Differential Gene Expression Analysis in a Transgenic Mouse Model of Metastatic Breast Cancer.

Thesis submitted in accordance with the requirements of the University of Liverpool for degree of Doctor in Philosophy by Peter Thomas Simpson.

September 2000.

Abstract.

Increased S100A4 has been implicated in the metastatic spread of breast cancer cells in several i*n vitro* and i*n vivo* model systems, and more recently in a series of human breast cancers. In transgenic mice expressing elevated levels of an S100A4 transgene, no tumour phenotype was observed yet the co-ordinated expression of S100A4 and the activated *neu* oncogene resulted in an enhanced metastatic capability of carcinomas that were induced, in part, by *neu* over-expression. Unlike *neu* transgenic mice, *neu/S100A4* transgenic mice exhibited a significant incidence of lung metastases. Despite this phenotypic difference the primary breast tumours in all mice are very similar histopathalogically and resemble human infiltrating ductal carcinomas associated with the over-expression of c-*erb*B-2.

Gene expression analysis has been performed on tumours exhibiting different metastatic capabilities to identify genes that may be responsible, in addition to S100A4, for the different tumour phenotypes between the two strains of transgenic mice. The expression of 588 cDNAs between a primary mammary tumour and a lung metastasis from the same bitransgenic mouse were compared using a mouse Atlas cDNA array. In a parallel experiment, the expression of these cDNAs were investigated between five pooled *neu* mammary tumours and five pooled *neu/S100A4* mammary tumours. The same cDNAs were abundant in each tumour sample and included several members of the cathepsin family, glutathione S-transferase, clusterin and HR21spA (involved in DNA repair). From 1176 gene expression comparisons, from the two experiments, only one cDNA was shown to be differentially expressed. 'Defender against cell death 1' (DAD 1) exhibited 2.3-fold higher expression in the metastasis relative to the primary tumour. The similar expression profiles of this set of genes may reflect the histological similarities of these tumours.

A combination of Suppression Subtractive Hybridisation and Reverse Northern screening was used to directly identify differentially expressed sequences between these tumours. Two subtractions were performed, creating four subtracted libraries, between the same bitransgenic primary tumour and metastasis as used in the array screening, and also between a bitransgenic primary tumour and a *neu* primary mammary tumour. A total of 768 subtracted cDNA clones, from the four subtracted libraries, were sequenced. 192 (25% of those screened) sequences were identified as being differentially expressed by > 2-fold. These included known genes previously implicated to be involved in breast cancer, ESTs and some unknown sequences. 21 of these differentially expressed clones demonstrated coordinated expression with S100A4.

The differential expression of a panel of candidate genes, including osteopontin, Doc-1, BRP 39, an MMTV-derived sequence, 2 EST sequences and 2 potentially novel sequences was confirmed by Virtual Northern analysis. The expression of these sequences was further investigated, by Northern hybridisation, in additional tumours, normal tissues and cell lines derived from these transgenic mice. These experiments confirmed differential expression of these sequences is evident between other tumours, however this also demonstrated that tumours of matched histology and grade exhibit some degree of heterogeneity.

Acknowledgements.

During the course of the last few years a number of people have suffered at my expense, not least my family and close friends. It has often been the case of being present in body but not in mind, or for simply not being present at all. I am sorry for my absence, but I suppose that is a Ph.D. for you!! Thank you all for your patience and encouragement, it was greatly needed and very much appreciated.

I would like to acknowledge many people for their scientific input during the course of these four years. I am particularly grateful to Mike Davies for his continual guidance, encouragement and expertise in just about everything. Many thanks Mike. Thank you also to Roger Barraclough, Philip Rudland and Ross Sibson for your help and assistance along the way, to Bal Shoker for histopathalogical analysis, to Nigel Halliwell for handling my sequences and to Luda Diatchenko (of Clontech) for providing me with the opportunity to screen the Atlas arrays. Thanks also to all who provided useful insights into the world of scientific research and to Cancer & Polio Research Fund and to Clatterbridge Cancer Research Trust for funding the research in this project.

On a personal note, I would like to show my appreciation to all who have encountered me during my time on the 5th floor, at Clatterbridge, and on a social level. These very special people include Simon, Gareth, Geoff, Chris, Mark B., Mark H., Jane, Becca, Rhi, Sue, Ian, Rachel, Dave, Vikki, Mike, Fiona, Shirley, Nigel A. and Nigel H. I would like to mention everyone here, but there are just too many people who have provided great entertainment during this period in my life. Collectively, you were all great, I wish you well for the future and I hope our paths continue to cross because knowing you has been a lot of fun.

Contents.

Abbreviations.

Amp	Ampicillin
BLAST	Basic local alignment search tool
bp	base pairs
cDNA	complementary DNA
cpm	counts per minute
dbj/....	Database of Japan accession number
DD	differential display
DNA	deoxyribonucleic acid
ECM	extracellular matrix
emb/....	European molecular biology database accession number
ER	estrogen receptor
EST	expressed sequence tag
gb/....	Genbank accession number
H&E	haematoxylin & eosin
IDC	infiltrating ductal carcinoma
IPTG	isopropyl-b-D-thiogalactoside
kb	kilobase pairs
kDa	kilodaltons
LB	Luria-Bertani medium
LTR	long terminal repeat
MMLV-RT	Moloney murine leukemia virus reverse transcriptase
MMP	matrix metalloproteinase
MMTV	mouse mammary tumour virus
mRNA	messenger RNA
NCI	National Cancer Institute
NHSBSP	National Health Service Breast Screening Programme
poly(A)$^+$	polyadenylation tail
RNA	ribonucleic acids
RT	reverse transcriptase
RT-PCR	Reverse transcription-polymerase chain reaction
SAGE	serial analysis of gene expression
SMART	Switch mechanism at the 5' of RNA templates
SSH	Suppression Subtractive Hybridisation
TIMP	tissue inhibitor of metalloproteinases
uPA	urokinase-type plasminogen activator
X-Gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside

Chapter 1

Introduction.

1.1 Breast cancer.

Breast cancer is the major malignancy to affect women in the world (Pisani *et al.*, 1999). A recent report from the National Cancer Institute (NCI) estimates that 1 in 8 women in the United States will suffer from the disease (Feuer and Wun, 1999). This is a serious health problem that warrants considerable effort in improving awareness, early detection and diagnosis, effective treatment and ultimately patient survival. Such measures may have contributed to recent decreases in breast cancer incidence and patient mortality in the United Kingdom, Sweden and the United States (Beral *et al.*, 1995; Chu *et al.*, 1996; Herman and Beral, 1996; Garne *et al.*, 1997; Peto *et al.*, 2000). The precise etiology of breast cancer is unknown although one major risk factor involved in developing the disease is the related family history, as genetic predisposition accounts for between 5-10% of breast cancer cases (Madigan *et al.*, 1995). Incidence also increases with age, and has been associated with risk factors such as reproductive lifestyle and hormone replacement therapy (Harris *et al.*, 1992; Madigan *et al.*, 1995; Beral, 1999).

The most dangerous aspect of breast cancer is the metastatic spread of the primary tumour to distant sites in the body where secondary tumours develop. At the time of diagnosis and removal of the primary tumour it is possible, and likely, that many breast tumours have already undergone initial stages of secondary spread. Thus it is particularly important to further our understanding of the mechanisms involved in the spread of cancer so that inhibiting aspects of metastasis and defining treatment of established secondary tumours are effective in treating the disease.

1.2 Multistep nature of breast cancer.

A widely accepted aspect of cancer development and metastatic progression is that multiple steps are required, each reflecting a sequential acquisition of a new cellular phenotype that is necessary for the progressive conversion of a normal cell to a population of metastatic cells. It is believed that a minimum of 4-7 sequential changes is sufficient to confer metastatic capability (Renan, 1993; and reviewed in

Hanahan and Weinberg, 2000). These are summarised below and represent stages of progression that are likely to be common to all types of cancer, including breast cancer.

Cell proliferation during normal circumstances exists as a balance between signals that stimulate and inhibit cell proliferation, for example, paracrine and autocrine growth factors and inhibitors, and signals that promote programmed cell death (apoptosis). For the generation of a population of tumour cells this controlled balance is usually disturbed in favour of an overriding ability of cells to proliferate in response to self-stimulatory signals whilst avoiding growth inhibitory signals. The rate of growth must also be higher than the rate of apoptosis, which is usually higher in a tumour than in normal tissues. This leads to a sustained state of cell growth only when the finite replicative limit of a cell is broken and the cells become immortalised with an unlimited replicative potential, which apparently results from the expression of telomerase (Nakamura and Cech, 1998). The well documented angiogenic switch is necessary for a tumour to maintain growth beyond a size of 1-2 mm (Folkman, 1985). The recruitment/induction of new blood vessels provides the required nutrients and oxygen for the maintenance of this growth and removes waste and toxic products. At some point a proportion of cancers undergo metastasis, whereby some cells acquire the ability to invade the local tissue, penetrate blood vessel or lymphatic systems and travel to distant sites in the body where they may colonise as a growing secondary tumour. This 'metastatic cascade' (Hart *et al.*, 1989) is discussed in more detail below.

1.3 Genetic aspect of cancer.

The multistep nature of cancer development and metastasis coincides with an accumulation of changes to the genome that are associated with changes in gene expression patterns or altered activity of proteins. These have been extensively reviewed elsewhere with specific reference to breast cancer (Devilee and Cornelisse, 1994; Beckmann *et al.*, 1997) or to cancer in general (Hart *et al.*, 1989; Bishop, 1991; Weinberg, 1991; Fearon, 1992; Hanahan and Weinberg, 2000). Abnormal protein expression occurs as a consequence, resulting in the acquisition of new cellular properties that may govern the changing cellular phenotype.

Several lines of evidence suggest that multiple and sequential genetic changes are required for tumourigenesis and progression to occur. Firstly, patients who inherit

a mutation in, for example, p53 or the retinoblastoma gene tend to develop tumours but not immediately, suggesting that a single mutation, carried from birth is insufficient for tumourigenesis to occur and that an additional genetic change is necessary for tumour formation (Knudson, 1971). Molecular analysis of different precursor stages of colorectal tumourigenesis identified several key genetic aberrations that must be acquired, some at preferential stages, for disease progression to occur (Vogelstein *et al.*, 1988). Experimentally, transfection of at least two co-operating oncogenes is also required for the transformation of primary rat embryonal fibroblasts in culture (Rassoulzadegan *et al.*, 1982; Land *et al.*, 1983; Ruley, 1983; Weinberg 1991; Fearon 1992), and 3-4 hits are required to transform, *in vitro*, normal human epithelial cells or fibroblasts for the formation of tumours *in vivo* (Hahn *et al.*, 1999). These were: hTERT, the catalytic subunit of telomerase which overcomes telomere shortening during DNA replication, as means of limiting the replicative potential of a normal cell (Nakamura and Cech, 1998); H-rasV12, an activated oncogene; and Simian virus 40 large-T antigen, a viral oncoprotein which inhibits the growth control factors and tumour suppressors p53 and retinoblastoma. The use of transgenic animals as models of tumourigenesis with the addition of activated oncogenes, usually results in the formation of sporadic primary tumours, thus not all cells carrying and expressing the activated oncogene are transformed, a second or third genetic 'hit' is often required.

Molecular abnormalities can be subdivided into two main groups determined by the effect produced on the genes; some induce a gain in function to proto-oncogenes, others result in the loss of function of proteins characterised as possessing tumour suppressor qualities. Oncogenic activation can arise via gene amplification, chromosomal rearrangement, or indirect elevation in expression, or via mutational activation leading to constitutive activity. In breast cancer, the amplification of one of three chromosomal regions is often observed, 8q24, 17q12 and 17q13, these have been found to harbour the proto-oncogenes c-*myc*, c-*erb*B-2 and *int*-2 respectively (Beckmann *et al.*, 1997). The amplification or over-expression of c-*erb*B-2 is one of the most commonly identified genetic alterations in breast cancer, occurring in 25-30% of all breast carcinomas (Slamon *et al.*, 1987, 1989; Winstanley *et al.*, 1991; Press *et al.*, 1993; Pauletti *et al.*, 1996; Revillion *et al.*, 1998; Pollack *et al.*, 1999). Gene amplification and overexpression of this protein is of clinical significance and so immunodetection of c-*erb*B-2 can be used as a

prognostic marker in breast tumours. Detection of this alteration is associated with a particularly aggressive tumour phenotype and subsequently a reduced overall survival (Slamon *et al.*, 1987, 1989; Press *et al.*, 1993; Pauletti *et al.*, 1996, Winstanley *et al.*, 1991). The clinical implications of c-*erb*B-2 have led to extensive research into the role of this protein, an epidermal growth factor receptor tyrosine kinase, in tumour development and progression. Numerous mouse strains transgenic for this oncogene have been developed (Bouchard *et al.*, 1989; Guy *et al.*, 1992, Muller *et al.*, 1988; Bargmann & Weinberg, 1988; Andrechek *et al.*, 2000) indicating that overexpression of this oncogene is sufficient for the induction of an aggressive tumour phenotype. Clinical trials have taken place involving the use of monoclonal antibody therapy for the treatment of women with c-*erb*B-2 overexpressing metastatic breast cancer as a single agent and in combination with chemotherapy (Shak, 1999). The antibody, Herceptin, is directed to the extracellular domain of the c-*erb*B-2 oncoprotein and has been shown to lead to decreased tumour cell proliferation.

Tumour suppressor genes are associated with loss of function caused by mutation or chromosomal loss. Consistent in many breast tumours and in other tumours is the mutational inactivation of the tumour suppressors *p53* and the retinoblastoma-susceptibility gene *pRb*. Similarly 40-70% of hereditary breast cancers contain mutations to the BRCA1 or BRCA2 breast cancer genes (Shattuck-Eidens *et al.*, 1995; Tavtigian *et al.*, 1996). The mutation state of these proteins might also be used as indicators of susceptibility to developing breast cancer.

Although few genetic changes (4-7) are necessary for tumour initiation and metastasis to occur (Renan, 1993; Hanahan and Weinberg, 2000), many more molecular abnormalities are identified in cancer cells. Genetic changes fall into categories determined by the effect they have on the phenotype of the cells, for example, some have a causative role in initiating or maintaining tumour growth or in promoting the metastatic process. Alternatively genetic changes can occur as a consequence of the changing cellular phenotype, occur independently of the neoplastic physiological state and which do not have a significant functional influence on the progression. Thus, the genetic aspects of cancer, and more specifically, the precise genetic changes which occur that induce a phenotypic response are critically important to understand. Considerable research time and investment has been applied, in many different types of cancer, to pin point the

particular genetic anomalies which have a role in inducing a tumour phenotype or which enhance metastatic progression. The identification of such molecules will undoubtedly further our understanding of the molecular and biological processes occurring in these circumstances, and in addition provide new prognostic markers for specific types of cancer or different stages of the disease. Some molecules will possibly also become new targets for therapeutic intervention.

1.4 Metastasis.

The severity of the metastatic process is such that the majority of deaths of patients with cancer are attributable to the formation of secondary tumours at site(s) in the body distant to the primary tumour growth (Fidler *et al.*, 1978). Dissemination of cells from the primary tumour cell mass requires the completion of several complex steps: (1) detachment from the primary tumour cell mass; (2) migration; (3) invasion through the basement membrane and local extracellular matrix (ECM); (4) intravasation (penetration) of blood vessels or lymphatics; (5) survival in the circulatory system; (6) re-attachment to blood vessels; (7) extravasation; (8) adaptation to the new local tissue microenvironment; and (9) subsequent colonisation as a proliferating metastasis (Fidler, 1991). Alternatively, cells may arrest and grow in the lymphatic system. The biological mechanisms of individual stages of this cascade of events, and the types of molecular changes and proteins implicated are becoming better understood.

