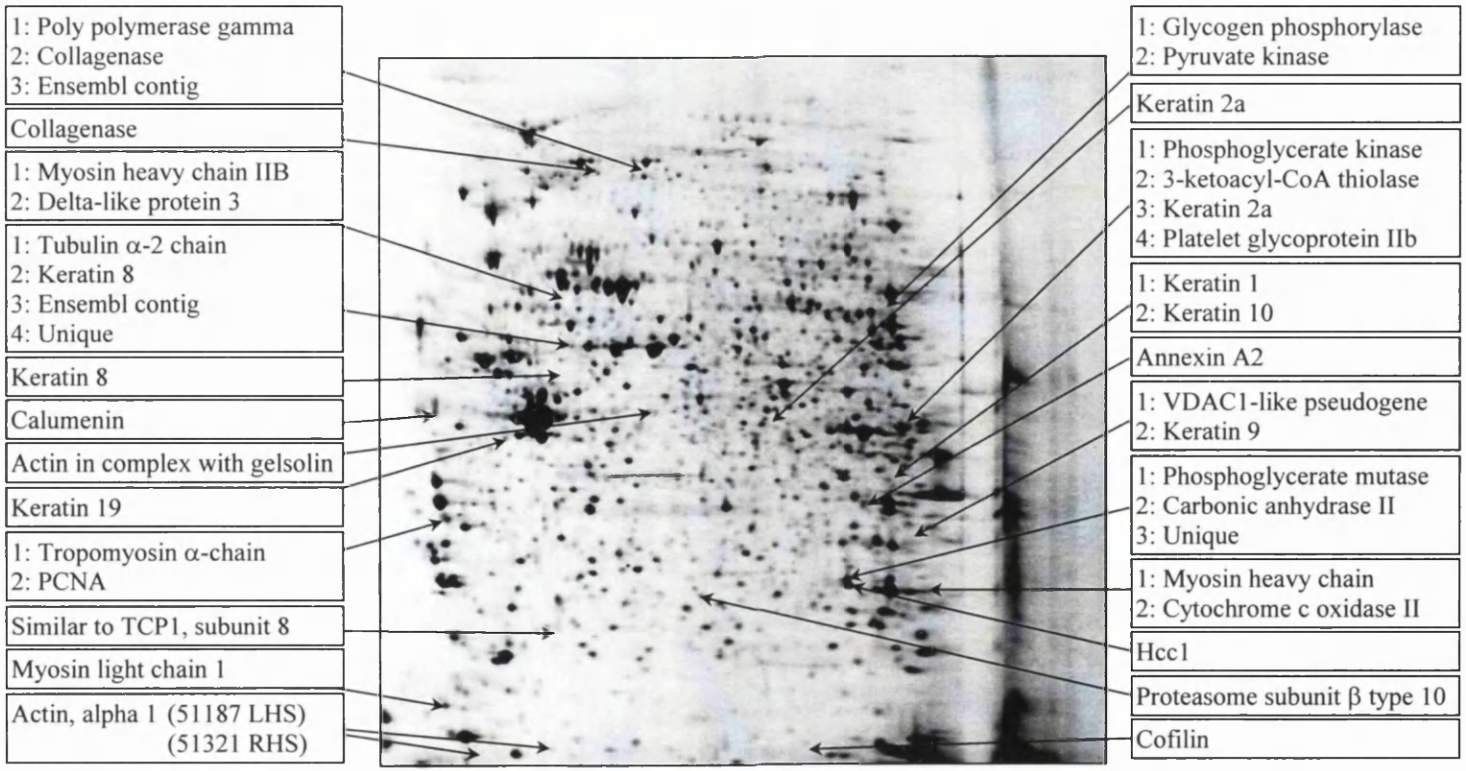


Figure 4.8 Location of annotations identified in the TEB database. Annotations are located according to their pI and Mr on a 2-D gel image. All annotations that clustered have been group together.

(actin alpha 1, gelsolin, tubulin, tropomyosin, cofilin and myosin). As with all the databases, quite a number of annotations were identified as being involved in energy production. A total of eight annotations were classified as ME/MP. Nine annotations were classed as 'other'. Two of the annotations in this group were identified as being collagenase which are most likely to be contaminants from the digestion procedure. Three functional categories remain in this group, namely protein turnover, RNA processes and signalling proteins. Calumenin, proteasome subunit beta type 10 and 'similar to' chaperonin containing *t* complex polypeptide 1 (TCP1) were grouped as protein turnover annotations. The latter annotation was also identified in the Lactation/Involution database. Both the pI and Mr were altered which suggests that the protein has different isoforms (Figures 4.7 and 4.8). An increase in weight suggests a glycosylation change and a shift in pI indicates a change in the phosphorylation state of the protein.

Two annotations were classed under RNA processes and one was classed as a signalling protein (annexin A2) which had also been determined in the developmental database. On comparing annexin A2 from the two databases they both shared the same Mr but their isoelectric focusing points were different. Annexin A2 from the developmental database was identified in the basic end of the gel, whereas in the TEB database it was more neutral (pI 7.48). This indicates that annexin A2 has different isoforms and the shift in location of annexin A2 on the gel is probably due to changes in the state of phosphorylation.

4.3.5 Selection criteria for candidate protein annotations

The selection criteria used at this stage of the project for candidate follow-up proteins was based on the biological interest of the annotation in terms of its expression profile over mammary gland development, the novelty of the protein to mammary gland development, relevance to cancer, the quality of mass spectrometry sequence data and the resources and materials available for that protein. Annotations which consistently appeared in other proteomic projects at Oxford GlycoSciences such as ME/MP were not selected for further work. This approach was taken in order to maximise the chances of selecting a potentially important annotation in mammary gland development.

The quality of the sequences was rated as good hits only if the annotations had either tandem sequence data from Sequest and/or Oxford GlycoSciences internal programs. Mass matches from MALDI mass spectrometry alone without tandem sequence data were not considered of sufficient quality to support confidence in the annotation determined. The algorithms set with these programs in determining a protein annotation have different

thresholds. The algorithm used by the Oxford GlycoSciences internal program is more stringent in determining a protein annotation (details not available) compared to Sequest. Therefore annotations with an Oxford GlycoSciences internal tandem match are considered with greater confidence. Eight annotations from the developmental and Lactation/Involution databases were selected for further work based on reasons outlined in the discussion of this Chapter, and a further three were chosen as they had been identified as unique or hypothetical proteins. Table 4.4 summarises the annotations chosen for further analysis. All of these annotations had at least a Sequest and/or Oxford GlycoSciences internal tandem peptide sequence alignment (Table 4.5).

Table 4.4 Candidate selection from the developmental and Lactation/Involution databases

MCI	Annotation	Expression	General Function	Cluster	Sequence Quality
49502	Contrapsin	Pregnancy	Serum protein	Yes	MALDI/Tandem Sequest
49502	Lumican	Pregnancy	Other	Yes	MALDI/Tandem Sequest
47015	Rab11a/b	Pregnancy	Transport protein	No	MALDI/Tandem Sequest
46776	Unique	Pregnancy	Other	No	Tandem Sequest
45950	WDR1	L7/Inv1	Signalling protein	No	MALDI/Tandem Sequest
48838	Unique/ Hypothetical	L7/Inv1	Other	Yes	Tandem Sequest no mass match
48474	Annexin A2	Involution	Signalling protein	Yes	MALDI/Tandem OGS
46330	MCM3	Involution	Other	Yes	Tandem Sequest no mass match
46036	Perilipin	Involution	Adipose associated	No	Tandem OGS
49021	PTRF	Involution	RNA processes	No	MALDI/Tandem OGS

PTRF (Polymerase I and transcript release factor); OGS (Oxford GlycoSciences)

4.4 Discussion

The discussion that follows first considers each of the annotations identified in the developmental and Lactation/Involution databases in relation to their functional attributes and then gives the reasons for including or rejecting them for future analysis. The TEB annotations are then discussed separately in section 4.4.11.

Although some proteins have been selected because of their association with breast cancer, the other area of focus of this project was to determine novel proteins regulating mammary gland development. Therefore not every protein that has been shown to be associated with breast cancer cell lines or tissue samples has been selected as a follow-up. Some of the protein annotations described in this discussion have similar functions but, as a limited number of annotations were selected for follow-up, not all could be chosen. In these

Table 4.5 Peptide sequence data matched to selected follow-up protein annotations

MCI	Annotation	MALDI Mass Match	Tandem (OGS)	Tandem (Sequest)	Tandem (Sequest-no mass match)
48474	Annexin A2	(1) LSLEGDHSTPPSAYGSVK (2) RAEDGSVIDYELIDQDAR (3) SYSPYDMLESIKK	(1) GVDEVTIVNILTNR (2) TPAQYDASELK (3) DIISDTSGDFR		
49502	Contrapsin	(1) DLQILAEFHEK (2) EVFTEQADLSGITETK		(1) MQQVEASLQPETL R	
49502	Lumican	(1) SVPMVPPGIKYLYLRNNQIDHIDEK (2) SLEYLDLSFNQMSK (3) FSALQYLR (4) SIPTVNENLENYYLEVNELEK		(1) ISNIPDEYFK	
46330	MCM3				(1) SQDSMSSDTAR
46036	Perilipin		(1) IPALQYPPEK (2) LSLMEPESEFR (3) PSVMEPILGR		
49021	PTRF	(1) IIGAVDQIQLTQAQLEERQAEMEGAVQSIQGELSK (2) VMIYQDEVK (3) LVPVER (4) KSFTPDHVVYAR (5) VPPFTFHVK (6) ATEMVEVGPEDDEVGAER	(1) KLEVNEAELLR		
47015	Rab11a/b	(1) (a/b) AQIWDTAGQER (2) (a) RENDMSPSNNVVPIHVPPTTENKPK (3) (b) NILTEIYR		(1) (a/b) STIGVEFATR	
45950	WDR1	(1) DLAWTEDSK (2) FTIGDHSR		(1) VFASLPQVER	
46776	Unique			(1) CVNIQMLQGVK	
48838	Unique				(1) TGIIIILSEVK
48838	Hypothetical				(1) KPVEDCPR

The peptide sequences identified are from MADLI mass matched data followed by tandem sequence data which mapped the sequence data to the protein annotation by using SEQUEST and OGS internal programs which searched the sequence databases available (private and public).

circumstances the annotation with the best sequence quality was favoured. This also created a list of candidate proteins with a diverse range of functions.

4.4.1 Identification of milk proteins

A number of the proteins identified were associated with milk and were grouped as milk proteins. All except for one were caseins. Caseins comprise a group of acidic proline-rich phosphoproteins that incorporate and provide a rich source of calcium in milk. Their expression is regulated by hormones at both the transcriptional and post-transcriptional level (Schmitt-Ney *et al.,* 1991). Low levels of caseins are expressed in the virgin mouse, but their synthesis increases dramatically during pregnancy. Progesterone represses casein production in pregnancy. However, during lactation the removal of progesterone together with the induction of prolactin causes the levels of these milk proteins to peak. Cessation of the suckling stimulus caused by the removal of the pups from the lactating mother leads to a decline in the circulating levels of prolactin and hence a decrease in casein levels.

It was thought that the casein data could be used as a good positive control to validate the proteomic data, as their expression plots across mammary development were already known. Therefore none were selected for follow-up work. However as the casein data were examined in greater detail several discrepancies were found. The two beta caseins identified in the developmental database were present only in late involution and not in pregnancy. The clustered annotation of alpha casein in the developmental database showed expression only in P17.5. It would have been helpful if these MCIs could have included the lactation and early involution samples to determine their overall expression profile across development. However, without locating the MCI within these additional gel images in lactation and involution it would not be possible to determine whether the MCI resided in the smeared regions of the gel. Additionally it is possible that alpha casein was not the correct annotation for MCI 51886 as it was a cluster protein. After expanding the data for the alpha casein (48449) found in the Lactation/Involution database, it did not resemble the expected profile across mammary gland development.

Confidence grew with the data when plotting the graph for gamma casein 47001 (Figure 4.6). This showed the predicted expression profile for this protein across mammary gland development. Gamma casein was also found in the developmental database during the later stage of pregnancy.

In summary, not all the caseins showed the expected expression profiles. In some of the cases in the developmental database, it would have been helpful to have had the expression data across all of development. In some of the cases found in the Lactation/Involution database, the expanded data did not reveal the expected expression profile. One possible explanation for these unexpected expression profiles is that proteomics has detected different isoforms of these proteins which have not previously been identified.

Lactadherin, one of the major glycoproteins, was detected twice in the Lactation/Involution database. Lactadherin is a component of the human milk fat globule (Ceriani *et al.,* 1983; Peterson *et al.,* 1990). It is expressed in the epithelial cells during lactation but undergoes a 10-20 fold increase in expression during involution (Collins *et al.,* 1997). The expression of this annotation determined by proteomics further supported the validity of this technique as it confirmed previously reported work. This annotation was not selected for further interrogation as this protein was already well known in mammary gland development.

4.4.2 Identification of metabolic enzymes and mitochondrial proteins (ME/MP)

The mammary gland is a complex tissue to study in terms of the regulatory processes that take place over its development. It was therefore not surprising that a large number of the proteins identified were associated with cellular metabolism. These have been listed as ME/MP. The expression of these enzymes will have fulfilled the high energy requirements of the mammary cells in order to perform these developmental processes. As they are usually expressed at high levels, the probability of detecting these proteins by mass spectrometry analysis was increased. Thus, to minimise their selection for mass spectrometry identification, the expression data of this project was propagated against other proteomic projects. Propagation is the comparison of different project data. Protein annotations from previous studies that have for example been identified in cell respiration can be avoided for mass spectrometry selection. This technique can rule out proteins which are not regarded as interesting follow-ups. As these cellular metabolic enzymes were not priority, no further investigations were made with these annotations.

4.4.3 Identification of cytoskeletal proteins

The cytoskeletal proteins consisted of cytokeratins, actin, gelsolin and tubulin. Keratins are structural proteins which are classified as two families: type I (acidic) and type II (basic). One member of each class forms a heterodimer which is required for forming a mature keratin intermediate filament. Keratins were detected in both databases and for a few of these proteins the keratin pair was annotated, for example keratin 1 and 10. They were also

expressed during the same stages of development when expanding the information from the Lactation/Involution database to include all of the time points. However, it was decided that the keratin data from the two databases would not be taken any further. This was mainly because of the common problem of contamination that arises with proteomics due to the shedding of keratins from the mouse epidermis or even human. Keratin contamination was kept to the minimum during sample preparation and gel electrophoresis, but it was never completely absent.

F-actin capping protein beta subunit (CapZ beta) was found to be present in L7 and absent in Inv1. Capping proteins are major determinants of the cytoskeleton. They assemble, but do not sever, actin filaments and stabilise them by binding to their ends and thereby blocking the exchange of subunits at these ends (Provost *et al.,* 2001; Rohrig *et al.,* 1995). If this protein had been identified earlier it could have been investigated further, as it is known that structural changes do begin to take place early in involution. It seems that the start of structural tissue remodelling (involution) caused the structural protein F-actin to be inactivated.

Another capping protein detected was 'similar to' gelsolin. It was assumed that this protein performed the same functions as gelsolin. This actin-binding severing protein has a major role in the reorganisation and function of the actin cytoskeleton. The deletion of this gene results in failure of ductal elongation at puberty and a lack of terminal branching after the epithelium has filled the fat pad (Crowley *et al.,* 2000). Gelsolin was not selected for further research as a considerable amount of work has been invested in mammary gland development.

Tubulin alpha chain was found to increase in expression from L7 to Inv1. Tubulin is a constituent of the cytoskeleton and is the major component of the microtubules. Changes in the cytoskeleton during mammary gland development were not forefront for investigation, and therefore tubulin was not studied any further.

4.4.4 Identification of signalling proteins

Signalling proteins appeared in both databases. The eps15 homology (EH) domain containing protein 2 (EHD2) was expressed from early to mid pregnancy. These proteins are often associated with the regulation of intracellular protein transport/sorting membrane trafficking (Salcini *et al.,* 1997) as well as with endocytosis (Carbone *et al.,* 1997; Mayer, 1999). Their apparent function in signal transduction downstream of the receptor tyrosine

kinases indicates that these proteins may participate in a variety of signalling cascades. Although EHD2 is potentially interesting to mammary development, it was not selected for further investigations because of poor quality sequence data.

The clustering MCI identifying apoptosis stimulating protein of p53 (ASPP)1 belongs to the ASPP family of proteins interact which with and enhance p53 induced apoptosis. The inhibition of endogenous ASPP causes the suppression of the apoptotic function of endogenous p53 as a result of an apoptotic stimulus. ASPP is frequently down-regulated in human breast cancers that express wild-type p53 (Samuels-Lev *et al.,* 2001). Despite ASPP1's novelty to mammary gland development, it was not chosen as a candidate annotation because of poor quality sequence data.

Annexin A2, another clustering protein, was discovered in the developmental database. Annexins are a family of proteins that are able to bind with membrane phospholipids in a calcium dependent manner (Crumpton and Dedman, 1990; Geisow and Walker, 1986). Annexin A2 has been found to be present in low levels across pregnancy and lactation (Handel *et al.,* 1991), and is increased in several human cancers (Schwartz-Albiez *et al.,* 1993). As proteomics had found annexin A2 to be present in late involution it was interesting to confirm this pattern of expression. The role of annexin A2 in breast cancer studies was also an important consideration in the decision to follow it up as a key protein.

WDR1 was identified in the Lactation/Involution database. WD repeat proteins are important in a variety of cellular processes such as cytoskeletal organisation, signal transduction, transcriptional regulation, vesicular trafficking and cell division (Neer *et al.,* 1994; Smith *et al.,* 1999; Neer *et al.,* 2000). This appeared to be another interesting protein to work with because of its role as an actin severing binding protein. It is known that there is much tissue remodelling taking place at the start of involution and therefore it seemed plausible that this protein could have an important role with the onset of apoptosis.

Elongation factor 2 (EEF2) was expressed throughout lactation, although the identification of this MCI was not absolute as it was a cluster. Protein synthesis consumes a large amount of metabolic energy of mammalian cells which is mainly due to peptide chain elongation. EEF2 mediates the translocation step of elongation, the rate of which can be modulated by the phosphorylation of EEF2 (Merrick and Nyborg, 2000). The function of this protein with regard to mammary gland development was not particularly intriguing, as during lactation there is much energy consumption due to the vast production of milk proteins.

Contrapsin was a clustering protein. Contrapsin is a serine protease inhibitor and belongs to the serpin superfamily. Serpins control many physiological functions including blood coagulation, fibrinolysis and aspects of the inflammatory response. Luminal epithelial cells are known to be involved in the phagocytosis of neighbouring apoptotic epithelial cells during involution which is probably where these immune response genes are expressed (Stein *et al.,* 2004). Although contrapsin is expressed during pregnancy which may be caused by an immune response to the embryo, it has not been linked to mammary gland or breast development. Contrapsin was therefore selected as a candidate protein.

Another protein that has been classified as a signalling protein is disabled homolog 2 (DOC2). Expression of DOC2 has been demonstrated in the developing rat brain, suggesting a role in proliferating cells and this could therefore also imply a possible role in mammary gland development (Korteweg *et al.,* 2000). Unfortunately this annotation was added to the database at a later date when candidate proteins had been chosen.

The signalling lymphocytic activation molecule (SLAM) detected throughout lactation is a protein that is involved in T-cell stimulation (Cocks *et al.,* 1995). Expanding its expression profile across all of development showed that it increased at the onset of involution. Although this was a clustering protein, an infiltration of immune cells into the mammary gland has been reported in many animals during involution, including mice (Stein *et al.,* 2004). This evidence supports the possibility of the MCI being this annotation. SLAM was not investigated because of its known function in mammary gland development.

The last signalling protein to be discussed is caspase-like apoptosis regulatory protein (CLARP). This protein had been selected for mass spectrometry because it increased in expression during involution. It is an apoptosis regulator protein that could function as a vital link between cell survival and cell death pathways (Han *et al.,* 1997). Although this was a clustering protein, it would have been an important protein to select for further investigations. However, this was another late edition to the proteomics data.

4.4.5 Identification of transport proteins

The Ras-related protein Rab11a/b was identified as a single annotation but the peptide sequence which linked the feature to Rab11 could not distinguish between Rab11a or b. Rab11a modulates transport through recycling endosomes (Ullrich *et al.,* 1996) and Rab11b is essential for the transport of internalised transferrin (Schlierf *et al.,* 2000).

Rab11 was chosen as a priority follow-up because there had been no previous link of this protein to mammary or breast biology. Rab11's expression during pregnancy (P12.5-P17.5) and its link to vesicle transport suggest that Rab11 plays a role in milk secretion.

The adapter-related protein complex 3 mu 1 subunit (Mu-adaptin 3A) was identified later in the project. This protein facilitates the budding of vesicles from the golgi membrane and may be involved in the trafficking to lysosomes (Dell' Angelica *et al.,* 1999). Earlier knowledge of the protein would have been helpful in deciding whether to investigate it further.

'Similar to' vesicle amine transport protein 1 ('similar to' VAT-1) was classed as a transport protein. VAT-1 is a major component of synaptic vesicles that regulate neurotransmitters in nerve terminals (Matteoli and De Camilli, 1991). This protein was not taken forward for more research because of its late submission to the project, despite the fact that VAT-1 had been localised to a region encompassing the breast cancer gene *BRCA1* (Friedman *et al.,* 1995). The detection of 'similar to' VAT1 could also imply a different role to VAT1.

ADP ribosylation factors (ARFs) are part of the Ras GTP-binding proteins and their activation is controlled by GTP-ase activating proteins (GAP) (Rothman, 1994). Protein tyrosine kinase 2 (PYK)-associated protein β (PAPβ) contains ARF-GAP domains. Despite the obvious link with Rab11, it was not selected because of its late submission to the project. It appears that vesicle transport plays a part in mammary gland development, as five annotations have been linked to this role.

The final protein to discuss under the classification of transport proteins is sodium bicarbonate cotransporter 3/2b (NCB3/NCB2b). This protein plays an important role in transepithelial transport and pH regulation (Jacobson, 1981; Boron and Boulpaep, 1983). This protein was not taken through to the next stage of the project because of prioritisation.

4.4.6 Identification of protein turnover proteins

Only a few proteins were classed as having a role in protein turnover, one of which was 'similar to' chaperonin containing TCP1. Chaperonin containing TCP1 proteins are important in the folding of newly translated cytoskeletal proteins (Gao *et al.,* 1993; 1992). Assuming that the 'similar to' chaperonin containing TCP1 functions in a closely related manner to TCP1, one explanation for its absence during involution could be due to the rapid loss of epithelial cells via apoptosis.

Two proteasome subunit alpha proteins were detected (subunits 1 and 7). Proteasomes appear to be involved in the ATP/ubiquitin-dependent non-lysosomal proteolytic pathway (Orlowski, 1990). These proteins were not selected for further work, since their functional relevance in mammary gland development was not clear.

4.4.7 Identification of RNA processing proteins

A few proteins have been classed as RNA processing proteins, one of which was the heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP A2/B1). These proteins appear to influence pre-mRNA processing, mRNA processing and transport (Kamma *et al.,* 1999; Matunis *et al.,* 1994). Reports suggest that it may be involved in growth regulation and cancer development (Zhou *et al.,* 1996). This annotation was not selected further as it had poor sequence data.

PTRF was a single cluster identification. It is involved in the transcription termination of the RNA polymerase I reaction (Jansa *et al.,* 1998). Due to the unique expression profile of this protein, its novel biological interest and good sequence data from mass spectrometry, this protein was selected for further work.

Proteomics identified polyadenylate-binding protein. These proteins are required for poly(A) shortening and translation initiation (Deo *et al.,* 1999). However, the peptide sequence matching to this protein could not distinguish between protein 1 and 2. For this reason and because of the weak peptide sequence data, this protein was not investigated.

The last protein to discuss in this category is heterogenous nuclear ribonucleoprotein D (hnRNP D0). Its function is unknown but it has been implicated in the regulation of RNA instability (Kiledjian *et al.,* 1997). Again this protein was not selected for further work because other annotations had better sequence data from mass spectrometry.

4.4.8 Identification of proteins adipose associated proteins

Two proteins were classed as adipose associated proteins, one of which was perilipin. Perilipins are proteins which are associated with the surface of lipid storage droplets. Although they are expressed in adipocytes during lactation (Blanchette-Mackie *et al.,* 1995, Neville *et al.,* 1998), their profile is unknown throughout mammary development. Therefore this protein was taken to the next stage of the project.

The other protein in this category was fatty acid binding protein (adipocyte). Its function is to transport lipids in adipocytes (Matarese and Bernlohr, 1988). This protein showed an increase in expression during involution which is probably due to replenishing adipocytes. Based on this hypothesis it was decided not to select this protein as a follow-up.

4.4.9 Identification of proteins classed as 'other'

The final classification of proteins made was those whose function did not come under any other classes named and these were called 'other' proteins. Quite a number of hypothetical and unique proteins were determined in the two databases and these were classed as 'other'. These were of significant interest because of the novelty to mammary gland development and to scientific literature. Therefore these were selected for further research, although it should be noted that because of time constraints research was focused on the hypothetical and unique proteins detected earlier in the project and not those detected from the second sequence database search.

The peptide sequences that had been linked to part of a chromosome, RIKEN cDNA sequences and ensembl contigs were not taken forward for more detailed analysis, as more information about the sequences was required and the majority had been determined in the second round of sequence database searches.

Ubiquitin conjugating enzyme E2M was another protein classed as 'other'. This protein catalyses the attachment of ubiquitin-like protein NEDD8 to other proteins (Osaka *et al.*, 1998). Very little information and resources were available for this protein. These factors contributed to the decision not to follow-up this protein further.

Lumican was a clustering protein which was detected in the developmental database. Lumican is a small leucine-rich proteoglycan. Lumican may be involved in the maintenance of tissue stromal structure as it can organise collagen fibrils (Scott, 1992; 1996). However it was the association of this protein with breast tissue, plus the over expression of lumican in invasive carcinomas which highlighted it for further investigation.

DNA ligase IV was expressed only in Inv20. A breast cancer case control study revealed that a polymorphism in DNA IV ligase was associated with a decrease in breast cancer risk (Kuschel *et al.*, 2002). Despite this link with breast cancer, DNA IV ligase was not chosen as a protein for further research as other proteins had taken precedence.

DNA replication licensing factor MCM3 (MCM3) ensures that DNA replication occurs only once per cell cycle (reviewed by Stillman, 1996). MCM3 consistently labels a higher proportion of tumour cells compared to conventional proliferation markers (Stoeber *et al.*, 2001; 1999), and may in fact be detecting cells with proliferative potential. MCM3 was selected as a candidate protein because, its expression during normal mammary development had not been previously reported. Interestingly MCM3 had been detected during involution in the mammary gland and reports have shown MCM3 to be cleaved during apoptosis which is thought to prevent continuation of mitotic events (Schwab *et al.*, 1998).

4.4.10 Artefacts of contamination

The serum albumin and immunoglobulin annotations are most certainly artefacts of serum contamination. Although serum albumin appears to be differentially expressed, experience at Oxford GlycoSciences has found that serum albumin annotations are often not the correct annotations for the expression profile in question. These proteins are able to adhere to gel surfaces and gel plates, and therefore they can contaminate gel segments cut for mass spectrometry. As serum albumin produces good spectral data, often the correct protein is not detected.

4.4.11 TEB protein annotations

The TEB database was based on present absent calls. Differential expression was not used as a method of selection, as this change in expression may have only arisen because the TEB sample is a concentrated epithelial cell population whereas the adult virgin sample is a mixture of epithelial and stromal cells. If a sufficient collection of purified ducts had been possible at the time the start of this part of the study, it would have been preferable to compare TEB and ducts rather than the whole mammary gland. This would have identified proteins that were specific to TEBs. Some of the proteins identified in the TEBs were pursued and the data were published by Morris and colleagues (Morris *et al.*, 2004). However, the main focus of this thesis concentrated on proteins expressed in pregnancy and the lactation involution switch.

The majority of the 24 annotations identified were either ME/MP or structural proteins. Most of the cytoskeletal annotations from the TEB database have been associated with axonal growth cones, namely actin alpha 1, gelsolin, cofilin, myosin IIB, myosin light chain 1, tropomyosin alpha and tubulin alpha. The family of proteins known as tropomyosins are components of the actin based cytoskeleton which stabilise actin

microfilaments (Broschat *et al.,* 1989; Kojima *et al.,* 1994). Actin alpha 1 and two of its binding proteins, gelsolin and cofilin, were identified in this database. Actin binding proteins and gelsolin are important in growth cones as they are actin severing binding and depolymerising proteins (Lu *et al.,* 1997). Cofilin is also an actin severing protein; it enables actin filaments to elongate (Chan *et al.,* 2000; Zebda *et al.,* 2000).

The interaction of actin and myosin filaments enables forward movement of growth cones. Traction force is generated from peripheral actin binding to a substrate and studies have shown that myosin IIB is required for traction force (Brown and Bridgman, 2003; Bridgman, 2002; Bridgman *et al.,* 2001). Myosin light chain was detected in the TEB database and is known to phosphorylate myosin II (Schmidt *et al.,* 2002). The cytoskeletal microtubule alpha tubulin was also detected and is important in growth cone turning (Jockusch and Jockusch, 1981; Miller *et al.,* 1987; Tanaka and Kirschner, 1991). The annexin A2 annotation was also identified in the developmental database. It has been shown to be involved in changes of the cytoskeletal structure (Ma *et al.,* 1994).

Six keratins were identified in this database, four of which were expressed as pairs (keratin 1, 2a, 9 and 10). These keratins are epidermal specific proteins. Keratins have previously been reported to be preferentially expressed in TEBs (Grimm *et al.,* 2002; Rosen, 2004 in preparation). However, this data must be approached cautiously as it may be a result of contamination.

4.4.12 Selection of candidate proteins annotations from the developmental and Lactation/Involution databases

The following annotations were selected for further investigations: annexin A2, contrapsin, lumican, MCM3, perilipin, PTRF, Rab11a/b, WDR1 and three annotations from two MCIs (46776 and 48838) which were unique/hypothetical annotations.

Chapter 5 presents data accumulated to confirm the proteomics results. The techniques used to achieve this were western blotting, IHC, TaqMan and microarray analysis.