Gain in motility of normally confined cells and altered cell-cell and cell-matrix interactions are key aspects of metastasis and enable detachment of tumour cells from the tumour cell mass and subsequently the potential to invade the surrounding tissue. Expression of S100A4, a breast cancer metastasis-related gene (which is discussed in detail later and is reviewed by Barraclough (1998) and Sherbert and Lakshmi (1998)), is observed in normally motile lymphocytic cells and appears abnormally high as normally non-motile epithelial carcinoma cells acquire this phenotype. This is thought to occur via interactions with components of the cytoskeleton. Conversely, loss of expression or function of E-cadherin, a cell-cell adhesion molecule, or associated members of the catenin family, which are involved in anchoring E-cadherin to the cytoskeleton, has been observed in many human tumours and is also associated with increased metastatic capability (Christofori and Semb, 1999).

It is believed that complex interactions between tumour cells and the surrounding 'normal' host tissue is particularly important in determining the efficiency of the sequential stages of this progression. For example, paracrine factors released from the tumour cell population augment a stromal reaction which includes the release of the cytokine vascular endothelial growth factor (VEGF) to stimulate endothelial cell growth and angiogenesis in response to various factors including metabolic stress, mechanical stress or through genetic mutation (reviewed in Carmeliet, 1999; Kerbel, 2000). The degree of angiogenesis exhibited by a tumour, measured by immunohistochemical identification of microvessels correlates with metastatic potential and thus has provided a potential prognostic indicator for the metastasis of human breast cancer (Folkman 1971; Weidner, 1991). ECM degradation is an intrinsic aspect of invasion since the ECM is one of the first barriers tumour cells encounter following detachment from the tumour cell mass. Concerted effort by tumour cells and host tissue, including fibroblasts and activated macrophages, conspire to degrade ECM by the secretion of matrix degrading enzymes. For example, urokinase-type plasminogen activator (uPA), a marker of breast cancer invasion (Duffy, 1996) was shown to be expressed in fibroblast-like cells at the invasive edge of breast carcinomas (Carriero *et al.*, 1994). Likewise cathepsin D and metalloproteinase (MMP) stromelysin-1, other markers of tumour invasion, have been shown to be expressed by both carcinoma cells and reactive stromal cells (Tetu *et al.*, 1999; Sternlicht *et al.*, 2000). An additional target for cleavage by MMP stromelysin-1 is E-cadherin (Lochter *et al.*, 1997), which provides another means for proteolytic enzymes to promote tumour cell invasion.

Cells are presumably able to undergo the later stages of the metastatic cascade using cellular properties acquired during the early stages of metastasis, since these mechanisms are essentially the same (tissue invasion, motility and proliferation). However, there are several lines of evidence suggesting that this is not the case, and that cells that have entered the circulatory systems do not always have the capability to form metastases. This is of potential clinical importance. Whilst invasion of the primary tumour and detection of tumour cells in the lymph nodes are currently used as indicators of the risk of secondary disease, a similar role has been suggested for immunodetection of blood vessels around the primary tumour and detection of circulating tumour cells in the blood. However, detection of tumour cells in the circulatory system or in secondary sites in patients has been shown to not

always correlate with metastatic propensity (Tarin *et al.*, 1984a, 1984b). Similarly, cells that are able to complete certain stages of the process are not necessarily metastatic (Fidler and Radinsky, 1990). This has been demonstrated by injecting cells into experimental animals and measuring the metastatic capability by evidence of growing metastases. The metastatic capability of both human melanoma and rat prostate carcinoma cells was suppressed by the addition of chromosome 6 or 17, respectively. Cells carrying the additional chromosome, suspected of containing metastasis suppressing genes, demonstrated an ability to spread to secondary sites yet an inability to grow in this new environment (Chekmareva *et al.*, 1998; Goldberg *et al.*, 1999). Interestingly, cells appearing as colonised micrometastases that failed to proliferate in the new tissue were able to grow if removed and injected into the same organ to which the primary tumour arose, suggesting that adjusting to a new microenvironment is an important rate limiting step in the formation of a metastatic tumour (Goldberg *et al.*, 1999).

Many aspects of metastasis have been investigated using sophisticated tumour cell imaging technology such as *in vivo* videomicroscopy (reviewed in Chambers *et al.*, 2000) or confocal *in vivo* imaging (Farina *et al.*, 1998). Cells grown in culture, derived from different cancers and exhibiting different *in vivo* metastatic capabilities, were fluorescently labelled and injected into experimental animals, either directly into the circulatory system or orthotopically into a target organ as a primary tumour. Epifluorescence illumination permits cancer cells to be followed as they undertake the sequential steps of metastasis. These experiments demonstrated that processes of tumour cell survival in the circulation, cell arrest and extravasation were efficient and occurred independent of the different metastatic capacity of the cell lines. In contrast processes of intravasation and survival and sustained growth in the secondary tissue were inefficient and were only possible by cells that were highly metastatic (Chambers *et al.*, 2000; Farina *et al.*, 1998).

That the metastatic process is an inefficient one has been known for some time (Weiss, 1990), but we still can not differentiate histologically between primary tumours, or tumour cells elsewhere in the body, that possess high or low metastatic capability. Several molecules have been identified which provide indicators of metastatic capacity, and more are being continuously identified. Loss of molecules known to suppress metastasis, such as E-cadherin and nm23 provide good indicators of metastatic capability (Christofori and Semb, 1999; Hennessy *et al.*, 1991).

Similarly, S100A4 has been described as a metastasis-related gene whose elevated expression in breast carcinomas correlates well with enhanced metastatic potential (Davies *et al.*, 1993) and patient demise (Platt-Higgins *et al.*, 2000; Rudland *et al.*, 2000).

The co-ordinated events required for growth of a secondary tumour in specific types of cancer, including in breast cancer, is still poorly understood. For example, are the metastatic events in specific subtypes of breast cancer (for example infiltrating lobular or ductal carcinoma associated with either c-*erb*B-2 or c-*myc* or *int*-2 amplification) consistently associated with alterations of specific genes? Thus, it is of particular importance to elucidate the molecular and biological events that occur during cancer progression in different classes of breast cancer.

1.5 S100A4

One molecule of particular interest is the metastasis-related protein p9Ka (Barraclough *et al.*, 1987), a member of the S100 family of small calcium binding proteins. Rat p9Ka is the homologue to the human CAPL (Englekamp *et al.*, 1992), and mouse *mts*-1 (Ebralidze *et al.*, 1989) proteins, and has also been referred to as calvasculin (Watanabe *et al.*, 1992), pEL98 (Goto *et al.*, 1988), 18A2 (Linzer and Nathans, 1983), 42A (Masiakowski and Shooter, 1988) and fsp (Strutz *et al.*, 1995). The nomenclature has now been standardised as S100A4. S100A4 has recently been the subject of several reviews (Barraclough, 1998; Sherbert and Lakshmi, 1998) where experimental evidence linking the expression of this protein to metastasis was discussed. Some of the relevant information will be presented here alongside an updated review of the literature and current views of S100A4 in the metastasis of human breast cancer.

The rat protein p9Ka was first identified as being differentially expressed between two rat mammary (Rama) tumour cell lines by 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE; Barraclough *et al.*, 1982). Rama 25 cells were a cloned, cuboidal cell line derived from a rat mammary tumour of epithelial cell origin (Bennett *et al.*, 1978). These cells underwent spontaneous conversion, *in vitro*, to elongated cells of myoepithelial-like morphology (Warburton *et al.*, 1985, Ormerod and Rudland, 1982). One elongated cell clone, Rama 29, derived from the cell line Rama 25 (Bennett *et al.*, 1978) produces high levels of p9Ka (Barraclough *et al.*, 1982). p9Ka was so named because it has an apparent

molecular weight of 9.0 ± 0.5 kDa and an acidic isoelectric point of 5.5 ± 0.3 (Barraclough *et al*., 1984).

1.5.1 *In vitro* models of S100A4 in metastasis.

Elevated expression of rat S100A4 was observed as a normal rat mammary cell line underwent spontaneous conversion to elongated, myoepithelial-like cells (Barraclough *et al*., 1984), during serum-induced increases in growth rate of mouse fibroblasts (Jackson-Grusby *et al*., 1987), and by the oncogenic or carcinogenic transformation of mouse fibroblasts (Goto *et al*., 1988), or rat kidney cells (De Vouge and Mukherjee, 1992). It was possible that these changes in growth and differentiation were associated with elevated S100A4. Conversely changes in S100A4 expression may also occur as a consequence of such changes in phenotype.

Increased expression of S100A4 was seen in a metastatic rat mammary tumour cell line (Rama 800; Dunnington *et al*., 1984) relative to non-metastatic epithelial cells from which they were derived (Dunnington *et al*., 1983) and to cells of the normal rat mammary gland (Barraclough *et al*., 1984). This suggested that a progressive increase in S100A4 expression coincided with an increase in metastatic potential, as determined by injection of cells into the mammary fat pads of syngeneic rats. Similarly cell lines derived from a spontaneous mouse mammary carcinoma, exhibited increases in metastatic potential correlating with increases in expression of mouse S100A4 (Ebralidze *et al*., 1989). Yet, as with the spontaneous conversion of cells in culture it was not known whether S100A4 expression was directly responsible for these differences in metastatic capability.

Transfection experiments of the rat, mouse and human S100A4 gene into rodent cell lines and subsequent assessment of the *in vivo* metastatic capacity demonstrated that elevated S100A4 could induce metastatic progression. The first of these independent experiments demonstrated that stable transfection of rat S100A4 genomic sequence into Rama 37 cells induced, on injection into syngeneic Wistar-Firth rats, lung and lymph node metastases that were immunopositive for S100A4 (Davies *et al*., 1993). Untransfected Rama 37 cells, a benign rat mammary epithelial cell line, induced primary mammary gland tumours only, and Rama 37 cells transfected with the oncogene EJ-*ras* also only produced primary tumours (Davies *et al*., 1993). Elevation of S100A4 by transfection of human S100A4 into the benign Rama 37 cells also induced lung metastases (Lloyd *et al*., 1998). In similar

experiments mouse S100A4 was able to induce metastases when transfected into normally non-metastatic human breast carcinoma cell line MCF-7 and injected in to immunodeficient nude mice (Grigorian *et al.*, 1996). These experiments demonstrated the metastasis-inducing properties of human, rat and mouse S100A4 in both human and rodent cells.

1.5.2 Physiological role of S100A4.

Despite the implications of S100A4 in metastasis, the physiological role of this protein in normal circumstances and in a tumourigenic environment is still unknown. Immunofluoresence studies of cultured rodent cells indicated that S100A4 was located to cytoskeletal elements, a pattern of staining that was very similar to that observed for phalloidin-localised actin filaments (Takanaga *et al.*, 1994; Gibbs *et al.*, 1994; Davies *et al.*, 1993; Watanabe *et al.*, 1993). A direct interaction between S100A4 (calvasculin) with F-actin was demonstrated, *in vitro*, by its ability to co-sediment with F-actin, this also caused bundling of F-actin filaments in a calcium (Ca^{2+}) dependent manner (Watanabe *et al.*, 1993). Interactions between S100A4 and non-muscle myosin have also been demonstrated (Kriajevska *et al.*, 1994; Ford and Zain, 1995; Ford *et al.*, 1997; Kriajevska *et al.*, 1998) and again these were dependent on Ca^{2+} ions. S100A4 appeared to destabilise myosin filaments (Ford *et al.*, 1997) and inhibit the actin-activated ATPase activity of myosin via S100A4 binding to the C-terminal end of the non-muscle myosin heavy chain region (Ford *et al.*, 1997; Kriajevska *et al.*, 1998). Furthermore, this interaction inhibits phosphorylation of myosin heavy chain by protein kinase C (Kriajevska *et al.*, 1998), an event that is thought to control assembly and disassembly of myosin filaments, thus suggesting a mechanism for S100A4 in modulating cytoskeletal dynamics. Further association of S100A4 with components of the cytoskeleton arose through binding studies of mouse fibroblast non-muscle tropomyosin with NIH 3T3 cell extracts. S100A4 was identified as a binding partner and was subsequently co-localised with cytoskeletal tropomyosin by immunofluorescence (Takanaga *et al.*, 1994). In addition, Lakshmi *et al.*, (1993) demonstrated an association between increases in S100A4 expression and alterations in the polymerisation state of tubulin accompanied changes in cellular morphology and metastatic behaviour.

In addition to the discovery of these *in vitro* binding partners, S100A4 was also shown to form homodimers, *in vitro*, in interactions that were independent of calcium (Pedrocchi *et al*., 1994; Ilg *et al*., 1996) and heterodimers with S100A1 *in vitro* and in a yeast 2-hybrid *in vivo* system (Wang *et al*., 2000). Site directed mutagenesis of conserved amino acid residues in S100A4, that were shown to be associated in the dimer interface of S100A6 (Potts *et al*., 1995), prevented both homo- and S100A1-interactions (Wang *et al*., 2000). Co-expression of S100A1 and S100A4 (Mandinova *et al*., 1998; Wang *et al*., 2000) and localisation of these two proteins to cytoskeletal structures (Davies *et al*., 1993; Gibbs *et al*., 1994; Mandinova *et al*., 1998; Wang *et al*., 2000) has been reported in mammalian cells. The biological function of these interactions between S100A1 and S100A4 are unknown yet they may represent means for regulating S100A4-related metastatic capability (Wang *et al*., 2000).

S100A4 is expressed in a wide variety of specific cell types of numerous normal tissues in the rat, mouse and human (Davies *et al*., 1995; Gibbs *et al*., 1995; Takanaga *et al*., 1994a, 1994b). Expression has also been detected in components of the immune system identified in the spleen, lymph nodes, bone marrow and blood, and in particular in cells of a normally highly motile nature such as lymphocytes, monocytes/macrophages and polymorphonuclear leukocytes/neutrophils (Gibbs *et al*., 1995; Takanaga *et al*., 1994a, 1994b). In line with this distribution and the observations that S100A4 interacts with components of the cytoskeleton and subsequently affects cytoskeletal dynamics, it is thought that abnormal S100A4 expression is able to induce metastasis by enhancing the motility of normally confined cells.

Several other *in vitro* cell systems have been used to show that elevated S100A4 expression is associated with alterations to the cytoskeletal structure. These changes were also often associated with changes in the morphological appearance, motility and/or invasive capacity of the cells. This evidence suggests a possible mechanism for elevated S100A4 expression enhancing the metastatic capability of cancer cells. For example, mouse S100A4 expression was analysed in human promyelocytic leukaemia HL-60 cells in response to phorbol 12-myristate 13-acetate (PMA) and dimethylsulfoxide (DMSO), treatment which induces macrophage and granulocytic differentiation of myeloid cells. Elevated levels of S100A4 were induced in response to PMA and DMSO which coincided with an increase in the

motility of these cells in semi-solid medium. These cells did not demonstrate altered cell adhesion or phagocytic properties in response to increased S100A4, suggesting the elevated S100A4 was related to gain in cellular motility (Takanaga *et al*., 1994).

The non-metastatic mouse mammary carcinoma CSMLO cell line expresses very low levels of mouse S100A4 whereas the metastatic variant cell line CSML100 expresses high S100A4 levels (Ebralidze *et al*., 1989). CSMLO cells transfected with S100A4 demonstrated a change in morphology to closely resemble the elongated CSML100 cells (Ford *et al*., 1995), a similar change in morphology associated with elevated S100A4 to that previously described (Bennett *et al*., 1978; Barraclough *et al*., 1984). The transfected CSML0 cells were analysed for cell motility and invasiveness in *in vitro* assays and for metastasis, *in vivo,* relative to non-transfected cells. Increased motility in Boyden chemotaxis chambers was observed with increased S100A4 expression, yet unlike CSML100 cells, no change in invasion through matrigel or metastatic capacity was observed (Ford *et al*., 1995). Thus, although CSML0-S100A4 and CSML100 cells express high S100A4 and share similar morphology they demonstrate different phenotypes. It was proposed that this difference is related to the co-ordinated expression and activation of metalloproteases MMP 2 and MMP 9 in CSML100 cells that was not observed in CSML0 cells either with or without transfection of S100A4. Thus, the co-ordinated, elevated expression of S100A4 and proteases is potentially required for the metastatic phenotype of these cells (Ford *et al*., 1995).

S100A4 expression was further associated with matrix metalloproteinase (MMP) activity when a highly metastatic human osteosarcoma cell line, expressing high levels of S100A4, was transfected with a ribozyme directed specifically against S100A4 mRNA (Maelandesmo *et al*., 1996; Bjornland *et al*., 1999). Transfected cells exhibited decreased S100A4, relative to control cells transfected with vector only, reduced motility across un-coated filters in a transwell chamber assay and reduced invasive capacity through matrigel-coated filters (Bjornland *et al*., 1999). In both cases the extent of reduced S100A4 expression, as measured by Northern blotting, reflected the extent of reduced motility and invasiveness (Bjornland *et al*., 1999). It also correlated with reduced MMP 2, membrane type-1 (MT-1) MMP and tissue inhibitor of metalloproteinase-1 (TIMP-1) mRNA and protein. Conversely, TIMP-2 expression was up regulated in all cells exhibiting decreased S100A4 (Bjornland *et al*., 1999). Together, altered expression of these genes may provide an

explanation for the reduced invasive properties of the ribozyme-transfected cells. Injection of the ribozyme-transfected cells into nude mice resulted in reduced metastatic potential of these cells relative to non-transfected cells. In all cases the proliferative rate and tumourigenicity of cells, in culture and *in vivo*, were not affected by the ribozyme (Maelandesmo *et al.*, 1996). Interestingly, on occasions when metastases were identified following injection of ribozyme-transfected cells, these metastases were immunoreactive for S100A4, suggesting a possible clonal selection of cells with incomplete ribozyme activity in the developing primary tumour may have taken place (Maelandesmo *et al.*, 1996). These experiments demonstrated, for the first time, that reducing S100A4 levels can reverse the high metastatic capability of cultured cells.

1.5.3 *In vivo* models of S100A4 in metastasis.

The utility of many experiments in culture as models of metastasis is incomplete since they do not demonstrate the ability of cells to complete all stages of the metastatic process, only the abilities to invade or be motile. However, injection of these cells into animals does help to corroborate their propensity to metastasise. Such experiments also have technical limitations in that cells released directly into the circulatory system during the injection process have not had to perform the processes of invasion, motility and intravasation. Therefore, if cells are not carefully injected into the tissue (e.g. the mammary gland) they might escape directly to the secondary site. Transgenic animals can be created as completely *in vivo* models of tumour progression to metastasis, whereby cells must acquire the ability to complete all stages of the metastatic process for secondary tumour formation to occur.

Transgenic mouse technology has been applied to help assess the role of S100A4 in inducing metastasis. Transgenic mice were developed which contained multiple copies of the rat S100A4 transgene integrated into the mouse genome. The gene was under the control of its own promoter and expression of mRNA and protein was detected in a transgene copy-number-dependent manner that was independent of the position of its integration into the mouse genome (Davies *et al.*, 1995). Despite elevated S100A4 expression, the mice developed normally displaying no detectable neoplastic phenotype (Davies *et al.*, 1995), thus discounting the possibility that S100A4 acted as a transforming oncogene. Interestingly, expression of the rat S100A4 mRNA and protein exhibited the same tissue distribution as that observed

for endogenous rat S100A4, presumably due to the rat transgene being under the control of its own promoter (Davies *et al*., 1995). Furthermore this tissue distribution was distinct from that observed for the endogenous mouse S100A4, as determined by Northern blot analysis. This indicates that whilst mice possess the transcriptional control machinery required for the expression of the rat transgene, the transcription of the rat and mouse S100A4 genes is controlled by different regulatory mechanisms. Expression of the rat transgene was detected by Northern hybridisation in a wide variety of tissues. This was highest in the lung, low in the brain and of an intermediate level in other tissues such as mammary gland, kidney, spleen and heart. Mouse S100A4 was detected mainly in lymphoid tissue (spleen, thymus and lymph node) but also to a lower degree in other organs such as lung, mammary gland, brain and kidney (Davies *et al*., 1995).

To more directly assess the role of S100A4 in inducing metastasis, S100A4 transgenic mice were mated with transgenic mice that were susceptible to mammary tumour formation due to the expression of *neu*, the activated rodent homologue to the human c-*erb*B-2 oncogene, in the mammary epithelium (Bouchard *et al*., 1989). The importance of c-*erb*B-2 in transgenic models of breast cancer was described earlier. In these mice, *neu* contains a point mutation in the transmembrane domain of the protein, which confers constitutive activation of its tyrosine kinase activity. Expression of this transgene is controlled by the mouse mammary tumour virus (MMTV) long terminal repeat (LTR) regulatory sequences which are glucocorticoid activated (Ross and Solter, 1985). *Neu* expression was thus targeted to epithelial cells of the mammary glands of these transgenic mice. Activated *neu* expression correlated with the stochastic formation of primary mammary tumours in female mice between 7-14 months of age. Mammary epithelial tumour cells contained high levels of Neu protein, detected by immunohistochemistry (Bouchard *et al*., 1989; Davies *et al*., 1996).

Introduction of the S100A4 transgene into these MMTV-*neu* transgenic mice, by mating, produced further evidence that S100A4 was able to induce metastasis in a now fully *in vivo* model of metastatic breast cancer (Table 1.1). Transgenic offspring that inherited the *S100A4* transgene only, displayed no detectable neoplastic phenotype, as observed for the S100A4 parental transgenic mice. Female offspring containing the MMTV-*neu* transgene developed primary mammary gland tumours, also as exhibited by the MMTV-*neu* parent strain. Female bitransgenic offspring that

inherited and expressed both transgenes (MMTV-*neu/S100A4* bitransgenic mice) developed sporadic mammary gland tumours at a slightly earlier age than their MMTV- *neu* littermates. Fifty percent of MMTV-*neu/S100A4* mice had developed palpable mammary tumours by 11.4 months, whereas 50% of MMTV- *neu* mice had developed palpable mammary tumours by 14.4 months. Thus suggesting that the presence of elevated levels of S100A4 enhanced the rate of tumour growth in these mice (Davies *et al*., 1996). Unlike MMTV-*neu* mice, bitransgenic mice had clear evidence of metastatic spread to the lungs (Davies *et al*., 1996).

Transgene	Number of mice	Mice with breast tumours		Mice with lung lesions		Mean number of lung lesions	Mean % of lung area
		12 months	14 months	Microscopic deposits	Macroscopic metastases		
None	8	0	0	-	-	-	-
S100A4	12	0	0	-	-	-	-
MMTV-*neu*	16	4	7	3	0	4.4	0.07
Both	24	14	17	4	6	19.0	5.10
P value		0.03 [a]	0.06 [a]			0.055 [b]	0.021 [b]

Table 1.1 Incidence of mammary tumours and lung metastases in transgenic mice.

P values: a = Fisher Exact test; b = Student T-test. Table based on data from Davies *et al*., 1996.

1.5.3.1 Histopathology of transgenic mouse tumours.

The histology of 162 breast tumours was examined from 24 transgenic mice, 42 of these were from MMTV-*neu* mice and 120 from MMTV-*neu/S100A4* mice (B. Shoker, personal communication; Simpson *et al*., manuscript in preparation). The tumours were assessed according to guidelines recommended by the National Health Service Breast Screening Programme (NHSBSP, 1997). All the breast tumours resembled human infiltrating ductal carcinomas (IDC; Figure 1.1), a form of breast carcinoma that is often associated with the over-expression of c-*erb*B-2 in humans (Van de Vijver *et al*., 1988) and that has previously been identified in other strains of transgenic mice containing the *neu* transgene (Jolicoeur *et al* 1996). Thus, it appears

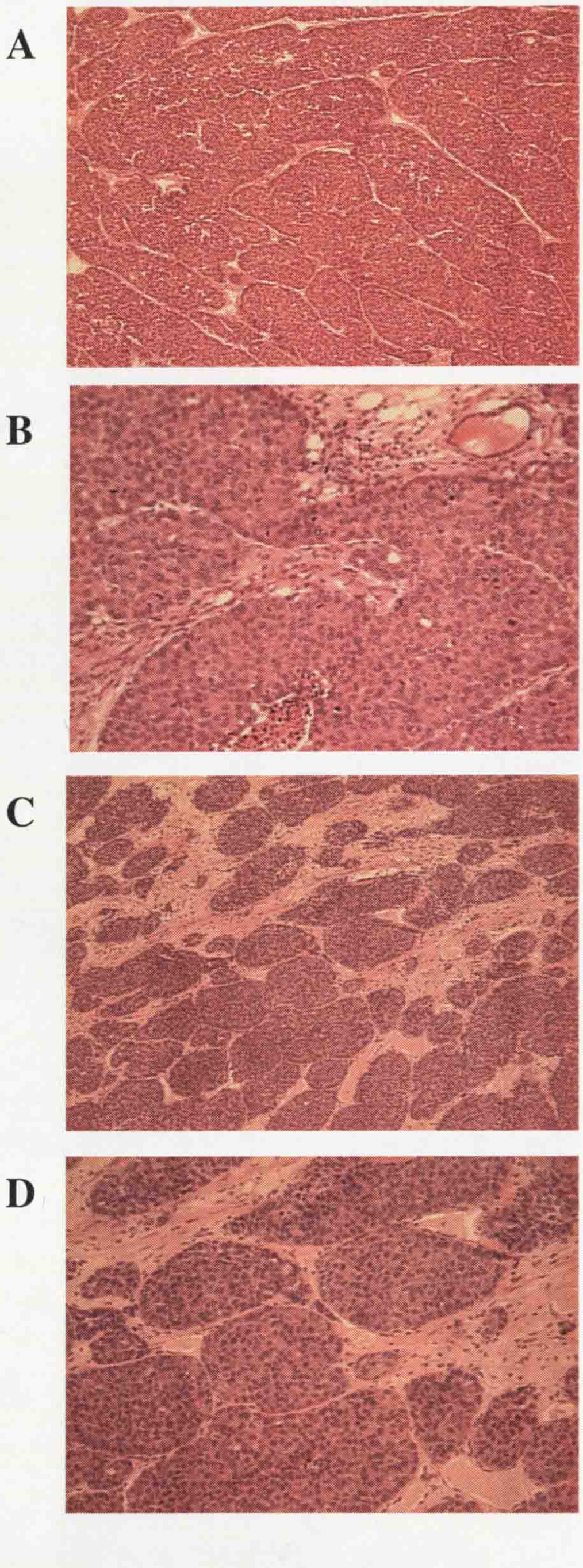


Figure 1.1 Histology of paraffin embedded sections of primary mammary gland tumours from an MMTV-*neu* mouse and an MMTV-*neu*/*S100A4* mouse.

Low (A and C) and high (B and D) power views of mammary gland tumour sections stained with H & E from an MMTV-*neu* mouse (A and B) and from an MMTV-*neu*/*S100A4* mouse (C and D).

that elevated expression of this protein is responsible for initiating the tumours in these transgenic mice and also in defining their histopathology. The majority of the tumours had a mainly circumscribed solid growth pattern with expansive margins, the solid areas frequently formed an insular/alveolar pattern, and exhibited regions of tumour necrosis. Some tumours exhibited other architectural patterns including papillary and focal cystic change. The tumours were graded according to extent of tubule formation, nuclear pleomorphism and mitotic rate. All the tumours showed either moderate or poor tubule formation, the former of these often created a cribriform pattern. Carcinomas showing good tubule formation were not identified. The majority of carcinomas also showed marked nuclear pleomorphism with few exhibiting moderate and none exhibiting mild nuclear pleomorphism. A high mitotic count was present in all but three of the tumours. The overall tumour grade was 3 for 157 (97%) tumours and 2 for 5 (3%) tumours. There was no difference in overall grade or any of its components (tubule formation, nuclear pleomorphism and mitotic count) when MMTV-*neu* mice, MMTV-*neu/S100A4* mice with no metastases and MMTV-*neu/S100A4* mice with metastases were compared (B. Shoker, personal communication; Simpson *et al*., manuscript in preparation).

The breast tissue surrounding many of the tumours contained "dysplastic" lesions. These showed a spectrum of changes from lobules in which occasional acini were lined by a layer of large pleomorphic cells with vesicular nuclei to lobules grossly distended by cells with high or intermediate grade nuclei usually cytologically similar to the invasive tumours. The former lesions probably represent the earliest morphologically detectable change in the progression to neoplasia whilst the latter were similar to high or intermediate nuclear grade ductal carcinoma *in situ* (DCIS). Some of the larger lesions were difficult to distinguish from invasive carcinoma.

Histological examination of tissues other than the mammary gland identified tumour cells in the lungs of some mice. These cells were immunopositive for neu protein suggesting that they originated from the primary mammary gland tumours. No other tissues were found to contain detectable tumour cells. Three out of seven MMTV-*neu* mice that developed mammary tumours contained groups of tumour cells in the lungs, these were often missed by haematoxylin and eosin staining because they were so small and were only detected following immunohistochemical staining of histological sections for neu protein. These cells existed as small clumps

that were encapsulated within intact vascular basement membrane. It would appear, therefore that, although tumour cells from some of these MMTV-*neu* transgenic mice were able to undertake initial stages of metastasis, these cells were non-metastatic because they were unable to colonise the new microenvironment by invading the lung tissue and re-initiating sustained cell proliferation. Ten bitransgenic mice, out of 17 analysed, also contained neu-staining tumour cells in the lungs. These tumours were like the majority of primary tumours from the transgenic mice in that they demonstrated an invasive capacity. In contrast to MMTV-*neu* mice, bitransgenic mice contained multiple micrometastases and macrometastases with as much as 24% of the histological section of the lungs consisting of metastases (Figure 1.2).

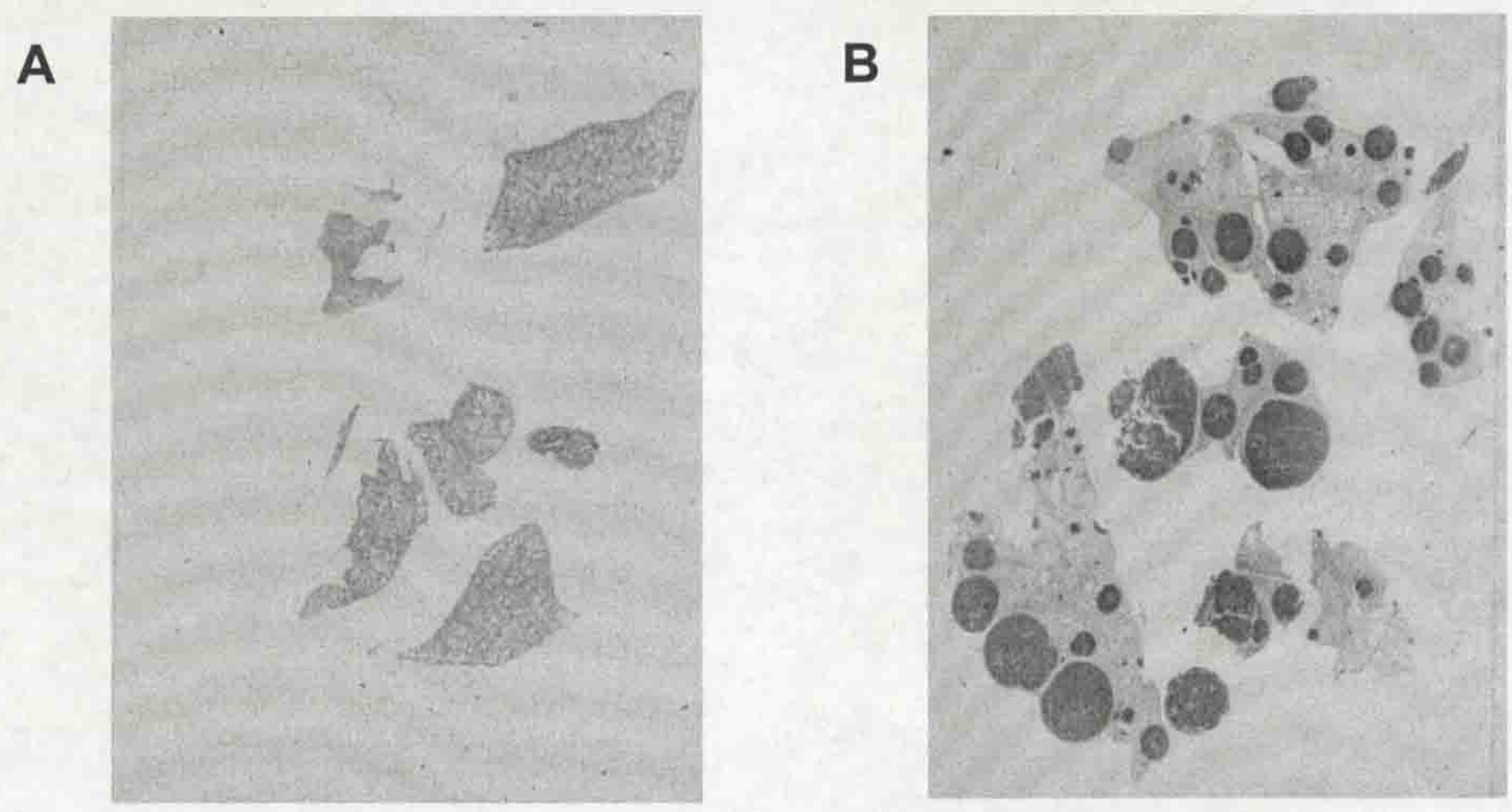


Figure 1.2 Histology of lung tissue from transgenic mice.