Table 4.1 Functional classification of annotations identified in the developmental database

MCI	Protein Annotation	Mr	pI	Av10	P4.5	P8.5	P12.5	P14.5	P17.5	Inv5	Inv10	Inv20
Cytoskeletal proteins												
50964	Keratin 10	48913	6.06	1	1	1	3	1	1	0	1	0
49144	Keratin 1	22465	6.23	0	1	1	2	2	3	1	1	0
51886	**Cluster C:** Keratin 1	44440	6.11	0	0	0	0	0	3	0	0	0
45715	Similar to gelsolin	95584	5.68	0	0	1	0	1	3	0	0	0
45623	**Cluster D:** Keratin 10	147965	9.37	1	1	0	0	0	0	3	0	0
45623	**Cluster D:** Keratin 2a	147965	9.37	1	1	0	0	0	0	3	0	0
48474	**Cluster F:** Keratin 6C/Keratin 6E	35426	10.39	1	0	0	1	0	1	1	3	1
48474	**Cluster F:** Keratin 17	35426	10.39	1	0	0	1	0	1	1	3	1
48687	**Cluster H:** Keratin 9	94977	5.07	1	1	1	1	1	0	1	1	3
Adipose associated proteins												
46036	Perilipin	56453	9.15	1	0	1	1	1	1	1	1	3
Metabolic enzymes and mitochondrial proteins (ME/MP)												
48413	**Cluster A:** ATP synthase alpha chain, mitochondrial	50483	7.39	1	2	3	3	1	1	1	2	1
48413	**Cluster A:** Carbonic anhydrase III	50483	7.39	1	2	3	3	1	1	1	2	1
49502	**Cluster B:** Lung carboxylesterase	86075	4.61	1	0	0	0	3	0	1	0	1
48996	Long-chain-fatty-acid--CoA ligase 2	74046	7.82	0	1	1	0	0	0	3	1	1
51590	Superoxide dismutase, mitochondria	18519	7.61	0	0	1	3	0	0	0	0	0
48789	Fatty acid synthase	34927	5.07	0	1	1	1	1	1	2	1	3
48431	Trifunctional enzyme, beta subunit	45169	10.04	0	0	1	0	0	0	0	3	1
48474	**Cluster F:** Malate dehydrogenase, mitochondrial	35426	10.39	1	0	0	1	0	1	1	3	1
46203	ATP synthase alpha chain, mitochondrial	50522	8.89	1	0	0	0	0	1	1	1	3

MCI	Protein Annotation	Mr	pI	Av10	P4.5	P8.5	P12.5	P14.5	P17.5	Inv5	Inv10	Inv20
Metabolic enzymes and mitochondrial proteins (ME/MP)												
46389	**Cluster G:** Phosphoglycerate kinase	42661	8.00	1	0	0	0	0	0	0	1	3
46389	**Cluster G:** Glutamine synthetase	42661	8.00	1	0	0	0	0	0	0	1	3
48807	Cytochrome C1, heme	29287	6.34	0	0	1	0	1	1	0	1	3
Milk proteins												
51886	**Cluster C:** Alpha casein	44440	6.11	0	0	0	0	0	3	0	0	0
51934	Gamma casein	22377	5.53	0	0	0	0	0	3	0	0	0
48547	Beta casein	17535	7.36	1	0	1	0	0	0	2	3	2
48550	Beta casein	16749	6.85	0	0	1	0	0	1	3	3	1
Other												
51175	Ubiquitin-conjugating enzyme E2M	15088	7.81	1	2	3	1	2	1	0	0	1
50118	Immunoglobulin lambda-chain VJ region	16736	4.61	0	2	3	1	2	3	1	1	0
49502	**Cluster B:** Lumican	86075	4.61	1	0	0	0	3	0	1	0	1
49502	**Cluster B:** Chromosome 21 segment HS21C001	86075	4.61	1	0	0	0	3	0	1	0	1
46776	Incyte Unique	32179	6.35	1	0	0	0	1	3	0	1	0
46619	**Cluster E:** Hypothetical 35.6 kDa/Unknown**	36003	9.87	0	0	0	1	1	1	3	1	0
46389	**Cluster G:** DNA ligase IV	42661	8.00	1	0	0	0	0	0	0	1	3
48687	**Cluster H:** BC009928, clone MGC**	94977	5.07	1	1	1	1	1	0	1	1	3
RNA processes												
46619	**Cluster E:** Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1)	36003	9.87	0	0	0	1	1	1	3	1	0
49021	Polymerase I and transcript release factor (PTRF)	53480	5.15	1	0	0	0	0	0	3	0	0
48370	Polyadenylate-binding protein 1 or 2	69011	9.98	0	0	0	0	1	1	0	3	1

MCI	Protein Annotation	Mr	pI	Av10	P4.5	P8.5	P12.5	P14.5	P17.5	Inv5	Inv10	Inv20
Serum proteins												
45925	Serum albumin	64152	4.40	1	1	2	1	3	2	1	0	0
Signalling proteins												
48383	EH-domain containing protein 2 (EHD2)	60695	6.75	1	2	3	2	3	0	1	1	1
49502	**Cluster B:** Contrapsin	86075	4.61	1	0	0	0	3	0	1	0	1
45623	**Cluster D:** ASPP1	147965	9.37	1	1	0	0	0	0	3	0	0
48474	**Cluster F:** Annexin A2	35426	10.39	1	0	0	1	0	1	1	3	1
Transport proteins												
47015	Ras-related protein Rab-11 a/b	22514	5.47	0	1	1	3	2	3	1	1	0

An MCI was present if identified in at least 2/3 triplicate gels and absent if it was absent in none or 1/3 triplicate gels. The annotated information of each MCI was identified with the aid of SEQUEST and Oxford GlycoSciences internal programs which interrogated both the public NCBI (Released ID 244) and private Incyte sequence databases. These databases were searched on 27/10/01 and 04/10/02. Red coloured numbers highlight features which were classed as being present across mammary gland development, annotations with two red asterisks (**) are annotations that have been altered due to the second round of sequence database searches. Details of isoelectric focusing point (pI), molecular mass (Mr in Daltons) and protein identity as confirmed by mass spectrometry are given. All annotations have been classed according to their general function. The classes are cytoskeletal protein (cytoskeletal), fat protein, metabolic enzyme/mitochondrial protein (ME/MP), milk protein, other, protein turnover, RNA processes, serum protein, signalling protein (signalling), and transport protein (transport).

Table 4.2 Functional classification of annotations identified in the Lactation/Involution database

MCI	Protein Annotation	Mr	pI	Description of expression profile
Cytoskeletal proteins				
50787	**Cluster C:** F-actin capping protein beta subunit (CapZ beta)**	31451	5.57	Present in L7 and absent in Inv1
47953	**Cluster D:** Tubulin alpha chain*	52004	5.43	Increased expression from L7 to Inv1 (2.14 fold change)
45878	**Cluster I:** Keratin 6b*	68042	6.00	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
46660	**Cluster J:** Keratin 1*	35758	5.89	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
46242	**Cluster M:** Keratin 10*	48875	5.94	Increased from early to late involution
46242	**Cluster M:** Keratin 1	48875	5.94	Present in all stages of Involution (increased from early to late involution)
Adipose associated proteins				
48143	**Cluster N:** Fatty acid-binding protein, adipocyte*	26714	9.37	Present in all stages of Involution (Increased from Inv1 to Inv20)
Metabolic enzymes and mitochondrial proteins (ME/MP)				
46197	**Cluster A:** Glutamate dehydrogenase, mitochondrial*	49464	7.01	Present in L7 and absent in Inv1
49037	Aspartyl aminopeptidase	48217	7.07	Present in L7 and absent in Inv1
50027	6-phosphogluconate dehydrogenase, decarboxylating (PDG)	37021	8.09	Absence in L7 and presence in Inv1
47953	**Cluster D:** Glycerol kinase*	52004	5.43	Increased expression from L7 to Inv1 (2.14 fold change)
46103	**Cluster F:** Cytosol aminopeptidase*	52516	6.81	Present in all stages of lactation (increased expression from L1 to L7)
46103	**Cluster F:** Dihydrolipoamide dehydrogenase, mitochondrial (DLD)*	52516	6.81	Present in all stages of lactation (increased expression from L1 to L7)
45878	**Cluster I:** Glycerol phosphate dehydrogenase 1, mitochondrial (GPD1)*	68042	6.00	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
48027	**Cluster G:** Isocitrate dehydrogenase, mitochondrial*	43155	9.24	Present in all stages of lactation (decreased expression from L1 to L7)
46380	**Cluster K:** Isocitrate dehydrogenase cytoplasmic*	43427	7.06	Increased from early to late involution
46380	**Cluster K:** Glutamate-ammonia ligase/Glutamine synthetase*	43427	7.06	Increased from early to late involution
46380	**Cluster K:** Acyl-CoA dehydrogenase, long-chain specific, mitochondrial*	43427	7.06	Increased from early to late involution

MCI	Protein Annotation	Mr	pI	Description of expression profile
Metabolic enzymes and mitochondrial proteins (ME/MP)				
46380	**Cluster K:** Multifunctional protein ADE2*	43427	7.06	Increased from early to late involution
46820	**Cluster L:** Similar to 3-hydroxyisobutyrate dehydrogenase (Similar to HIBADH)*	30474	6.15	Increased from early to late involution
48143	**Cluster N:** Succinate dehydrogenase iron-sulphur protein, mitochondrial*	26714	9.37	Increased from early to late involution
48143	**Cluster N:** GTP:AMP phosphotransferase mitochondrial*	26714	9.37	Increased from early to late involution
48143	**Cluster N:** Malate dehydrogenase, cytoplasmic*	26714	9.37	Increased from early to late involution
48143	**Cluster N:** 3-hydroxyacyl-CoA dehydrogenase type II (Type II HADH)*	26714	9.37	Increased from early to late involution
46305	Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial (UQCRC1)*	46462	5.36	Increased from early to late involution
46330	**Cluster O:** Similar to Fumarate hydratase (Similar to fumarase)	45017	7.75	Increased from early to late involution
46638	Aldose reductase*	35515	7.00	Increased from early to late involution
Milk proteins				
48449	Alpha casein*	40428	9.43	Absence in L7 and presence in Inv1
45957	Lactadherin*	66142	4.89	Increased expression from L7 to Inv1 (2.49 fold change)
46213	Lactadherin	50012	6.55	Increased expression from L7 to Inv1 (3.49 fold change)
47011	Gamma casein*	23218	5.75	Present in all stages of lactation (increased expression from L1 to L7)
Other				
47811	Unique*	88706	6.06	Present in L7 and absent in Inv1
46197	**Cluster A:** Unique*	49464	7.01	Present in L7 and absent in Inv1
46197	**Cluster A:** AP000511 genomic DNA, chromosome 6p21.3, HLA Class I region, section 10/20*	49464	7.01	Present in L7 and absent in Inv1
46197	**Cluster A:** Unique*	49464	7.01	Present in L7 and absent in Inv1
48013	**Cluster B:** Unique*	44314	6.90	Present in L7 and absent in Inv1

MCI	Protein Annotation	Mr	pI	Description of expression profile
Other				
48838	**Cluster E:** Unique	18035	9.37	Increased expression from L7 to Inv1 (2.69 fold change)
48838	**Cluster E:** Hypothetical protein DJ845O24.2	18035	9.37	Increased expression from L7 to Inv1 (2.69 fold change)
48027	**Cluster G:** Ensembl Contig*	43155	9.24	Present in all stages of lactation (decreased expression from L1 to L7)
47692	Hypothetical 116.4 kDa/AAH13670 Unknown*	115657	6.30	Present in all stages of lactation (increased expression from L1 to L7)
47752	**Cluster H:** AK025781 cDNA	100432	6.86	Present in all stages of lactation (increased expression from L1 to L7)
46660	**Cluster J:** Hypothetical 71.5 kDa*	35758	5.89	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
46660	**Cluster J:** BC007573, ribosomal protein, large P2*	35758	5.89	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
46660	**Cluster J:** AAH16078 Unknown/Hypothetical 35.0 kDa*	35758	5.89	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
48143	**Cluster N:** Ensembl Contig*	26714	9.37	Present in all stages of Involution (Increased from Inv1 to Inv20)
48143	**Cluster N:** 0610041L09Rik/AAH21941 RIKEN cDNA *	26714	9.37	Present in all stages of Involution (Increased from Inv1 to Inv20)
46330	**Cluster O:** DNA replication licensing factor MCM3 (MCM3)	45017	7.75	Present in all stages of Involution (increased from early to late involution)
Protein turnover				
46646	Similar to chaperonin containing TCP1, subunit 8 (TCP1)**	34712	9.58	Present in L7 and absent in Inv1
46820	**Cluster L:** Proteasome subunit alpha type 1*	30474	6.15	Present in all stages of Involution (increased from early to late involution)
48143	**Cluster N:** Proteasome subunit alpha type 7*	26714	9.37	Present in all stages of Involution (Increased from Inv1 to Inv20)
RNA processes				
48013	**Cluster B:** heterogeneous nuclear ribonucleoprotein d/A+U-rich RNA-binding (hnRNP D0)*	44314	6.90	Present in L7 and absent in Inv1
Serum proteins				
46968	Serum albumin	23718	6.70	Absence in L7 and presence in Inv1
47953	**Cluster D:** Serum albumin*	52004	5.43	Increased expression from L7 to Inv1 (2.14 fold change)

MCI	Protein Annotation	Mr	pI	Description of expression profile
Signalling proteins				
50787	**Cluster C:** Disabled homolog 2 (DOC2)*	31451	5.57	Present in L7 and absent in Inv1
45950	WD-repeat protein 1 (WDR1)	62707	6.40	Increased expression from L7 to Inv1 (2.71 fold change)
Signalling proteins				
46820	**Cluster L:** Caspase-like apoptosis regulatory protein (CLARP)*	30474	6.15	Present in all stages of Involution (increased from early to late involution)
47752	**Cluster H:** Elongation factor 2*	100432	6.86	Present in all stages of lactation (increased expression from L1 to L7)
45878	**Cluster I:** Signalling lymphocytic activation molecule (SLAM)	68042	6.00	Present in all stages of lactation (decreased L1 to L3 and increased from L3 to L7)
Transport proteins				
48013	**Cluster B:** Adapter-related protein complex 3 mu 1 subunit (Mu-adaptin 3A)*	44314	6.90	Present in L7 and absent in Inv1
45967	Sodium bicarbonate cotransporter 3 (NCB3)/2b (NCB2b)*	62508	6.02	Present in L7 and absent in Inv1
48527	Crystal Structure of the Arf-Gap Domain and Ankyrin Repeats of Papbeta (PAP β)*	22437	9.32	Increased expression from L7 to Inv1 (3.01 fold change)
46242	**Cluster M:** Similar to vesicle amine transport protein 1 (Similar to VAT1)*	48875	5.94	Present in all stages of Involution (increased from early to late involution)

Selection of MCIs in this database was based on MCIs which were present in all three triplicate gels or absent in all three triplicate gels. The annotated information of each MCI was identified with the aid of SEQUEST and Oxford GlycoSciences internal programs which interrogated both the public NCBI (Released ID 244) and private Incyte sequence databases. These databases were searched on 27/10/01 and 04/10/02. Annotations with one red asterisk (*) were determined after the second round of sequence database searches and two red asterisks (**) are annotations that have been altered due to the second round of sequence database searches. Details of isoelectric focusing point (pI), molecular mass (Mr in Daltons) and protein identity as confirmed by mass spectrometry are given. All annotations have been classed according to their general function. The classes are cytoskeletal protein (cytoskeletal), fat protein, metabolic enzyme/mitochondrial protein (ME/MP), milk protein, other, protein turnover, RNA processes, serum protein, signalling protein (signalling), and transport protein (transport).

Table 4.3 Functional classification of annotations identified in the TEB database

Presence in TEBs and absence in adult virgin 10 weeks			
MCI	**Protein Annotation**	**Mr**	**pI**
Cytoskeletal proteins			
45949	**Cluster:** Myosin heavy chain IIB	63487	5.29
50381	**Cluster:** Tubulin alpha-2 chain/Tubulin alpha-6 chain/Tubulin alpha-1 chain, brain-specific*	52799	5.42
50381	**Cluster:** Keratin 8 (Endo' A)*	52799	5.42
49033	Keratin 8 (Endo' A)	49470	5.31
52127	Actin in complex with gelsolin segment 1	43690	5.73
48442	Keratin 2a*	41927	6.31
46588	**Cluster:** Keratin 10/bovine VIB*	36955	7.89
46588	**Cluster:** Keratin 1*	36955	7.89
52339	**Cluster:** Tropomyosin alpha chain, skeletal and cardiac muscle	34327	4.65
48800	**Cluster:** Keratin 9	31814	8.63
50494	Myosin light chain 1, skeletal muscle isoform*	15632	4.70
49580	Cofilin, muscle isoform	12533	7.03
51321	Actin, alpha 1, skeletal muscle/uncomplexed actin /Actc1	12466	5.15
51187	Actin, alpha 1, skeletal muscle/uncomplexed actin (Actc1)*	12364	4.87
49394	**Cluster:** Keratin 2a*	42409	8.35
46428	Keratin 19*	41975	4.96
52506	**Cluster:** Myosin heavy chain, skeletal muscle, adult 1*	25188	8.91
Metabolic enzymes and mitochondrial proteins (ME/MP)			
46022	**Cluster:** Glycogen phosphorylase, muscle form*	56500	8.13
46022	**Cluster:** Pyruvate kinase, M2 isozyme*	56500	8.13

MCI	Protein Annotation	Mr	pI
Metabolic enzymes and mitochondrial proteins (ME/MP)			
48800	**Cluster:** VDAC1-like pseudogene, voltage-dependent anion channel 1	31814	8.63
46909	**Cluster:** Phosphoglycerate mutase, brain form*	26853	6.97
46909	**Cluster:** Carbonic anhydrase II*	26853	6.97
49394	**Cluster:** Phosphoglycerate kinase*	42409	8.35
49394	**Cluster:** 3-ketoacyl-CoA thiolase, mitochondrial*	42409	8.35
52506	**Cluster:** Cytochrome c oxidase II*	25188	8.91
Other			
53219	**Cluster:** I40805 collagenase - Clostridium histolyticum	120176	5.69
53219	**Cluster:** Ensembl Contig	120176	5.69
52746	Collagenase	112975	5.41
45949	**Cluster:** Delta-like protein 3	63487	5.29
50381	**Cluster:** Ensembl Contig*	52799	5.42
50381	**Cluster:** Unique*	52799	5.42
52339	**Cluster:** Proliferating cell nuclear antigen (PCNA)	34327	4.65
46909	**Cluster:** Unique*	26853	6.97
49394	**Cluster:** Platelet glycoprotein IIb*	42409	8.35
Protein turnover			
50398	Calumenin*	45508	4.46
46943	Proteasome subunit beta type 10*	25492	5.90
50484	Similar to chaperonin containing TCP1, subunit 8**	21144	5.24

MCI	Protein Annotation	Mr	pI
RNA processes			
53219	**Cluster:** Poly polymerase gamma	120176	5.69
49107	Nuclear protein Hcc-1*	30995	6.55
Signalling proteins			
46650	Annexin A2	35297	7.48

An MCI was present if identified in 2/2 duplicate gels and absent if it was absent in none of the duplicate gels. The annotated information of each MCI was identified with the aid of SEQUEST and OGS internal programs which interrogated both the public and private Incyte sequence databases. These databases were searched on 27/10/01 and 04/10/02. Annotations with one red asterisk (*) were determined after the second round of sequence database searches and two red asterisks (**) are annotations that have been altered due to the second round of sequence database searches. Details of isoelectric focusing point (pI), molecular mass (Mr in Daltons) and protein identity as confirmed by mass spectrometry are given. All annotations have been classed according to their general function. The classes are cytoskeletal protein (cytoskeletal), fat protein, metabolic enzyme/mitochondrial protein (ME/MP), milk protein, other, protein turnover, RNA processes, serum protein, signalling protein (signalling), and transport protein (transport).

Chapter 5

Candidate protein annotations

5.1 Summary

Eleven annotations were identified in the previous Chapter as candidate proteins for further analysis. This Chapter describes the criteria used for further selection and the evidence produced to make these decisions. Several different techniques were used in an attempt to validate the proteomic data: TaqMan, western blotting, microarray and IHC analyses. In some cases the proteomic data was not confirmed by the follow-up experiments conducted in this Chapter. Reasons for this have been outlined throughout the Chapter.

Two proteins, MCM3 and annexin A2, were investigated in more detail in this Chapter compared to the other candidate proteins selected for follow-up. Further investigations of MCM3 using TaqMan and western blot analyses revealed that its expression peaked during pregnancy. The results obtained by IHC for MCM3 and MCM2 during normal mammary gland development were compared to a known proliferation marker Ki67. Both MCM3 and MCM2 antibodies stained a greater proportion of epithelial cells than the Ki67 antibody. A possible explanation for this could be that MCM proteins are expressed in cells which have proliferative potential, whilst Ki67 stains nuclei in all stages of the cycle other than G_0.

The additional results obtained for annexin A2 in normal mammary gland development showed that its expression peaked during pregnancy and involution. This was confirmed by TaqMan, microarray and western blot analyses. The results obtained by IHC revealed that the annexin A2 antibody stained the stroma more strongly than the myoepithelial and luminal epithelial cells. The altered levels of staining found in the different cell types has not previously been reported. Annexin A2 expression was investigated in normal breast tissue and breast carcinomas and revealed that annexin A2 is up-regulated in breast tumours.

5.2 Introduction

5.2.1 Candidate protein investigations

Fifty nine annotations were identified in the development and Lactation/Involution databases. However, only a small proportion of these could plausibly be followed-up within the duration of the project. Eleven proteins were selected as candidate proteins and the specific reasons for their selection have been stated in Chapter 4. Functional candidate proteins were selected based on their sequence quality, expression profile and biological interest. The priority of this Chapter was to confirm the proteomic expression data and this was attempted using a variety of techniques. TaqMan analysis was used to determine the mRNA expression of the genes of interest to try to identify the correct protein annotation from a clustered MCI; western blot analysis was used to confirm the protein expression profile across development (where reagents were available); and where necessary the mouse mammary gland microarray expression profile was compared to both the TaqMan expression data and protein expression profiles obtained for the selected candidate proteins. IHC was used for tissue localisation purposes where antibodies were available.

5.2.2 Clustering annotations

2-D proteomics can resolve simultaneously thousands of proteins, including their modifications. The most appropriate method for verifying the proteomics data is to use western blot analysis with specific antibodies. However, antibodies are not always available for analysing the protein annotations of interest. In this situation, techniques that analyse mRNA expression can be used to correlate general trends in proteomic data with mRNA trends. However, there are limitations to the conclusions that can be drawn when comparing mRNA levels with protein levels.

One such technique which analyses mRNA levels is TaqMan. The advantage of using TaqMan analysis is that the mRNA expression profiles of all the annotations within a cluster can be determined quantitatively. These expression profiles can then be used and compared to the proteomic profile of the clustered MCI in order to eliminate the most unlikely annotation. However, a limitation with TaqMan analysis is that potential PTMs are not considered at the mRNA level, and such PTMs may be functionally important in different stages of mammary gland development. Additionally, at the protein level, the antibodies that are available for certain proteins may not be specific enough to distinguish between different isoforms resolved by 2-D gel electrophoresis, and so can only detect total protein levels. Unfortunately, at present, the isoform-specific resources required to support the complex data produced by proteomics are rarely available. Figure 5.1

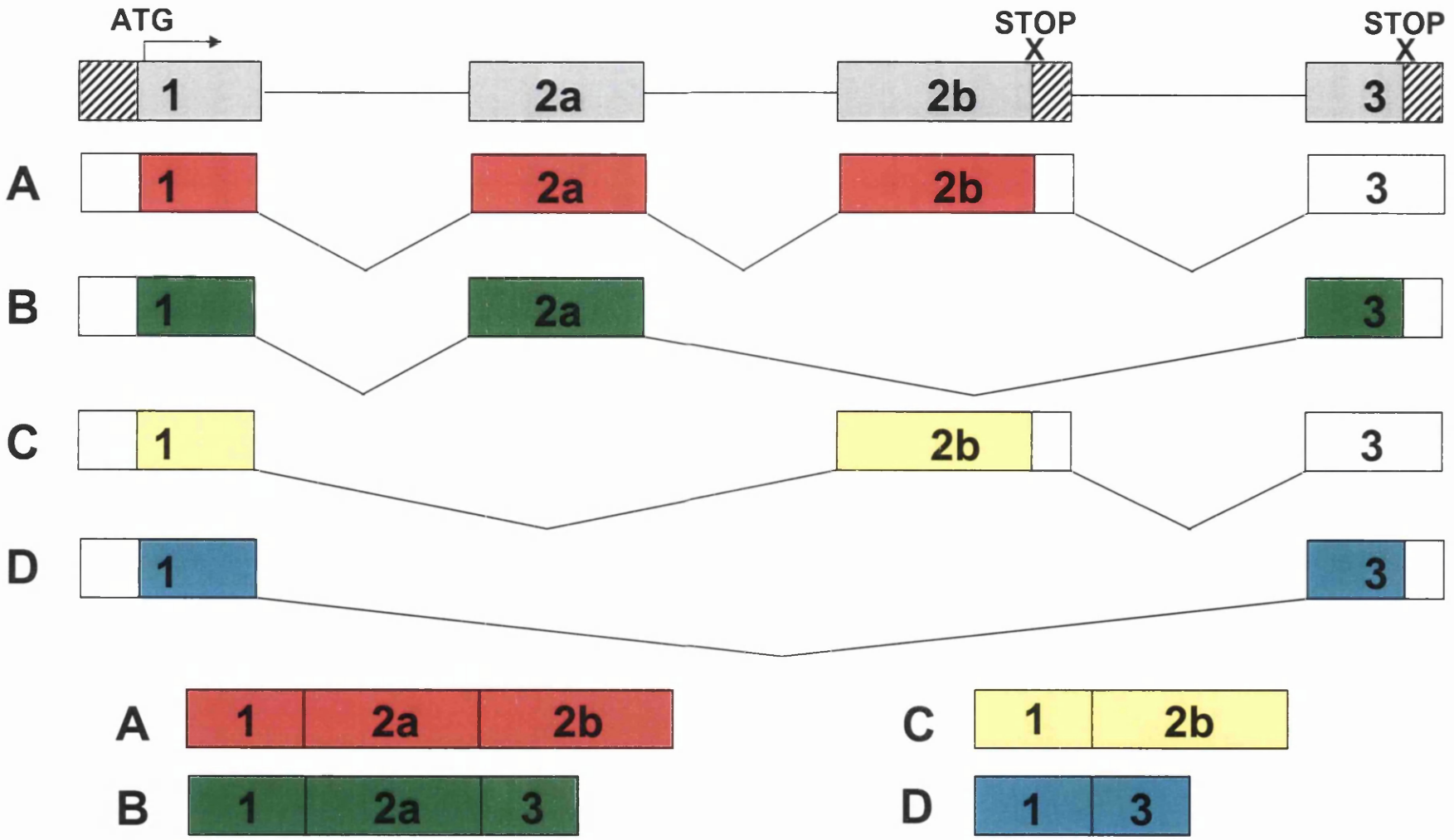


Figure 5.1 A great deal of diversity can be created from one gene by the process of alternative splicing. This example displays how four individual and distinct transcripts can be generated by the alternative splicing of only one exon. The inclusion of exon 2a allows the transcript to read-through to the stop codon in exon 3. However, if exon 2b is included a premature stop codon prevents read-through to exon 3.

demonstrates the multiple isoforms that can result from alternative splicing of a single gene, and the need for specific antibodies for the detection of each individual isoform.

Eleven candidate proteins were selected from Chapter 4 and, of these, six were in clusters that had more than one protein identification. All candidate clustered annotations were analysed by TaqMan across mammary gland development in order to determine the correct identity of the clustered MCIs. This was attempted by comparing the TaqMan expression profiles with the appropriate proteomic profiles. The TaqMan expression profiles, which were most similar to the MCI expression profile, were considered the most likely identification.

5.2.3 Clustering proteins and post-translational modifications

When studying a complex cellular heterogeneous protein extract, apparent gains or losses of a feature at a particular time point can be due to a number of factors. One explanation is a change in cellular composition where one cell type either individually changes in expression of a particular protein, or there is a large change in the number of cells of a specific cell type each producing the same amount of protein. The latter is an unlikely explanation where present/absent calls are determined at very close time points that have no observable morphological change in cellularity, for example L7 and Inv1.

The second factor is an alteration of the location of a protein on a 2-D gel as a result of PTMs. The state of a protein may change during different time points resulting in a presence call at one time point and an absence in the next or visa versa.

5.2.4 2-D gel electrophoresis and post-translational modifications

Phosphorylation is an important PTM process in cell regulation and metabolism. It is estimated that more than a third of proteins are modified by phosphorylation in mammalian cells. Serine phosphorylation represents about 90% of cellular phosphorylation, threonine about 10% and an estimated 0.1% tyrosine phosphorylation (Adamczyk *et al.*, 2001). The level of phosphorylation is modulated by phosphatases (enzymes that catalyse the removal of phosphate) and kinases (enzymes that catalyse phosphorylation). The resolution achieved by 2-D gel electrophoresis is often sufficient to separate protein modification states directly. When a protein becomes phosphorylated, the charge is altered causing a shift of the protein on the gel towards the basic end. This change is often shown on a 2-D gel by a horizontal trail of protein features. Phosphorylated proteins and their non-phosphorylated counterparts migrate closely together; however, the phosphorylated form

usually migrates at a slightly slower rate. To identify conclusively an isolated phosphoprotein, the protein can be detected by autoradiography, phosphoimaging or western blotting after radio-labelling with ^{32}P. Investigating protein modifications is more difficult than determining the identity of a protein. These studies require more material, as phosphorylated proteins are present in small amounts in cells. Also determining protein modifications by mass spectrometry requires all the peptides that do not have the expected mass to be analysed further.

5.2.5 Overview of candidate protein annotations

5.2.5.1 Perilipin

Perilipin appears to be localised to the periphery of lipid storage droplets in adipocytes (white and brown) and cholesterol ester-rich steroidogenic cells (Greenberg *et al.,* 1993; Servetnick *et al.,* 1995; Blanchette-Mackie *et al.,* 1995). There are four protein isoforms of perilipin A, B, C and D, (Londos *et al.,* 1995; 1999). It is hypothesised that perilipins function as regulators of lipolysis (Londos *et al.,* 1999), a process mediated by HSL in adipocytes. Protein kinase A (PKA) phosphorylates perilipin and HSL (Anthonsen *et al.,* 1998), and the increase of PKA causes the lipase to change from the cytosol to the lipid droplet surface (Egan *et al.,* 1992, Brasaemle *et al.,* 2000). Phosphorylated perilipin may act as a docking protein to HSL or it may alter the surface of the lipid droplet, thus allowing lipase translocation (Londos *et al.,* 1995). The data available suggests that non-phosphorylated perilipin inhibits lipolysis. Little is understood of the expression patterns of perilipins and research regarding lipid droplets is revealing that they are not merely used for lipid storage.

5.2.5.2 Contrapsin

Contrapsin is a plasma glycoprotein which is able to inhibit trypsin-like proteases (Takahara and Sinohara, 1982; 1983). Contrapsin (Spi2) is an extracellular protein and its human homologue alpha 1 antichymotrypsin is known to be associated with pro-inflammatory responses such as acute inflammation (Das and Potter, 1995; Hong *et al.,* 1995; Kanemaru *et al.,* 1996; Lieb *et al.,* 1996). The main function of most serpin proteins is involved in the regulation of proteolytic events using a variety of biochemical pathways, for example extracellular remodelling and cell differentiation. Second to human serine protease inhibitors, the mammalian serpins are most well characterised at the gene level, although understanding the biological function of these proteins remains elusive.

5.2.5.3 Lumican

Lumican is a major keratan sulphate proteoglycan (Axelsson and Heinegard, 1978) which consists of a core protein and glycosaminoglycan side chains. Keratan sulphate proteoglycans are important in cell migration and proliferation during embryonic development, tissue repair and tumour growth, in addition to matrix assembly and structure (Iozzo, 1997; Fullwood *et al.,* 1996; Wight *et al.,* 1992). Lumican is present in the extracellular matrix of many tissues and exists as two distinct forms with respect to its tissue localisation. Lumican is abundant in the cornea where it exists as a highly sulphated proteoglycan, whereas in the other connective tissues it is a poorly/non-sulphated glycoprotein (Funderburgh *et al.,* 1991a; 1991b). An alteration in expression of a gene involved in normal cellular processes, such as stromal-epithelial interactions, may contribute towards the change to a cancerous state (Kinzler and Vogelstein, 1998). This has been suggested for lumican, which is expressed in normal stroma breast tissue and is increased in invasive carcinomas (Leygue *et al.,* 1998).

5.2.5.4 PTRF

PTRF is involved in transcriptional termination. Transcription termination by RNA polymerase I (Pol I) is achieved by pausing elongating transcription complexes, and occurs at specific terminator elements downstream of the pre-rRNA (Grummt *et al.,* 1985; Bartsch *et al.,* 1987). The Pol I terminator elements known as the 'Sal box' are recognised by a specific DNA-binding protein (the transcription termination factor TTF-I) that stops elongating Pol I (Evers *et al.,* 1995; Grummt *et al.,* 1986). However, DNA-bound TTF-I requires a cellular factor to release the transcript. Dissociation is achieved by PTRF, which interacts with Pol I, TTF-I and the 3' end of pre-rRNA (Mason *et al.,* 1997; Jansa *et al.,* 1998). Cellular PTRF is phosphorylated at multiple sites and its regulation of phosphorylation may control the 3' end formation of pre-rRNA or the recycling of Pol I.

5.2.5.5 Rab11

Rab proteins are GTP-binding proteins which are involved in the formation, targeting, and/or fusion of transport vesicles (Novick and Zerial, 1997). The two isoforms of Rab11, a and b, differ in the last 30 amino acids of the C-terminal, a region which interacts with target membranes (Brennwald and Novick, 1993; Chavrier *et al.,* 1991). Rab11a is sub-apically located in epithelial cells (Goldenring *et al.,* 1996) and is associated in vesicle recycling through the pericentriolar recycling endosome compartment (Ullrich *et al.,* 1996). Rab11b is required for the transport of transferrin from the recycling compartment to the plasma membrane, but little is known about its localisation and function (Schlierf *et*

al., 2000). Rab proteins function by cycling between an active GTP-bound membrane form and an inactive GDP-bound cytosolic form (Bourne *et al.*, 1990). The exact function of these proteins is still poorly understood but it is thought that each Rab protein has a defined role since different Rab proteins localise to distinct vesicle compartments.

5.2.5.6 WDR1

WDR1 is a WD-repeat protein that contains nine repeat motifs to mediate protein-protein interactions (Baillat *et al.*, 2001; Tieu and Nunnari, 2000). WD-repeats have a conserved core of approximately 40 amino acids. Of the many known WD-repeat proteins, several have known functions which span a range of important regulatory mechanisms (Yu *et al.*, 2000). Some WD-repeat proteins have undefined functions (Smith *et al.*, 1999; Neer *et al.*, 1994) such as WDR1. WDR1 in yeast causes depolymerisation of actin filaments, but this is only achieved in the presence of cofilin, an actin binding protein (Rodal *et al.*, 1999). Actin filaments are regulated by actin-binding proteins which are necessary for the actin cytoskeleton to mediate endocytosis, exocytosis, cell motility, polarity and cytokinesis.

5.2.5.7 MCM3

MCM (mini-chromosome maintenance) proteins were originally identified by genetic studies in yeast (Maine *et al.*, 1984). The family of MCM proteins contains at least six evolutionary conserved members from MCM2 to MCM7 which are essential in the initiation of replication (reviewed by Tye, 1999). The origin of recognition complex (ORC) is believed to be the initiator of DNA replication (Bell and Dutta, 2002; Dutta and Bell, 1997). This binds to a specific site on chromatin with additional initiation factors, Cdc6, Cdt1 and MCM proteins in G_1 of the cell cycle, which results in the formation of a pre-replicative complex (pre-RC). The assembly of the pre-RC makes chromatin 'licensed' for replication (reviewed by Chavlier and Blow, 1996). Prior to the cell cycle entering S phase, MCM proteins bind to chromatin to make it competent for replication (Chong *et al.*, 1996; Kearsey *et al.*, 1996). The MCM proteins remain bound to chromatin until they gradually dissociate with phosphorylated members as S phase progresses (Kimura *et al.*, 1994; Schulte *et al.*, 1995; Todorov *et al.*, 1995). The replicated chromatin without bound MCM proteins ensures that DNA is replicated once during a single cell cycle division. Studies have indicated that MCM proteins could potentially be used as prognostic markers in breast cancer as they are able to give a more accurate insight into the cell cycle state in tissue compared with conventional markers (Gonzalez *et al.*, 2003; Stoeber *et al.*, 2001).

5.2.5.8 Annexin A2

Annexin proteins are widely expressed in a range of tissues and cell types. Annexin A2 is a phospholipid binding peripheral membrane protein. It contains four annexin repeats (Geisow *et al.,* 1986) which enable the protein to shuttle between water-soluble and membrane compartments as a result of fluctuating calcium levels. The core domain of annexin A2 possesses calcium binding sites (Crumpton and Dedman, 1990; Geisow and Walker, 1986), and it has been reported to be a calcium regulated protein on endothelial cells, macrophages and tumour cells. The N-terminal tail has two phosphorylation sites, tyrosine (Tyr)-23 and serine (Ser)-25 which are phosphorylated by protein tyrosine kinases and protein kinase C respectively (Glenney and Tack, 1985; Gould *et al.,* 1986). It has recently been shown that annexin A2 regulates the glycoprotein tissue plasminogen activator (t-PA), which is synthesised by and binds to endothelial cells (Hajjar and Menell, 1997). The glycoprotein activates plasminogen to form plasmin, a fibrinolytic protease, and increases in expression in promyelocytic leukaemia cells due to abnormally high levels of annexin A2 (Menell *et al.,* 1999).

Annexin A2 is developmentally expressed throughout pregnancy and lactation of the mouse mammary gland, outlining the apical membrane of alveoli (Handel *et al.,* 1991). It appears to be present in both large and small vesicles of actively secreting epithelial cells, consistent with a possible structural and/or functional role in membrane trafficking within these cells (Creutz, 1992). Conflicting results, published by Lozano *et al.* in 1989, found comparable levels of annexin A2 in both non-secretory and secretory mammary epithelial cells grown on collagen gels (Handel *et al.,* 1991; Jamieson, *et al.,* 1990).

Breast tissue studies of annexin A2 have shown that it is the most predominant annexin in normal and malignant mammary epithelial cells. In contrast to normal tissue higher levels of annexin A2 were expressed in stromal cells of mammary tumours. The differential expression of annexin A2 in normal and malignant human breast tissue suggests an important role for these proteins in the mammary gland (Schwartz-Albiez *et al.,* 1993). Furthermore, it has been suggested that the association of annexin A2 with the plasma membrane in primary renal carcinoma cells may lead to pro-metastatic effects, such as preventing coagulation, oxidative stress and immunological surveillance (Tanaka *et al.,* 2004). A variety of malignant tumours, except human prostate cancer (Chetcuti *et al.,* 2001), have shown an increase in expression of annexin A2, such as human hepatocellular (Frohlich *et al.,* 1990) and gastric carcinoma (Emoto *et al.,* 2001).

5.3 Results

Eight annotations were selected for further follow-up work with an additional three hypothetical/unique proteins.

One of the techniques used in this Chapter was TaqMan analysis and Appendix 4 lists the primers and probe sequences for all the annotations interrogated. Not all annotations within a clustered MCI were analysed by TaqMan because other resources were available for confirmation of the proteomics results or because of cost issues. The results obtained from real-time RT-PCR are represented as relative fold change using the comparative C_T method as described in Materials and Methods. The same calibrator (reference sample) was used for annotations within the same cluster in order to compare the results. However, the calibrator varied between different MCIs as these did not need to be compared.

As data was available from a separate mouse mammary gland microarray study, the expression profiles of all clustered annotations were obtained wherever necessary. The microarray data was used as an adjunct to the real-time PCR results. The results are displayed as scaled signal intensities (see Materials and Methods for details). It should be noted that the same RNA samples for the stages that are common to microarray analysis and TaqMan were used. Additional stages of mammary gland development were used for the microarrays. These were Av6, Av12 (adult virgin 6 weeks and 12 weeks), P1, P2 and P3 (pregnancy 24 h, 48 h and 72 h respectively). Pregnancy 4.5, Inv5 and Inv10 were not used in the microarray study. Furthermore, the protein samples used for proteomics were obtained from the left gland of a mouse, and the RNA samples used for TaqMan and microarray analyses were obtained from the right gland of the same mouse.

Western blot analysis was performed where possible, but as there was a restriction in the number of loading wells present per polyacrylamide gel, all stages of development were used except for Inv20. The exception to this rule was made for perilipin as this protein was detected during Inv20 in the proteomic data and therefore P4.5 was substituted for Inv20.

Once the expression profiles of the candidate proteins were determined using TaqMan, microarray and western blot analyses, the next stage was to identify their localisation within mammary gland tissue sections. Where possible IHC was performed with the antibodies used for western blotting to determine the tissue localisation of the proteins. In some cases the data collected using western blotting, real-time RT-PCR and microarrays had not confirmed the proteomic data. However, from the data gathered at this stage of the

project (experimental and literature based) it was interesting to determine the regions of the gland in which these proteins were expressed.

Unfortunately not all the antibodies used for western blotting were suitable for IHC. This was true for PTRF, Rab11 and WDR1; but annexin A2, MCM3 and perilipin were successfully used for IHC.

5.3.1 Expression of β-casein

For reference an MCI found in 2/3 triplicate samples was counted as a feature presence and 1/3 was counted as a feature absence.

The expression profile of β-casein across mammary gland development has recently been published, using the same stages of development as in this study (Stein *et al.*, 2004). The antibody to this protein was used only as a positive control to test the protein samples used for western blot analysis. The western blot profile for β-casein showed the expected up-regulation during mid pregnancy and down-regulation during involution (Figure 5.2). β-casein migrates at 26 kDa, and from the western blot multiple bands can be seen which may correspond to the other isoforms of casein.

5.3.2 Analysis of perilipin

Perilipin was not a cluster protein but mass spectrometry could not distinguish between the different isoforms of this protein. Perilipin is known to have four isoforms and the antibody used for western blotting detected isoforms A, B and C. Each of these three isoforms have different molecular weights, 57 kDa, 46 kDa and 38 kDa respectively (Londos *et al.*, 1995; 1999). A slight change in the migration of perilipin C was observed on the western blot (Figure 5.3), but this can be attributed to having loaded the control in the last lane of the gel which occasionally alters the migration pattern of a protein. The western blot for perilipin showed that all three isoforms had strong expression during pregnancy, with a decrease during lactation. Their patterns of expression differed to each other during involution. All isoforms showed an increase in expression during involution, but this occurred at a later time point for perilipin B. None of these profiles matched that determined by proteomics which had shown perilipin to be expressed in Inv20.

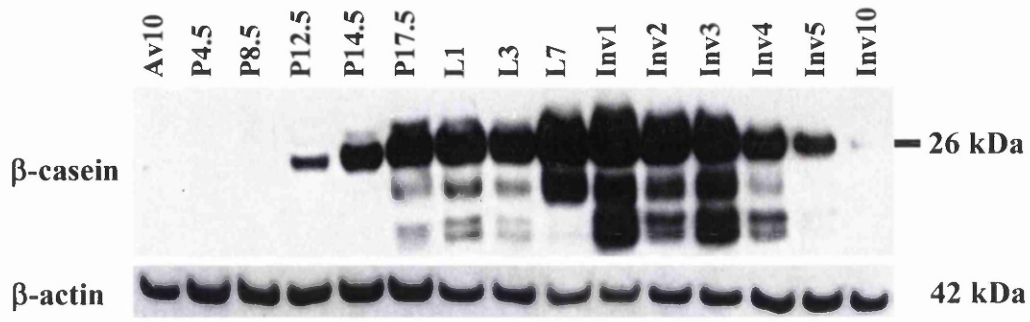


Figure 5.2 Western blot profile of β-casein during mouse mammary gland development. All stages of development were used (40 μg protein) except for Inv20. The blot was subsequently probed with β-actin for load verification.

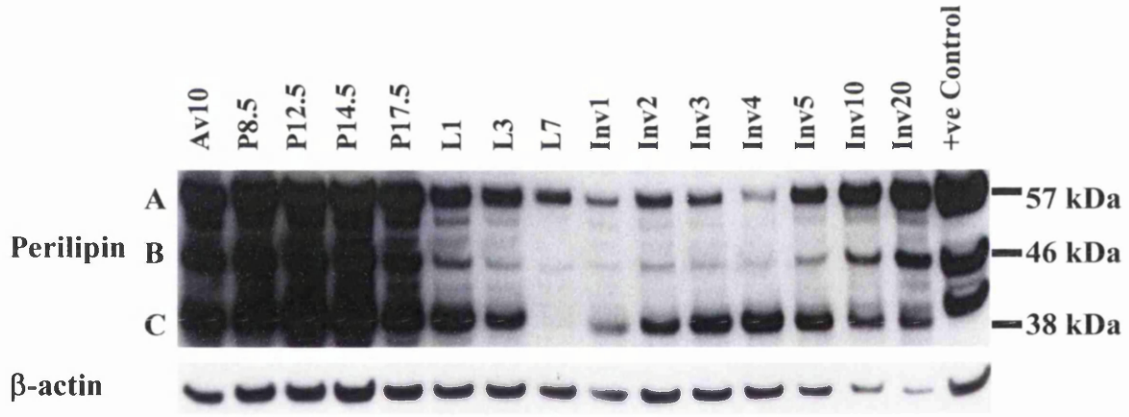


Figure 5.3 Western blot analysis of perilipin during mouse mammary development. The Mr to which the proteins migrated have been indicated. All stages of development were used (40 μg) except for P4.5 due to the number of wells present on the gel. The blot was subsequently probed with β-actin for load verification. The adipocytes cell line 3T3-L1 cells were used as a positive control.

IHC demonstrated that perilipin's expression was localised to the surface of adipocytes in the mammary gland as expected (Figure 5.4). Although this technique is not quantitative, clear changes in staining could be seen in extreme time points. This was due to the altering numbers of adipocytes during development rather than a change in the intensity of staining. A decrease in the number of cells stained with the perilipin antibody could be seen as the number of adipocytes decreased during pregnancy (Figure 5.4A), and this continued into lactation (Figure 5.4B). An increase in the number of cells stained with the perilipin antibody occurred during involution, as adipocytes are replenished during this phase of development (Figure 5.4C). The western blot for perilipin showed different expression profiles during involution for the three different isoforms detected. IHC did not distinguish the different isoforms of perilipin but detected all three; hence the discrepancy between the two techniques. If the isoforms of perilipin A, B and C were merged on the western blot image, an increase in expression would be observed from early to late involution as seen by IHC. No staining was observed in the negative control used for IHC which had had no primary antibody added to the section (Figure 5.4D). These expression profiles appear to have resulted from the altering ratio of epithelial to stromal cells during development.

5.3.3 Analysis of contrapsin and lumican

Contrapsin, lumican, carboxylesterase and HS21C001 (segment of chromosome) all clustered together (MCI 49502). The chromosome segment was not investigated by real-time RT-PCR; however, the other three were. Carboxylesterase was eliminated as being the potential identification of MCI 49502, as no signal was detected in all 40 cycles of the PCR reaction, whereas a signal was produced for the positive control. The expression profiles produced for contrapsin and lumican were very different to each other (Figure 5.5). Contrapsin showed virtually no change to the calibrator during pregnancy and lactation, but an increase was observed during mid involution. The signal intensity of contrapsin from the microarray data was below a detectable level and was classed as absent; therefore no comparisons could be made to the TaqMan data. However lumican showed a similar microarray expression profile to the TaqMan results (Figure 5.5), as both showed a decrease in expression after mid pregnancy through to the start of involution. The expression plots varied during involution, but both showed an increase during mid involution. These profiles were not comparable to the proteomics results which had shown this MCI to be present only during P14.5. It is most probable that these differences are due to TaqMan and microarray analyses detecting total mRNA expression, whereas proteomics detects multiple isoforms.

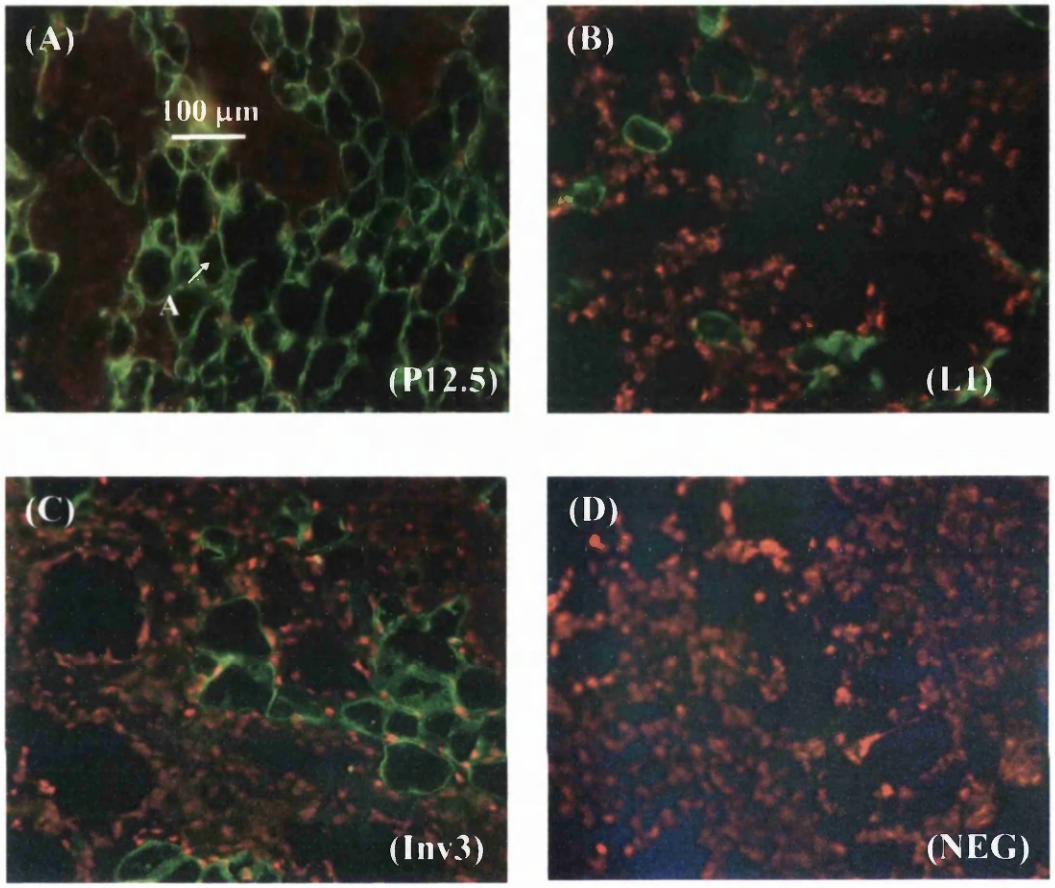


Figure 5.4 Immunohistochemical localisation of perilipin in various stages of mouse mammary gland development. Frozen tissue sections were stained with a guinea pig polyclonal anti-perilipin antibody. A FITC-conjugated secondary antibody was used to visualise staining and propidium iodide to counter stain. Sections were view at 250x magnification using a Wang Biomedical microscope and a fluorescein filter. Images were processed using Adobe Photoshop 7.0 software. (A) to (C) range from P12.5 to Inv3 and (D) is the negative control. Adipocytes (A).

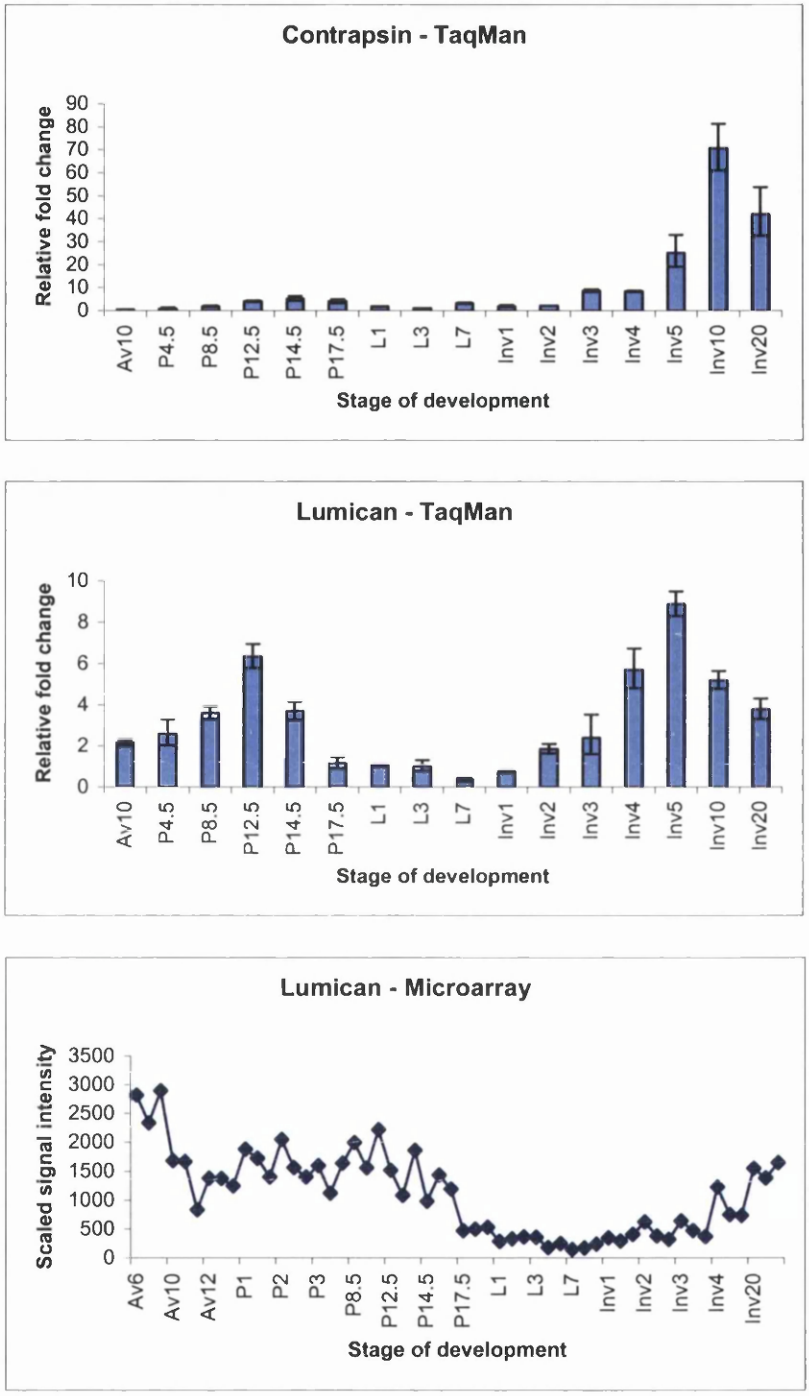


Figure 5.5 Expression profiles of contrapsin and lumican during mouse mammary gland development. The expression profiles of contrapsin and lumican are demonstrated by TaqMan and/or microarray analysis. TaqMan results are presented as fold change relative to a stage within the set of samples tested. Microarray results are presented by signal intensity which has been scaled to an arbitrary value. Liver was used as the positive control tissue for contrapsin and heart was used for lumican.

Unfortunately no antibody was available for contrapsin and the lumican antibody obtained failed to detect this protein within the mammary gland and the positive control. Therefore an attempt was made to localise these proteins in the mammary gland using *in situ* mRNA hybridisation. Several probes were designed and experimented on control tissue, but none of the probes successfully hybridised to the sections.

5.3.4 Analysis of PTRF

PTRF was a single MCI identification from the developmental database and western blotting was used to confirm the proteomics results. This protein migrated at 50 kDa on a western blot. When comparing the PTRF western blot expression profile (Figure 5.6) to the β-actin profile, they showed similar band concentrations suggesting that PTRF did not change in expression across development. However, cellular PTRF can be phosphorylated and as the antibody did not distinguish between phosphorylated and non-phosphorylated PTRF, it is unclear whether the phosphorylated form of this protein altered in expression across development. As PTRF was only detected during Inv5 in the proteomics data, it is possible there was only a change in phosphorylation state during this stage. The shift of pI caused by phosphorylation would have excluded the non-phosphorylated form from being detected on the 2-D gel. At this stage of the project attention was focused on particular candidate proteins. As resources were not available to distinguish between the phosphorylation states of PTRF, the analysis of this protein was not taken further.

5.3.5 Analysis of Rab11

Rab11 was another single MCI identification from the developmental database. The western blot expression profile of Rab11 appeared to fluctuate across development but showed a distinct strong expression during the early stages of lactation and again from early to mid involution (Figure 5.6). However, the proteomics data suggested Rab11 was expressed from mid to late pregnancy. The antibody used for Rab11 did not distinguish between Rab11 a or Rab11b. Neither could the mass spectrometry distinguish between the two isoforms. Hence, the discrepancy between the western blot and proteomic results could be due to the different isoforms being present. Rab11 was not investigated further because of the lack of resources available for discriminating between its different isoforms.

5.3.6 Analysis of WDR1

WDR1 was investigated by western blotting. Unfortunately no positive control was obtained for WDR1 (chick basilar papilla). However, the protein migrates at 67 kDa (Adler *et al.*, 1999) and this was approximately the size of the bands detected by western blot

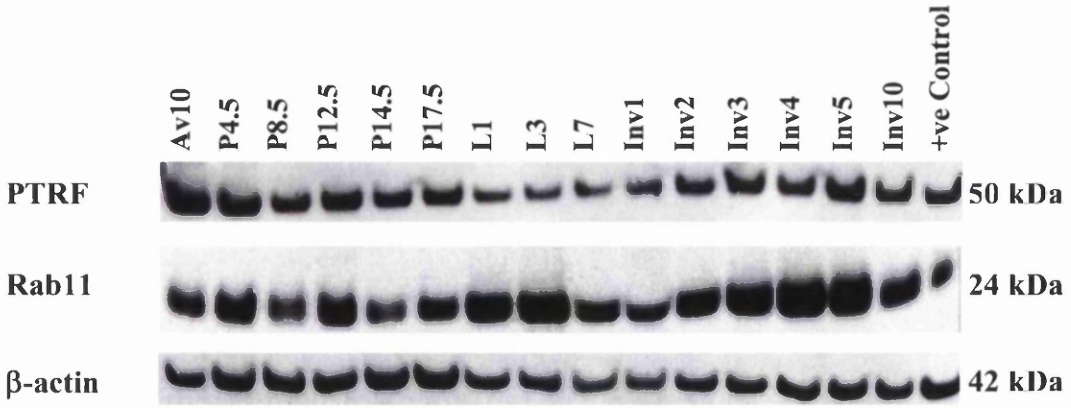
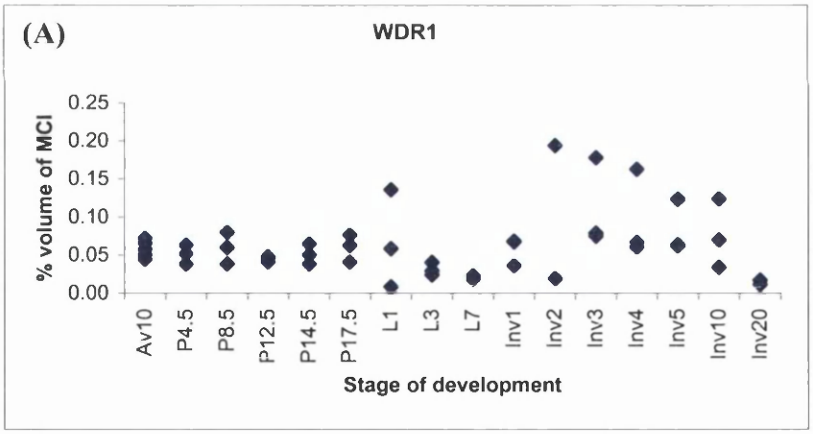


Figure 5.6 Western blot analysis of candidate annotations during mouse mammary gland development. The Mr to which the proteins migrated has been indicated. All stages of development were used (40 µg) except for Inv20 due to the number of wells present on the gel. All blots were subsequently probed with β-actin for load verification. NIH-3T3 fibroblast cells were used as a positive control for western blot analysis of PTRF, and Madin Darby Canine kidney (MDCK) cells were used for western blot analysis of Rab11.



(B)

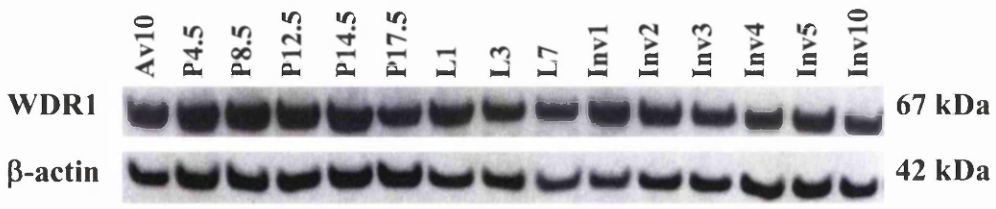


Figure 5.7 Expanded proteomic expression profile of WDR1 during mouse mammary gland development. (A) extended proteomic expression profile has been determined by the percentage volume of the MCI within the 2-D gel. (B) Western blot analysis of WDR1. No positive control was used. All stages of development were used (40 µg protein) except for Inv20 due to the number of wells present on the gel. Blot was subsequently probed with β-actin for load verification.

analysis (Figure 5.7B). No change in expression was observed across development when comparing the expression profile to β-actin. Expanding the expression profile of the proteomics results for WDR1 (Figure 5.7A) revealed that its expression was not restricted to L7 and Inv1 but was in fact present during all stages of development. The proteomics profile showed that its expression was quite constant during Av10 and pregnancy but for lactation and involution WDR1 showed no alteration in expression.

5.3.7 Analysis of MCM3

MCM3 and fumarate hydratase clustered together and both were analysed by real-time PCR. MCM3 and fumarate hydratase TaqMan expression profiles were overall quite similar, as they showed a temporary decrease in expression during lactation (Figure 5.8B and C). When comparing these results to the expanded proteomic data (Figure 5.8A), neither could be eliminated from being the annotation for MCI 46330. MCM3 and fumarate hydratase TaqMan profiles did not have exact matches to the proteomic expression profile. However, both were similar enough to it that they could not be discounted.

An antibody was not available for fumarate hydratase, but one was obtained for MCM3. Overall the western blot and TaqMan expression profiles for MCM3 resembled each other (Figure 5.8D), as expression was highest during pregnancy and dropped during lactation. All experiments showed MCM3 increasing in expression during involution, although the point at which this occurred varied between the three different techniques.

IHC was performed with the same MCM3 antibody which localised to nuclei of the epithelial cells of the mammary gland (Figure 5.9). The intensity and number of cells detecting MCM3 appeared greater during pregnancy (Figure 5.9B and C) compared to lactation (Figure 5.9D), the latter showed a heterogeneous level of staining. This pattern of expression was comparable with the results obtained by western blot analysis. A slight increase in the number of cells stained with MCM3 was observed after the first few days of involution (Figure 5.9E) compared to lactation and this persisted through to late involution (Figure 5.9F). Overall these changes in expression observed by IHC appeared to be in agreement with the other results produced for MCM3. No staining was detected in the negative control (Figure 5.9G).

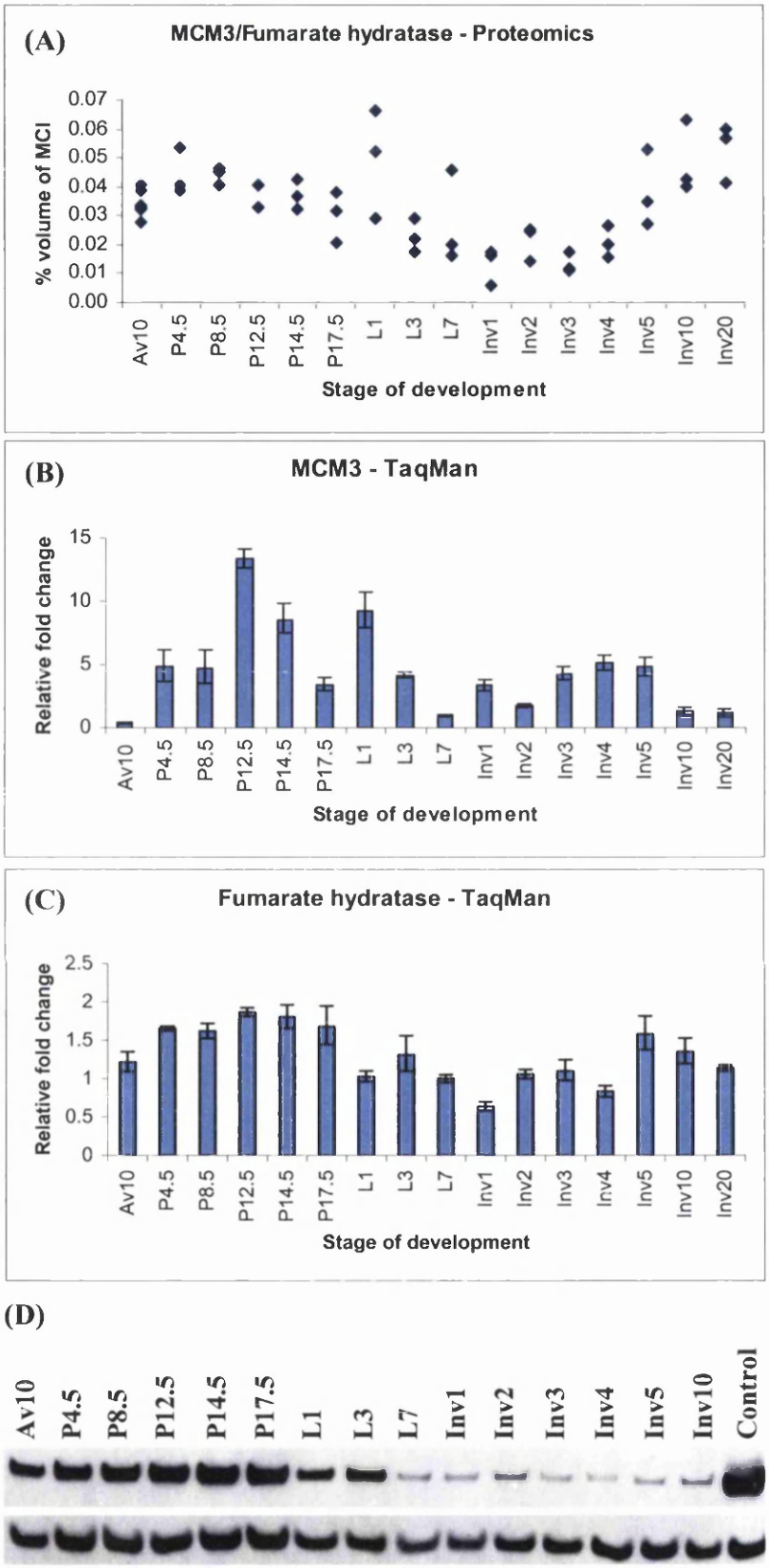


Figure 5.8 Expression profile of MCM3 and fumarate hydratase during mouse mammary gland development. (A) The proteomic profile could be MCM3 or fumarate hydratase and is represented as the percentage volume of the MCI within the 2-D gel. (B) and (C) TaqMan analysis of MCM3 and fumarate hydratase, respectively. Results are presented as fold change relative to a stage within the set of samples tested. (D) Western blot analysis of MCM3 using NIH-3T3 fibroblast cells as a positive control. All stages of development were used (40 µg protein) except for Inv20 due to the number of wells present on the gel. Blot was subsequently probed with β-actin for load verification.