Lung tissue of an MMTV-*neu* transgenic mouse that had developed primary mammary gland tumours (A). Lung tissue of an MMTV-*neu/S100A4* transgenic mouse that had developed primary mammary tumours and subsequently multiple micrometastases and macrometastases (B). Both tissues stained with haematoxylin and eosin. Figure B taken from Davies *et al.*, 1996.

The metastases were too small to undergo formal grading according to the same NHSBSP guidelines, yet they demonstrated the same histological characteristics as the majority of primary tumours (Figure 1.3). In particular, they demonstrated regions of necrosis, similar proportions of tumour cells undergoing mitosis, a high

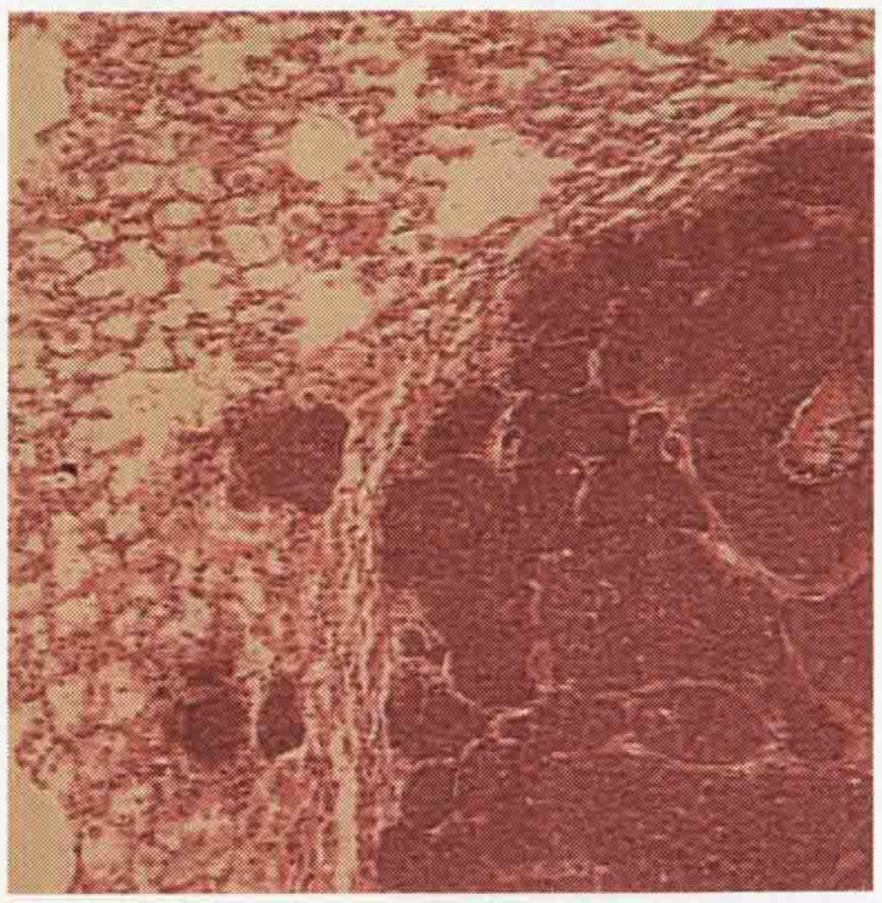
degree of nuclear pleomorphism and a lack of tubule formation. S100A4 immunohistochemical staining of the normal lung tissue was high in MMTV-*neu/S100A4* transgenic mice, as expected following histological characterisation of this protein in the S100A4 transgenic mice. S100A4 immunopositive staining of the metastatic tumour cells was also evident, particularly in regions where cells were invading the local lung tissue (Figure 1.3). Laminin immunohistochemical staining demonstrated that these cells had penetrated the basement membrane of alveoli or blood vessels. Thus the additional presence of S100A4 in these tumours seems to confer an enhanced ability of cells to undergo metastasis *in vivo* (Davies *et al.*, 1996).

In a completely independent *in vivo* analysis of S100A4 in metastasis, transgenic mice were generated to contain multiple copies of mouse *mts*-1 under the control of the MMTV-LTR. As with S100A4 (p9Ka) transgenic mice, these mice were phenotypically normal confirming, again, that S100A4 was not oncogenic *per se* (Ambartsumian *et al.*, 1996). This strain of transgenic mice was mated with GRS/A mice which spontaneously developed mammary gland tumours due to MMTV proviral integration and subsequent activation of the *wnt*-1 and *int*-2 oncogenes (Mester *et al.*, 1987); these tumours rarely metastasise (Van der Valk, 1981). Expression of *mts*-1 was detected in stromal cells surrounding primary mammary tumours of GRS/A with no effect on the ability of tumours to undergo metastatic spread. However, mouse S100A4 transgene expression in the mammary tumour cells themselves did confer an increased metastatic capability to these primary tumours (Ambartsumian *et al.*, 1996).

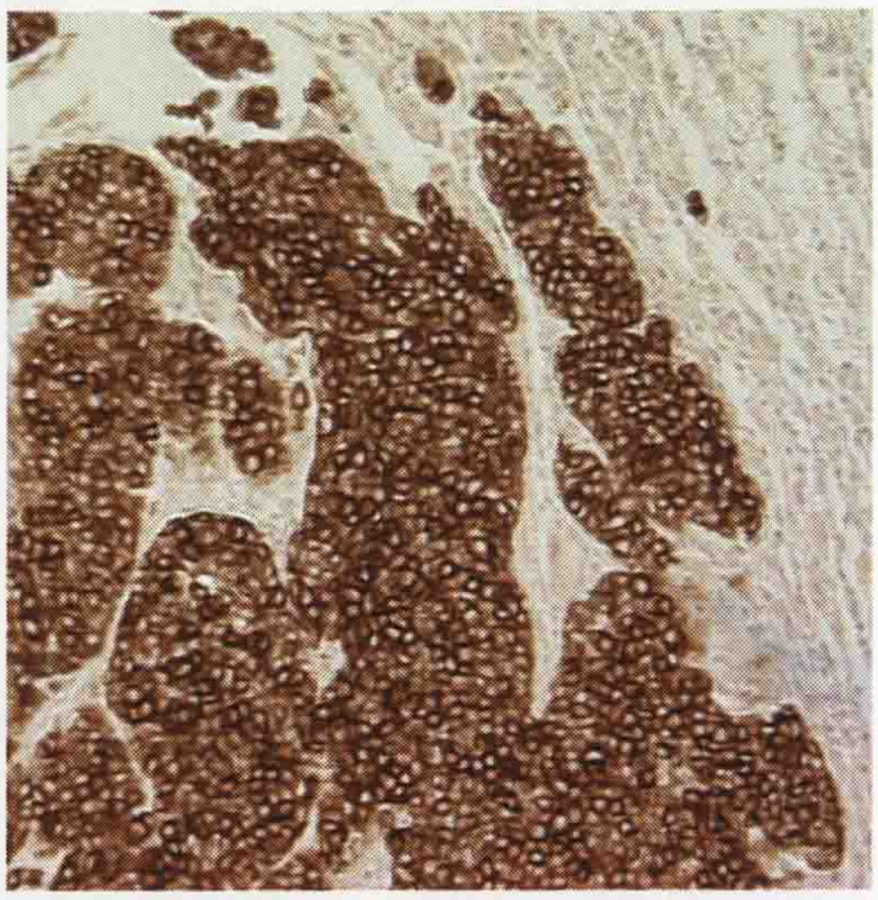
1.5.4 S100A4 in human tumours.

Studies in human tumours have sought to corroborate data obtained in *in vitro* and *in vivo* models of cancer to ascertain if a role for S100A4 in metastasis is apparent in the human disease. Initial evidence suggests that S100A4 expression does correlate with an increased susceptibility of the tumour to secondary disease. Immunohistochemical analysis of S100A4 in human colorectal neoplastic tissue confirmed that, as in cell lines, elevated S100A4 expression correlated with an increased metastatic potential (Takenaga *et al.*, 1997). Normal colonic epithelia and 12 adenomas were immunonegative for S100A4, whereas expression was detected in lymphocytes, macrophages and in smooth muscle of normal colorectal tissue,

A



B



C

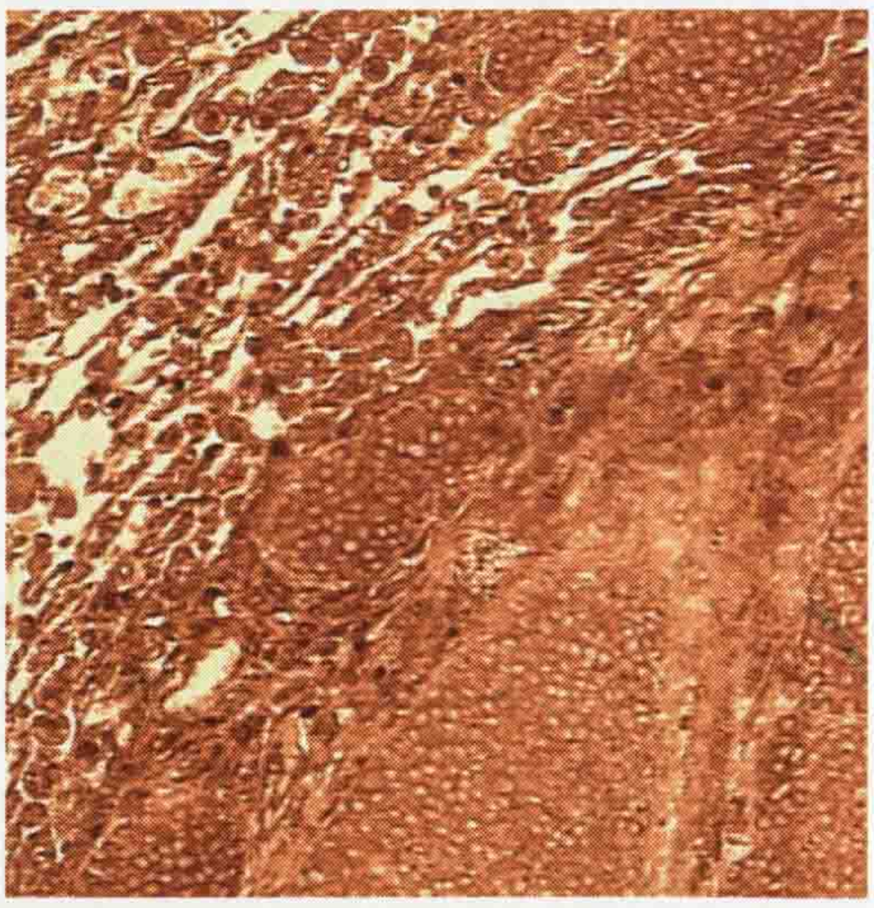


Figure 1.3 Histology of paraffin embedded sections of an MMTV-*neu*/*S100A4* lung metastasis.

Low power view of lung sections stained with H & E (A); high power views of same sample stained with immunohistochemistry for c-*erb*B-2 (brown stain; B) and for p9Ka (brown stain; C).

consistent with that observed previously in the rat (Gibbs *et al.*, 1995). S100A4 expression was also detected in 8/18 focal carcinomas within adenoma and in 50/53 adenocarcinoma specimens, suggesting abnormal expression of S100A4 occurs in these epithelial-derived tumour cells. Expression was most prominent in regions of tumour invasion and significantly almost all of the carcinoma cells found in the lymphatics, blood vessels and as liver metastases stained positive for S100A4 (Takenaga *et al.*, 1997). These results demonstrate a positive correlation between abnormally high S100A4 expression and enhanced metastatic potential of human colorectal adenocarcinoma cells.

This investigation has recently been confirmed by S100A4 detection, by immunohistochemical staining, *in situ* hybridisation and quantitative RT-PCR in human colorectal adenocarcinomas and matched liver metastases from the same patients (S. Taylor, personal communication). S100A4 expression was observed in the primary tumours and this was elevated in the subsequent metastases suggesting that a clonal selection of S100A4-positive primary tumour cells have metastasised.

Northern hybridisation experiments have shown that S100A4 mRNA was generally higher in cell lines derived from human malignant breast tumours relative to benign breast tumours or a normal human mammary cell line immortalised with the SV40 virus (Nikitenko *et al.*, 2000). Albertazzi *et al.* (1998) demonstrated, by similar means, that S100A4 mRNA was detected in 19 out of 34 IDC, and that 13 (68%) of these patients showed metastatic spread to the lymph nodes. Only 4 out of 15 patients with S100A4 negative IDC were found to have lymph node metastasis. Pedrocchi *et al.* (1994) observed high S100A4 in aggressive breast cancer cell lines, by Northern hybridisation, and in breast cancer biopsy samples, by Western blotting. Incidentally, these samples also demonstrated high levels of a marker of tumour cell invasion uPA (Pedrocchi *et al.*, 1994). However, these experiments do not distinguish between cell type specific expression of S100A4 which may be due to infiltrating lymphocytes or fibroblasts that are also high in S100A4 (Gibbs *et al.*, 1995; Takanaga *et al.*, 1994a, 1994b). I*n situ* hybridisation using an S100A4 oligonucleotide antisense probe demonstrated S100A4 mRNA was localised to both carcinoma cells and stroma cells of human breast lesions, however, S100A4 mRNA was present at higher levels in carcinomas relative to benign breast tumour specimens (Nikitenko *et al.*, 2000).