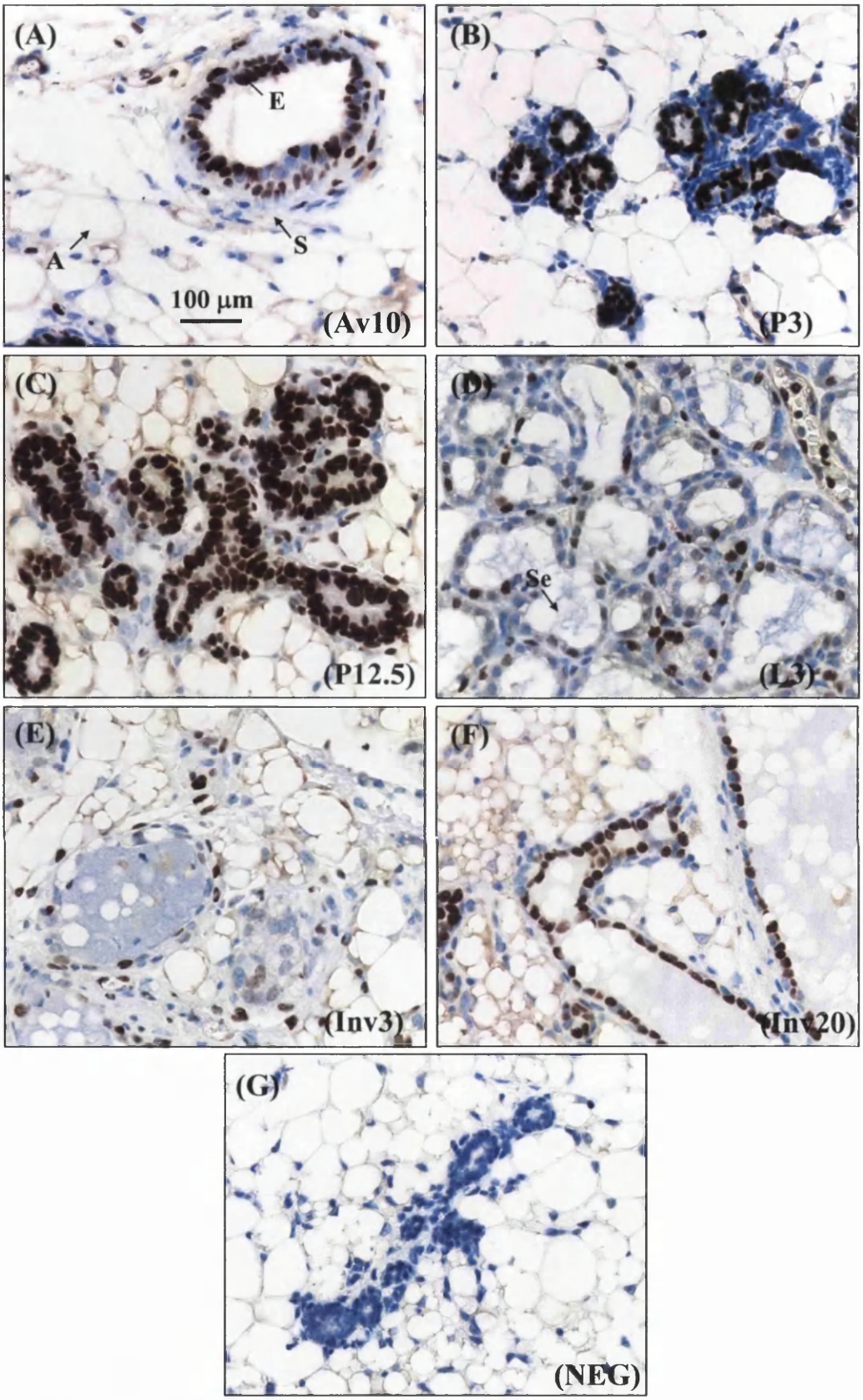


Figure 5.9 Immunohistochemical localisation of MCM3 in various stages of mouse mammary gland development. Paraffin embedded tissue sections were stained with a goat polyclonal anti-MCM3 antibody. An HRP-conjugated secondary antibody was used to visualise staining with DAB. Sections were view at 400x magnification using a Zeiss Axiophot microscope and images were processed using Axiovision 3.1 software. (A)-(F) range from Av10 to Inv20 and (G) is the negative control. Stroma (S); epithelial cells (E); secretions (Se); adipocytes (A).

5.3.8 Immunohistochemical staining of MCM2

All MCM proteins (2-7) complex together to 'license' cell division. As an antibody was available for another MCM protein (MCM2) this was investigated by IHC to support the data obtained for MCM3. MCM2 localised to the epithelial cells, and staining produced across development reflected the pattern seen with MCM3 IHC staining (Figure 5.10). A decrease in the intensity of staining and the number of cells that were positive for MCM2 occurred from pregnancy (Figure 5.10B and C) to lactation (Figure 5.10D) and heterogeneous staining was observed during lactation. As with MCM3 IHC staining, after the first couple of days of apoptosis the number of cells expressing MCM2 increased (Figure 5.10E) and this progressed through to the later stages of involution (Figure 5.10F). No staining was observed in the negative controls for either of these proteins although slight background staining could be seen (Figure 5.10G).

5.3.9 Ki67

Ki67 is a conventional proliferation marker and as MCM2 and MCM3 have been reported to detect higher levels of cycling cells in human breast tissue (Stoeber *et al.,* 2001), Ki67 was used as a comparison against the MCM proteins in mouse mammary gland tissue. Each comparison showed that the MCM antibodies consistently stained a greater number of epithelial cells than the Ki67 antibody (Figure 5.11). In the case of the MCM proteins and Ki67, no quantitative analysis was carried out as clear differences could be seen by visual inspection.

5.3.10 Analysis of annexin A2

Annexin A2 was investigated by TaqMan analysis since it was a cluster MCI (see Table 4.1). Annexin A2 clustered with two keratin annotations and malate dehydrogenase. As annexin A2 had the most interesting potential biology in the mammary gland it was decided that this protein had priority over other annotations which clustered with annexin A2. Therefore annexin A2 was the only annotation in the cluster that was analysed by TaqMan and subsequently by western blotting, IHC and microarray analyses.

The expression profiles determined from TaqMan (Figure 5.12A), microarray (Figure 5.12B) and western blot (Figure 5.12C) analyses for annexin A2 were similar; a decrease in expression was observed throughout pregnancy to lactation, and an increase during involution. These results contrasted to the proteomics data which showed annexin A2 to be present only in Inv10. The most likely explanation for this discrepancy is that western

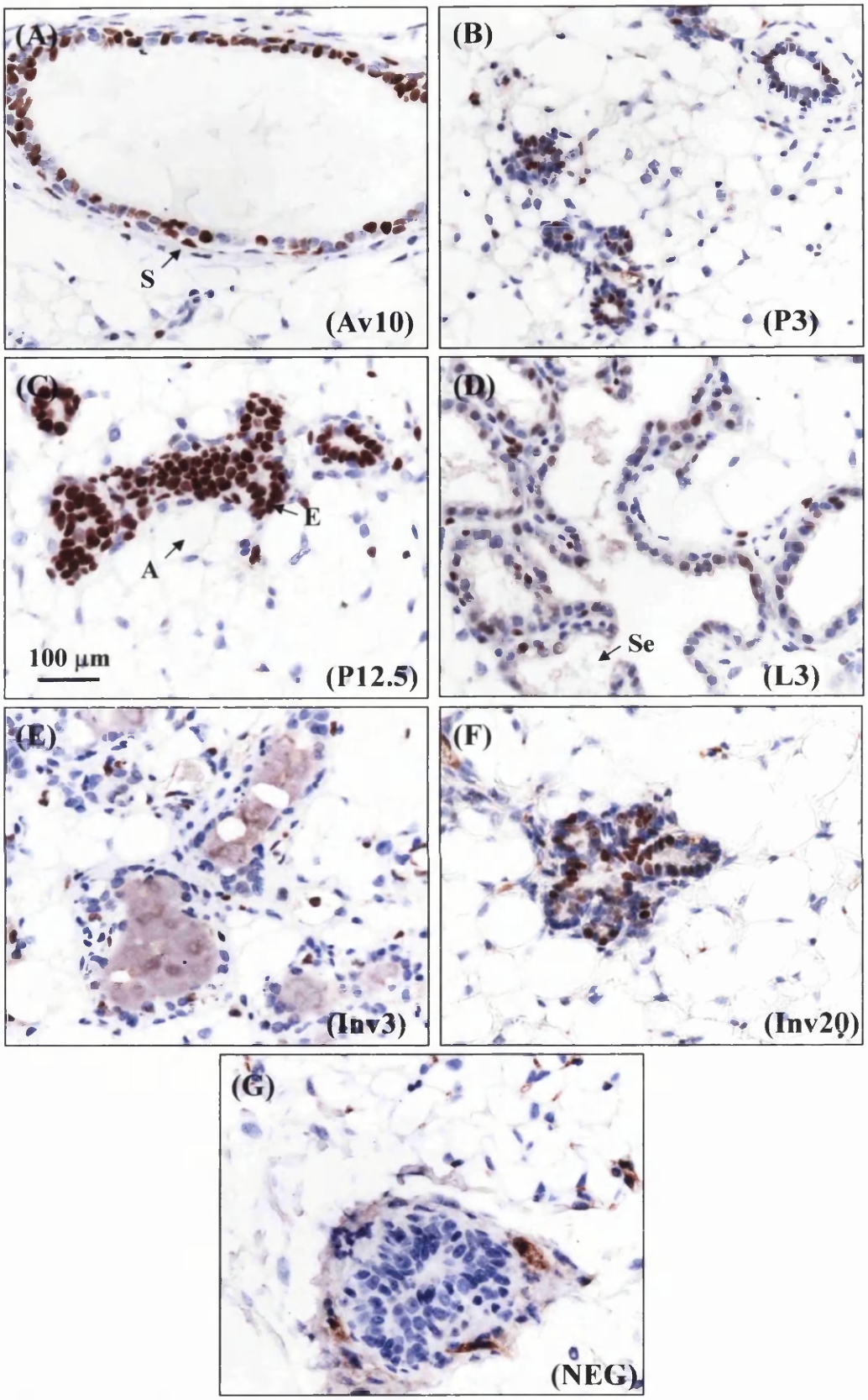


Figure 5.10 Immunohistochemical localisation of MCM2 in various stages of mouse mammary gland development. Paraffin embedded tissue sections were stained with a mouse monoclonal anti-MCM2 antibody. Envision detection system was used and staining was visualised with DAB. Sections were view at 400x magnification using a Zeiss Axiophot microscope and images were processed using Axiovision 3.1 software. (A)-(F) range from Av10 to Inv20 and (G) is the negative control. Stroma (S); epithelial cells (E); secretions (Se); adipocytes (A).

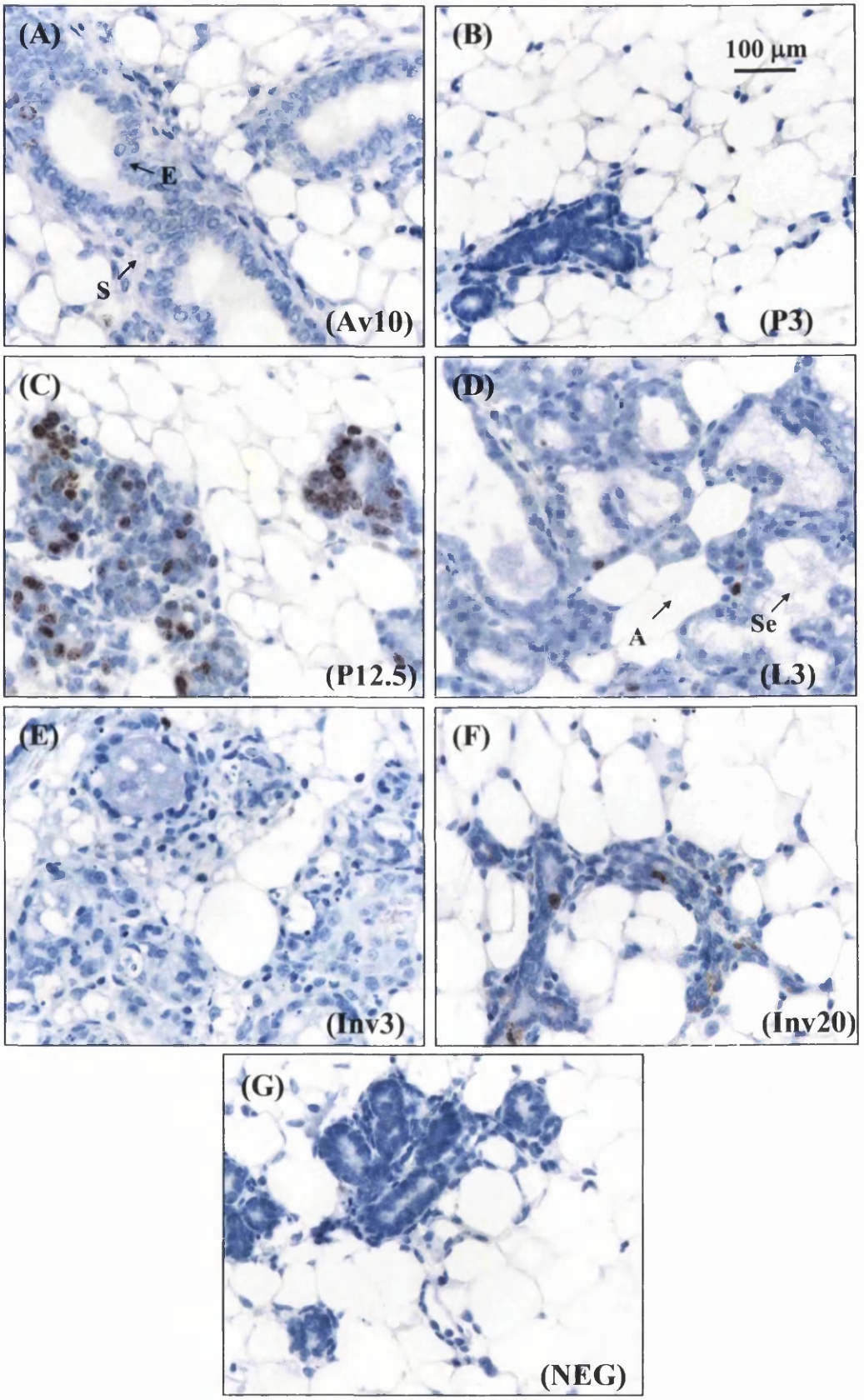


Figure 5.11 Immunohistochemical localisation of Ki67 in various stages of mouse mammary gland development. Paraffin embedded tissue sections were stained with a rat monoclonal anti-Ki67 antibody. An HRP-conjugated secondary antibody was used to visualise staining with DAB. Sections were view at 400x magnification using a Zeiss Axiophot microscope and images were processed using Axiovision 3.1 software. (A)-(F) range from Av10 to Inv20 and (G) is the negative control. Stroma (S); epithelial cells (E); secretions (Se); adipocytes (A).

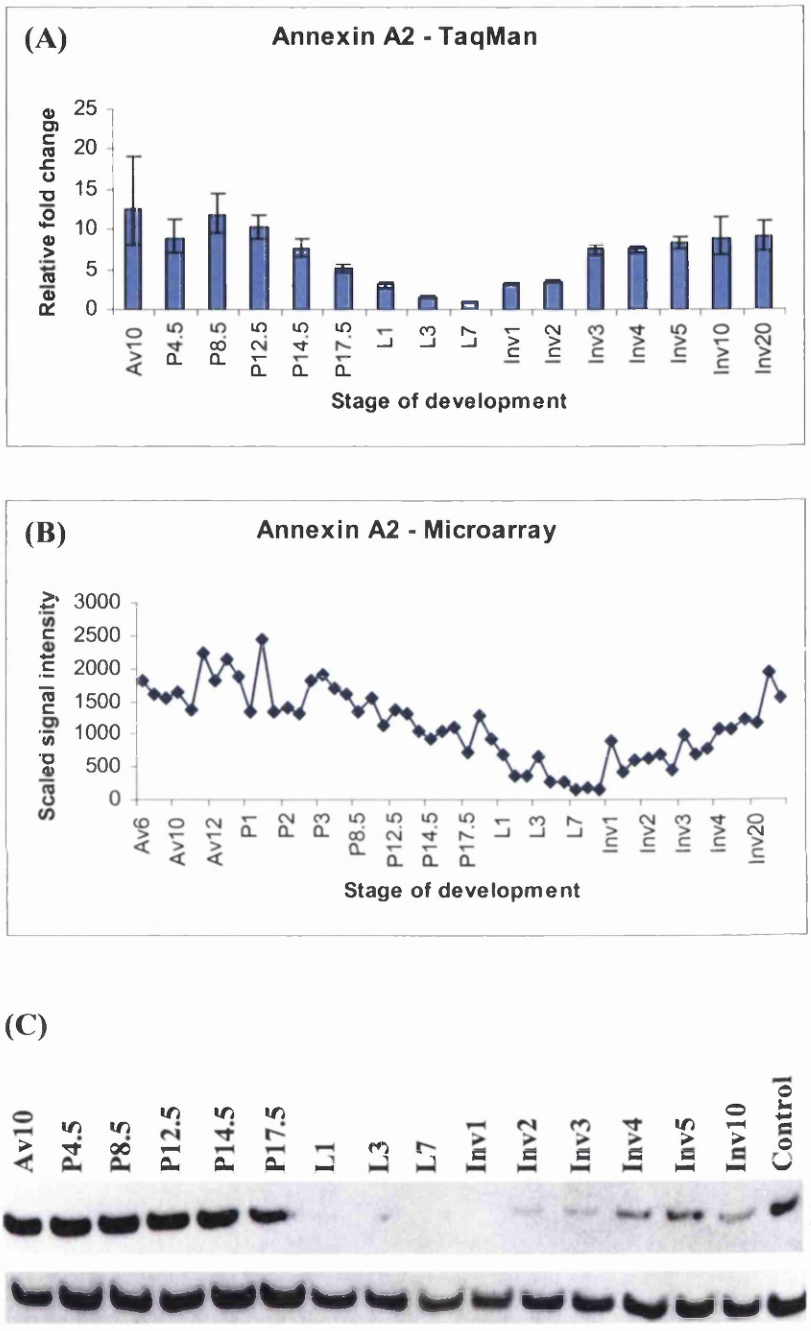


Figure 5.12 Expression profiles of annexin A2 during mouse mammary gland development. (A) TaqMan results are presented as fold change relative to a stage within the set of samples tested. (B) Microarray results are presented by signal intensity which has been scaled to an arbitrary value. (C) Western blot analysis of annexin A2 using human epidermoid carcinoma cells as a positive control. All stages of development were used (40 μg protein) except for Inv20 due to the number of wells present on the gel. Blot was subsequently probed with β-actin for load verification.

blotting, TaqMan and microarray analyses can detect total amounts of protein or mRNA whereas proteomics identifies different isoforms.

IHC analysis of annexin A2 was performed on mouse mammary tissue sections. A serial dilution of the antibody on Av12 sections revealed that different cell types stained at different intensities. At a 1:50 dilution, staining was seen in the stroma, myoepithelial and luminal epithelial cells (Figure 5.13A). Annexin A2 has previously been reported in the stroma of normal and tumour breast tissue (Schwartz-Albiez, *et al.* 1993), and in mouse mammary epithelial cells (Handel *et al.,* 1991; Creutz, 1992 and Lozano *et al.,* 1989). Smooth muscle cells and adipocytes were consistently negative and endothelial cells were positive at higher concentrations. The antibody was used at a 1:100 dilution across mammary development to determine quantitative changes in expression between different cell types. The stromal fibroblast cells showed striking changes in the intensity of staining, and appeared to be both membrane and cytoplasmic. The epithelial cells were consistently negative.

Additional stages of mammary gland development were analysed with annexin A2. An increasing number of fibroblasts stained positive for annexin A2 from Av6 to Av10, as well as the intensity of the stain. At Av6 only occasional cells were positive around the TEBs. The fibroblasts stained strongly positive from Av12 (Figure 5.13B) through to the early stages of pregnancy (P1, P2 and P3) (Figure 5.13C), however this gradually decreased from P4.5 to P17.5 (Figure 5.13D). Although the proportion of stromal to epithelial cells decreased during pregnancy, the intensity per cell also decreased. The low level of staining continued through lactation (Figure 5.13E) and early involution (Figure 5.12F), but as the stromal to epithelial cell ratio increased, the intensity of the stain returned (Inv4-Inv20) (Figure 5.13G). At all stages of development, the myoepithelial and luminal cells were consistently negative, and no staining was observed in the negative control (Figure 5.13H).

5.3.11 IHC staining of annexin A2 in normal human breast tissue

Differential staining of annexin A2 was determined in the mouse mammary gland after using an appropriate antibody dilution. This result has not previously been reported as other studies used the antibody at too high a concentration. As this may have been the case with the breast tissue reports on annexin A2, further investigations were performed to identify whether differential staining could be seen in different cell populations of the breast. Strong IHC staining was seen in the myoepithelial and stromal cells, when using a

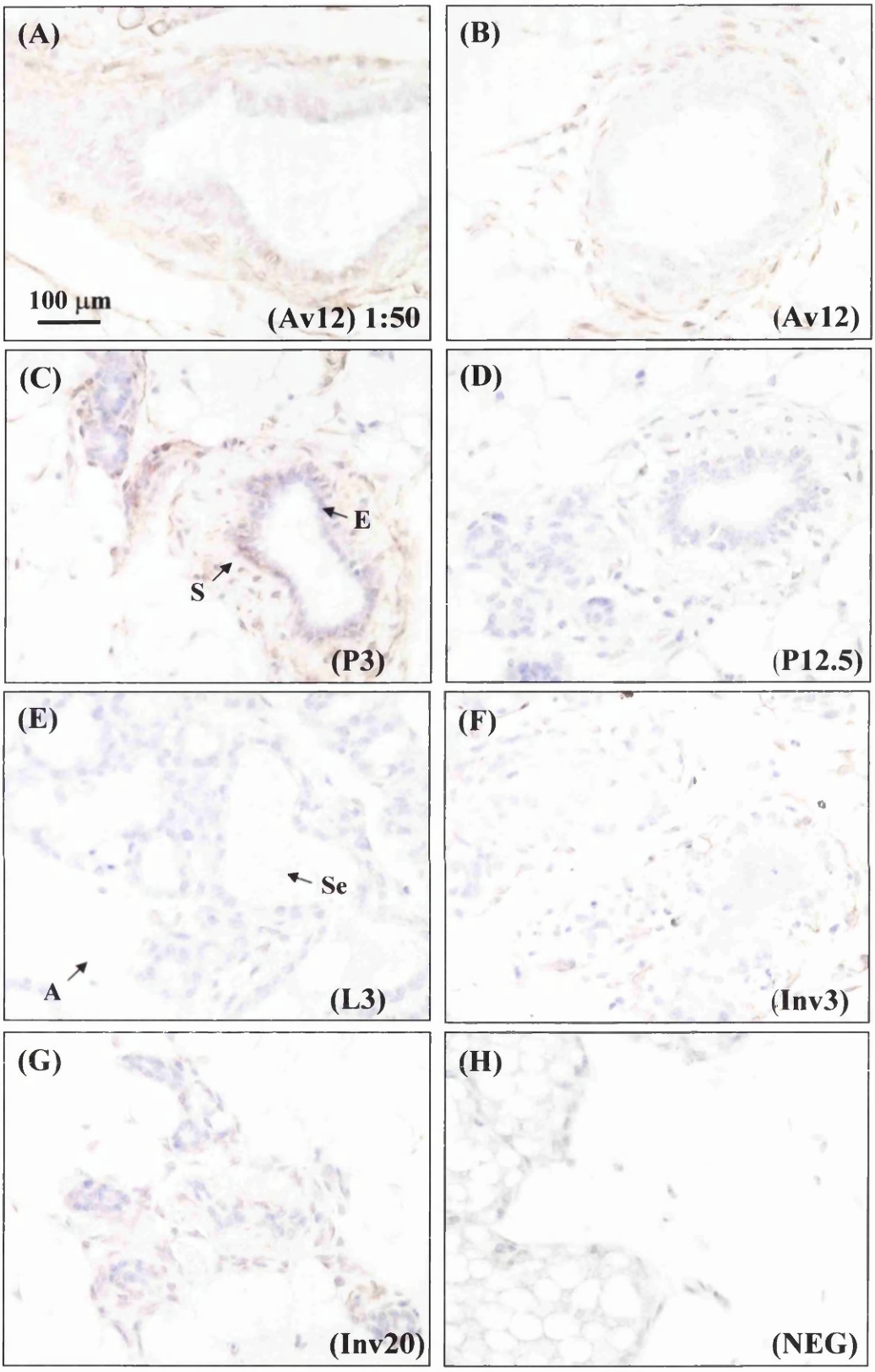


Figure 5.13 Immunohistochemical localisation of annexin A2 in various stages of mouse mammary gland development. Paraffin embedded tissue sections were stained with a mouse monoclonal anti-annexin A2 antibody. An HRP-conjugated secondary antibody was used to visualise staining with DAB. Sections were viewed at 400x magnification using a Leitz Orthoplan microscope. Images were processed using Axiovision 3.1 software. (A, 1:50) adult virgin gland, (B-F, 1:100) range from Av12 to Inv20, and (H) is the negative control. Stroma (S); epithelial cells (E); secretions (Se); adipocytes (A).

1:50 dilution of the annexin A2 antibody, whereas weak staining was observed in the luminal cells (Figure 5.14A). Annexin A2 was predominantly localised to the stromal and myoepithelial cells at a 1:100 dilution however, the luminal cells were rarely positive (Figure 5.14B). Luminal staining of the apical membrane was identified in some ducts, and a variable staining pattern could be seen in the myoepithelial cells. In some areas the staining appeared to be vesicular (Figure 5.14C) and in others annexin A2 was located to the membrane (Figure 5.14D). At high power the staining appeared to be granular however, these could be very small vesicles which can not be not resolved at the light microscope level (Figure 5.14D). Positive staining was identified at the myoepithelial basal/stromal interface, and it was often difficult to deduce whether the staining was associated with the myoepithelial cells or the epithelial/stromal interface.

5.3.12 IHC staining of annexin A2 in benign hyperplasia and papillomas

Five cases of epithelia hyperplasia and two benign papillomas were examined. Both types of lesions are associated with the coordinated proliferation of luminal and myoepithelial cells. In the benign hyperplasia (Figure 5.14E) and papilloma cases (Figure 5.14F), the stain localised to the membrane and vesicular pattern of the myoepithelial cells but the luminal cells were relatively negative.

5.3.13 IHC staining of annexin A2 in *in situ* breast cancer

Twenty cases of ductal carcinoma *in situ* (DCIS) were examined, together with eight cases of lobular carcinoma *in situ* (LCIS). The epithelial cells that form these lesions had a similar pattern of staining as was seen in the myoepithelial cells in normal breast. The staining was strong in most of the cases, but it was stronger in all cases when compared to the predominantly negative luminal cells in the adjacent normal breast tissue. The staining pattern was variably granular and membrane, with small to large vesicles within the cytoplasm. See Figure 5.14G for membrane staining. There was very variable staining within DCIS, both in the intensity and staining pattern. Two tumours were negative (Figure 5.14H), others displayed membrane staining (Figure 5.14I) and some showed clear intracytoplasmic vesicles (Figure 5.14J). The attenuated myoepithelial cells were difficult to identify in the majority of cases of DCIS, but occasionally clear positivity was seen (Figure 5.14K)

5.3.14 IHC staining of annexin A2 in invasive lobular and invasive ductal carcinoma

Ten cases of invasive lobular carcinoma and 39 cases of invasive ductal carcinoma of various grades were examined. The staining pattern of infiltrating lobular carcinoma cases

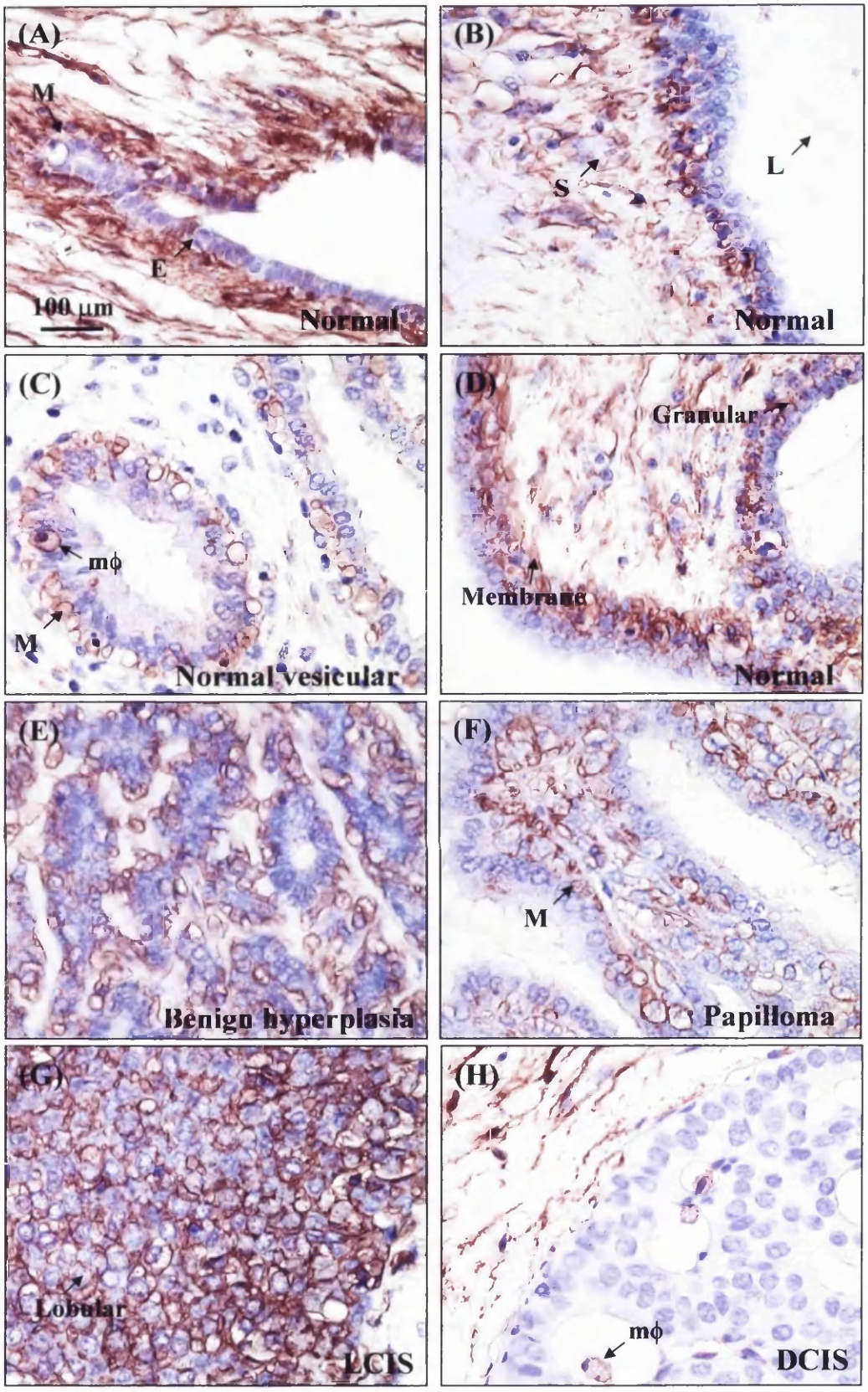


Figure 5.14 Immunohistochemical localisation of annexin A2 in breast tissue. Paraffin embedded tissue sections were stained with a mouse monoclonal anti-annexin A2 antibody. An HRP-conjugated secondary antibody was used to visualise staining with DAB. Sections were view at 400x magnification using a Leitz Orthoplan microscope and images were processed using Adobe Photoshop 7.0 software. (A) 1:50 dilution; (B-H) 1:100. Lobular *in situ* carcinoma (LISC); ductal *in situ* carcinoma (DCIS); epithelial cells (E); lumen (L); myoepithelial cells (M); macrophages (mϕ); stroma (S).

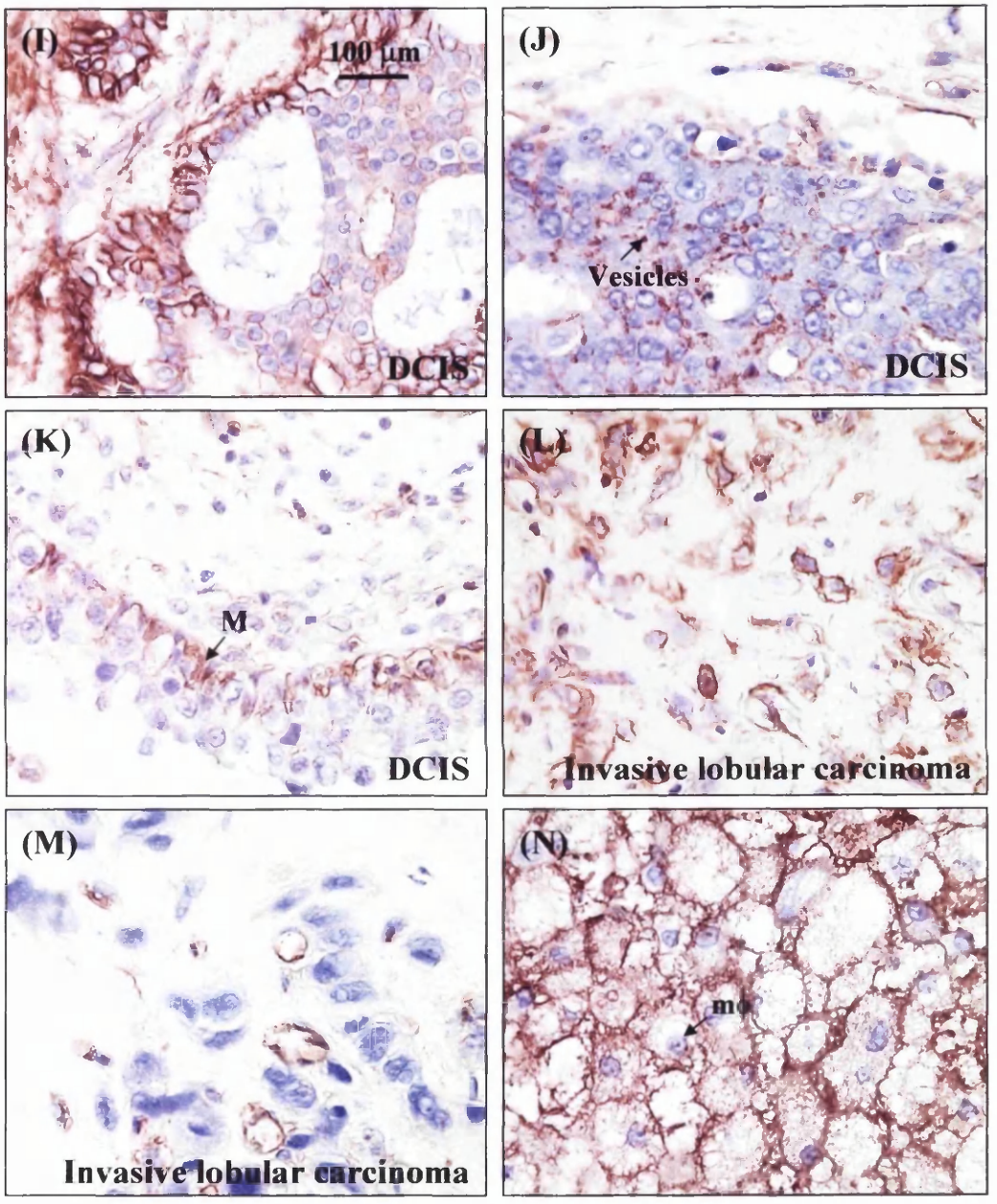


Figure 5.14 Immunohistochemical localisation of annexin A2 in breast tissue. Paraffin embedded tissue sections were stained with a mouse monoclonal anti-annexin A2 antibody. An HRP-conjugated secondary antibody was used to visualise staining with DAB. Sections were view at 400x magnification using a Leitz Orthoplan microscope and images were processed using Adobe Photoshop 7.0 software. (I-N) 1:100 dilution. Ductal *in situ* carcinoma (DCIS); myoepithelial cells (M); macrophages (mϕ).

were difficult to interpret, as 9/10 cases showed strong staining on the membrane of the infiltrating carcinoma cells or at the epithelial/stromal interface. This was difficult to interpret under a light microscope, but the similar staining in LCIS suggests that it is membrane staining (Figure 5.14L). One infiltrating lobular carcinoma case was negative (Figure 5.14M). The staining of the infiltrating ductal carcinoma cases varied from strong to weak and one case was negative. The staining patterns were similar to those seen in the DCIS cases. There was not a consistent staining pattern within a tumour and all patterns of staining could be seen in different lesions. Weak staining tended to be focal, whereas the 15 cases that had strong staining showed that most of the tumour cells were positive. Within the *in situ* and invasive carcinomas, macrophages were strongly positive (Figure 5.14N), which is important when interpreting focal staining.

The tumours were graded according to the Nottingham grading classification (Table 5.1). The intensity of the staining was graded from zero (negative) to three (strongly positive) and no association was found between the strength of staining and the tumour grade. Chi-square test was used for statistical analysis which compared the number of cases that were positive for annexin A2 to the grade of the tumour. The results were not significant.

Table 5.1 Annexin A2 staining intensities of invasive carcinoma cases

	Staining intensity				
	0	**1**	**2**	**3**	**Total**
Grade 1	0	3	1	2	6
Grade 2	0	6	3	6	15
Grade 3	2	5	4	7	18

5.3.15 2-D western blot analysis of annexin A2

To demonstrate the problems caused by PTMs, a protein which was known to be modified by phosphorylation was detected by western blotting after being separated by 2-D gel electrophoresis. The protein investigated by this method was annexin A2 which had already been shown to be modified in this project, as it had two different pI values in the developmental and TEB databases (see Table 4.1 and 4.3). Three stages of development were analysed, Av10, L7 and Inv10 in duplicate (Figure 5.15). Non-specific staining was detected above the located area of annexin A2 which had migrated at 36 kDa. The duplicate images obtained from the Av10 samples were very similar to each other. A distinct horizontal trail of the protein was detected across the blot which suggested that the protein had been phosphorylated (Figure 5.15A and B). The duplicate samples from L7

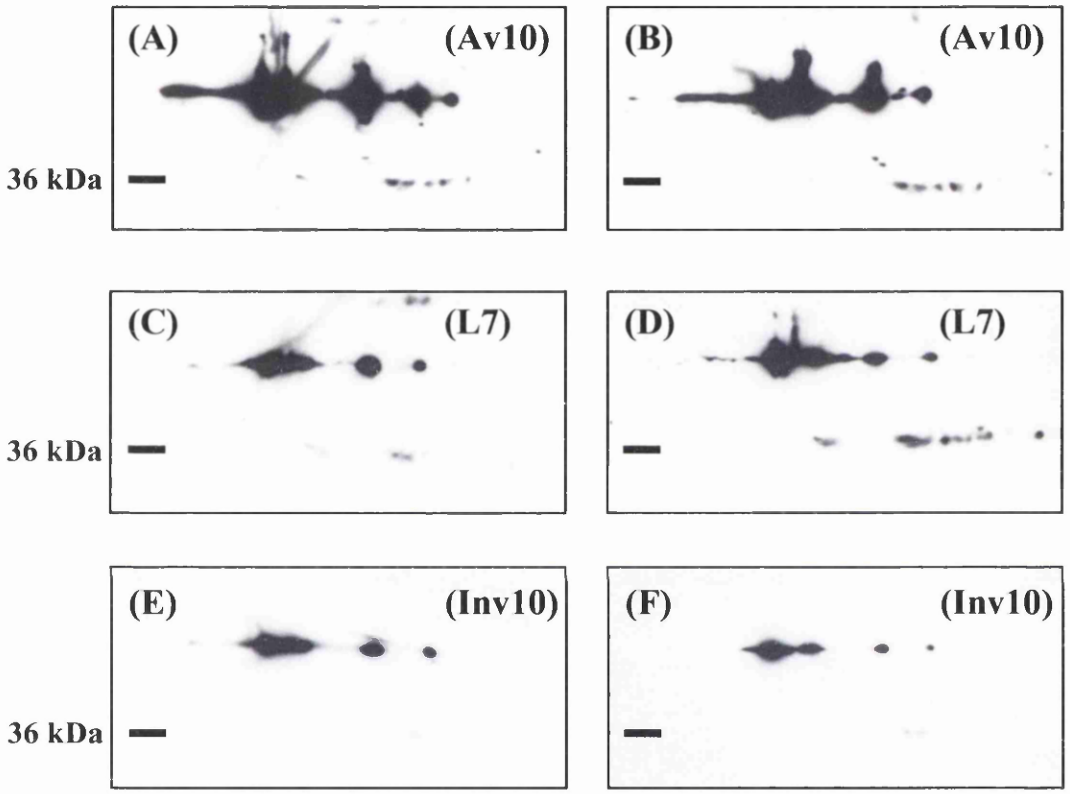


Figure 5.15 2-D western blot analysis of annexin A2. Annexin A2 migrated at 36 kDa. All stages indicated were repeated in duplicate, loading 80 μg of protein per 2-D gel.

were not as similar to each other; however, the horizontal trail of annexin A2 was again evident (Figure 5.15C and D). The last stage analysed (Inv10) showed a decrease in the level of annexin A2 staining which was almost undetectable. Although both gels showed weak staining, one shown in Figure 5.15F was slightly stronger than the other (Figure 5.15E). The horizontal trail produced by this protein was less prominent compared with the other two stages. Duplicating the results using pre-cast mini 2-D gels was difficult to achieve as can be seen from the results, which also illustrate the problems caused by PTMs in 2-D gel analysis. In this proteomic project the shift in pI caused by phosphorylation would have meant that the altered form had a different MCI to the one selected and would not have been detected as the same protein. This may explain the discrepancy between the proteomic expression profile for annexin A2 compared to the 1-D western blot profile.

5.3.16 Analysis of hypothetical and Incyte unique annotations

Two annotations which were classified as unique proteins and one classed as a hypothetical protein were selected for further follow-up work. The hypothetical protein, DJ845O24.2 clustered with one of the unique proteins, MCI 48838. These annotations were of particular interest as they are uncharacterised annotations with no known function. The release of the mouse genome sequence into the public domain enabled the peptide sequences which were matched to these annotations to be reanalysed to determine whether they were linked to a characterised protein. Mouse NCBI, Celera_16 and Sanger databases were searched using protein basic local alignment search tool (BLAST) and theoretically predicted protein sequences (TBLASTN and GENSCAN). The unique sequence identified for MCI 46776 CVNIQMLQGVK revealed no hit alignment even when changing leucine to isoleucine as they have the same mass values. The peptide sequence determined for the unique protein MCI 48838 TGIIIILSEVK was quite ambiguous, as it contained many leucine and/or isoleucines and these created many possible combinations of the protein sequence. All combinations were searched but no matches were found. KPVEDCPR from the hypothetical protein MCI 48838, the last sequence searched, also revealed no hit alignment within the mouse sequence. However, two hits were matched to the human database NCBI, accession numbers O60810 and XP_088816. The former sequence had previously been identified by Oxford GlycoSciences which was hypothetical protein DJ845O24.2 and the latter sequence identified the protein 'similar to' hypothetical protein DJ845O24.2 which has since been removed from the NCBI database. In conclusion no additional information about these three sequences was revealed by these searches.

5.4 Discussion

5.4.1 Identifying post-translational modifications

Before proceeding further with the selected annotations, several different techniques were used in order to support the proteomics results. As antibodies were not available for all annotations, mRNA techniques were used to determine the identity of a clustered MCI. However, in most cases these methods did not confirm the proteomics results. Some annotations which were detected during only one stage of mammary gland development by proteomic analysis, appeared to be expressed in most stages of development using the other follow-up techniques. This does not mean that the proteomics results are incorrect. As has been explained in the introduction of this Chapter, using mRNA techniques is not an ideal method for confirming the proteomics results, because protein and mRNA trends do not always correlate. However, as no other resources were available, this was the only option to confirm the proteomic data, as for most candidate proteins antibodies were not available. Furthermore, even when antibodies are obtainable, they detect total protein levels rather than individual isoforms.

The most obvious explanation for the inconsistencies between the proteomic results and the follow-up data is the presence of PTMs and sensitivity of 2-D gel proteomics. Many changes are regulatory and reversible, such as protein phosphorylation. Modifications to the phosphorylation state of a protein cause a shift horizontally across a 2-D gel and glycosylation increases the Mr of the protein. Protein modifications can be detected by mass spectrometry but they were not identified in this project because of time restrictions and limitations on sample size. Many phosphorylated proteins are present in such small quantities that it is difficult to isolate them *in vivo* and therefore many phosphorylation studies are performed *in vitro*.

5.4.2 Expression of β-casein during mammary gland development

In order to check the protein sample preparation, the samples were tested at the protein level by western blotting with β-casein. β-casein was tested, because the expression profile during mammary gland development was well known. The western blot proved that the protein samples were of good integrity.

5.4.3 Expression and localisation of perilipin in mammary gland development

Western blot analysis and IHC staining of perilipin did not agree with the proteomics data which may be due to PTMs. The western blot data did reveal differential expression of the different isoforms across mammary gland development. The patterns of expression

observed by IHC and western blotting were typical of those expected as a result of diminishing amounts of adipocyte cells in pregnancy and the replenishment of these cells during the latter stages of involution. For perilipin C, however, a high level of expression was seen during the earlier stages of involution rather than the latter suggesting it may be important during the early stages of tissue remodelling with the return of adipocytes. Perilipins are thought to be important in regulating lipolysis (Londos *et al.,* 1999) and therefore the expression of this protein would be expected to change with the changing energy requirements needed over mammary gland development.

5.4.4 Expression of lumican and contrapsin during mammary gland development

TaqMan analysis ruled out carboxylesterase as being the possible identity of MCI 49502, as no expression was detected across mammary gland development. Therefore lumican, contrapsin and a segment of chromosome remained as potential identifiers. Neither expression profile obtained by TaqMan for contrapsin or lumican matched the proteomic expression profile, which may be due to PTMs or due to the fact that the fourth annotation identified was the correct annotation for MCI 49502. Lumican has been reported to be expressed in the stroma of breast tissue and the mouse mammary gland. Therefore the expression profiles obtained by TaqMan and microarray analysis for lumican may only be a mirror image of the change in epithelial to stromal ratio. During pregnancy there is an increase in the epithelial to stromal ratio and this is reversed during involution. The increased expression of lumican reported in breast stroma of normal and cancerous tissue (Leygue *et al.,* 1998) suggested that it could play a regulatory role in mammary gland development.

In contrast the expression profile of contrapsin does not appear to result from a change in the stromal to epithelial ratio. The TaqMan expression profile for contrapsin only increased towards the end of involution and virtually no change was found in the other stages studied. Further confirmation of these results is required and although the results do not agree with the proteomic data, the data obtained by TaqMan analysis implicate a role for contrapsin during involution of the mammary gland. The expression of contrapsin during involution is particularly interesting as it is a stage where much remodelling of tissues, including the ECM, is taking place. This theory is further supported as contrapsin is known to be an extracellular matrix protein and serpins have previously been linked to extracellular matrix remodelling (Pei *et al.,* 1994; Cao *et al.,* 2004).

5.4.5 Expression of PTRF, Rab11 and WDR1 during mammary gland development

Western blot analysis of PTRF did not confirm the proteomic results and again this may be a possible consequence of phosphorylation. The PTRF antibody did not distinguish between phosphorylated and non-phosphorylated PTRF and it is known that PTRF is phosphorylated at multiple sites. Further investigations into the phosphorylation state of PTRF are required, since an antibody which detects only phosphorylated PTRF may show expression changes on a western blot profile. Its involvement in transcriptional termination by Pol I may still be regulated during mammary gland development (Mason *et al.,* 1997; Jansa *et al.,* 1998). However, its expression profile remains unclear.

No conclusion can be drawn from the Rab11 results. Further investigations into the expression profile of the different isoforms is required during mammary gland development because neither mass spectrometry nor western blotting differentiated between Rab11a or Rab11b. It would appear from the western blot that the overall expression of Rab11 changes during pregnancy with a strong increase in lactation which may indicate that it is important in this stage of development. This is further supported by its involvement in vesicle transport as it may play a role in the transportation of milk components during lactation (Novick and Zerial, 1997).

The western blot profile of WDR1 showed no change in expression across development. After expanding the proteomic results, a constant level of expression was also observed during Av10 and pregnancy. However, the expression levels of each sample varied considerably in involution, making it difficult to draw conclusions about its overall expression. It was therefore surmised from these experiments that WDR1 did not play a regulatory role during mammary gland development.

5.4.6 Expression and localisation of MCM proteins

Overall the results collected in this study for MCM3 complimented each other. The TaqMan data was generally similar to the proteomic profile except for the latter stages of involution. This may be due to discrepancies which are known to occur between RNA and protein. A tentative explanation for these discrepancies may be because MCM3 is cleaved during apoptosis (Schwab *et al.,* 1998). The six members of the MCM family contain several caspase recognition sequences (Hu *et al.,* 1993). Investigations into MCM proteins during apoptosis of proliferating cells revealed that MCM3 was cleaved early in apoptosis. It has been proposed that cleavage of MCM3 prevents untimely DNA replication during

apoptosis (Schwab *et al.*, 1998). Cleavage of a protein would alter the location of a protein on a 2-D gel which may be why the TaqMan results differed during involution.

The results obtained from IHC for both MCM proteins investigated generally had the same expression profile to each other and to all the other data obtained for MCM3. The MCM antibodies appeared to stain only epithelial cells, and because of their involvement in cell replication the staining obtained by IHC for MCM2 and MCM3 was compared to the Ki67 proliferation marker. The MCM antibodies stained a higher proportion of luminal epithelial cells than the Ki67 antibody, as has been demonstrated in human breast tissue (Stoeber *et al.*, 2001). One suggested theory for this is that the MCM proteins are expressed in cells which have proliferative potential as well as in actively proliferating cells (Stoeber *et al.*, 2001). MCM proteins have therefore been investigated as potential prognostic markers for breast cancer. These proteins may be useful to assess proliferation in the mouse mammary gland as it is thought that Ki67 provides a limited assessment of the cell cycle state.

No conclusion was made about the correct identity of MCI 46330 and further investigations are needed to access the fumarate hydratase annotation.

5.4.7 Expression and localisation of annexin A2 in mammary gland development

The patterns of expression of annexin A2 determined by western blotting, TaqMan and microarray analyses all complimented each other. However, the results did not confirm the proteomic data. There are a number of possible explanations for this. Firstly annexin A2 clustered with two keratin isoforms and malate dehydrogenase. No investigations were performed to exclude these annotations, as the identity to MCI 48474 and therefore annexin A2 may not be the correct one. Secondly, annexin A2 is phosphorylated within the amino-terminal region. Tyr-23 is phosphorylated by v-Src (Glenney, 1985) and Ser-25 is phosphorylated by protein kinase C (Gould *et al.*, 1984). Two isoforms of annexin A2 were determined from the developmental and TEB databases which probably have different states of phosphorylation due to their differing pI values. It is therefore also plausible that, if the correct identification of MCI 48474 was annexin A2, a change in its phosphorylation state during mammary gland development would result in a different proteomic expression profile when compared to a western blot. 1-D western blots do not distinguish between different phosphorylation changes in the way that 2-D gels can and this has been demonstrated with annexin A2 in this Chapter.

A 2-D western blot of annexin A2 revealed the typical pattern changes seen by a modification in the phosphorylation state of a protein. However the reproducibility of the results was questionable, since variations between duplicate samples could be seen. Immobilised pH gradient gels used have superior resolution and reproducibility compared to past techniques performed with tube gels and ampholytes. Reproducibility in tube gels was difficult to achieve due to cathodic drift and fragility of the gels. However, optimising the focusing conditions may have improved the reproducibility of the gels. The images produced with mini 2-D gels were sufficient to demonstrate the effect of protein phosphorylation changes and therefore no further investigations were performed.

The results obtained for annexin A2 confirmed previous data concerning the distribution of this protein in the mammary gland during pregnancy and lactation, and provided additional information for involution. However the altered levels of expression between different cell types, and changes in expression during mammary development, have not previously been reported. Annexin A2 localised to stromal, myoepithelial and epithelial cells of the mammary gland. Staining in the epithelial cells was not identified until a higher concentration of antibody was used. The altering level of expression of annexin A2 may have resulted from the altering ratio of epithelial to stromal cells during development. In relative terms there appears to be less stroma during pregnancy and lactation, and more during mid to late involution. However, the IHC results showed a clear decrease in intensity per cell during pregnancy and lactation and an increase during involution.

As previously mentioned, conflicting results have been published on the expression of annexin A2 in the mammary gland. Annexin A2 is reported to play a role in DNA synthesis and proliferation (Jindal and Vishwanatha, 1990; Jindal *et al.,* 1991; Kumble and Vishwanatha, 1991) however, annexin A2 is not expressed in the highly proliferative cap cells of the TEBs (Schwartz-Albiez *et al.,* 1993). The results in this study found very few TEBs stained with the annexin A2 antibody. In addition, unless a high concentration of antibody was used, annexin A2 was not detectable in the epithelial and myoepithelial cells during the proliferative phase of pregnancy. Annexin A2 was highly expressed in the stroma in the virginal and early pregnant gland, and numerous studies have demonstrated the importance of the stroma during pregnancy and involution in mammary gland development (Neville *et al.,* 1998; Elias *et al.,* 1973). Previous investigations have shown that annexin A2 binds to tPA which has a role degrading ECM proteins (Hajjar *et al.,* 1994; Mignatti *et al.,* 1986). A tentative explanation for its role in early pregnancy is that

annexin A2 is initially required for remodelling of the ECM during this early proliferative phase of development.

As the IHC data demonstrated different cell staining patterns with different antibody dilutions, it was thought that the discrepancies between previous reports may have been due to the concentration of the antibody being used. It is important to determine the appropriate dilution of an antibody since changes in expression of a protein may be masked with low dilutions. This may have been the case with previous annexin A2 breast tissue and cancer studies. Therefore, careful analysis of the expression of annexin A2 in human breast and a range of benign and malignant breast diseases was performed.

5.4.8 Expression of annexin A2 in breast tissue

In the normal breast, annexin A2 was differentially expressed in stromal and myoepithelial cells, compared with the luminal cells. The luminal cells were rarely positive when a 1:100 dilution of antibody was used. Differential staining was observed between the basal (myoepithelial cells) versus luminal cells in the benign breast lesions. In the *in situ* breast cancer cases (LCIS and DCIS) annexin A2 was up-regulated in the tumour cells. This was also seen in the majority of the infiltrating carcinomas (lobular and ductal).

This pilot study was performed on a limited number of samples, and statistical analysis showed that the up-regulation of annexin A2 was independent of the grade of tumour. However, further investigations on a larger number of samples may show significance between staining intensity of annexin A2 and tumour grade. A more extensive study linked to clinical outcome is necessary to assess whether the up-regulation of annexin A2 is related to any clinico-pathological correlates such as survival and metastatic spread.

A theoretical consideration for annexin A2 being over-expressed in breast cancer, may be linked to this protein being a cell surface receptor to the t-PA and plasminogen (Hajjar *et al.,* 1994). The plasminogen proteolytic system plays a key role in tumour cell invasion through its interaction with secreted proteases (Mignatti *et al.,* 1986). In addition to its role in degrading ECM proteins, plasmin can activate several latent metalloproteases relevant for malignant cell invasion (He *et al.,* 1989; Knauper *et al.,* 1996). However, the invasive process of t-PA is not completely clear, and the precise functions of annexin A2 are unknown. A possible explanation for the over expression of annexin A2 in several tumour types could be related to it binding to t-PA, which is important in tumour invasion.

5.4.9 Hypothetical and unique protein annotations

No further information was acquired for the hypothetical and unique proteins of interest. Reasons for not finding these sequences in the various databases may be (1) due to different isoforms being present because of alternative splicing, (2) the sequence has not been predicted or (3) it is missing in the genome. If these peptide sequences had matched to a protein based on the assumption that sequence domains reflect functional aspects of a protein, it may have been possible to predict the functions of these annotations by sequence similarity comparisons. This could have been achieved using the SWISS-PROT database and BLAST searches (NCBI, http://www.ncbi.nlm.nih.gov). However, there can be errors with computational predictions necessitating experimental evidence to ascertain their biological role.

In conclusion, the major limitation of following-up the proteomic data was the lack of specific antibody resources with which to both identify individual proteins and also to distinguish between potential isoforms. The discrepancies which were found with the follow-up experiments could be due to PTMs. The annexin A2 2-D gel western blot demonstrates the problem caused in proteomics by a shift in protein phosphorylation state. If PTMs had been determined at the mass spectrometry stage of this proteomic project, it would have aided the subsequent follow-up studies of candidate proteins. It is possible to detect PTMs by mass spectrometry. However, this requires a larger sample size, greater time, expertise and expense. Therefore, for these reasons this additional analysis was not performed in this study.

Despite the lack of specific resources available to confirm the proteomic results, some of the candidate annotations displayed interesting expression profiles across mammary gland development and it would be interesting to determine their precise role in this field of research.

Chapter 6 presents data on MCIs which were reanalysed by mass spectrometry at the University of Glasgow in July 2003. These MCIs had previously been analysed at Oxford GlycoSciences in October 2001 and 2002, and no protein annotations had been identified. MCIs were selected for reanalysis based on their proteomic expression profiles across mammary gland development.

Chapter 6

Reanalysis of selected MCIs by mass spectrometry

6.1 Summary

This Chapter describes the reanalysis of selected MCIs which been analysed by mass spectrometry at Oxford GlycoSciences but had revealed no protein annotation. Twenty eight MCIs were chosen for reanalysis by mass spectrometry from the developmental and Lactation/Involution databases. The majority of MCIs selected from the developmental database for reanalysis had present calls in more than one stage of pregnancy. MCIs which were chosen for reanalysis from the Lactation/Involution database were selected on the basis of their pattern of expression after expanding the profile to the rest of the developmental stages. Included in the selection were the hypothetical protein MCI 48838 and the unique identification MCI 46776.

Fourteen MCIs out of 28 which previously had no annotations identified by the first round of sequence database searches at Oxford GlycoSciences now had an identification after reanalysis by mass spectrometry at the University of Glasgow. The second sequence database search performed by Oxford GlycoSciences, provided additional annotated information on all MCIs investigated by mass spectrometry (see Chapter 4 Tables 4.1 and 4.2). This data provided a direct comparison to the results produced by the University of Glasgow. In most cases the data was complimentary.

Prioritising the newly acquired data was necessary due to time constrictions, and therefore only one of these annotations was followed up, namely Transcription factor BTF3 (BTF3).

6.2 Introduction

Twenty eight MCIs were reanalysed by mass spectrometry in an attempt to determine a protein annotation. The first MCIs had been selected at Oxford GlycoSciences based on a presence/absence assessment at specific stages in development. Although this was an absolute change it did not yield results which were readily confirmed. The discrepancy found between the proteomics data and follow-up experiments was mainly attributed to PTMs. Additionally, proteomics is able to identify specific protein isoforms whereas the follow-up techniques used only identified total protein or mRNA levels. A presence/absence assessment could not have avoided these modifications, but choosing MCIs based on their expression pattern across development would perhaps minimise their selection. Reanalysis of 28 MCIs based on their expression pattern, introduced new protein annotations to the existing data. One particular annotation is the main focus of this Chapter.

The initiation of gene transcription by RNA polymerase II requires the assembly of a number of promoter elements which bind to the polymerase in a multiprotein complex (Sawadogo and Sentenac, 1990; Greenblatt, 1991; Drapkin *et al.,* 1993). BTF3 associates with RNA polymerase II, although its function remains unclear. BTF3 was first purified in HeLa cells (Zheng *et al.,* 1987) and two isoforms BTF3a and BTF3b originate from a single gene as a result of alternative splicing (Zheng *et al.,* 1990; Kanno *et al.,* 1992). BTF3b encodes a shorter protein which lacks the first 44 N-terminal amino acid residues of BTF3a and, despite its ability to bind to RNA polymerase II, it is transcriptionally inactive. BTF3 is a 27 kDa protein, and is not abundantly expressed in cells.

A 2-D gel study investigating changes in protein patterns between apoptotic and non-apoptotic human Burkitt lymphoma cells identified BTF3. The protein feature on the 2-D gel identified as BTF3 revealed a decrease in intensity in the apoptotic compared with non-apoptotic cells, although further investigations were needed to differentiate between (a) and (b) isoforms. The report concluded that BTF3 had been identified as an apoptotic associated protein (Brockstedt *et al.,* 1999).

BTF3 was selected for further follow-up work because of its apparent developmental regulation in the mammary gland which was displayed by its proteomic expression profile.

6.3 Results

6.3.1 Reanalysis of unidentified MCIs by mass spectrometry

Similar to the initial mass spectrometry analysis performed at Oxford GlycoSciences, cuts were taken from the gel containing the most abundant amount of the MCI of interest. Reanalysis of selected MCIs was performed by the Sir Henry Wellcome Functional Genomics Facility (SHWFGF) at the University of Glasgow. After tryptic digestion, the samples were analysed by nano-LC electrospray tandem mass spectrometry. Tandem mass spectrometry was performed on QStar Pulsar i (Applied Biosystems) after having separated the tryptic peptides by reversed phase (PepMap C18 column) liquid chromatography on an LC Packings Ultimate LC system (Dionex) directly interfaced to the Qstar via a nanoflow source (Protana) in order to maximise sensitivity. The data obtained from mass spectrometry were used to search publicly available sequence databases using an in-house Mascot (Matrix Science) search engine to determine the protein identifications of the MCIs analysed.

Mascot calculates the probability of the peptide match being a random event. Mascot scores above 59 are significant ($p<0.05$), although this value produces a relatively high error rate. Therefore in this study protein annotations were only counted as valid if the Mascot score was above 99 with at least two peptides matched to the annotation, each with significant scores.

6.3.2 Protein annotations identified

The annotated MCI results obtained by SHWFGF are listed in Tables 6.1 and 6.2. These tables list the protein annotations according to the databases from which they were chosen and from their general function respectively. The majority of the annotations identified have been classified as ME/MP (Table 6.2). Four newly identified annotations were determined from the developmental database. All were selected for reanalysis by mass spectrometry due to their presence in several stages of pregnancy, indicating a possible role during this phase of development. Histone H2A and membrane associated progesterone receptor component have been classed under 'other', and keratin 5 and actin as cytoskeletal proteins. No cytoskeletal proteins were identified from MCIs taken from the Lactation/Involution database.

The remaining 10 annotations were identified from the Lactation/Involution database. However, not all the MCIs chosen produced novel identifications to the project due to the second update by Oxford GlycoSciences. Although this second update was performed in

Table 6.1 Annotations identified after reanalysis of MCIs by mass spectrometry

MCI	Reanalysed annotation	OGS annotation
Selection from developmental database		
48877	Histone H2A	No ID
49022	Keratin 5	No ID
49048	Actin	No ID
50090	Membrane associated progesterone receptor component 1	No ID
Selection from expanded Lactation/Involution database		
48199	Transcription factor BTF3 (RNA polymerase B transcription factor 3)	No ID
45878	Glycerol-3-phosphate dehydrogenase	**Cluster 1:** Glycerol phosphate dehydrogenase 1, mitochondrial **Cluster 2:** Signaling lymphocytic activation molecule **Cluster 3:** Keratin 6b
46790	2,4-dienoyl CoA reductase	No ID
46820	**Cluster 1:** 3-hydroxyisobutyrate dehydrogenase, **Cluster 2:** Proteasome subunit, alpha type 1	**Cluster 1:** Similar to 3-hydroxyisobutyrate dehydrogenase **Cluster 2:** Proteasome subunit alpha type 1 **Cluster 3:** Caspase-like apoptosis regulatory
47692	Oxoglutarate dehydrogenase	Hypothetical 116.4 kDa
46380	Isocitrate dehydrogenase	**Cluster 1:** Isocitrate dehydrogenase cytoplasmic **Cluster 2:** Glutamate-ammonia ligase/Glutamine synthetase **Cluster 3:** Acyl-CoA dehydrogenase, long-chain specific, mitochondrial **Cluster 4:** Multifunctional protein ADE2
46638	**Cluster 1:** Aldose reductase **Cluster 2:** Annexin A2 **Cluster 3:** Polymerase delta interacting protein 38	Aldose reductase
48527	Glutathione transferase	Crystal structure of the Arf-Gap domain and ankyrin repeats of Papbeta
47752	**Cluster 1:** Elongation factor 2 **Cluster 2:** Muscle glycogen phosphorylase	**Cluster 1:** Elongation factor 2 **Cluster 2:** AK025781 cDNA
47953	**Cluster 1:** Glycerol kinase **Cluster 2:** Serum albumin precursor	**Cluster 1:** Serum albumin **Cluster 2:** Tubulin alpha chain **Cluster 3:** Glycerol kinase

The annotated information of each MCI was identified using Mascot (Matrix Science) which interrogated the public sequence databases. The table shows the annotations determined by the proteomics unit and Oxford GlycoSciences. MCIs are ordered according to the databases that they were determined from in Chapter 4 (see Tables 4.1 and 4.2).