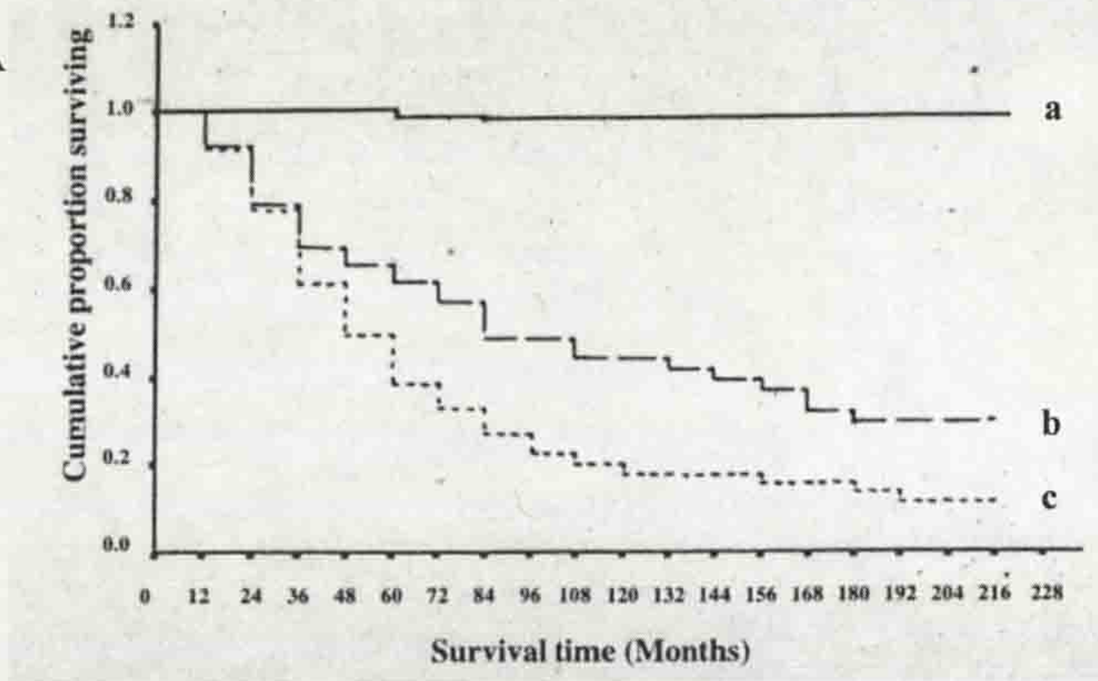
The most comprehensive study to date of S100A4 expression in human tumours investigated primary breast tumours from a group of 349 patients with operable breast cancer with a 19 year follow up period (Rudland *et al.*, 2000; Platt-Higgins *et al.*, 2000). The significance of immunocytochemically detectable S100A4 expression was compared with other clinically relevant prognostic markers in terms of patient survival: tumour size, grade, lymph node status and immunological detection of proteins involved in cellular proliferation (p53, c-*erb*B-2, c-*erb*B-3, estrogen receptor (ER), progesterone receptor and pS2), and cellular invasion (cathepsin D).

Positive histological staining of S100A4 in ≥5% of carcinoma cells correlated closely with patient demise, demonstrated clearly in figure 1.4a. In combination with other markers such as lymph node status and c-*erb*B-2 positivity (Figure 1.4), survival rates were associated most closely with S100A4 staining and S100A4 negative tumours indicated good outcome despite positivity for the other marker. Together S100A4 positivity and lymph node involvement or c-*erb*B-2 positivity was related to a worse outcome than S100A4 positivity alone. This data is significant because it supports the differences in tumour phenotype observed between MMTV-*neu* and MMTV-*neu/S100A4* mice. Although primary mammary gland tumours from both strains of transgenic mice exhibit high grade and a histologically aggressive appearance, a considerably higher metastatic capability in bitransgenic mice is clearly evident (Davies *et al.*, 1996). Combined elevated expression of both neu and S100A4 presents as a worse outcome for these mice relative to single *neu* transgenic mice, in the same way that S100A4 positive/c-*erb*B-2 positive tumours are indicative of a worse prognosis relative to S100A4 negative/c-*erb*B-2 positive tumours. Cumulative data from these studies revealed that, in this group of patients, S100A4 provided the most significant marker of patient mortality of all the prognostic markers used (Rudland *et al.*, 2000; Platt-Higgins *et al.*, 2000).

1.6 Differential Gene Expression analysis.

The discovery of cancer related genes is an important aspect in furthering our understanding of the molecular and biological aspects of tumour development and metastatic progression. Various experimental means have been applied to identify gene expression changes that are associated with these processes. The candidate gene approach is often useful in studying a particular gene or its protein that has

A

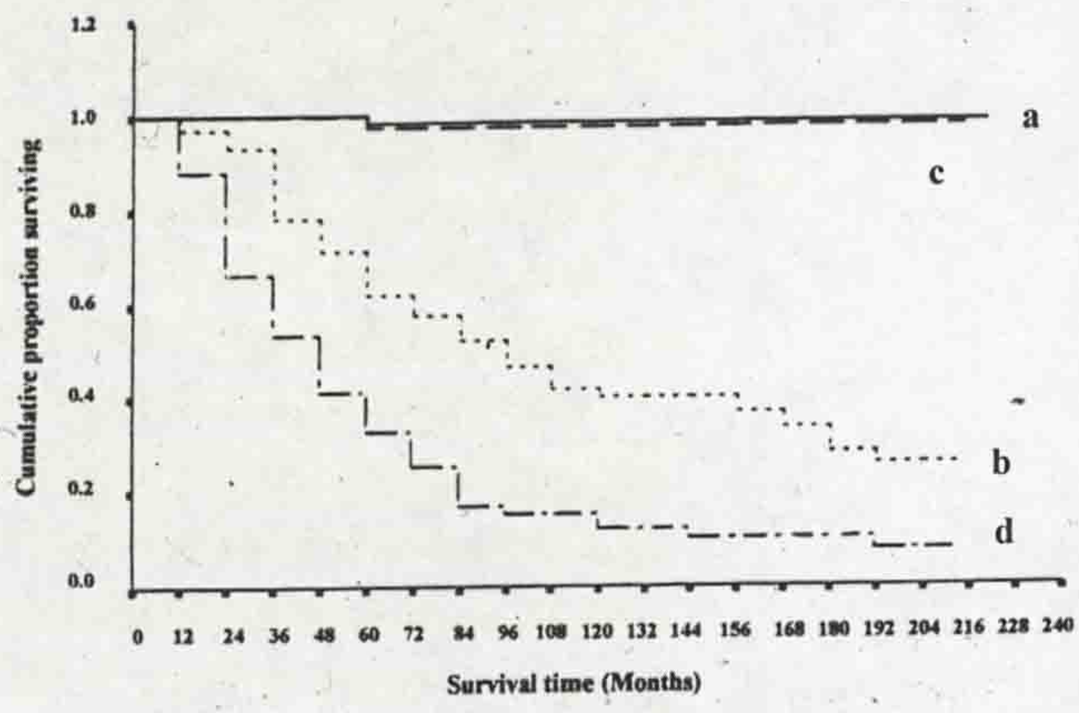


a = S100A4 negative carcinomas (<1% of carcinoma cells stain positive for S100A4)

b = S100A4 borderline carcinomas (1-5% of carcinoma cells staining positive for S100A4)

c = S100A4 positive carcinomas (>5% of carcinoma cells staining positive for S100A4)

B



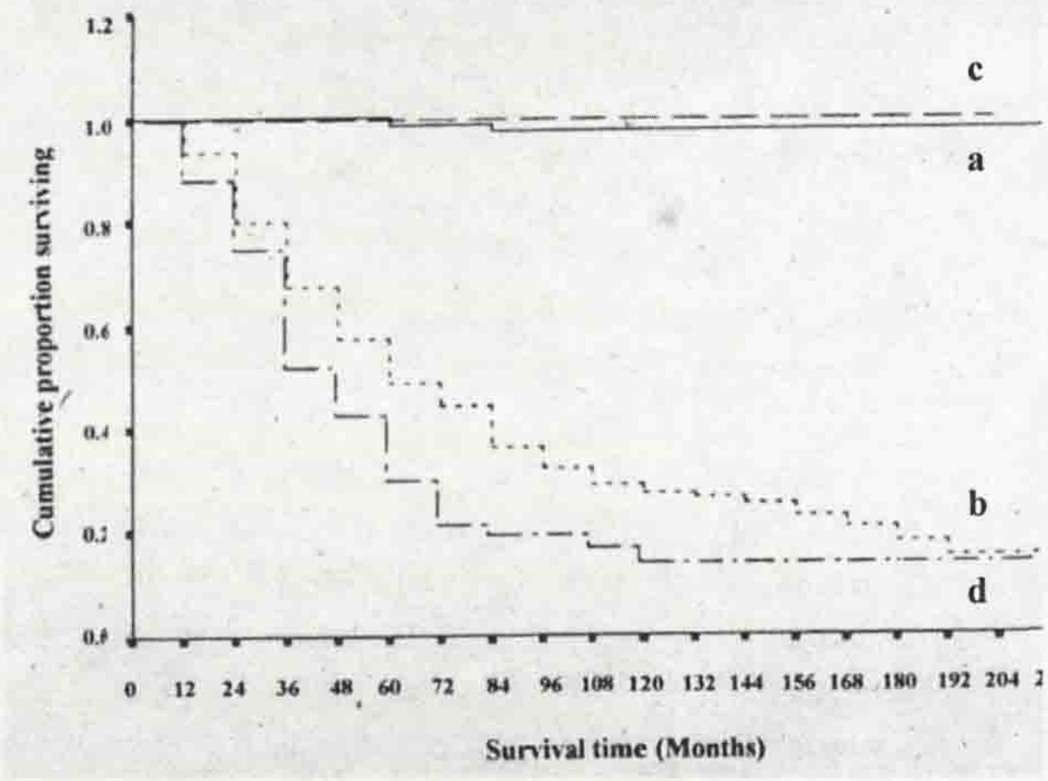
a = S100A4 negative/lymph node negative

b = S100A4 positive/lymph node negative

c = S100A4 negative/lymph node positive

d = S100A4 positive/lymph node positive

C



a = S100A4 negative/c-erbB-2 negative

b = S100A4 positive/c-erbB-2 negative

c = S100A4 negative/c-erbB-2 positive

d = S100A4 positive/c-erbB-2 positive

Figure 1.4 Association of levels of immunocytochemical staining for S100A4 with overall patient survival.

The cumulative proportion of patients surviving as a percentage of the total for each year after presentation with operable breast carcinoma. (A) demonstrates patient survival associated with S100A4 immunopositivity, (B) demonstrates patient survival associated with combined S100A4 immunopositivity and lymph node involvement and (C) demonstrates patient survival associated with combined S100A4 immunopositivity and c-erbB-2 immunopositivity. Figure (A) was taken from Rudland *et al.*, 2000; figures (B) and (C) were taken from Platt-Higgins *et al.*, 2000.

previously been implicated to be involved in cancer, for example mutation detection of p53 or expression monitoring of c-*erb*B-2, E-cadherin, vascular endothelial growth factor (VEGF) or cathepsin-D. Such molecules have provided useful diagnostic markers in breast and in other cancers and would therefore provide interesting targets for other studies. However such targeted experiments do not facilitate the discovery of new genes or previously identified genes that have not been associated with the disease but which may play important roles in determining tumour phenotype. To this end, recent molecular technological developments, alongside the ever increasing sequencing efforts of human and other mammalian genomes is providing a wealth of information which is expected to reveal many novel genes and potential targets for therapy and diagnosis for many diseases. The application of this emerging technology in cancer research is discussed below.

1.6.1 High throughput DNA sequencing.

Large scale sequencing of partial cDNA fragments (expressed sequence tags, ESTs) from cDNA libraries has been effectively applied to identify genes, including novel sequences that are expressed in experimental samples (Adams *et al.*, 1991). One problem with this form of expression profiling for identifying differential expression is the large number of ESTs that must be cloned, sequenced and identified by extensive sequence-alignment searching. This is due to the complexity of the mRNA profiles expressed in a cell, as was initially reported to exist within HeLa cells by Bishop *et al.* (1974) and in chicken oviduct and liver by Axel *et al.* (1976). An estimated 33,000 distinct mRNA species were transcribed in each HeLa cell to produce a total of 35,000 mRNA transcripts. The distribution of individual mRNAs determined that three abundance classes existed, 22% of the mRNA transcriptome (mRNA profile) comprised only 17 different mRNAs (high abundant sequences), 28% comprised 370 mRNAs (medium abundant) and 50% comprised approximately 33,000 mRNAs (low abundant). Comparable mRNA prevalence data was determined more recently following EST sequencing of HeLa cells before and after interferon-γ (IFN- γ) treatment (Wan *et al.*, 1996) and from a liver carcinoma cell line (Okubo *et al.*, 1992). Thus, even sequencing of thousands of ESTs is expected to identify predominantly highly redundant (high and medium abundant) sequences, and therefore only differential expression between these relatively few sequences is likely to be identified with reliability (Wan *et al.*, 1996).

Serial analysis of gene expression (SAGE; Velculescu *et al.*, 1995, 1997; Zhang *et al.*, 1997) has been developed and applied to circumvent exhaustive EST sequencing on the basis that a small 'tag' of 9 base pairs (bp) of mRNA sequence from a defined position, is sufficient to determine the identity of the message. Double stranded cDNA is digested with a frequently cutting restriction endonuclease and ligated with an adaptor containing a type IIS restriction enzyme recognition site. Further digestion of DNA with this specific enzyme, which cuts 12 bp downstream (3') of the recognition site, creates a small tag containing 3 bp of adaptor sequence and 9 bp of cDNA sequence. Concatamerisation of these small tags, and subsequent high throughput automated sequencing enables the identification and abundance determination of many thousands of different messages.

The Cancer Gene Anatomy Project (CGAP; Strausberg *et al.*, 1997) database of the National Cancer Institute (NCI) contains 1.5 million ESTs sequenced from normal, precancerous and tumour tissues derived from, for example, breast, ovary, lung, colon and prostate. A database of SAGE cDNA fragments (SAGEmap; Lal *et al.*, 1999) has also been generated at CGAP which contains over 3.5 million transcripts sequenced from cDNA libraries derived from numerous different normal and tumour tissues, such as colon, breast, lung and brain (refer to Velculescu *et al.*, 1999 for a list of tissues and tumours, the number of transcripts in the database and an example of data-mining this gene expression data). Public access to such databases of ESTs derived from cDNA libraries (http://cgap.gov or http://ncbi.nlm.nih.gov/SAGE respectively) for depositing, retrieving and analysing gene expression information enhances the ability of researchers to further characterise the gene expression profiles of these human tissues and tumours. The benefits of using this technology are that the data, or sequence prevalence, is stored electronically and can therefore be used to compare the mRNA abundance between multiple normal and tumour tissues that have been determined at independent times and by independent research groups. Digital Differential Display, a data mining tool for the CGAP database, was used for the discovery of genes differentially expressed in tumours from breast, colon, lung, ovary, prostate, and pancreas (Scheurle *et al.*, 2000). 200 known genes and 500 novel ESTs were shown to be differentially expressed by >10-fold, including ribosomal proteins, enzymes, cell surface receptors, secretory proteins and cell adhesion molecules.

1.6.2 Differential display (DD).

Parallel gene expression comparisons of multiple independent experimental samples can be achieved using differential display technology (Liang and Pardee, 1992, Liang *et al.*, 1992) and has subsequently become the basis of numerous techniques encompassing technical variations and improvements to enhance the reproducibility and success of the method (Matz and Lukyanov, 1998; Martin and Pardee, 1999). The general principle of DD is the reverse transcription and amplification of partial cDNA fragments using an oligonucleotide that anchors to the polyadenylated [poly(A)] tail plus 2 additional 3' bases of mRNA species and arbitrary primers that anneal to different regions of different mRNAs. Specific subsets of the mRNA population are amplified using different oligo(dT) primers that select mRNAs based on differences in the last two 3' bases and different sets of arbitrary primers. Polyacrylamide gel electrophoresis (PAGE) of amplified fragments displays an expression profile specific to each experimental sample. Different banding patterns between different samples constitute differences in the abundance of specific mRNA fragments that can be purified from the gel, sequence identified and further characterised to confirm differential expression. DD methods have suffered from technical difficulties which include high rate of false positives being identified, lack of reproducibility and the identification of multiple cDNA fragments per extracted DNA band (Debouck, 1995; Wan *et al.*, 1996).

Kirschmann *et al.*, (1999) have applied this technology to compare the expression profiles of three cell lines demonstrating different invasive and metastatic capabilities when injected into susceptible mice. MCF-7 cells, which are poorly invasive and non-metastatic, were compared with MoVi cells (MCF-7 cells transfected with the mouse vimentin gene), with an invasive yet poorly metastatic phenotype and MDA-MB-231 cells that exhibit an invasive and metastatic phenotype. Forty four out of 88 PCR products analysed were differentially expressed between these cell lines, as assessed by Northern analysis. The expression of 5 of these sequences were consistently associated with invasive and/or metastatic phenotype in multiple comparable breast carcinoma cell lines.