Table 6.2 Functional classification of reanalysed MCIs

MCI	Protein Annotation	Database
Cytoskeletal proteins		
49022	Keratin 5	Developmental
49028	Actin	Developmental
Metabolic enzymes and mitochondrial proteins (ME/MP)		
47953	**Cluster:** Glycerol kinase	Lactation/Involution
45878	Glycerol-3-phosphate dehydrogenase	Lactation/Involution
46790	2,4-dienoyl CoA reductase	Lactation/Involution
47692	Oxoglutarate dehydrogenase	Lactation/Involution
46380	Isocitrate dehydrogenase cytoplasmic	Lactation/Involution
48527	Glutathione transferase	Lactation/Involution
47752	**Cluster:** Muscle glycogen phosphorylase	Lactation/Involution
46820	**Cluster:** 3-hydroxyisobutyrate dehydrogenase	Lactation/Involution
46638	**Cluster:** Aldose reductase	Lactation/Involution
Other		
50090	Membrane associated progesterone receptor component 1	Developmental
48877	Histone H2A	Developmental
46638	**Cluster:** Polymerase delta interacting protein 38	Lactation/Involution
Protein turnover		
46820	**Cluster:** Proteasome subunit alpha type 1	Lactation/Involution
RNA processes		
48199	Transcription factor BTF3 (RNA polymerase B transcription factor 3)	Lactation/Involution
Serum proteins		
47953	**Cluster:** Serum albumin Serum albumin precursor	Lactation/Involution
Signalling proteins		
46638	**Cluster:** Annexin A2	Lactation/Involution
47752	**Cluster:** Elongation factor 2	Lactation/Involution

The annotated information of each MCI was identified using Mascot (Matrix Science) which interrogated the public sequence databases. All annotations have been classed according to their general function. The classes are cytoskeletal protein (cytoskeletal), metabolic enzyme/mitochondrial protein (ME/MP), other, protein turnover, RNA processes, serum protein, and signalling protein (signalling).

October 2002, the data was not released until after the analysis had been completed at SHWFGF. Two of the 10 MCIs had no previous annotations identified by Oxford GlycoSciences. Eight of these MCIs had identified annotations both at Oxford GlycoSciences and SHWFGF, the majority of which were complimentary (Table 6.1).

6.3.3 Increased expression during pregnancy

Expanding the proteomic expression profile of MCI 48199 revealed an increase in expression from Av10 to P17.5 and a subsequent decrease for the remaining stages of development (Figure 6.1A). This MCI was selected as a priority, firstly because of its interesting expression profile which could indicate it having a regulatory role during pregnancy and secondly, because the triplicate values for each stage of development were very close to each other and showed a distinct and similar profile across development. The identification of this MCI was transcription factor BTF3 and is the only annotation classed under RNA processes (Table 6.2). No previous identification was made at Oxford GlycoSciences.

The other annotation which had not been identified at Oxford GlycoSciences was 2,4-dienoyl CoA reductase (MCI 46790). Figure 6.1B shows its expanded proteomic expression profile. This MCI had been selected for reanalysis, as its expression had peaked early in pregnancy, and only increased again mid involution. This was also the reason for selecting MCI 45878 and 46820 (Figures 6.1C and D). Both mass spectrometry analyses determined MCI 45878 as glycerol phosphate dehydrogenase however, Oxford GlycoSciences had identified it as a cluster (Table 6.1). MCI 46820 was identified as a cluster by Oxford GlycoSciences and the SHWFGF although the number of annotations in the cluster differed. The SHWFGF identified this MCI as 3-hydroxyisobutyrate dehydrogenase and a proteasome subunit.

6.3.4 Decreased expression from pregnancy to lactation

The expanded profiles of the following five MCIs all appeared to decrease in expression from Av10 through to mid involution, and following mid involution an increase was observed. This was true for all except MCI 47692 which only showed an increase in expression during the latter stages of involution (Figure 6.1E). MCI 47692 was identified as a single annotation oxoglutarate dehydrogenase, although this differed to the result obtained by Oxford GlycoSciences (Table 6.1). MCI 46242 revealed no annotation by the SHWFGF but MCI 46380 revealed a single identification, namely isocitrate dehydrogenase (Figure 6.1F). Oxford GlycoSciences also determined the same annotation

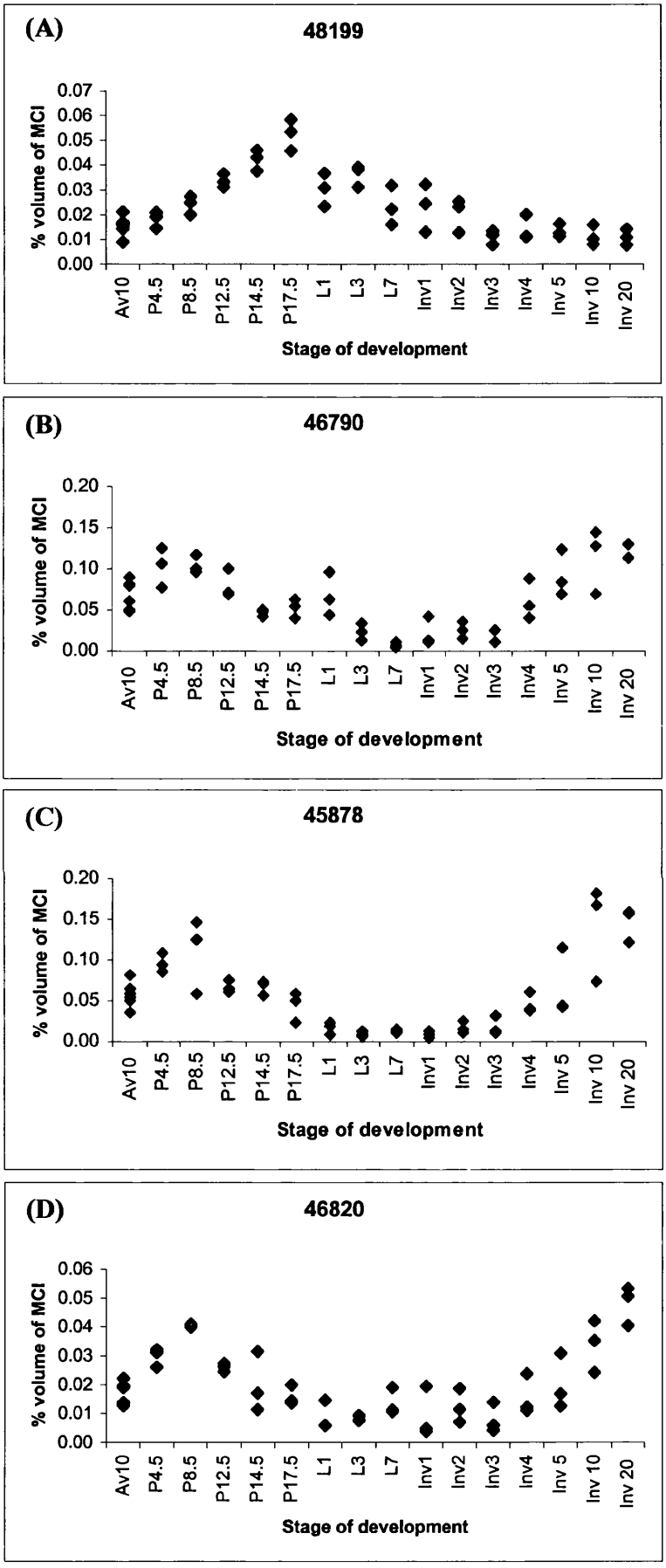


Figure 6.1 (A-D) Expanded proteomic expression profiles. These MCIs were selected for reanalysis by mass spectrometry based on their expression profiles across development. Expression levels are represented as the percentage volume of the MCI.

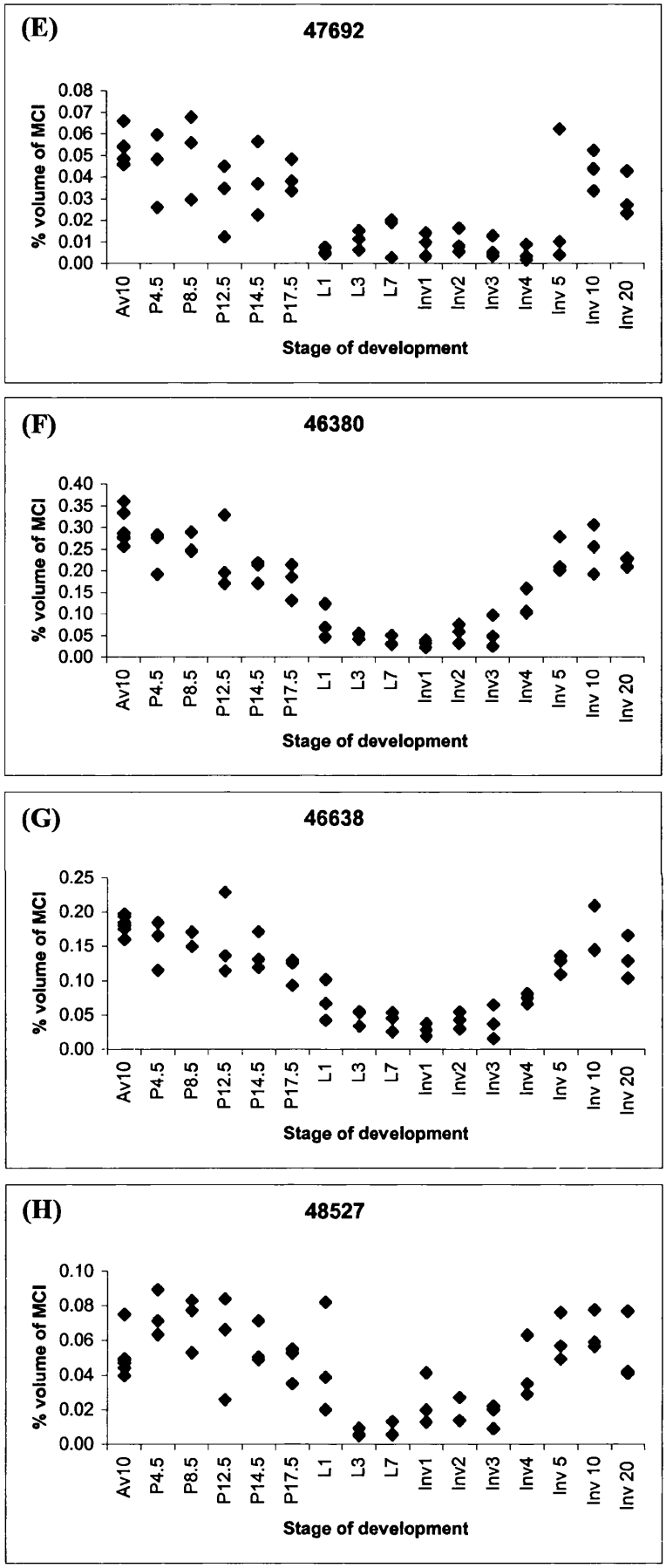


Figure 6.1 continued (E-H) Expanded proteomic expression profiles. These MCIs were selected for reanalysis by mass spectrometry based on their expression profiles across development. Expression levels are represented as the percentage volume of the MCI.

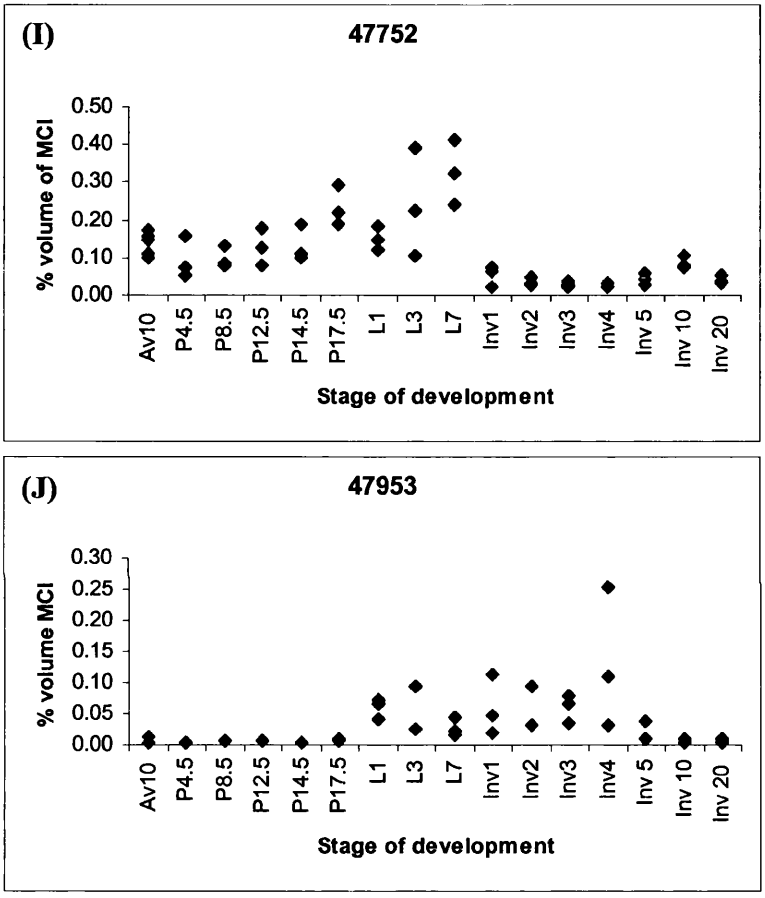


Figure 6.1 continued (I-J) Expanded proteomic expression profiles. These MCIs were selected for reanalysis by mass spectrometry based on their expression profiles across development. Expression levels are represented as the percentage volume of the MCI.

although it had more than one annotation. The reverse was true for MCI 46638 (Figure 6.1G) as three annotations were determined by the SHWFGF (aldose reductase, annexin A2 and a polymerase interacting protein) and a single cluster was determined by Oxford GlycoSciences (aldose reductase). MCI 48527 (Figure 6.1H) was identified as glutathione transferase by SHWFGF which differed to the annotation determined by Oxford GlycoSciences.

6.3.5 Altered expression during mammary development

The final two MCIs were identified as cluster annotations by the SHWFGF. The two annotations for MCI 47752 were elongation factor 2 and glycogen phosphorylase. Elongation factor 2 had also been found by Oxford GlycoSciences. This MCI had been selected because the expanded profile showed a general increase in expression from Av10 to L7 and down-regulation thereafter (Figure 6.1I). MCI 47953 was identified as glycerol kinase and serum albumin, and were classed under ME/MP and serum proteins respectively. Both annotations were also found by Oxford GlycoSciences. This MCI had originally been selected due to its presence during lactation and early to mid involution (Figure 6.1J).

Two of the MCIs selected for reanalysis were identified by Oxford GlycoSciences but not by the SHWFGF, these were MCI 46242 and 50787.

6.3.6 RT-PCR analysis of BTF3 in the mammary gland

BTF3a was of primary interest as this was the isoform that was transcriptionally active. However, the spectral data obtained for BTF3 did not distinguish between BTF3a or b because the two peptide sequences matching to this protein (VQASLAANTFTITGHAETK and QLTEMLPSILNQLGADSLTSLRR) were common to both isoforms. An antibody for BTF3 was obtained from Santa Cruz for western blotting but unfortunately it did not detect this protein in the mammary gland samples or in positive control tissue, and therefore RNA experiments were used for confirmation of the proteomics results. To ensure that BTF3 was expressed in the mammary gland, non-quantitative RT-PCR analysis was carried out using one stage of mammary gland development (Figure 6.2). Primers were designed to amplify BTF3a. Primers could not be designed to detect only BTF3b because all of BTF3b's sequence was present in BTF3a. However, primers were designed in a region common to both isoforms (BTF3ab). HeLa cells and kidney cDNA were used as positive controls, and automated sequencing at the MBSU confirmed that the amplified products from the controls and mammary gland samples were BTF3a and BTF3ab. The experiment

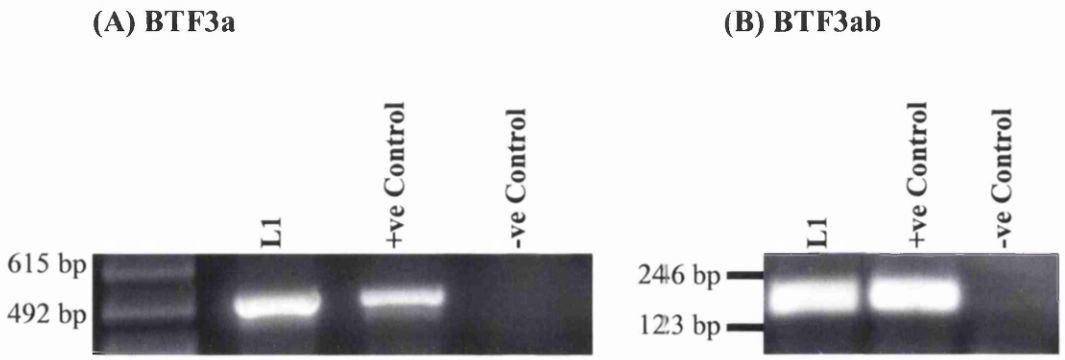


Figure 6.2 RT-PCR analysis of BTF3. Analysis was carried out with gene-specific intron-spanning primers for BTF3a and BTF3ab. cDNA was used as template. The expected product size for BTF3a was 491 bp and 233 bp for BTF3ab.

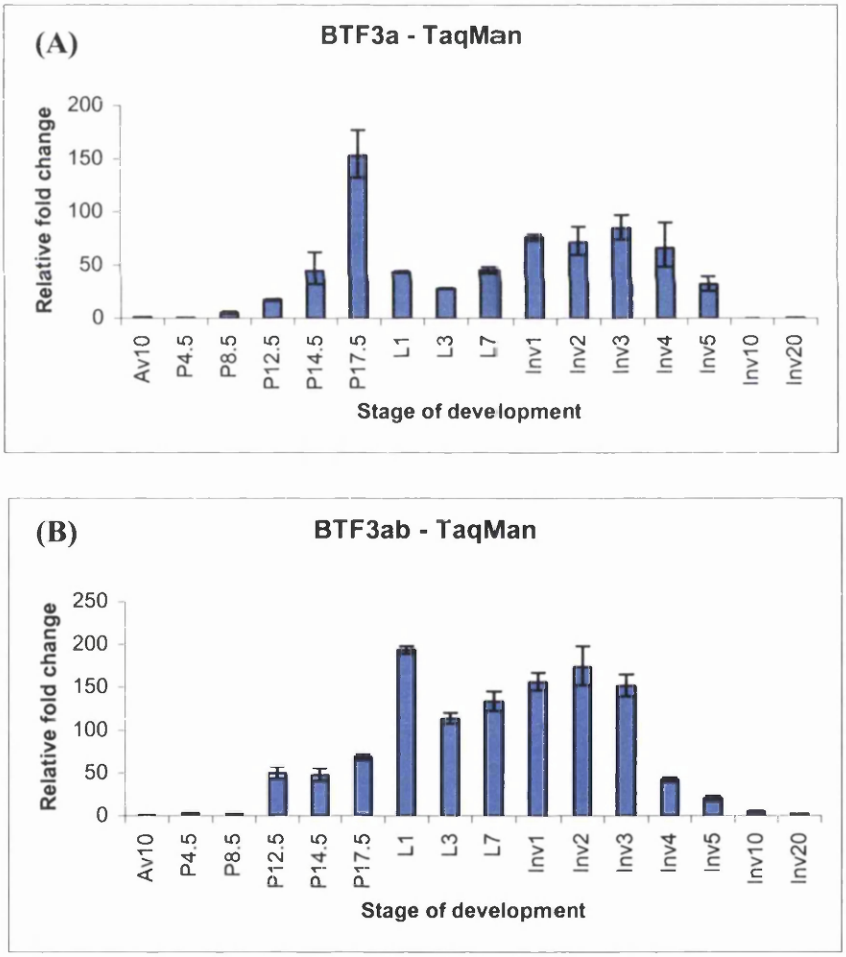


Figure 6.3 Expression profiles of BTF3 during mouse mammary gland development. The expression profile of BTF3a (A) and BTF3ab (B) produced by TaqMan analysis. TaqMan results are presented as fold change relative to a stage within the set of samples tested.

indicates that BTF3b is expressed, as the intensity of the product for BTF3ab is greater than BTF3a alone. The sequence of BTF3a has not previously been isolated in the mouse, and primers were designed from a predicted sequence. See Table 6.3 for primer sequences.

Table 6.3 Primer sequences used for BTF3

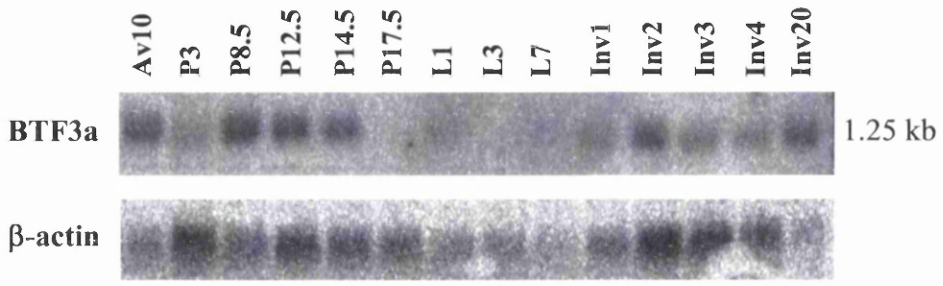
Name	Sequence 5' → 3'	Application
BTF3a forward	CCACCCAGGCGGACTCTC	RT-PCR analysis of BTF3a
BTF3a reverse	ATTGTTTGGGCAGAGCTTC	
BTF3ab forward	GGGTGAACAACATCTCTGG	RT-PCR analysis of BTF3ab
BTF3ab reverse	ATTGTTTGGGCAGAGCTTC	
BTF3ab forward	CATTGGTGGGAAAGGAACTG	Northern blot analysis BTF3ab
BTF3ab reverse	ATTGTTTGGGCAGAGCTTC	

6.3.7 TaqMan analysis of BTF3 in mammary gland development

To quantify the expression of BTF3a and BTF3ab across mammary gland development, TaqMan analysis was performed (Figure 6.3). Appendix 4 shows the primer and probe sequences used for TaqMan analysis of BTF3. As with the previous results obtained by TaqMan the data is presented as relative fold change to a reference sample. The results of the BTF3a (Figure 6.3A) experiment were very similar to the expanded proteomic profile (Figure 6.1A). The expression profile obtained from the primers which identified both BTF3a and BTF3b (i.e. BTF3ab) (Figure 6.3B), again indicates that BTF3b is expressed as the profile is different to BTF3a. The comparison of the BTF3a and BTF3ab TaqMan profiles suggest that the proteomic annotation was BTF3a and not BTF3b.

6.3.8 Northern blot analysis of BTF3 in mammary gland development

In order to confirm these results and to determine the transcript sizes of BTF3a and b in the mouse, northern blot analysis was performed. The probe was designed within the homologous region of the two isoforms and would distinguish between the two because of their differing size. A previous report had published that BTF3a and b differed by approximately 750 bp. The precise size of the first exon was not known and therefore only an approximation can be given (Kanno *et al.*, 1992). The intron-spanning probe used for northern blotting was 329 bp in length, see table 6.3 for primers used to design the probe. Automated sequencing at the MBSU confirmed that the probe was BTF3ab. Northern blotting revealed only one signal at 1.25 kb which is consistent with the size previously published for BTF3a by Kanno *et al.* in 1992 (Figure 6.4). No signal was identified for BTF3b. Overall the northern blot expression profile shows that BTF3a has a higher



6.4 Northern blot analysis of BTF3a across mammary gland development. Total RNA was probed with cDNA of BTF3ab. This probe did not distinguish between the two isoforms of this gene although only a single band at 1.25 kb was produced which appears to be BTF3a. The blot was subsequently probed with cDNA for β-actin for load verification.

expression level in pregnancy compared with the other stages analysed. Slight discrepancies can be seen between the northern blot and TaqMan analysis profiles for BTF3a. This may be as a result of this gene being expressed in low levels, which can lead to inaccuracies. Alternatively, equal transfer was not achieved across the northern blot membrane.

In summary the results obtained for BTF3 indicate that the proteomic expression profile for BTF3 is in fact BTF3a. The data generated from the different experiments performed for BTF3a show that its expression peaks during pregnancy.

6.4 Discussion

6.4.1 Additional MCI annotations

The second round of mass spectrometry analysis revealed 14 MCIs with protein identifications. Six of these MCIs had no previously annotated data. Some of the annotations identified had not been detected before in this project which included histone H2A, keratin 5 and membrane associated progesterone receptor. Histone H2A was found in several stages of pregnancy. A change in its expression may be caused by an increase in the number of proliferating cells or reorganisation of the chromatin structure associated with altered transcriptional activities (Jeong *et al.,* 2004). Actin and keratin 5 were found in multiple stages of pregnancy and although these MCIs were selected based on presence rather than expression level, the expectation would be that the relative amounts of these proteins would increase with the increasing number of epithelial cells during pregnancy. Membrane associated progesterone receptor was detected in multiple stages of pregnancy which would appear appropriate with respect to its function. Progesterone receptor plays a role in mediating the physiological effects of progesterone, a hormone which plays an essential role in establishing and maintaining pregnancy (Atwood *et al.,* 2000; Ichinose and Nandi, 1966). Finally annexin A2 had not previously been linked to MCI 46638. Although it was a clustering protein, its expanded expression profile (Figure 6.1G) was the same as that identified by TaqMan and microarray analysis in Chapter 5 (Figure 5.12). This supports the theory that different isoforms of annexin A2 are present in the mammary gland.

6.4.2 Identification of BTF3 in mammary gland development

BTF3 was the most interesting annotation out of all those detected by the second round of mass spectrometry and was selected for follow-up for several reasons. BTF3 is down regulated during apoptosis (Brockstedt *et al.,* 1999), and this agreed with the down

regulation in involution on the mammary gland development proteomic expression profile. BTF3 may also have an effect on transcription of the oestrogen receptor which plays an important clinical role in breast cancer treatment (el-Tanani and Green, 1998). BTF3 had not previously been identified in this project, but the triplicate values obtained for each stage were very close and displayed a distinct regulatory role in pregnancy.

Transcription factors play important roles in mammary gland development such as the signal transduction and activation of transcription 3 (Stat3) transcription factor which is essential during involution (Chapman *et al.,* 1999). Transcription factors have also been shown to be important during alveolar expansion and differentiation of the mammary gland. A greater understanding of how transcription factors regulate epithelial proliferation and differentiation, requires the identification of downstream target genes.

In conclusion, the data that has been generated in this Chapter suggests that BTF3a is regulated during mouse mammary gland development, a result not previously reported. Specifically, the experiments performed for BTF3a show that its expression peaks during pregnancy. It would appear that like other transcription factors (Chapman *et al.,* 1999), BTF3a may be important during alveolar expansion and differentiation. Further analyses of BTF3a are required to determine its distinct role in mammary gland development.

Chapter 7
Further discussion

7.1 Further discussion and future work

Proteomics was used as a method for determining differentially expressed proteins during mammary gland development. Chapter 4 described the generation of the two databases created from the mouse mammary gland proteomes. Two databases had to be formed due to the large presence of milk proteins in the lactation and early involution samples. The use of columns to remove these milk proteins from the samples had been considered during their preparation for 2-D gel electrophoresis, however it was decided not to use them. This decision was made based primarily on the preservation of the protein sample and the desire to create a proteomic database which captured the *in vivo* situation of the mammary gland and not a database from a 'manipulated' sample preparation.

7.1.1 Open access proteomics

Although 169 MCIs had been selected for mass spectrometry analysis in Chapter 4, only 83 revealed annotated information after the second round of sequence database searches. The annotated data acquired by mass spectrometry is dependent on the sequence databases searched. Due to the sequence similarities shared between different species, the mouse, rat and human sequence databases were interrogated in this project. This increased the number of annotated hits obtained, since at the time of this study, the human genome sequence was more complete than the mouse genome. The more recent release of the mouse genome will be of great benefit to future mass spectrometry based proteomic projects. The effect of the increasing amounts of genome sequence data becoming available was demonstrated in this project by the additional number of MCIs found with annotated information after the second round of sequence database searches. Owing to the costs incurred and the time taken to identify annotations from 2-D gels there is a need to corroborate proteomic data in order that duplication of results can be avoided. Propagation of data from different databases was already used at Oxford GlycoSciences but a more open access facility to other proteomic data would result in faster and cheaper advances in proteomics.

7.1.2 e-Science

Although not used in this project, one step further than propagation is the use of e-Science, which hopes to capture global collaboration in key areas of science. Already many areas of science are reliant on collaboration and multidisciplinary work. e-Science is a vision which if successful will enable different communities to collaborate and create a robust, flexible, secure, co-ordinated resource-sharing organisation to approach complex problems with a wide variety of distributed resources. Global collaborations enabled by the internet will require access to a wide range of data collections, computing resources, control facilities

and tools for visualisation, data mining and analysis. Its application will be a significant development to collaborative science and to proteomics (Hey and Trefethen, 2003).

7.1.3 Resources available for confirmation of proteomic results

Chapter 5 was based on experiments performed to validate the proteomics results obtained in Chapter 4. Unfortunately for many of the candidate proteins selected the proteomics results were not confirmed. This was attributed to PTMs, as discussed in Chapter 5, as this project did not search for these alterations. The proteomic profiles of well known proteins in mammary gland development such as β-casein did not show typical expression profiles. It is known that proteomics is capable of detecting different isoforms of the protein detected and that the traditional methods used to determine their expression profiles such as IHC cannot differentiate between them, unless specific antibodies are created to do so.

In hindsight, basing MCI selection on pattern changes across development (as demonstrated in Chapter 6) rather than an absolute change of presence/absence as was chosen in Chapter 4 proved to be a better method of analysis. The follow-up results obtained for BTF3 in Chapter 6 were able to confirm the proteomics results, whereas the majority of follow-up work based on presence/absence was not. Using pattern changes of an MCI across development provided a better understanding of the MCI's involvement in mammary gland development.

2-D gel electrophoresis is a technique which is able to separate different isoforms of proteins which are then identified by the sensitive technique of mass spectrometry. Equally important is confirmation of these results; however, the lack of suitable antibodies which are specific to these isoforms hinders this sophisticated technique. This has been the case in this project with annexin A2, as the antibodies available did not distinguish between phosphorylated and non-phosphorylated annexin A2. There is a high demand for antibody collections which are specific to different isoforms and readily available to improve the speed of confirmation of results. An alternative approach would be to 'tag' the proteins of interest with a sequence which is easily recognised by an antibody specific to the tag.

Due to the lack of antibody resources available mRNA expression profiles were turned to for distinguishing between clustered proteins. Microarray and TaqMan data were used, although there are known discrepancies between mRNA and protein expression. Generally where mRNA and western blotting expression profiles were available for the same gene, the data concurred.

7.1.4 Microarrays and Proteomics

Over the past few years the improvements made to microarrays has developed them into a valuable tool for understanding gene expression. Their use in identifying pathways and cellular processes has been demonstrated numerous times in the mammary gland (Stein *et al.,* 2004; Clarkson *et al.,* 2004; Master *et al.,* 2002). Whilst proteomics could potentially analyse all proteins revealed from a proteome image, this would require huge financial resources and time. In reality as with this project a limited number of protein annotations are analysed which have been chosen due to the specific conditions set. Also the results of each protein identification are not produced simultaneously as with microarray profiling. A clear advantage of proteomics over microarrays is that it studies proteins, the level where most regulatory processes take place and where most drug targets are found.

7.1.5 Protein arrays

Protein arrays are another alternative approach to analysing the mouse mammary gland proteome. This system allows high-throughput analysis of protein-protein, protein-substrate and protein-small molecule interactions to be detected (Zhu *et al.,* 2000, MacBeath and Schreiber 2000). However, a major disadvantage of this technology is the long development times and labour intensive nature of the process. Conversely array technologies are advantageous due to smaller sample volumes, efficient analysis and high-throughput.

7.1.6 Complex protein samples

The samples analysed by 2-D gel electrophoresis were of whole protein extracts. The development of the extraction technique used was described in Chapter 3 and has the advantage of being a quick and reproducible sample preparation technique. In theory 2-D gel electrophoresis of whole tissue protein extracts provides a complete overview of all the proteins in the sample, based on protein feature patterns. Attention has now been turned to reducing the complexity of tissue samples prior to analysis in order to detect low abundant proteins which would otherwise be masked by those that are highly abundant. Most regulatory proteins such as phosphatases and kinases exist in low copy numbers but are very specific to subcellular localisation. Due to the low resolution of protein separation technologies currently used in proteomics, additional fragmentation steps are needed prior to 2-D gel electrophoresis and mass spectrometry. Prefractionation techniques such as subcellular organelle fractionation (Pasquali *et al.,* 1999), affinity purification (Gavin *et al.,* 2002) and zoom gels (Gorg *et al.,* 2000) are processes which enrich the proteins in the

separated fractions. Thus they improve the ability to investigate low abundant proteins such as those that are post-translationally modified. Focusing on certain compartments of the cell and restricting the focusing range of the mammary gland samples may have revealed a more comprehensive view of the regulatory processes of the mammary gland compared to the method that was used in this project.

Although proteomic technologies have made rapid progress in recent years, membrane proteins are still under-represented in datasets. Very few membrane proteins are resolved on 2-D gels due to their hydrophobic properties. Their key roles in signal transduction, cell adhesion and ion transport make it important to resolve this problem (Santoni *et al.*, 2000). Also, membrane proteins are hugely important to the drug industry, as they represent a large percentage of pharmaceutical drug targets. Most refinements to improve their isolation have been directed at sample preparation such as prefractionation and enrichment techniques and solubilisation with organic solvents (Ferro *et al.*, 2000; 2002). Advances in mass spectrometric methods to analyse proteins without the use of 2-D gels are being developed, such as ICAT labelling (Gygi *et al.*, 1999a) and multidimensional protein identification technology (MudPIT) (Link *et al.*, 1999) and have been applied to analysing membrane preparations.

Again these two techniques can be used to reduce the complexity of the samples analysed and are an alternative proteomic approach to interrogating the mammary gland without the use of 2-D gels. MudPIT separates complex peptide mixtures by multidimensional capillary chromatography and analyses them by the connecting ion trap mass spectrometer. However the reproducibility of this technique is unclear due to the limited number of studies performed by this technology. ICAT labelling has been described in the main introduction, but briefly it differentiates between two populations of proteins using probes which differ in their isotopic labelling. The limitation of this technique is the number of samples can be analysed. Therefore this technique would not be useful for differential analysis of mammary gland development, although ICAT labelling would be useful for analysing the switch from lactation to involution. In retrospect, limiting analysis to these two stages of development may have been more profitable in producing a comprehensive picture of the regulatory processes involved in this switch in development. The restricted number of MCIs for mass spectrometry analysis would have been concentrated on two stages, rather than using the same number across a wider range of samples, as was performed in this project.

7.1.7 Functional assays

Due to the extensive confirmatory results carried out on the candidate proteins no functional assays were performed. The future direction of the project could have been in gene manipulation studies. If knockout mice had already been generated for these candidate proteins, then investigations could have been made into whether their deletion caused any defects in mammary gland morphogenesis. If none were available, then functional investigations would have focused on manipulating the gene of interest by reconstituting mammary gland development in tissue culture. When culturing EpH4 mouse mammary epithelial cells on matrigel complex (basement membrane) three-dimensional structures can be generated which resemble those formed during mammary gland development. This system represents the *in vivo* situation of the mammary gland more closely than 2-D tissue culture. Three-dimensional cultures are able to reproduce the morphogenic processes in the developing mammary gland, as branched tubules can form in the presence of hepatocyte growth factor/scatter factor and alveolar structures in the presence of neuregulin (Niemann *et al.,* 1998). The matrigel system enables the investigation of signalling cascades that are activated by morphogenic factors such as ECM components. The interactions which are exchanged between stroma and epithelium play key roles in the different stages of development of the mammary gland. In summary, this system is a sophisticated method of gaining functional insights into the roles of proteins in mammary gland development before efforts are spent on manipulations *in vivo*.

7.1.8 Protein Interactions

Methods used to deduce protein function aim at identifying the proteins with which they interact, as they are generally involved in the same cellular processes. The yeast two-hybrid system and fluorescence resonance energy transfer (FRET) analysis are methods which are used to identify protein-protein interactions. These techniques could have produced important insights into role of the candidate proteins selected in mammary gland development.

In summary a number of different approaches and improvements have been described for analysing the mouse mammary gland proteome. This study has identified many protein annotations which are regulated in mammary gland development and has highlighted the necessity of identifying PTMs. For confirmation purposes, reagents which are specific to the isoforms detected need to be developed in order to create confidence in the results obtained. Subsequently investigations into the precise role of these proteins in mammary gland development can begin to be explored.

Appendices

Appendix 1: Buffers used for 2-D gel electrophoresis

2-D lysis buffer	4% (w/v) CHAPS
	8 M urea
	2 M thiourea
	65 mM DTT
	0.8% (w/v) Resolytes 3.5-10 (Bio-Rad)
	trace bromophenol blue
Equilibration buffer	per 100 ml of dH_2O
	2% (w/v) SDS
	36 g Urea (6 M)
	30% (v/v) glycerol
	2% (w/v) DTT
	5 ml Tris (50 mM) pH 6.8
	trace bromophenol blue
10x Tris-glycine running buffer	per 1000 ml of dH_2O
	144 g glycine (1.92 M)
	29 g Tris Base (250 mM)
	1% (w/v) SDS
	1x buffer should be pH 8.3
Tris-glycine transfer buffer	per 1000 ml of dH_2O
	1.45 g Tris base (12 mM)
	7.2 g glycine (96 mM)
	200 ml methanol

Appendix 2: Buffers used for SDS-PAGE and western blot analysis

4x NuPage sample buffer	per 10 ml of dH_2O
	0.42 M Tris-HCl, pH 6.8
	8% (w/v) SDS
	40% (w/v) glycerol
	2.04 mM EDTA
	0.88 mM SERVA® Blue G250
	1x buffer should be pH 8.5
20x MES-SDS Running Buffer	per 500 ml of dH_2O
	97.6 g MES (1 M) pH 7.2
	60.6 g Tris base (1 M)
	1% (w/v) SDS
	3 g EDTA (20 mM)
	1x buffer should be pH 7.3
20 x Transfer Buffer	per 125 ml of dH_2O
	10.2 g bicine (0.5 M)
	13.1 g bis-tris (0.5 M)
	0.75 g EDTA (3 mM)
	0.025 g chlorobutanol (1 mM)
	1x buffer should be pH 7.2
(Invitrogen)	
Ponceau S staining solution	per 500 ml of dH_2O
	1.5 g trichloroacetic acid
	0.5 g Ponceau S stain
1x PBS	137 mM NaCl
	2.7 mM KCl
	10 mM Na_3PO_4
	2 mM KH_2PO_4, pH 7.3

From (Sambrook and Russell, 2001)

Appendix 3: Solutions for formaldehyde gel and buffers used for Northern blot analysis

1% formaldehyde-agarose gel	147 ml DEPC dH_2O
	2 g agarose
	20 ml 10x MOPS
	33 ml formaldehyde
RNA loading buffer	234 µl DEPC dH_2O
	500 µ1 50% (v/v) formamide
	166 µl formaldehyde (2.2M)
	100 µl 1x MOPS
	(prepare fresh)
Sample loading dye	1 ml glycerol
	1 ml 10x MOPS
	trace bromophenol blue
20x MOPS buffer	per 500 ml DEPC dH_2O
	41.9 g MOPS
	6.8 g NaAc
	2.6 g EDTA
	Adjust to pH 7.0 with NaOH
20X SSC	per 2 l DEPC dH_2O
	350.64 g NaCl (3M)
	176.46 g $C_6H_8O_7$ trisodium citrate (0.3M)

Appendix 4: TaqMan primer and probe sequences

Name	F/R	Sequence 5' → 3'	Probe 5' → 3'
Annexin II	F	GCCTGGAGGGTGATCATTCTAC	CCAAGTGCCTACGGGTCA GTCAAACC
	R	CTCTCAGCATCGAAGTTGGTGTA	
Carboxylesterase	F	TCCCGGCTGTGCTCTTGT	TGGACACTCCAGCATCTC TGAGGCTACG
	R	GGCGATACCGAAACTCATACATG	
Contrapsin	F	TCACAGAAACCAAGAAACTGAGTGT	CCTTGTGGACCACCTG
	R	TGTGCCTGTCTCAGCCACAT	
Fumarate hydratase	F	TCTTGGACAGGAATTCAGTGGTT	TCACCATCGCATACTGGA CTTGCTGAA
	R	TTGGCATGGCGGCTTT	
Lumican	F	CTTGAAAAGTTTGATGTGAAGACCTT	TCCTGGGACCACTGTCTT ACTCCAAGATCA
	R	TGCCATCCAAGCGCAGAT	
MCM3	F	ACCCTTACGACTTCAGTGAAGCA	ACGCAGATGCCTCAAGTG CACACC
	R	CTCTTGGGAATCGTCAGTCTTTG	
BTF3a	F	AGCCGAGGACGGGAGTCT	TTCTCCTGGTTCATGATT GTTTCTTTCATCT
	R	CACTTGTGCCTGCAGTTTGG	
BTF3	F	GGTGCAGACAGCCTGACTAGTTT	TTGGGCAGAGCTTCAGCC AGTCTCCT
	R	GTGCTTTTCCATCCACAGATTG	
GAPDH	F	ATGTGTCCGTCGTGGATCTGA	TGCCGCCTGGAGAAACCT GCC
	R	TCACCACCTTCTTGATGTCATCA	

F/R refers to forward and reverse primers respectively.

References

Adamczyk, M., Gebler, J. C. and Wu, J. Selective analysis of phosphopeptides within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry. *Rapid Commun Mass Spectrom* **15**, 1481-8 (2001).

Adams, J. C. and Watt, F. M. Regulation of development and differentiation by the extracellular matrix. *Development* **117**, 1183-98. (1993).

Adler, H. J., Winnicki, R. S., Gong, T. W. and Lomax, M. I. A gene upregulated in the acoustically damaged chick basilar papilla encodes a novel WD40 repeat protein. *Genomics* **56**, 59-69 (1999).

Anderson, L. and Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-7 (1997).

Anderson, N. L., Taylor, J., Scandora, A. E., Coulter, B. P. and Anderson, N. G. The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin Chem* **27**, 1807-20 (1981).

Anthonsen, M. W., Ronnstrand, L., Wernstedt, C., Degerman, E. and Holm, C. Identification of novel phosphorylation sites in hormone-sensitive lipase that are phosphorylated in response to isoproterenol and govern activation properties in vitro. *J Biol Chem* **273**, 215-21 (1998).

Appel, R. D., Bairoch, A. and Hochstrasser, D. F. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem Sci* **19**, 258-60 (1994).

Atwood, C. S., Hovey, R. C., Glover, J. P., Chepko, G., Ginsburg, E., Robison, W. G. *et al.* Progesterone induces side-branching of the ductal epithelium in the mammary glands of peripubertal mice. *J Endocrinol* **167**, 39-52. (2000).

Axelsson, I. and Heinegard, D. Characterization of the keratan sulphate proteoglycans from bovine corneal stroma. *Biochem J* **169**, 517-30 (1978).

Baillat, G., Moqrich, A., Castets, F., Baude, A., Bailly, Y., Benmerah, A. *et al.* Molecular cloning and characterization of phocein, a protein found from the Golgi complex to dendritic spines. *Mol Biol Cell* **12**, 663-73 (2001).

Bairoch, A. and Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Res* **2**, 31-6 (1997).

Barcellos-Hoff, M. H., Aggeler, J., Ram, T. G. and Bissell, M. J. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. *Development* **105**, 223-35. (1989).

Bartsch, I., Schoneberg, C. and Grummt, I. Evolutionary changes of sequences and factors that direct transcription termination of human and mouse ribsomal genes. *Mol Cell Biol* **7**, 2521-9 (1987).

Bell, S. P. and Dutta, A. DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**, 333-74 (2002).

Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. *et al.* Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**, 317-39 (1982).

Blanchette-Mackie, E. J., Dwyer, N. K., Barber, T., Coxey, R. A., Takeda, T., Rondinone, C. M. *et al.* Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *J Lipid Res* **36**, 1211-26 (1995).

Boron, W. F. and Boulpaep, E. L. Intracellular pH regulation in the renal proximal tubule of the salamander. Basolateral HCO3- transport. *J Gen Physiol* **81**, 53-94 (1983).

Borrebaeck, C. A., Ekstrom, S., Hager, A. C., Nilsson, J., Laurell, T. and Marko-Varga, G. Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry. *Biotechniques* **30**, 1126-30, 1132 (2001).

Boudreau, N., Sympson, C. J., Werb, Z. and Bissell, M. J. Suppression of ICE and apoptosis in mammary epithelial cells by extracellular matrix. *Science* **267**, 891-3. (1995).

Bourne, H. R., Sanders, D. A. and McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125-32 (1990).

Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-54 (1976).

Brasaemle, D. L., Levin, D. M., Adler-Wailes, D. C. and Londos, C. The lipolytic stimulation of 3T3-L1 adipocytes promotes the translocation of hormone-sensitive lipase to the surfaces of lipid storage droplets. *Biochim Biophys Acta* **1483**, 251-62 (2000).

Brennwald, P. and Novick, P. Interactions of three domains distinguishing the Ras-related GTP-binding proteins Ypt1 and Sec4. *Nature* **362**, 560-3 (1993).

Bridgman, P. C. Growth cones contain myosin II bipolar filament arrays. *Cell Motil Cytoskeleton* **52**, 91-6 (2002).

Bridgman, P. C., Dave, S., Asnes, C. F., Tullio, A. N. and Adelstein, R. S. Myosin IIB is required for growth cone motility. *J Neurosci* **21**, 6159-69 (2001).

Brockstedt, E., Otto, A., Rickers, A., Bommert, K. and Wittmann-Liebold, B. Preparative high-resolution two-dimensional electrophoresis enables the identification of RNA polymerase B transcription factor 3 as an apoptosis-associated protein in the human BL60-2 Burkitt lymphoma cell line. *J Protein Chem* **18**, 225-31 (1999).

Broschat, K. O., Weber, A. and Burgess, D. R. Tropomyosin stabilizes the pointed end of actin filaments by slowing depolymerization. *Biochemistry* **28**, 8501-6 (1989).

Brown, M. E. and Bridgman, P. C. Retrograde flow rate is increased in growth cones from myosin IIB knockout mice. *J Cell Sci* **116**, 1087-94 (2003).

Brush, M. Dye hard; protein gel staining products. *Scientist* **12**, 16-22 (1998).

Cao, M., Sahmi, M., Lussier, J. G. and Price, C. A. Plasminogen Activator and Serine Protease Inhibitor-E2 (Protease Nexin-1) Expression by Bovine Granulosa Cells In Vitro. *Biol Reprod* (2004).

Carbone, R., Fre, S., Iannolo, G., Belleudi, F., Mancini, P., Pelicci, P. G. *et al.* eps15 and eps15R are essential components of the endocytic pathway. *Cancer Res* **57**, 5498-504 (1997).

Carrell, R. W. and Lomas, D. A. Conformational disease. *Lancet* **350**, 134-8 (1997).

Celis, J. E., Gesser, B., Rasmussen, H. H., Madsen, P., Leffers, H., Dejgaard, K. *et al.* Comprehensive two-dimensional gel protein databases offer a global approach to the analysis of human cells: the transformed amnion cells (AMA) master database and its link to genome DNA sequence data. *Electrophoresis* **11**, 989-1071 (1990).

Ceriani, R. L., Peterson, J. A., Lee, J. Y., Moncada, R. and Blank, E. W. Characterization of cell surface antigens of human mammary epithelial cells with monoclonal antibodies prepared against human milk fat globule. *Somatic Cell Genet* **9**, 415-27 (1983).

Chan, A. Y., Bailly, M., Zebda, N., Segall, J. E. and Condeelis, J. S. Role of cofilin in epidermal growth factor-stimulated actin polymerization and lamellipod protrusion. *J Cell Biol* **148**, 531-42 (2000).

Chapman, R. S., Lourenco, P. C., Tonner, E., Flint, D. J., Selbert, S., Takeda, K. *et al.* Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. *Genes Dev* **13**, 2604-16 (1999).

Chavrier, P., Gorvel, J. P., Stelzer, E., Simons, K., Gruenberg, J. and Zerial, M. Hypervariable C-terminal domain of rab proteins acts as a targeting signal. *Nature* **353**, 769-72 (1991).

Chetcuti, A., Margan, S. H., Russell, P., Mann, S., Millar, D. S., Clark, S. J. *et al.* Loss of annexin II heavy and light chains in prostate cancer and its precursors. *Cancer Res* **61**, 6331-4 (2001).

Chevalier, S. and Blow, J. J. Cell cycle control of replication initiation in eukaryotes. *Curr Opin Cell Biol* **8**, 815-21 (1996).

Chong, J. P., Thommes, P. and Blow, J. J. The role of MCM/P1 proteins in the licensing of DNA replication. *Trends Biochem Sci* **21**, 102-6 (1996).

Clarkson, R. W., Wayland, M. T., Lee, J., Freeman, T. and Watson, C. J. Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. *Breast Cancer Res* **6**, R92-R109 (2004).

Cocks, B. G., Chang, C. C., Carballido, J. M., Yssel, H., de Vries, J. E. and Aversa, G. A novel receptor involved in T-cell activation. *Nature* **376**, 260-3 (1995).

Cole, H. A. The Mammary Gland of The Mouse, During the Oestrous Cycle, Pregnancy and Lactation. *Proc Roy Soc B* **114**, 136-160 (1933).

Collin, C., Moll, R., Kubicka, S., Ouhayoun, J. P. and Franke, W. W. Characterization of human cytokeratin 2, an epidermal cytoskeletal protein synthesized late during differentiation. *Exp Cell Res* **202**, 132-41 (1992).

Collins, C., Nehlin, J. O., Stubbs, J. D., Kowbel, D., Kuo, W. L. and Parry, G. Mapping of a newly discovered human gene homologous to the apoptosis associated-murine mammary protein, MFG-E8, to chromosome 15q25. *Genomics* **39**, 117-8 (1997).

Creutz, C. E. The annexins and exocytosis. *Science* **258**, 924-31 (1992).

Crowley, M. R., Head, K. L., Kwiatkowski, D. J., Asch, H. L. and Asch, B. B. The mouse mammary gland requires the actin-binding protein gelsolin for proper ductal morphogenesis. *Dev Biol* **225**, 407-23. (2000).

Crumpton, M. J. and Dedman, J. R. Protein terminology tangle. *Nature* **345**, 212 (1990).

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

Cunha, G. R., Young, P., Hom, Y. K., Cooke, P. S., Taylor, J. A. and Lubahn, D. B. 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. (1997).

Daniel, C. W., Berger, J. J., Strickland, P. and Garcia, R. Similar growth pattern of mouse mammary epithelium cultivated in collagen matrix in vivo and in vitro. *Dev Biol* **104**, 57-64 (1984).

Daniel, C. W. and Silberstein, G. B. in *The mammary gland: Development, Regulation and Function.* (eds. Neville, M. C. & Daniel, C. W.) 3-36 (Plenum Press, New York, 1987).

Das, S. and Potter, H. Expression of the Alzheimer amyloid-promoting factor antichymotrypsin is induced in human astrocytes by IL-1. *Neuron* **14**, 447-56 (1995).

Delehedde, M., Deudon, E., Boilly, B. and Hondermarck, H. Heparan sulfate proteoglycans play a dual role in regulating fibroblast growth factor-2 mitogenic activity in human breast cancer cells. *Exp Cell Res* **229**, 398-406 (1996).

Dell'Angelica, E. C., Shotelersuk, V., Aguilar, R. C., Gahl, W. A. and Bonifacino, J. S. Altered trafficking of lysosomal proteins in Hermansky-Pudlak syndrome due to mutations in the beta 3A subunit of the AP-3 adaptor. *Mol Cell* **3**, 11-21 (1999).

Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-45 (1999).

DeOme, K. B., Faulkin, L. J., Jr., Bern, H. A. and Blair, P. B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res* **19**, 515-20 (1959).

Dove, A. Proteomics: translating genomics into products? *Nat Biotechnol* **17**, 233-6 (1999).

Drapkin, R., Merino, A. and Reinberg, D. Regulation of RNA polymerase II transcription. *Curr Opin Cell Biol* **5**, 469-76 (1993).

Dulbecco, R., Allen, W. R., Bologna, M. and Bowman, M. Marker evolution during the development of the rat mammary gland: stem cells identified by markers and the role of myoepithelial cells. *Cancer Res* **46**, 2449-56. (1986).

Dulbecco, R., Henahan, M. and Armstrong, B. Cell types and morphogenesis in the mammary gland. *Proc Natl Acad Sci U S A* **79**, 7346-50. (1982).

Dulbecco, R., Unger, M., Armstrong, B., Bowman, M. and Syka, P. Epithelial cell types and their evolution in the rat mammary gland determined by immunological markers. *Proc Natl Acad Sci U S A* **80**, 1033-7. (1983).

Dunn, M. J. in *Advances in electrophoresis* (eds. Chrambach, A., Dunn, M. J. & Radola, B. J.) 1-109 (Weinheim, 1987).

Dutta, A. and Bell, S. P. Initiation of DNA replication in eukaryotic cells. *Annu Rev Cell Dev Biol* **13**, 293-332 (1997).

Edman, P. and Begg, G. A protein sequenator. *Eur J Biochem* **1**, 80-91 (1967).

Edwards, P. A., Hiby, S. E., Papkoff, J. and Bradbury, J. M. Hyperplasia of mouse mammary epithelium induced by expression of the Wnt-1 (int-1) oncogene in reconstituted mammary gland. *Oncogene* **7**, 2041-51 (1992).

Edwards, P. A., Ward, J. L. and Bradbury, J. M. Alteration of morphogenesis by the v-myc oncogene in transplants of mammary gland. *Oncogene* **2**, 407-12. (1988).

Egan, J. J., Greenberg, A. S., Chang, M. K., Wek, S. A., Moos, M. C., Jr. and Londos, C. Mechanism of hormone-stimulated lipolysis in adipocytes: translocation of hormone-sensitive lipase to the lipid storage droplet. *Proc Natl Acad Sci U S A* **89**, 8537-41 (1992).

Elias, J. J., Pitelka, D. R. and Armstrong, R. C. Changes in fat cell morphology during lactation in the mouse. *Anat Rec* **177**, 533-47 (1973).

el-Tanani, M. K. and Green, C. D. Transcription factor, BTF3, and the AF-1 function of the estrogen receptor. *Biochem Soc Trans* **26**, S252 (1998).

Emerman, J. T. and Vogl, A. W. Cell size and shape changes in the myoepithelium of the mammary gland during differentiation. *Anat Rec* **216**, 405-15 (1986).

Emoto, K., Sawada, H., Yamada, Y., Fujimoto, H., Takahama, Y., Ueno, M. *et al.* Annexin II overexpression is correlated with poor prognosis in human gastric carcinoma. *Anticancer Res* **21**, 1339-45 (2001).

Eng, J. K., McCormack, A. L. and Yates, J. R. I. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).

Evers, R., Smid, A., Rudloff, U., Lottspeich, F. and Grummt, I. Different domains of the murine RNA polymerase I-specific termination factor mTTF-I serve distinct functions in transcription termination. *Embo J* **14**, 1248-56 (1995).

Fadok, V. A., Bratton, D. L., Konowal, A., Freed, P. W., Westcott, J. Y. and Henson, P. M. Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-beta, PGE2, and PAF. *J Clin Invest* **101**, 890-8 (1998).

Farrelly, N., Lee, Y. J., Oliver, J., Dive, C. and Streuli, C. H. Extracellular matrix regulates apoptosis in mammary epithelium through a control on insulin signaling. *J Cell Biol* **144**, 1337-48. (1999).

Faulkin, L. J., Jr and DeOme, K. B. Regulation of Growth and Spacing of Gland Elements in the Mammary Fat Pad of the C3H Mouse. *J Nat Cancer Inst* **24**, 953-969 (1960).

Fekete, E. A comparative morphologic study of the mammary gland in a high and low tumour strain of mice. *Am J Pathol* **14**, 557-563 (1938).

Feldman, M., Ruan, W., Cunningham, B. C., Wells, J. A. and Kleinberg, D. L. Evidence that the growth hormone receptor mediates differentiation and development of the mammary gland. *Endocrinology* **133**, 1602-8. (1993).

Fendrick, J. L., Raafat, A. M. and Haslam, S. Z. Mammary gland growth and development from the postnatal period to postmenopause: ovarian steroid receptor ontogeny and regulation in the mouse. *J Mammary Gland Biol Neoplasia* **3**, 7-22 (1998).

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

Ferguson, J. E., Schor, A. M., Howell, A. and Ferguson, M. W. Changes in the extracellular matrix of the normal human breast during the menstrual cycle. *Cell Tissue Res* **268**, 167-77 (1992).

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D. *et al.* Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. *Proc Natl Acad Sci U S A* **99**, 11487-92 (2002).

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J. *et al.* Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *Electrophoresis* **21**, 3517-26 (2000).

Fey, M. F., Moffat, G. J., Vik, D. P., Meisenhelder, J., Saris, C. J., Hunter, T. *et al.* Complete structure of the murine p36 (annexin II) gene. Identification of mRNAs for both the murine and the human gene with alternatively spliced 5' noncoding exons. *Biochim Biophys Acta* **1306**, 160-70 (1996).

Franke, W. W., Luder, M. R., Kartenbeck, J., Zerban, H. and Keenan, T. W. Involvement of vesicle coat material in casein secretion and surface regeneration. *J Cell Biol* **69**, 173-95 (1976).

Friedman, L. S., Ostermeyer, E. A., Lynch, E. D., Welcsh, P., Szabo, C. I., Meza, J. E. *et al.* 22 genes from chromosome 17q21: cloning, sequencing, and characterization of mutations in breast cancer families and tumors. *Genomics* **25**, 256-63 (1995).

Frohlich, M., Motte, P., Galvin, K., Takahashi, H., Wands, J. and Ozturk, M. Enhanced expression of the protein kinase substrate p36 in human hepatocellular carcinoma. *Mol Cell Biol* **10**, 3216-23 (1990).

Fullwood, N. J., Davies, Y., Nieduszynski, I. A., Marcyniuk, B., Ridgway, A. E. and Quantock, A. J. Cell surface-associated keratan sulfate on normal and migrating corneal endothelium. *Invest Ophthalmol Vis Sci* **37**, 1256-70 (1996).

Funderburgh, J. L., Funderburgh, M. L., Mann, M. M. and Conrad, G. W. Arterial lumican. Properties of a corneal-type keratan sulfate proteoglycan from bovine aorta. *J Biol Chem* **266**, 24773-7 (1991a).

Funderburgh, J. L., Funderburgh, M. L., Mann, M. M. and Conrad, G. W. Physical and biological properties of keratan sulphate proteoglycan. *Biochem Soc Trans* **19**, 871-6 (1991b).

Gao, Y., Herndon, J. M., Zhang, H., Griffith, T. S. and Ferguson, T. A. Antiinflammatory effects of CD95 ligand (FasL)-induced apoptosis. *J Exp Med* **188**, 887-96 (1998).

Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H. and Cowan, N. J. A cytoplasmic chaperonin that catalyzes beta-actin folding. *Cell* **69**, 1043-50 (1992).

Gao, Y., Vainberg, I. E., Chow, R. L. and Cowan, N. J. Two cofactors and cytoplasmic chaperonin are required for the folding of alpha- and beta-tubulin. *Mol Cell Biol* **13**, 2478-85 (1993).

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-7 (2002).

Geisow, M. and Walker, J. New proteins involved in cell regulation by Ca2+ and phospholipids. *Trends in Biochemical Sciences* **11**, 420-423 (1986).

Glenney, J. R., Jr. Phosphorylation of p36 in vitro with pp60src. Regulation by Ca2+ and phospholipid. *FEBS Lett* **192**, 79-82 (1985).

Glenney, J. R., Jr. and Tack, B. F. Amino-terminal sequence of p36 and associated p10: identification of the site of tyrosine phosphorylation and homology with S-100. *Proc Natl Acad Sci U S A* **82**, 7884-8 (1985).

Glukhova, M., Koteliansky, V., Sastre, X. and Thiery, J. P. Adhesion systems in normal breast and in invasive breast carcinoma. *Am J Pathol* **146**, 706-16 (1995).

Goldenring, J. R., Smith, J., Vaughan, H. D., Cameron, P., Hawkins, W. and Navarre, J. Rab11 is an apically located small GTP-binding protein in epithelial tissues. *Am J Physiol* **270**, G515-25 (1996).

Gonzalez, M. A., Pinder, S. E., Callagy, G., Vowler, S. L., Morris, L. S., Bird, K. *et al.* Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer. *J Clin Oncol* **21**, 4306-13 (2003).

Gordon, J. R. and Bernfield, M. R. The basal lamina of the postnatal mammary epithelium contains glycosaminoglycans in a precise ultrastructural organization. *Dev Biol* **74**, 118-35 (1980).

Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. *et al.* The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-53. [pii] (2000).

Gorg, A., Postel, W. and Gunther, S. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531-46 (1988).

Gould, K. L., Cooper, J. A. and Hunter, T. The 46,000-dalton tyrosine protein kinase substrate is widespread, whereas the 36,000-dalton substrate is only expressed at high levels in certain rodent tissues. *J Cell Biol* **98**, 487-97 (1984).

Gould, K. L., Woodgett, J. R., Isacke, C. M. and Hunter, T. The protein-tyrosine kinase substrate p36 is also a substrate for protein kinase C in vitro and in vivo. *Mol Cell Biol* **6**, 2738-44 (1986).

Gould, M. N., Biel, W. F. and Clifton, K. H. Morphological and quantitative studies of gland formation from inocula of monodispersed rat mammary cells. *Exp Cell Res* **107**, 405-16 (1977).

Gouon-Evans, V. and Pollard, J. W. Unexpected deposition of brown fat in mammary gland during postnatal development. *Mol Endocrinol* **16**, 2618-27 (2002).

Gouon-Evans, V., Rothenberg, M. E. and Pollard, J. W. Postnatal mammary gland development requires macrophages and eosinophils. *Development* **127**, 2269-82 (2000).

Greenberg, A. S., Egan, J. J., Wek, S. A., Moos, M. C., Jr., Londos, C. and Kimmel, A. R. Isolation of cDNAs for perilipins A and B: sequence and expression of lipid droplet-associated proteins of adipocytes. *Proc Natl Acad Sci U S A* **90**, 12035-9 (1993).

Greenblatt, J. Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell* **66**, 1067-70 (1991).

Grimm, S. L., Seagroves, T. N., Kabotyanski, E. B., Hovey, R. C., Vonderhaar, B. K., Lydon, J. P. *et al.* Disruption of steroid and prolactin receptor patterning in the mammary gland correlates with a block in lobuloalveolar development. *Mol Endocrinol* **16**, 2675-91 (2002).

Grummt, I., Maier, U., Ohrlein, A., Hassouna, N. and Bachellerie, J. P. Transcription of mouse rDNA terminates downstream of the 3' end of 28S RNA and involves interaction of factors with repeated sequences in the 3' spacer. *Cell* **43**, 801-10 (1985).

Grummt, I., Rosenbauer, H., Niedermeyer, I., Maier, U. and Ohrlein, A. A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. *Cell* **45**, 837-46 (1986).

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-9 (1999a).

Gygi, S. P., Rochon, Y., Franza, B. R. and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-30. (1999b).

Haab, B. B., Dunham, M. J. and Brown, P. O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* **2**, RESEARCH0004 (2001).

Hajjar, K. A., Jacovina, A. T. and Chacko, J. An endothelial cell receptor for plasminogen/tissue plasminogen activator. I. Identity with annexin II. *J Biol Chem* **269**, 21191-7 (1994).

Hajjar, K. A. and Menell, J. S. Annexin II: a novel mediator of cell surface plasmin generation. *Ann N Y Acad Sci* **811**, 337-49 (1997).

Han, D. K., Chaudhary, P. M., Wright, M. E., Friedman, C., Trask, B. J., Riedel, R. T. *et al.* MRIT, a novel death-effector domain-containing protein, interacts with caspases and BclXL and initiates cell death. *Proc Natl Acad Sci U S A* **94**, 11333-8 (1997).

Handel, S. E., Rennison, M. E., Wilde, C. J. and Burgoyne, R. D. Annexin II (calpactin I) in the mouse mammary gland: immunolocalization by light- and electron microscopy. *Cell Tissue Res* **264**, 549-54 (1991).

Harris, R. A., Yang, A., Stein, R. C., Lucy, K., Brusten, L., Herath, A. *et al.* Cluster analysis of an extensive human breast cancer cell line protein expression map database. *Proteomics* **2**, 212-23 (2002).

Haslam, S. Z. Mammary fibroblast influence on normal mouse mammary epithelial cell responses to estrogen in vitro. *Cancer Res* **46**, 310-6 (1986).

Haslam, S. Z. Cell to cell interactions and normal mammary gland function. *J Dairy Sci* **71**, 2843-54 (1988).

Haslam, S. Z. and Counterman, L. J. Mammary stroma modulates hormonal responsiveness of mammary epithelium in vivo in the mouse. *Endocrinology* **129**, 2017-23 (1991).

Hassner, A., Birnbaum, D. and Loew, L. M. Charge-shift probes of membrane potential. Synthesis. *J Org Chem* **14**, 2546-51 (1984).

He, C. S., Wilhelm, S. M., Pentland, A. P., Marmer, B. L., Grant, G. A., Eisen, A. Z. *et al.* Tissue cooperation in a proteolytic cascade activating human interstitial collagenase. *Proc Natl Acad Sci U S A* **86**, 2632-6 (1989).

Hewick, R. M., Hunkapiller, M. W., Hood, L. E. and Dreyer, W. J. A gas-liquid solid phase peptide and protein sequenator. *J Biol Chem* **256**, 7990-7 (1981).

Hey, T. and Trefethen, A. e-Science and its implications. *Philos Transact Ser A Math Phys Eng Sci* **361**, 1809-25 (2003).

Hollmann, K. H. in *Lactation: A comprehensive treatise* (ed. Larson, B. L.) 3-95 (Academic Press, New York, 1974).

Hong, J. H., Chiang, C. S., Campbell, I. L., Sun, J. R., Withers, H. R. and McBride, W. H. Induction of acute phase gene expression by brain irradiation. *Int J Radiat Oncol Biol Phys* **33**, 619-26 (1995).