1.6.3 Subtractive hybridisation (SH).

Elimination or removal of sequences common to two experimental samples, and the subsequent selection and cloning of sequences unique to or preferentially

expressed in one of the samples is the basis for a wide variety of subtractive hybridisation techniques (Hedrick *et al.*, 1984). These types of experiments seek to overcome the problem of abundant sequences that are common to two cell populations and therefore to identify differentially expressed sequences of lower abundance. The sample containing target cDNAs of interest is termed the 'tester' cDNA and this is mixed with a molar excess of a control or 'driver' cDNA population. Single stranded cDNAs from both samples undergo liquid hybridisation, the resulting double stranded sequences representing common mRNAs that can be removed using, for example, streptavidin-biotin interactions. The remaining single stranded sequences represent differentially expressed mRNAs. Several rounds of subtraction can be performed to further enrich for differentially expressed sequences that can then be cloned to create subtracted cDNA libraries. Subtractive hybridisation can be applied in both orientations, so that control cDNA used in the first experiment can be used as tester cDNA in the second experiment. Thus, both up and down regulated genes can be identified between two experimental samples.

These types of methods have been utilised for the identification of many sequences demonstrating differential expression. For example, osteopontin was identified as a metastasis-related gene by its up-regulation in a metastatic rat mammary cell line (Ca2-5-LT1) relative to a benign parental cell line (Rama 37) from which it was derived (Oates *et al.*, 1996). Subtractive hybridisation between freshly isolated breast cancer cells and normal mammary tissue identified several genes, including a novel sequence termed breast cancer associated gene-1 (bcg-1), preferentially expressed in the carcinoma cells but not in normal tissue or in breast cancer cell lines (Kurt *et al.*, 2000).

A recently described subtractive hybridisation technology, termed Suppression Subtractive Hybridisation (SSH; figure 1.6; Diatchenko *et al.*, 1996; Gurskaya *et al.*, 1996), has been developed that supersedes previously described methods and has already been utilised for the identification of differentially regulated genes in numerous biological systems (Von Stein *et al.*, 1997; Jin *et al.*, 1997; Thompson and Weigal, 1998; Akopyants *et al.*, 1998; Zuber *et al.*, 2000; Huften *et al.*, 1999). SSH includes an equalisation step during hybridisation which normalises mRNA abundance and therefore enables the cloning of both high and rarely transcribed differentially expressed sequences (Gurskaya *et al.*, 1996). This enhances the identification of potentially important and novel genes that are often missed

during EST sequencing or standard SH without normalisation of abundance. An effect known as suppression-PCR (Siebert *et al.*, 1995; Chenchik *et al.*, 1996) is also employed to selectively enrich for tester specific sequences whilst suppressing the amplification of driver cDNA sequences, this negates the need to physically remove sequences common to both experimental samples. The benefits of SSH allow for the rapid cloning of differentially expressed cDNAs that can be of low abundance and of potentially novel sequence.

1.6.4 Differential screening of arrayed cDNA clones.

Hybridisation of complex cDNA probes, representative of the whole mRNA population, to cDNA targets immobilised to solid supports, such as nylon membrane (Von Stein *et al.*, 1997) or glass slides (Schena *et al.*, 1995), has provided the next generation of large scale cDNA expression screening experiments. The expression of many thousands of cDNAs can be screened in parallel experiments to determine the expression profile of these cDNAs in multiple experimental samples.

The power of this technology has been used in distinguishing different cell types of normal mammary gland tissue and breast carcinomas (Perou *et al.*, 1999, 2000) and in distinguishing between histopathologically indistinguishable lymphomas (Alizadeh *et al.*, 2000) on the basis of cell type specific gene expression profiles. This is possible, not by the identification of differences in expression of individual cDNAs but by hierarchical clusteral analysis (Eisen *et al.*, 1998), where groups of cDNAs are clustered together on the basis of co-ordinated expression patterns under different conditions or in different cell types. However, to generate sufficient data for these types of analysis the expression of many thousands of cDNAs in many experimental samples is required. For example, approximately 1.8 million gene expression measurements were made by Alizadeh *et al.*, (2000) during experiments investigating the molecular heterogeneity of diffuse large B-cell lymphomas (DLBCL). DLBCL exist as a clinically heterogeneous group of tumours, 60% of patients succumb to the disease whereas the remainder respond well to therapy. Gene expression screening of 17,856 cDNA clones per microarray in 94 normal and malignant lymphocyte samples demonstrated that subtypes of DLBCL existed exhibiting distinct gene expression patterns. The different subgroups, defined by gene expression profiling, demonstrated statistically significant differences in patient survival (Alizadeh *et al.*, 2000) suggesting that such large scale expression

screening can be utilised to distinguish between patients at greater risk of dying from this disease, when histological analysis of their tumours could not.

These types of experiments are not feasible for many laboratories, due to limited access to such volumes of clinical material and the costs and technical challenges involved in producing arrays with high levels of reproducibility. Nevertheless, screening arrays that have been purchased or produced in house provide means to assess the expression of specific cDNAs in different circumstances and therefore the identification of potentially important differentially expressed sequences.

The combined use of subtractive hybridisation or SAGE and array hybridisation expression screening is beginning to increase the usefulness of differential gene expression experiments. Subtracted cDNA libraries or selected, differentially expressed SAGE tags arrayed to a solid support can be screened, for example, with multiple tumours to identify cDNAs consistently differentially expressed in association with a particular phenotype. The nature of such experiments means that the expression pattern of increased numbers of cDNAs, including those of previously unknown sequence, can be analysed (relative to non-array methods of expression screening such as RT-PCR and Northern blotting), which in turn increases the numbers of differentially expressed sequences that can be identified.

The expression of subtracted cDNA clones, generated by SSH between estrogen receptor (ER) positive (MCF-7) and ER-negative (MDA-MB-231) breast cancer cell lines, have been investigated by cDNA microarrays for expression in the same two cell lines and other ER-positive (T47D) and ER-negative (HBL-100) cell lines (Yang *et al.*, 1999). At a differential expression ratio of 3-fold, 23% of subtracted cDNA clones were considered differentially expressed between the two cell lines being compared, 7% of cDNA clones demonstrated consistent differential expression patterns in the additional ER-positive versus ER-negative cell lines. Similarly, SSH between metastatic pancreatic adenocarcinoma cell line Bsp73-ASML and its non-metastatic counterpart Bsp73-1AS and high through put Reverse Northern cDNA expression screening of 5000 subtracted cDNA clones, identified 625 (12.5%) differentially regulated cDNAs (the differential expression ratio used to identify these sequences was not specified; Von Stein *et al.*, 1997). Also, SAGE profiles were generated for two breast tumour cell lines, 21PT and 21MT, and two normal mammary epithelial cell lines (Nacht *et al.*, 1999). A total of 228,652 tags

were sequenced and 539 tags were expressed at 5-fold higher levels in the cancer cell lines. A subset of 68 differentially expressed tags were arrayed and screened for expression in 17 breast tumour samples (7 primary breast tumours and 10 metastases) and four normal breast epithelial cell cultures. Overall, primary and metastatic breast tumours showed a similar expression profile for these sequences relative to normal; those that were over expressed in primary tumours were also over expressed in metastases, and similarly, those under expressed in primaries were often also under expressed in metastases (Nacht *et al.*, 1999).

1.7 Aims of Ph.D. project.

The primary aim of this Ph.D. project was to identify gene expression changes that may co-operate with elevated expression of S100A4 in inducing the metastatic phenotype that is observed in MMTV-*neu/S100A4* transgenic mice. The use of cDNA expression array hybridisation and subtractive hybridisation in combination with Reverse Northern hybridisation was applied to compare gene expression patterns of different tumours from the transgenic mice. Characterisation of these differentially regulated genes was also carried out to determine if the changes in expression are likely to play a functional role in metastatic progression or if their expression changes occur as a consequence of the changing phenotype.

1.8 Strategy for identifying S100A4-associated metastasis-related sequences.

Archival material from the two transgenic mouse lines, MMTV-*neu* and MMTV-*neu/S100A4*, was available for analysis. This was in the form of methacarn fixed, paraffin embedded tissue blocks representing the majority of tumours and normal tissues from all transgenic mice. These blocks can be utilised for histological characterisation of the tumours, immunohistochemical analysis of specific proteins, DNA extraction from microdissected tissue and possibly *in situ* hybridisation for the determination of cell-type specific expression analysis of specific mRNAs. During necropsy parts of many normal tissues and primary tumours and a limited number of metastases were also snap frozen in liquid nitrogen and have since been stored at –140°C. These tissues can be utilised for histological characterisation as suggested above and also for the extraction of DNA, RNA and protein from the whole tumour or from histologically defined tissue by microdissection of frozen tumour sections.

RNA extracted from these frozen tumours was used as the source of genetic material for analysing gene expression patterns in these tumours.

1.8.1 cDNA array hybridisation.

cDNA expression arrays were screened to determine the expression profiles of 588 arrayed cDNAs. Duplicate arrays were hybridised with probes derived from different tumours from the transgenic mice to identify gene sequences that were differentially regulated between these tumours. However, these experiments are limited to screening the expression of sequences that are present on the array, which is relatively few compared to the potential 100,000-120,000 genes contained in the mammalian genome and the possible 12,000-33,000 different transcripts expressed in a cell. Many of these cDNAs may not be expressed in these tumour cells, let alone be relevant to the metastatic progression of these tumours.

1.8.2 SMART PCR cDNA synthesis and Suppression Subtractive hybridisation (SSH).

SSH has been selected as the basis for identifying differentially expressed sequences between cells of progressively metastatic stages of breast cancer from the MMTV-*neu* and MMTV-*neu/S100A4* transgenic mice. This technique, as discussed previously, allows for the rapid cloning and identification of both rare and high abundant differentially expressed sequences that maybe important in determining cell phenotype. An initial amplification step can also be incorporated into the procedure which enables subtractive hybridisation to be performed from small amounts of tissue, such as that obtained from clinical biopsies or microdissected tissue, where the yield of RNA, total or poly(A)$^+$-containing RNA, is generally very small but of a defined histological nature. The pre-amplification step was incorporated into these experiments so that SSH can be performed between RNA derived from small amounts of histologically defined tumour tissue. These techniques contain some complex stages and so require some degree of explanation.

1.8.2.1 SMART PCR-generated cDNA.

Switch mechanism at the 5' end of RNA templates (SMART^{TM}, Clontech) was developed for the synthesis of cDNA from limiting amounts of template RNA, from as little as 25 ng poly(A)$^+$ RNA or from 50 ng total RNA which relates to

approximately 50,000 or 5,000 cells respectively (assuming 10 pg total RNA/cell and 5% of this is poly(A)$^+$ RNA). An overview of this method is presented in figure 1.5. First strand cDNA synthesis is initiated by a modified oligo(dT) sequence containing a PCR priming site sequence. The RNase H$^-$ mutant of Moloney murine leukaemia virus reverse transcriptase (MMLV-RT) is used to copy the mRNA sequence. Two intrinsic properties of MMLV-RT are then exploited to add the second PCR priming site to the 3' end of the cDNA molecule. Firstly, terminal transferase activity of MMLV adds 2-5 non-template nucleotides to the 3' end of the cDNA molecule; the preferential sequence of this extension includes three C residues. A second oligonucleotide included in the RT reaction contains a 3' string of G residues which anneals to the extended cDNA molecule; the 5' sequence of this oligonucleotide contains the same PCR primer sequence as found in the modified oligo(dT) sequence. The second intrinsic feature of MMLV RT is to switch templates and to continue copying, this occurs and the second oligonucleotide sequence is thus incorporated into the 3' end of the cDNA molecule (figure 1.5). A PCR primer specific to the two adapter sequences incorporated onto either end of each cDNA molecule is used to amplify the SMART cDNA.

1.8.2.2 SSH.

An overview of the SSH procedure is presented in figure 1.6. The procedure can be performed twice for each pair of experimental samples to identify sequences up and down regulated between the two mRNA populations. Amplified SMART cDNA from both tester and control samples are digested with *Rsa* I, a restriction enzyme with a 4 bp recognition site. This creates shorter blunt-ended cDNA fragments that have better hybridisation kinetics relative to full length cDNA transcripts. The tester cDNA is subdivided into two equal portions; each is ligated with a different double stranded DNA adaptor (adaptor 1 and 2R). Driver cDNA contains no adaptors. A 30 fold excess of driver is mixed, independently with each adaptor-ligated tester cDNA, double stranded tester and driver cDNA is denatured and single stranded cDNA molecules are allowed to re-anneal. Figure 1.6 demonstrates the types of molecules that form during this first hybridisation step. These include single stranded tester cDNA molecules (type a molecules, Figure 1.6), double stranded homoduplex tester molecules (type b molecules), tester-driver heteroduplexes (type c molecules), single stranded driver cDNA molecules and

double stranded homoduplex driver molecules (both type d molecules). Differentially expressed tester specific sequences are enriched because sequences common to both driver and tester form heterodimers. The two hybridisation reactions are combined with a fresh amount of excess denatured driver cDNA and a second hybridisation is performed to further enrich for differentially expressed sequences. New tester specific hybrids form which have different adaptor sequences on either end (type e molecules).

Two rounds of PCR amplify the tester-specific population, primary PCR utilises sequences on the outside of the adaptor sequences (black filled boxes, Figure 1.6), nested PCR follows using primers specific to internal adaptor sequences (white and shaded boxes, Figure 1.6). Single stranded molecules containing a single adaptor (type a molecules) fail to amplify since the adaptor sequence is the same, and not complementary, to the PCR primer sequence. Hybrids lacking adaptor sequences (type d molecules) fail to amplify, those with single adaptor sequences (type c molecules) are amplified but only linearly. Sequences with the same adaptor on both ends (type b molecules) experience a suppression-PCR effect (Siebert *et al.*, 1995; Chenchik *et al.*, 1996). Tester specific sequences, hybrids with different adaptor sequences on either end (type e molecules) undergo enrichment by exponential amplification. Subtracted cDNA libraries are created by cloning PCR products of this secondary PCR.

1.8.4 Characterisation of subtracted libraries and differentially expressed cDNA clones.

Characterisation of the subtracted libraries is useful to determine the efficiency of subtraction performed, the size and also the complexity of the subtracted cDNA libraries created. Extensive sequence characterisation of subtracted cDNA libraries, by automated DNA sequencing and the use of database-alignment search tools aids determination of these features of the libraries and provides information concerning the type of genes that are expressed in the tumours being compared. SSH subtracted libraries are often highly complex, meaning that there are a high number of different sequences present, this occurs due to the normalisation of abundance during hybridisation. High abundant sequences are therefore less prevalent than in a non-normalised cDNA library, unless they are also differentially expressed. The complexities of the two experimental samples being compared and

the number of differentially expressed sequences present between samples also governs library complexity. Expression patterns of the subtracted cDNA clones were also analysed by Reverse Northern hybridisation to confirm the identity of the truly differentially expressed sequences. This is necessary since complete suppression of non-differentially expressed sequences is unlikely to occur. In fact the number or proportion of differentially expressed sequences identified is dependent on the efficiency of SSH performed and on the number of messages that are differentially expressed between the experimental samples. The more closely related the two samples are, the fewer differentially expressed sequences are expected, resulting in a greater number of background clones being identified.

Clones possessing particularly interesting sequence identity and expression data were selected for further characterisation by Virtual Northern analysis, to confirm differential expression in the tumours used to perform subtractive hybridisation. The expression levels of these sequences in other tumours, normal mouse tissues and cell lines derived from the mouse tumours was also investigated, by Northern hybridisation, to determine if these demonstrate consistent expression correlating with S100A4 expression and therefore also with S100A4-related metastatic progression.

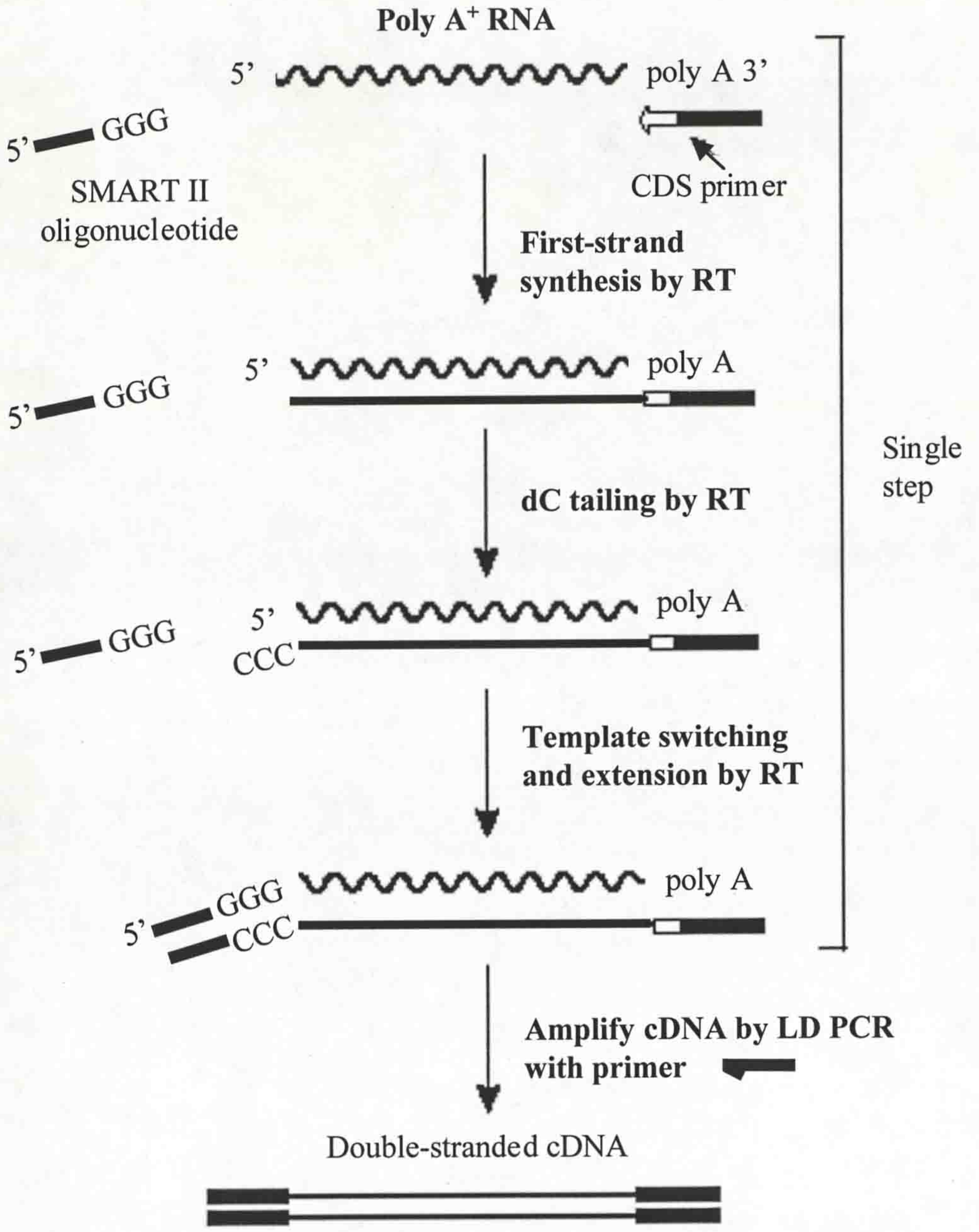


Figure 1.5 Overview of SMART PCR cDNA synthesis (Clontech).

SMART cDNA is synthesised using a combination of two oligonucleotides, the CDS primer and the SMART II oligonucleotide. Both contain a PCR priming site sequence that is subsequently used for long-distance PCR of all cDNA molecules. Figure taken from SMART PCR cDNA synthesis protocol handbook.

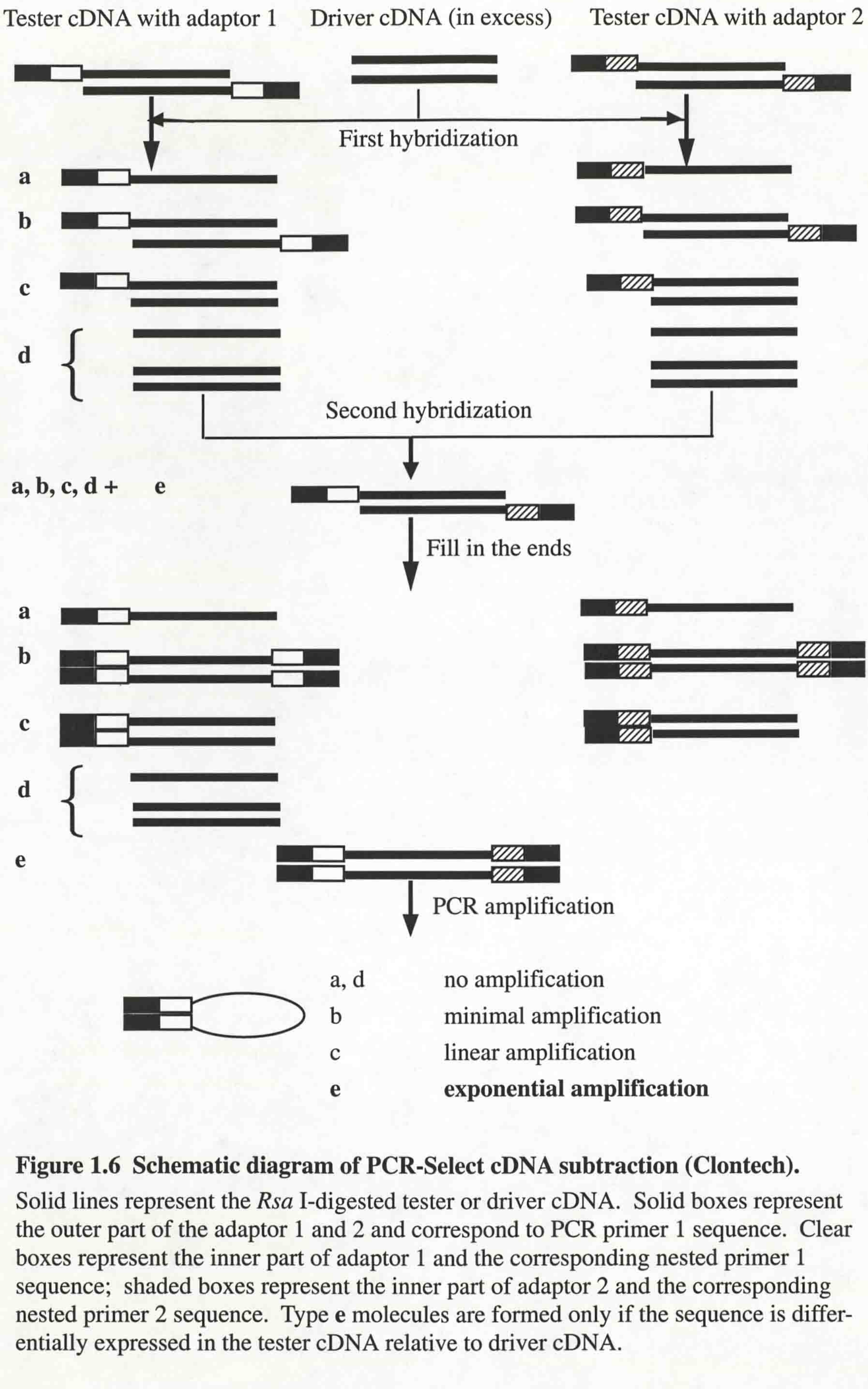


Figure 1.6 Schematic diagram of PCR-Select cDNA subtraction (Clontech).

Solid lines represent the *Rsa* I-digested tester or driver cDNA. Solid boxes represent the outer part of the adaptor 1 and 2 and correspond to PCR primer 1 sequence. Clear boxes represent the inner part of adaptor 1 and the corresponding nested primer 1 sequence; shaded boxes represent the inner part of adaptor 2 and the corresponding nested primer 2 sequence. Type **e** molecules are formed only if the sequence is differentially expressed in the tester cDNA relative to driver cDNA.

Chapter 2

Materials and Methods.

General molecular biology techniques.

2.1 Ethanol precipitation of DNA.

The solution containing the DNA to be precipitated was mixed with 2.5 volumes of absolute ethanol and 0.1 volumes of either 3 M sodium acetate or, for the precipitation of oligonucleotides, 10 M ammonium acetate and incubated at -20°C. Precipitated DNA was pelleted by centrifugation at 13,000 rpm in a microcentrifuge for 30 min. The supernatant was removed and the DNA pellet was washed in 70% (v/v) ethanol, re-centrifuged, air-dried and resuspended in an appropriate volume of sterile distilled water.

2.2 Spectrophotometric determination of concentrations of DNA, RNA and oligonucleotide solutions.

The yield of DNA, RNA or oligonucleotide was determined by measuring the absorbance at 235 nm, 260 nm, 280 nm and 320 nm in a spectrophotometer. The concentration of nucleic acid was given by the following formula:

$$\text{Concentration } (\mu g/\mu l) = \frac{(A_{260}-A_{320}) \times n \times \text{dilution factor}}{1000}$$

An optical density of 1 corresponds to approximately 50 µg/ml for DNA (n=50), 40 µg/ml for RNA (n=40) and approximately 27 µg/ml for oligonucleotides (n=27). The quality of the nucleic acid preparation was determined by measuring the A_{260}/A_{280} ratio and the A_{235}/A_{280} ratio that revealed if excessive protein or organic solvent, respectively, was carried through into the purified preparation. Absorbance at 320 nm was a measure of the baseline, and so was subtracted from the A_{260} reading to correct for variations in background absorbance.

2.3 Restriction enzyme digestion of DNA.

Cleavage of DNA was carried out using restriction enzymes, the appropriate buffer and reaction conditions as specified by the manufacturer. Typically, digests were

performed in a volume of 10 µl or 20 µl containing 1-10 µg DNA, 1 unit enzyme/µg DNA and by incubating at 37°C for 1-2 h. Heat inactivation of enzymes was carried out, when required and when specified by the manufacturers, by incubating at 65°C for 20 min.

2.4 Agarose gel electrophoresis of DNA.

DNA fragments were sized by electrophoresis through 0.8% (w/v) – 2.0% (w/v) agarose gels, using 1x TAE electrophoresis buffer [40 mM Tris-acetate, 1 mM EDTA (pH 7.6)]. DNA solution was added to one-sixth volume of DNA loading buffer [0.25% (w/v) bromophenol blue, 40% (w/v) glycerol]. Gels containing 0.5 µg/ml ethidium bromide were either visualised using an UV transilluminator (302 nm) or scanned at 610 nm in a FluorImager SI (Molecular Dynamics). Alternatively, gels were stained, post-electrophoresis, in 1:10,000 Sybr Gold solution (Molecular Probes) for 45 min and scanned at 530 nm. Scanned images were analysed using ImageQuaNT analysis software (Molecular Dynamics). To determine the size of DNA fragments analysed, comparison was made to DNA size markers including λ DNA digested with *Hind* III, 1 kb ladder and ϕx174 DNA digested with *Hae* III.

DNA purification techniques.

2.5 Purification of PCR amplified DNA.

Amplified DNA was purified from reaction constituents, in particular unincorporated primers, using a QIAquick PCR purification kit (Qiagen). The PCR reaction was mixed with 5 volumes of Buffer PB and applied to a QIAquick spin column. DNA was bound to the silica-gel matrix by centrifuging for 1 min at 13,000 rpm in a microcentrifuge. The DNA was washed by applying 750 µl Buffer PE to the column and repeating the centrifugation twice, and eluted by applying 30 µl sterile dH_2O to the centre of the membrane, incubating for 1 min and repeating the centrifugation. Purification of large numbers of PCR reactions was achieved as above yet using a QIAquick 96-PCR purification plate that was attached to a vacuum manifold to draw solutions through the column filters. Eluted DNA was quantified, by eye, by electrophoresis alongside 1 µg and 0.5 µg samples of λ DNA digested with *Hind* III and subsequent comparison to the intensity of staining of DNA fragments of known concentration.

2.6 Purification of DNA from agarose gels.

DNA fragments requiring purification by gel extraction were separated by electrophoresis on a 1x TAE 0.8-1.5% (w/v) normal or low-melting-point agarose gels (Seakem). Limited staining of DNA was either by including 0.5 μg/ml ethidium bromide in the 1x TAE electrophoresis buffer or by post-staining the gel in 1xTAE buffer containing 0.5 μg/ml ethidium bromide until DNA fragments were visible (5-10 min). Relevant DNA bands were excised from the gel and purified by one of the two following procedures. On each occasion, eluted DNA was quantified by electrophoresis alongside 1 μg and 0.5 μg λ DNA digested with *Hind* III.

2.7 QIAquick Gel Extraction (Qiagen).

The excised gel slice was mixed with 3 volumes Buffer QG (100 μl buffer/100 μ g of gel) and incubated at 50°C for 10 min with occasional vortex mixing. The melted agarose and DNA solution was applied to a QIAquick spin column and centrifuged for 1 min at 13,000 rpm in a microcentrifuge. DNA bound to the membrane was washed and eluted from the column as described above (2.5).

2.8 Geneclean Gel Extraction (Bio 101).

Alternatively DNA was extracted from agarose gel slices using a Geneclean II kit (Bio 101). The DNA was mixed with three volumes of NaI, and incubated at 55°C for 5 min to melt the agarose gel. 5 μl Glassmilk was added to DNA solution (unless >5 μg DNA was present, in which case 1 μl Glassmilk/μg DNA was added), vortex mixed and incubated at room temperature for 5 min. DNA bound to the Glassmilk was pelleted by centrifugation at 13,000 rpm for 30 sec in a microcentrifuge. The pellet was washed three times in 500 μl ice cold New Wash solution, for 10 min each with DNA being pelleted by centrifugation at 13,000 rpm for 15 sec each time in a microcentrifuge. DNA was eluted from Glassmilk by resuspending the pellet in 20 μl sterile H_2O, incubating at 55°C for 2-3 min and centrifuging for 1 min at 13,000 rpm in a microcentrifuge. Elution was repeated with another 20 μl and the eluates were combined.

2.9 Isolation of DNA, RNA and protein from frozen tumour specimens.

This method of RNA isolation was based on the protocol developed by Chirgwin *et al.* (1979) and Anandappa *et al.* (1994). The piece of frozen tissue (typically ≤ 1 cm^3) was taken from liquid nitrogen and placed immediately into 3.5 ml guanidine solution [0.5% (w/v) sodium lauryl sulphate, 4 M guanidine thiocyanate, 50 mM Tris-Cl (pH 7.5), 25 mM EDTA (pH 8.0), 10% (v/v) β-mercaptoethanol] in a 30 ml polypropylene SS34 centrifuge tube. The frozen tissue was homogenised using a Polytron PT 1200 homogeniser, and centrifuged at 8,000 rpm for 10 min at 4°C in an SS34 rotor. The supernatant was applied to the top of a 1 ml caesium chloride cushion (5.7 M caesium chloride, 0.1 M EDTA) in a 5 ml ultracentrifuge tube and centrifuged at 40,000 rpm for a minimum of 20 h at 20°C in an AH650 rotor of a Sorvall Combi Plus ultracentrifuge. The top layer, typically 2.5-3 ml of protein, was removed and stored at –80°C for subsequent use. The remaining liquid, containing DNA, was also removed and stored at –80°C. The centrifuge tubes were drained and the RNA pellet at the bottom of the tube was resuspended in 180-360 µl 0.1% (w/v) SDS. The RNA was precipitated by adding 0.1 volumes of 2 M NaCl and 2.5 volumes of absolute ethanol, and the mixture was stored at –20°C overnight. RNA was recovered by centrifugation at 13,000 rpm for 10 min at 4°C in a microcentrifuge, the supernatant was removed, the RNA pellets were freeze-dried under vacuum and resuspended in an appropriate volume of sterile dH_2O. The yield of RNA was determined using a spectrophotometer (2.2).