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

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

Howlett, A. R. and Bissell, M. J. The influence of tissue microenvironment (stroma and extracellular matrix) on the development and function of mammary epithelium. *Epithelial Cell Biol* **2**, 79-89 (1993).

Hu, B., Burkhart, R., Schulte, D., Musahl, C. and Knippers, R. The P1 family: a new class of nuclear mammalian proteins related to the yeast Mcm replication proteins. *Nucleic Acids Res* **21**, 5289-93 (1993).

Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H. *et al.* Crystal and molecular structure of human annexin V after refinement. Implications for structure, membrane binding and ion channel formation of the annexin family of proteins. *J Mol Biol* **223**, 683-704 (1992).

Humphreys, R. C. Programmed cell death in the terminal endbud. *J Mammary Gland Biol Neoplasia* **4**, 213-20. (1999).

Humphreys, R. C., Krajewska, M., Krnacik, S., Jaeger, R., Weiher, H., Krajewski, S. *et al.* Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development* **122**, 4013-22. (1996).

Humphreys, R. C., Lydon, J., O'Malley, B. W. and Rosen, J. M. Mammary gland development is mediated by both stromal and epithelial progesterone receptors. *Mol Endocrinol* **11**, 801-11. (1997).

Hynes, R. O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25. (1992).

Ichinose, R. R. and Nandi, S. Influence of hormones on lobulo-alveolar differentiation of mouse mammary glands in vitro. *J. Endocrinol.* **35**, 331-340 (1966).

Ilkbahar, Y. N., Thordarson, G., Camarillo, I. G. and Talamantes, F. Differential expression of the growth hormone receptor and growth hormone-binding protein in epithelia and stroma of the mouse mammary gland at various physiological stages. *J Endocrinol* **161**, 77-87 (1999).

Iozzo, R. V. The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. *Crit Rev Biochem Mol Biol* **32**, 141-74 (1997).

Jacobson, H. R. Effects of CO2 and acetazolamide on bicarbonate and fluid transport in rabbit proximal tubules. *Am J Physiol* **240**, F54-62 (1981).

Jamieson, A. M., Paschke, M. and Clegg, R. A. Mammary tissue lipocortins of the lactating rat. *Biochem Soc Trans* **18**, 1116-7 (1990).

Jansa, P., Mason, S. W., Hoffmann-Rohrer, U. and Grummt, I. Cloning and functional characterization of PTRF, a novel protein which induces dissociation of paused ternary transcription complexes. *Embo J* **17**, 2855-64 (1998).

Jeong, J., Adamson, L. K., Greenhalgh, D. G. and Cho, K. Injury-associated differential regulation of histone expression and modification in the thymus of mice. *Exp Biol Med (Maywood)* **229**, 327-34 (2004).

Jindal, H. K., Chaney, W. G., Anderson, C. W., Davis, R. G. and Vishwanatha, J. K. The protein-tyrosine kinase substrate, calpactin I heavy chain (p36), is part of the primer recognition protein complex that interacts with DNA polymerase alpha. *J Biol Chem* **266**, 5169-76 (1991).

Jindal, H. K. and Vishwanatha, J. K. Purification and characterization of primer recognition proteins from HeLa cells. *Biochemistry* **29**, 4767-73 (1990).

Jockusch, H. and Jockusch, B. M. Structural proteins in the growth cone of cultured spinal cord neurons. *Exp Cell Res* **131**, 345-52 (1981).

Kamma, H., Horiguchi, H., Wan, L., Matsui, M., Fujiwara, M., Fujimoto, M. *et al.* Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms. *Exp Cell Res* **246**, 399-411 (1999).

Kanemaru, K., Meckelein, B., Marshall, D. C., Sipe, J. D. and Abraham, C. R. Synthesis and secretion of active alpha 1-antichymotrypsin by murine primary astrocytes. *Neurobiol Aging* **17**, 767-71 (1996).

Kanno, M., Chalut, C. and Egly, J. M. Genomic structure of the putative BTF3 transcription factor. *Gene* **117**, 219-28 (1992).

Karas, M. and Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-301 (1988).

Kaye, J., Ross, M. H., Romrell, L. J. and Kaye, G. I. in *Histology: A text and Atlas* (eds. Ross, M. H., Romrell, L. J. & Kaye, K.) 678-738 (Williams and Wilkins, Maryland, 1995).

Kearsey, S. E., Maiorano, D., Holmes, E. C. and Todorov, I. T. The role of MCM proteins in the cell cycle control of genome duplication. *Bioessays* **18**, 183-90 (1996).

Kiledjian, M., DeMaria, C. T., Brewer, G. and Novick, K. Identification of AUF1 (heterogeneous nuclear ribonucleoprotein D) as a component of the alpha-globin mRNA stability complex. *Mol Cell Biol* **17**, 4870-6 (1997).

Kim, N. D. and Clifton, K. H. Characterization of rat mammary epithelial cell subpopulations by peanut lectin and anti-Thy-1.1 antibody and study of flow-sorted cells in vivo. *Exp Cell Res* **207**, 74-85 (1993).

Kimura, H., Nozaki, N. and Sugimoto, K. DNA polymerase alpha associated protein P1, a murine homolog of yeast MCM3, changes its intranuclear distribution during the DNA synthetic period. *Embo J* **13**, 4311-20 (1994).

Kinzler, K. W. and Vogelstein, B. Landscaping the cancer terrain. *Science* **280**, 1036-7 (1998).

Klaus, S. Functional differentiation of white and brown adipocytes. *Bioessays* **19**, 215-23 (1997).

Kleinberg, D. L., Ruan, W., Catanese, V., Newman, C. B. and Feldman, M. Non-lactogenic effects of growth hormone on growth and insulin-like growth factor-I messenger ribonucleic acid of rat mammary gland. *Endocrinology* **126**, 3274-6. (1990).

Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-43 (1975).

Knauper, V., Will, H., Lopez-Otin, C., Smith, B., Atkinson, S. J., Stanton, H. *et al.* Cellular mechanisms for human procollagenase-3 (MMP-13) activation. Evidence that

MT1-MMP (MMP-14) and gelatinase a (MMP-2) are able to generate active enzyme. *J Biol Chem* **271**, 17124-31 (1996).

Knight, C. H. and Peaker, M. Mammary cell proliferation in mice during pregnancy and lactation in relation to milk yield. *Q J Exp Physiol* **67**, 165-77 (1982).

Kojima, H., Ishijima, A. and Yanagida, T. Direct measurement of stiffness of single actin filaments with and without tropomyosin by in vitro nanomanipulation. *Proc Natl Acad Sci U S A* **91**, 12962-6 (1994).

Korach, K. S., Couse, J. F., Curtis, S. W., Washburn, T. F., Lindzey, J., Kimbro, K. S. *et al.* Estrogen receptor gene disruption: molecular characterization and experimental and clinical phenotypes. *Recent Prog Horm Res* **51**, 159-86 (1996).

Korteweg, N., Denekamp, F. A., Verhage, M. and Burbach, J. P. Different spatiotemporal expression of DOC2 genes in the developing rat brain argues for an additional, nonsynaptic role of DOC2B in early development. *Eur J Neurosci* **12**, 165-71 (2000).

Koukoulis, G. K., Virtanen, I., Korhonen, M., Laitinen, L., Quaranta, V. and Gould, V. E. Immunohistochemical localization of integrins in the normal, hyperplastic, and neoplastic breast. Correlations with their functions as receptors and cell adhesion molecules. *Am J Pathol* **139**, 787-99 (1991).

Kumble, K. D. and Vishwanatha, J. K. Immunoelectron microscopic analysis of the intracellular distribution of primer recognition proteins, annexin 2 and phosphoglycerate kinase, in normal and transformed cells. *J Cell Sci* **99 (Pt 4)**, 751-8 (1991).

Kuschel, B., Auranen, A., McBride, S., Novik, K. L., Antoniou, A., Lipscombe, J. M. *et al.* Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum Mol Genet* **11**, 1399-407 (2002).

Lee, E. Y., Lee, W. H., Kaetzel, C. S., Parry, G. and Bissell, M. J. Interaction of mouse mammary epithelial cells with collagen substrata: regulation of casein gene expression and secretion. *Proc Natl Acad Sci U S A* **82**, 1419-23 (1985).

Lefebvre, O., Wolf, C., Limacher, J. M., Hutin, P., Wendling, C., LeMeur, M. *et al.* The breast cancer-associated stromelysin-3 gene is expressed during mouse mammary gland apoptosis. *J Cell Biol* **119**, 997-1002. (1992).

Leygue, E., Snell, L., Dotzlaw, H., Hole, K., Hiller-Hitchcock, T., Roughley, P. J. *et al.* Expression of lumican in human breast carcinoma. *Cancer Res* **58**, 1348-52 (1998).

Li, F., Strange, R., Friis, R. R., Djonov, V., Altermatt, H. J., Saurer, S. *et al.* Expression of stromelysin-1 and TIMP-1 in the involuting mammary gland and in early invasive tumors of the mouse. *Int J Cancer* **59**, 560-8. (1994).

Li, M., Liu, X., Robinson, G., Bar-Peled, U., Wagner, K. U., Young, W. S. *et al.* Mammary-derived signals activate programmed cell death during the first stage of mammary gland involution. *Proc Natl Acad Sci U S A* **94**, 3425-30. (1997).

Li, M. L., Aggeler, J., Farson, D. A., Hatier, C., Hassell, J. and Bissell, M. J. Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. *Proc Natl Acad Sci U S A* **84**, 136-40 (1987).

Lieb, K., Fiebich, B. L., Schaller, H., Berger, M. and Bauer, J. Interleukin-1 beta and tumor necrosis factor-alpha induce expression of alpha 1-antichymotrypsin in human astrocytoma cells by activation of nuclear factor-kappa B. *J Neurochem* **67**, 2039-44 (1996).

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-82 (1999).

Lion, N., Gellon, J. O., Jensen, H. and Girault, H. H. On-chip protein sample desalting and preparation for direct coupling with electrospray ionization mass spectrometry. *J Chromatogr A* **1003**, 11-9 (2003).

Loboda, A. V., Krutchinsky, A. N., Bromirski, M., Ens, W. and Standing, K. G. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun Mass Spectrom* **14**, 1047-57 (2000).

Lollo, B. A., Harvey, S., Liao, J., Stevens, A. C., Wagenknecht, R., Sayen, R. *et al.* Improved two-dimensional gel electrophoresis representation of serum proteins by using ProtoClear. *Electrophoresis* **20**, 854-9 (1999).

Londos, C., Brasaemle, D. L., Gruia-Gray, J., Servetnick, D. A., Schultz, C. J., Levin, D. M. *et al.* Perilipin: unique proteins associated with intracellular neutral lipid droplets in adipocytes and steroidogenic cells. *Biochem Soc Trans* **23**, 611-5 (1995).

Londos, C., Brasaemle, D. L., Schultz, C. J., Segrest, J. P. and Kimmel, A. R. Perilipins, ADRP, and other proteins that associate with intracellular neutral lipid droplets in animal cells. *Semin Cell Dev Biol* **10**, 51-8 (1999).

Lozano, J. J., Silberstein, G. B., Hwang, S., Haindl, A. H. and Rocha, V. Developmental regulation of calcium-binding proteins (calelectrins and calpactin I) in mammary glands. *J Cell Physiol* **138**, 503-10 (1989).

Lu, M., Witke, W., Kwiatkowski, D. J. and Kosik, K. S. Delayed retraction of filopodia in gelsolin null mice. *J Cell Biol* **138**, 1279-87 (1997).

Lund, L. R., Romer, J., Thomasset, N., Solberg, H., Pyke, C., Bissell, M. J. *et al.* Two distinct phases of apoptosis in mammary gland involution: proteinase-independent and -dependent pathways. *Development* **122**, 181-93. (1996).

Ma, A. S., Bystol, M. E. and Tranvan, A. In vitro modulation of filament bundling in F-actin and keratins by annexin II and calcium. *In Vitro Cell Dev Biol Anim* **30A**, 329-35 (1994).

MacBeath, G. and Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760-3 (2000).

Maine, G. T., Sinha, P. and Tye, B. K. Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics* **106**, 365-85 (1984).

Mason, S. W., Sander, E. E. and Grummt, I. Identification of a transcript release activity acting on ternary transcription complexes containing murine RNA polymerase I. *Embo J* **16**, 163-72 (1997).

Master, S. R., Hartman, J. L., D'Cruz, C. M., Moody, S. E., Keiper, E. A., Ha, S. I. *et al.* Functional microarray analysis of mammary organogenesis reveals a developmental role in adaptive thermogenesis. *Mol Endocrinol* **16**, 1185-203 (2002).

Matarese, V. and Bernlohr, D. A. Purification of murine adipocyte lipid-binding protein. Characterization as a fatty acid- and retinoic acid-binding protein. *J Biol Chem* **263**, 14544-51 (1988).

Matteoli, M. and De Camilli, P. Molecular mechanisms in neurotransmitter release. *Curr Opin Neurobiol* **1**, 91-7 (1991).

Matunis, E. L., Kelley, R. and Dreyfuss, G. Essential role for a heterogeneous nuclear ribonucleoprotein (hnRNP) in oogenesis: hrp40 is absent from the germ line in the dorsoventral mutant squid. *Proc Natl Acad Sci U S A* **91**, 2781-4 (1994).

Mayer, B. J. Endocytosis: EH domains lend a hand. *Curr Biol* **9**, R70-3 (1999).

McGrath, C. M. Augmentation of the response of normal mammary epithelial cells to estradiol by mammary stroma. *Cancer Res* **43**, 1355-60 (1983).

Menell, J. S., Cesarman, G. M., Jacovina, A. T., McLaughlin, M. A., Lev, E. A. and Hajjar, K. A. Annexin II and bleeding in acute promyelocytic leukemia. *N Engl J Med* **340**, 994-1004 (1999).

Merchant, M. and Weinberger, S. R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* **21**, 1164-77 (2000).

Merrick, W. C. and Nyborg, J. in *Translational control of gene expression* (eds. Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) (Cold Spring Harbor, NY, 2000).

Mignatti, P., Robbins, E. and Rifkin, D. B. Tumor invasion through the human amniotic membrane: requirement for a proteinase cascade. *Cell* **47**, 487-98 (1986).

Miller, F. D., Naus, C. C., Durand, M., Bloom, F. E. and Milner, R. J. Isotypes of alpha-tubulin are differentially regulated during neuronal maturation. *J Cell Biol* **105**, 3065-73 (1987).

Miyamoto, S., Teramoto, H., Gutkind, J. S. and Yamada, K. M. Integrins can collaborate with growth factors for phosphorylation of receptor tyrosine kinases and MAP kinase activation: roles of integrin aggregation and occupancy of receptors. *J Cell Biol* **135**, 1633-42. (1996).

Mizoguchi, Y., Yamaguchi, H., Aoki, F., Enami, J. and Sakai, S. Corticosterone is required for the prolactin receptor gene expression in the late pregnant mouse mammary gland. *Mol Cell Endocrinol* **132**, 177-83 (1997).

Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J. C. *et al.* Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis* **19**, 837-44 (1998).

Monks, J., Geske, F. J., Lehman, L. and Fadok, V. A. Do inflammatory cells participate in mammary gland involution? *J Mammary Gland Biol Neoplasia* **7**, 163-76 (2002).

Moore, D. M., Vogl, A. W., Baimbridge, K. and Emerman, J. T. Effect of calcium on oxytocin-induced contraction of mammary gland myoepithelium as visualized by NBD-phallacidin. *J Cell Sci* **88 (Pt 5)**, 563-9 (1987).

Morris, J. S., Davies, C. R., Griffiths, M. R., Page, M. J., Bruce, J. A., Patel, T. *et al.* Proteomic analysis of mouse mammary terminal end buds identifies axonal growth cone proteins. *Proteomics* **4**, 1802-10 (2004).

Nandi, S. Endocrine control of mammarygland development and function in the C3H/ He Crgl mouse. *J Natl Cancer Inst* **21**, 1039-63 (1958).

Nechad, M. in *Brown adipose tissue* (eds. Trayhurn, P. & Nicholls, D. G.) 1-30 (Arnold Press Ltd, London, 1986).

Neer, E. J., Schmidt, C. J., Nambudripad, R. and Smith, T. F. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 297-300 (1994).

Neer, E. J. and Smith, T. F. A groovy new structure. *Proc Natl Acad Sci U S A* **97**, 960-2 (2000).

Neville, M. C., Medina, D., Monks, J. and Hovey, R. C. The mammary fat pad. *J Mammary Gland Biol Neoplasia* **3**, 109-16 (1998).

Niemann, C., Brinkmann, V., Spitzer, E., Hartmann, G., Sachs, M., Naundorf, H. *et al.* Reconstitution of mammary gland development in vitro: requirement of c-met and c-erbB2 signaling for branching and alveolar morphogenesis. *J Cell Biol* **143**, 533-45 (1998).

Novick, P. and Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr Opin Cell Biol* **9**, 496-504 (1997).

O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-21 (1975).

Orlowski, M. The multicatalytic proteinase complex, a major extralysosomal proteolytic system. *Biochemistry* **29**, 10289-97 (1990).

Ormerod, E. J. and Rudland, P. S. Regeneration of mammary glands in vivo from isolated mammary ducts. *J Embryol Exp Morphol* **96**, 229-43 (1986).

Osaka, F., Kawasaki, H., Aida, N., Saeki, M., Chiba, T., Kawashima, S. *et al.* A new NEDD8-ligating system for cullin-4A. *Genes Dev* **12**, 2263-8 (1998).

Page, M. J., Amess, B., Rohlff, C., Stubberfield, C. and Parekh, R. Proteomics: a major new technology for the drug discovery process. *Drug Discov Today* **4**, 55-62. (1999).

Pasquali, C., Fialka, I. and Huber, L. A. Subcellular fractionation, electromigration analysis and mapping of organelles. *J Chromatogr B Biomed Sci Appl* **722**, 89-102 (1999).

Patterson, S. D. and Aebersold, R. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* **16**, 1791-814 (1995).

Pei, D., Majmudar, G. and Weiss, S. J. Hydrolytic inactivation of a breast carcinoma cell-derived serpin by human stromelysin-3. *J Biol Chem* **269**, 25849-55 (1994).

Pennington, S., Wilkins, M., Hochstrasser, D. and Dunn, M. Proteome analysis: from protein characterization to biological function. *Trends Cell Biol* **7**, 168-173 (1997).

Perdew, G. H., Schaup, H. W. and Selivonchick, D. P. The use of a zwitterionic detergent in two-dimensional gel electrophoresis of trout liver microsomes. *Anal Biochem* **135**, 453-5 (1983).

Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-67 (1999).

Peterson, J. A., Zava, D. T., Duwe, A. K., Blank, E. W., Battifora, H. and Ceriani, R. L. Biochemical and histological characterization of antigens preferentially expressed on the surface and cytoplasm of breast carcinoma cells identified by monoclonal antibodies against the human milk fat globule. *Hybridoma* **9**, 221-35 (1990).

Pitelka, D. R. in *The mammary gland* 944-965 (Plenum Press, New York, 1980).

Pitelka, D. R. and Hamamoto, S. T. Form and function in mammary epithelium: the interpretation of ultrastructure. *J Dairy Sci* **60**, 643-54. (1977).

Plopper, G. E., McNamee, H. P., Dike, L. E., Bojanowski, K. and Ingber, D. E. Convergence of integrin and growth factor receptor signaling pathways within the focal adhesion complex. *Mol Biol Cell* **6**, 1349-65. (1995).

Provost, P., Doucet, J., Stock, A., Gerisch, G., Samuelsson, B. and Radmark, O. Coactosin-like protein, a human F-actin-binding protein: critical role of lysine-75. *Biochem J* **359**, 255-63 (2001).

Pullan, S., Wilson, J., Metcalfe, A., Edwards, G. M., Goberdhan, N., Tilly, J. *et al.* Requirement of basement membrane for the suppression of programmed cell death in mammary epithelium. *J Cell Sci* **109**, 631-42. (1996).

Quarrie, L. H., Addey, C. V. and Wilde, C. J. Programmed cell death during mammary tissue involution induced by weaning, litter removal, and milk stasis. *J Cell Physiol* **168**, 559-69. (1996).

Rabilloud, T. A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* **13**, 429-39 (1992).

Rabilloud, T. in *Methods in molecular biology* (ed. Link, A.) 297-305 (Humana Press, Totowa, 1999).

Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**, 307-16 (1997).

Richards, J., Guzman, R., Konrad, M., Yang, J. and Nandi, S. Growth of mouse mammary gland end buds cultured in a collagen gel matrix. *Exp Cell Res* **141**, 433-43 (1982).

Richardson, K. C. Contractile tissues in the mammary gland, with special reference to myoepithelium in the goat. *Proc Roy Soc* **136**, 30-45 (1947).

Richert, M. M., Schwertfeger, K. L., Ryder, J. W. and Anderson, S. M. An atlas of mouse mammary gland development. *J Mammary Gland Biol Neoplasia* **5**, 227-41. (2000).

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

Rodal, A. A., Tetreault, J. W., Lappalainen, P., Drubin, D. G. and Amberg, D. C. Aip1p interacts with cofilin to disassemble actin filaments. *J Cell Biol* **145**, 1251-64 (1999).

Rohrig, U., Gerisch, G., Morozova, L., Schleicher, M. and Wegner, A. Coactosin interferes with the capping of actin filaments. *FEBS Lett* **374**, 284-6 (1995).

Rosen, J. M., Welm, B. E., Tepera, S. B., Grimm, S. L., Venezia, T., Goodell, M. A. *et al.* Functional characterisation of mammary stem cells. *in preparation*, 1-4 (2003).

Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55-63 (1994).

Ruan, W. and Kleinberg, D. L. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. *Endocrinology* **140**, 5075-81. (1999).

Ruoslahti, E. and Pierschbacher, M. D. New perspectives in cell adhesion: RGD and integrins. *Science* **238**, 491-7 (1987).

Russo, I. H. and Russo, J. Developmental stage of the rat mammary gland as determinant of its susceptibility to 7,12-dimethylbenz[a]anthracene. *J Natl Cancer Inst* **61**, 1439-49 (1978).

Saacke, R. G. and Heald, C. W. in *Lactation: A comprehensive treatise* (eds. Larson, B. L. & Smith, V. R.) 147-189 (Academic Press, New York, 1974).

Sakakura, T., Sakagami, Y. and Nishizuka, Y. Dual origin of mesenchymal tissues participating in mouse mammary gland embryogenesis. *Dev Biol* **91**, 202-7 (1982).

Salcini, A. E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O. *et al.* Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. *Genes Dev* **11**, 2239-49 (1997).

Samuels-Lev, Y., O'Connor, D. J., Bergamaschi, D., Trigiante, G., Hsieh, J. K., Zhong, S. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* **8**, 781-94 (2001).

Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M. *et al.* Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis* **18**, 324-7 (1997).

Santoni, V., Molloy, M. and Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-70 (2000).

Sapino, A., Macri, L., Gugliotta, P. and Bussolati, G. Immunocytochemical identification of proliferating cell types in mouse mammary gland. *J Histochem Cytochem* **38**, 1541-7. (1990).

Savill, J., Dransfield, I., Hogg, N. and Haslett, C. Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis. *Nature* **343**, 170-3 (1990).

Sawadogo, M. and Sentenac, A. RNA polymerase B (II) and general transcription factors. *Annu Rev Biochem* **59**, 711-54 (1990).

Schlierf, B., Fey, G. H., Hauber, J., Hocke, G. M. and Rosorius, O. Rab11b is essential for recycling of transferrin to the plasma membrane. *Exp Cell Res* **259**, 257-65 (2000).

Schmidt, J. T., Morgan, P., Dowell, N. and Leu, B. Myosin light chain phosphorylation and growth cone motility. *J Neurobiol* **52**, 175-88 (2002).

Schmitt-Ney, M., Doppler, W., Ball, R. K. and Groner, B. Beta-casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. *Mol Cell Biol* **11**, 3745-55 (1991).

Schulte, D., Burkhart, R., Musahl, C., Hu, B., Schlatterer, C., Hameister, H. *et al.* Expression, phosphorylation and nuclear localization of the human P1 protein, a homologue of the yeast Mcm 3 replication protein. *J Cell Sci* **108 (Pt 4)**, 1381-9 (1995).

Schwab, B. L., Leist, M., Knippers, R. and Nicotera, P. Selective proteolysis of the nuclear replication factor MCM3 in apoptosis. *Exp Cell Res* **238**, 415-21 (1998).

Schwartz-Albiez, R., Koretz, K., Moller, P. and Wirl, G. Differential expression of annexins I and II in normal and malignant human mammary epithelial cells. *Differentiation* **52**, 229-37 (1993).

Scott, J. E. Supramolecular organization of extracellular matrix glycosaminoglycans, in vitro and in the tissues. *Faseb J* **6**, 2639-45 (1992).

Scott, J. E. Proteodermatan and proteokeratan sulfate (decorin, lumican/fibromodulin) proteins are horseshoe shaped. Implications for their interactions with collagen. *Biochemistry* **35**, 8795-9 (1996).

Servetnick, D. A., Brasaemle, D. L., Gruia-Gray, J., Kimmel, A. R., Wolff, J. and Londos, C. Perilipins are associated with cholesteryl ester droplets in steroidogenic adrenal cortical and Leydig cells. *J Biol Chem* **270**, 16970-3 (1995).

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D. *et al.* Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**, 76-85 (1985).

Smith, T. F., Gaitatzes, C., Saxena, K. and Neer, E. J. The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* **24**, 181-5 (1999).

Stein, T., Morris, J. S., Davies, C. R., Weber-Hall, S. J., Duffy, M. A., Heath, V. J. *et al.* Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. *Breast Cancer Res* **6**, R75-R91 (2004).

Stillman, B. Cell cycle control of DNA replication. *Science* **274**, 1659-64 (1996).

Stoeber, K., Halsall, I., Freeman, A., Swinn, R., Doble, A., Morris, L. *et al.* Immunoassay for urothelial cancers that detects DNA replication protein Mcm5 in urine. *Lancet* **354**, 1524-5 (1999).

Stoeber, K., Tlsty, T. D., Happerfield, L., Thomas, G. A., Romanov, S., Bobrow, L. *et al.* DNA replication licensing and human cell proliferation. *J Cell Sci* **114**, 2027-41 (2001).

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

Streuli, C. H., Bailey, N. and Bissell, M. J. Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. *J Cell Biol* **115**, 1383-95 (1991).

Streuli, C. H. and Gilmore, A. P. Adhesion-mediated signaling in the regulation of mammary epithelial cell survival. *J Mammary Gland Biol Neoplasia* **4**, 183-91. (1999).

Streuli, C. H., Schmidhauser, C., Bailey, N., Yurchenco, P., Skubitz, A. P., Roskelley, C. *et al.* Laminin mediates tissue-specific gene expression in mammary epithelia. *J Cell Biol* **129**, 591-603 (1995).

Takahara, H. and Sinohara, H. Mouse plasma trypsin inhibitors. Isolation and characterization of alpha-1-antitrypsin and contrapsin, a novel trypsin inhibitor. *J Biol Chem* **257**, 2438-46 (1982).

Takahara, H. and Sinohara, H. Inhibitory spectrum of mouse contrapsin and alpha-1-antitrypsin against mouse serine proteases. *J Biochem (Tokyo)* **93**, 1411-9 (1983).

Talhouk, R. S., Bissell, M. J. and Werb, Z. Coordinated expression of extracellular matrix-degrading proteinases and their inhibitors regulates mammary epithelial function during involution. *J Cell Biol* **118**, 1271-82 (1992).

Tanaka, E. M. and Kirschner, M. W. Microtubule behavior in the growth cones of living neurons during axon elongation. *J Cell Biol* **115**, 345-63 (1991).

Tanaka, T., Akatsuka, S., Ozeki, M., Shirase, T., Hiai, H. and Toyokuni, S. Redox regulation of annexin 2 and its implications for oxidative stress-induced renal carcinogenesis and metastasis. *Oncogene* **23**, 3980-9 (2004).

Tieu, Q. and Nunnari, J. Mdv1p is a WD repeat protein that interacts with the dynamin-related GTPase, Dnm1p, to trigger mitochondrial division. *J Cell Biol* **151**, 353-66 (2000).

Todorov, I. T., Attaran, A. and Kearsey, S. E. BM28, a human member of the MCM2-3-5 family, is displaced from chromatin during DNA replication. *J Cell Biol* **129**, 1433-45 (1995).

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J. *et al.* Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**, 377-96 (2001).

Topper, Y. J. and Freeman, C. S. Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol Rev* **60**, 1049-106. (1980).

Tye, B. K. MCM proteins in DNA replication. *Annu Rev Biochem* **68**, 649-86 (1999).

Ullrich, O., Reinsch, S., Urbe, S., Zerial, M. and Parton, R. G. Rab11 regulates recycling through the pericentriolar recycling endosome. *J Cell Biol* **135**, 913-24 (1996).

Unlu, M., Morgan, M. E. and Minden, J. S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-7 (1997).

Van Zwieten, M. J. in *The Rat As Animal Model in Breast Cancer Research* 53-134 (Martinus Nijhoff Publishing, Boston, 1984).

Voll, R. E., Herrmann, M., Roth, E. A., Stach, C., Kalden, J. R. and Girkontaite, I. Immunosuppressive effects of apoptotic cells. *Nature* **390**, 350-1 (1997).

Walden, P. D., Ruan, W., Feldman, M. and Kleinberg, D. L. Evidence that the mammary fat pad mediates the action of growth hormone in mammary gland development. *Endocrinology* **139**, 659-62. (1998).

Warburton, M. J., Mitchell, D., Ormerod, E. J. and Rudland, P. Distribution of myoepithelial cells and basement membrane proteins in the resting, pregnant, lactating, and involuting rat mammary gland. *J Histochem Cytochem* **30**, 667-76 (1982).

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).

Werb, Z. ECM and cell surface proteolysis: regulating cellular ecology. *Cell* **91**, 439-42 (1997).

Wight, T. N., Kinsella, M. G. and Qwarnstrom, E. E. The role of proteoglycans in cell adhesion, migration and proliferation. *Curr Opin Cell Biol* **4**, 793-801 (1992).

Wilkins, M. R., Gasteiger, E., Gooley, A. A., Herbert, B. R., Molloy, M. P., Binz, P. A. *et al.* High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* **289**, 645-57 (1999).

Wilkins, M. R., Hochstrasser, D. F., Sanchez, J. C., Bairoch, A. and Appel, R. D. Integrating two-dimensional gel databases using the Melanie II software. *Trends Biochem Sci* **21**, 496-7 (1996a).

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C. *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-5 (1996b).

Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews* **13**, 19-50 (1996c).

Williams, J. M. and Daniel, C. W. Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol* **97**, 274-90. (1983).

Wiseman, B. S., Sternlicht, M. D., Lund, L. R., Alexander, C. M., Mott, J., Bissell, M. J. *et al.* Site-specific inductive and inhibitory activities of MMP-2 and MMP-3 orchestrate mammary gland branching morphogenesis. *J Cell Biol* **162**, 1123-33 (2003).

Wiseman, B. S. and Werb, Z. Stromal effects on mammary gland development and breast cancer. *Science* **296**, 1046-9 (2002).

Yang, E. and Korsmeyer, S. J. Molecular thanatopsis: a discourse on the BCL2 family and cell death. *Blood* **88**, 386-401 (1996).

Yates, J. R., 3rd. Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**, 5-8 (2000).

Yu, L., Gaitatzes, C., Neer, E. and Smith, T. F. Thirty-plus functional families from a single motif. *Protein Sci* **9**, 2470-6 (2000).

Zebda, N., Bernard, O., Bailly, M., Welti, S., Lawrence, D. S. and Condeelis, J. S. Phosphorylation of ADF/cofilin abolishes EGF-induced actin nucleation at the leading edge and subsequent lamellipod extension. *J Cell Biol* **151**, 1119-28 (2000).

Zheng, X. M., Black, D., Chambon, P. and Egly, J. M. Sequencing and expression of complementary DNA for the general transcription factor BTF3. *Nature* **344**, 556-9 (1990).

Zheng, X. M., Moncollin, V., Egly, J. M. and Chambon, P. A general transcription factor forms a stable complex with RNA polymerase B (II). *Cell* **50**, 361-8 (1987).

Zhou, J., Mulshine, J. L., Unsworth, E. J., Scott, F. M., Avis, I. M., Vos, M. D. *et al.* Purification and characterization of a protein that permits early detection of lung cancer. Identification of heterogeneous nuclear ribonucleoprotein-A2/B1 as the antigen for monoclonal antibody 703D4. *J Biol Chem* **271**, 10760-6 (1996).

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101-5 (2001a).

Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G. *et al.* Analysis of yeast protein kinases using protein chips. *Nat Genet* **26**, 283-9 (2000).

Zhu, H. and Snyder, M. Protein arrays and microarrays. *Curr Opin Chem Biol* **5**, 40-5. (2001b).