2.10 Propagation of bacteria (and bacteria containing plasmids).

E.coli XL1-blue, *E. coli* JA221 (a derivative of *E. coli* HB101) and *E.coli* TOP10F' (Invitrogen) bacteria were cultured in LB-broth [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, pH 7.0], or on LB-agar plates [LB-broth containing 1.5% (w/v) bacto-agar] supplemented with either 50 µg/ml ampicillin or 50 µg/ml tetracycline (if bacteria contained a plasmid with the ampicillin or tetracycline resistance genes), at 37°C in a shaking incubator. Stocks of bacteria containing plasmid were made by mixing 850 µl bacterial culture with 150 µl sterilised glycerol (approximately 15% (v/v)), snap freezing in liquid nitrogen and storing at –80°C.

Genotype of *E. coli* strains used:

E.coli XL1-blue: *end*A1 *gyr*A96 *hsd*R17 *lac rec*A1 *rel*A1 *sup*E44 *thi*-1 F'[*pro*AB, *lac*I^qZΔM15, Tn10]

E. coli HB101: *ara*14 *gal*K2 *hsd*S20 *lac*Y1 *leu*B6 *mtl*-1 *pro*A2 *rec*A13 *rps*L20 *sup*E44 9 14 *thi*-1 *xyl*-5 Δ(*mcr*C-*mrr*) F^-

E.coli TOP10F': F'{*lac*I^qTn10(Tet^R)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *ga*/U *ga*/K *rps*L *end*A1 *nup*G

Isolation of plasmid DNA from bacteria.

Recombinant plasmid DNA containing a DNA sequence of interest was isolated from bacterial culture using the Qiagen miniprep or midiprep methods.

2.11 Qiagen miniprep.

A single colony containing the required plasmid was used to inoculate 5 ml LB-broth (2.10). The overnight culture was centrifuged at 6,000 rpm in a Sorval SS-34 rotor for 10 min at 4°C. The bacterial cell pellet was resuspended in 250 µl buffer P1 (50 mM Tris-Cl (pH 8.0), 10 mM EDTA, 100 µg/ml RNase A, 4°C) and mixed with 250 µl buffer P2 (200 mM NaOH, 1% (v/v) SDS), by inverting tubes 4-6 times. 350 µl buffer N3 was added and the bacterial suspension was mixed in the same way before centrifuging at 13,000 rpm for 10 min in a microcentrifuge. The supernatant was applied to a QIAprep spin column and DNA was bound to the silica matrix by centrifuging the column at 13,000 rpm for 1 min in a microcentrifuge. DNA was washed, by applying 750 µl buffer PE to the column and repeating the centrifuge step, and eluted by applying 50 µl sterile dH_2O to the centre of the column, incubating for 1 min before re-centrifuging at 13,000 rpm for 1 min in a microcentrifuge.

2.12 Qiagen midiprep.

A single colony containing the required plasmid was used to inoculate 100 ml LB-broth (2.10). The overnight culture was centrifuged at 6,000 rpm in a Sorval SS-34 rotor for 10 min at 4°C. The bacterial cell pellet was resuspended in 4 ml buffer P1 and mixed with 4 ml buffer P2 by inverting the centrifuge tube several times and incubating for 5 min. Four ml, ice-cold buffer P3 [3.0M potassium acetate (pH 7.5)]

was added and the bacterial suspension was mixed by inverting the centrifuge tube several times and incubated on ice for 15 min. The mixture was centrifuged at 13,000 rpm for 30 min at 4°C in an Sorvall SS-34 rotor and the clear supernatant was immediately removed and applied to a Qiagen-100 column pre-equilibrated with 4 ml buffer QBT [750 mM NaCl, 50 mM MOPS (pH 7.0), 15% (v/v) isopropanol, 0.15% (v/v) Triton X-100]. The column was washed with 20 ml buffer QC [1.0 M NaCl, 50 mM MOPS (pH 7.0), 15% (v/v) isopropanol] and the DNA was eluted in 5 ml buffer QF [1.25 M NaCl, 50 mM Tris-Cl (pH 8.5), 15% (v/v) isopropanol]. Eluted DNA was precipitated by adding 3.5 ml isopropanol and centrifuging at 11,000 rpm for 30 min at 4°C in a Sorvall SS-34 rotor. The supernatant was removed and the plasmid DNA pellet was washed in 2 ml 70% ethanol and recentrifuged for 15 min. The pellet was dried, resuspended in an appropriate volume of sterilised water and quantified (2.2).

Radioactive labelling of DNA.

DNA fragments to be labelled were purified from PCR (2.5) of from low-melting-point agarose gels (2.6). Radioactive labelling of double-stranded DNA was achieved using the random primed DNA labelling kit (Boehringer Mannheim) or the Strip-EZ PCR™ probe synthesis kit (Ambion).

2.13 Random Primed DNA labelling (Boehringer Mannheim).

Approximately 25 ng DNA in 12 µl volume, denatured by boiling for 10 min and immediate quenching on ice, was labelled by mixing with 1 µl of each 0.5 mM nucleotide solution (dGTP, dTTP and either dATP or dCTP), 2 µl reaction solution (containing hexanucleotide mixture and 10x reaction buffer), 20 µCi [α^{32}P] dATP or [α^{32}P] dCTP (3000Ci/mmol) and 1 µl Klenow enzyme (2 units). The reaction mixture was incubated at 37°C for 1 h.

2.14 Strip-EZ™ PCR DNA labelling (Ambion).

Strip-EZ technology was used to label probes for Northern hybridisation so that complete removal of probe from hybridised membranes could be obtained with less harsh stripping conditions than standard stripping methods. This was carried out to increase the re-use of Northern blots and therefore to improve the sensitivity of

repeated probing experiments. Asymmetric PCR was used for this labelling to increase sensitivity of mRNA detection by using less unlabelled double stranded template DNA and creating a bias in the amplification so that the antisense strand is preferentially amplified and labelled. Thus, decreasing the amount of competing, unlabelled DNA in the hybridisation and consequently increasing the specificity of the resulting probe.

Approximately 0.1 ng purified DNA was mixed with 1x PCR Buffer [10 mM Tris-HCl, (pH 8.3), 50 mM KCl; Perkin Elmer], 1x dNTP mix [-dATP/modified dCTP; Ambion], 5 mM $MgCl_2$, 1.0 μM of antisense strand PCR primer, 0.01 μM of sense strand PCR primer, 20 μCi [α^{32}P] dATP (3000 Ci/mmol) and 1 unit of AmpliTaq Gold DNA Polymerase (Perkin Elmer). The reaction was incubated at 94°C for 10 min with 30 cycles of 95°C for 30 sec, 55°C for 20 sec and 72°C for 60 sec with a final extension of 72°C for 5 min.

2.15 TCA precipitation of radioactively labelled probes.

In order to monitor the incorporation of radioactivity, 1 μl of the labelling reaction mixture was diluted in 99 μl 1x TEN buffer [0.1 M NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)]. Two-microlitre aliquots of this diluted mixture were spotted onto each of six 1 cm x 2 cm flags of Whatman No. 1 paper and air-dried. Three of the flags were washed twice in 100 ml 5% (w/v) TCA/1% (w/v) sodium pyrophosphate, twice in 100 ml ethanol and twice in 100 ml diethyl ether and then air dried. All six flags were counted in a scintillation counter for 1 min. The specific activity was calculated as per manufacturers instructions (Boehringer Mannheim random primed DNA labelling kit specifications sheet):

$$\% \text{ incorporation of radioactivity} = \frac{\text{total cpm of the three washed flags x100}}{\text{total cpm of the three unwashed flags}}$$

$$\text{Amount of incorporated} = \mu\text{Ci dCTP or dATP used} \times 2.2\times10^{4} \times \% \text{ incorporation radioactivity } (\mu\text{Ci})$$

$$\text{Amount of newly synthesised DNA (ng)} = \frac{\mu\text{Ci dCTP or dATP used x 13.2 x \% incorporation}}{\text{specific activity of dCTP or dATP (3000Ci/mmol)}}$$

$$\text{Specific activity of probe } (\mu\text{Ci}/\mu\text{g}) = \frac{\text{amount of incorporated radioactivity} \times 10^{3}}{\text{amount of DNA probe used in the reaction (ng)} + \text{amount of newly synthesised DNA (ng)}}$$

2.16 Purification of radiolabelled probes using Sephadex G50 spun columns.

Spun columns were prepared by applying sterile silica wool to the bottom of 1 ml syringe barrel and adding Sephadex G50 saturated 1x TEN buffer. Columns were centrifuged at 1000 rpm for 4 min in a MSE Centaur 2 centrifuge with a swing out rotor. Unincorporated-labelled nucleotides were removed from synthesised labelled probes by applying the probe to the spun column and centrifuging for 4 min at 1000 rpm.

2.17 Stripping hybridised membranes.

Radioactively labelled probe was removed from membranes using one of two methods. If Strip-EZ™ technology was utilised, membranes were incubated in 1x probe degradation dilution buffer (Ambion) and 1x DNA probe degradation buffer (Ambion) in 10 ml/membrane for 10 min at 68°C in a rotating hybridisation oven. Membranes were then incubated in 1x membrane reconstitution buffer and 0.1% (w/v) SDS also for 10 min at 68°C. Alternatively, stripping was attempted by boiling 0.5-0.1% (w/v) SDS and incubating with membranes in a shaking water bath at 80°C for 30 min. Membranes were sealed in Saranwrap™ and either exposed to autoradiographic film, to monitor success of stripping, or stored at 4°C.

Manipulation of frozen tumour specimens.

2.18 Cutting frozen tissue sections for RNA extraction.

Frozen histological sections were cut on a microtome (-20°C) at a thickness of 8 μm; groups of 2-4 sections were collected into 0.5 μl microcentrifuge tubes that were pre-cooled in liquid nitrogen. The sections were immediately snap frozen in liquid nitrogen and stored at -80°C. Intervening sections, every 3rd-5th section, were mounted onto microscope slides (stored at room temperature) for haematoxylin and eosin staining.

2.19 Haematoxylin and Eosin (H&E) staining of frozen tissue sections.

Slide-mounted frozen sections were fixed in 95% ethanol, pre-cooled to -20°C, for 2-5 min and allowed to air dry at room temperature. The sections were placed in Harris Haematoxylin for 1 min, washed in running tap water for 30 sec, dipped in acid-alcohol [1% (v/v) concentrated HCl in 70% (v/v) ethanol] 4 times and washed again in running tap water for 30 sec. Sections were stained for 1-5 sec in Scott's tap water [0.35% (w/v) sodium bicarbonate and 2% (w/v) magnesium sulphate in tap water] and washed for 30 sec in running tap water. Finally sections were stained in Eosin [1% (w/v) Eosin yellow shade in tap water] for 5 sec and washed in running tap water. The sections were dehydrated by incubating twice in absolute alcohol and twice in xylene for 2-5 min each. The slides were mounted in DPX mounting medium.

2.20 Immunohistochemistry of frozen sections.

Histological frozen sections for immunohistochemistry on plain glass microscope slides were allowed to air dry for 2 min at room temperature and then fixed in cold (4°C) acetone for 5-10 min. The sections were washed in running tap water and incubated in 3% (v/v) H_2O_2 in methanol for 20 min to destroy endogenous peroxidase activity in the tissue that may interfere with the subsequent detection process. The sections were washed again in running tap water and then rinsed in TBS buffer containing 0.1% (w/v) BSA (pH 7.6). The primary antiserum (rabbit anti-human c-*erb*B-2 oncoprotein, Dako) was applied to the sections at a dilution of 1:500 for a 1 h incubation at room temperature. Sections were washed three times for 2 min each in TBS and then treated with biotinylated swine anti-rabbit immunoglobulin (Dako) at a dilution of 1:300 for 45 min at room temperature. Sections were washed

three times for 2 min each in TBS and incubated, for 30 min at room temperature, in a complex of streptavidin-biotin-horseradish peroxidase (StreptABComplex/HRP, Dako). This was prepared 30 min before it was required in order to allow for the complex to form, and according to the manufacturer's instructions. The sections were washed three times for 2 min each in TBS and the bound antibody was visualised by incubating in 0.05% (w/v) diaminobenzidine, 0.1% (v/v) H_2O_2 in sterile distilled water for 5 min at room temperature. The sections were washed in running tap water for 30 sec and the cell nuclei counterstained in Harris Haematoxylin for 5-10 sec, the sections were again washed in running tap water and placed in Scott's tap water to blue the sections. The sections were washed in running tap water for 30 sec and dehydrated through 70% (v/v) alcohol, twice in absolute alcohol and three times in xylene, each incubation for 10 min. The sections were finally mounted in DPX mounting medium. Control staining was performed in which sections were incubated in TBS instead of the primary antibody, all the other conditions were identical to the normal staining procedure.

2.21 Microdissection (Luqmani *et al.*, 1992, 1994; Gong et al., 1992).

Histological frozen sections on plain glass microscope slides to be used for immediate microdissection were maintained at -20°C in the microtome, otherwise sections could be stored in a sealed box containing dessicant at -70°C. A single section was stained with haematoxylin and eosin (2.18) to identify areas for microdissection on unstained parallel sections. Each unstained section was placed under a stereodissecting microscope (Olympus SZ11) to air dry for 30 sec. A microdissection micromanipulator, with a disposable needle, was used to remove a small and defined area of the section. The needle was used to pick up the dislodged tissue and this was collected in 20 µl of DEPC-treated water containing 1 unit of PRIME RNase Inhibitor (5' Prime-3' Prime). The samples were briefly vortexed to release the mRNA, centrifuged in a microcentrifuge, snap-frozen in liquid nitrogen and stored at -80°C. Sections were stained after microdissection with either H&E (2.18) or with an antibody for the c-*erb*B-2 oncoprotein (2.19) to confirm the histology of the removed areas.

2.22 Poly(A)$^+$ RNA extraction.

Poly(A)$^+$ RNA was extracted from frozen tissue sections using the Dynabeads mRNA DIRECT kit (Dynal). Dynabeads-oligo $(dT)_{25}$ were prepared by resuspending the stock suspension of beads and transferring 10 µl (50 mg of beads)/extraction to a 1.5 ml microcentrifuge tube placed in a Dynal magnetic particle concentrator (MPC). After 30 sec the supernatant was removed and the beads resuspended in 20 µl lysis/binding buffer [100 mM Tris-HCl (pH 8.0), 500 mM LiCl, 10 mM EDTA (pH 8.0), 1% (w/v) lithium dodecyl sulphate (LiDS), 5 nM DTT]. Tissue sections were lysed by adding 300 µl lysis/binding buffer and repeated pipetting up and down with a P200 pipette tip. The tissue lysate and 20 µl conditioned beads were mixed by rotating on a roller for 2-5 min at room temperature. The beads-mRNA complex was purified from tissue extract by placing the vial in the Dynal MPC for 2 min and removal of the supernatant. The beads were resuspended and washed twice in 600 µl washing buffer containing LiDS [10 mM Tris-HCl (pH 8.0), 0.15 M LiCl, 1 mM EDTA, 0.1% (w/v) LiDS], once in 300 µl washing buffer [10 mM Tris-HCl (pH 8.0), 0.15 M LiCl, 1 mM EDTA] and twice in 300 µl 1x First Strand cDNA synthesis buffer [50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$; Gibco BRL]. Poly(A)$^+$ RNA was eluted by resuspending the beads in 10 µl DEPC-treated water, incubating at 65°C for 2 min, placing the vial in the Dynal MPC for 30 sec and transferring the supernatant containing RNA to a new tube. Poly(A)$^+$ mRNA was snap frozen in liquid nitrogen and stored at -80°C.

2.23 Reverse Transcription-Polymerase Chain Reaction (RT-PCR).

An RT-PCR assay was been developed to assess the integrity of RNA extracted from stored, frozen tumour specimens. Hypoxanthine phosphoribosyl transferase (HPRT) PCR primers were designed to amplify a region of the cDNA that spans multiple intron-exon boundaries at the 5' end of the mouse gene.

cDNA was synthesised by incubating 1-3 µl poly(A)$^+$ RNA with 0.5 µg oligo$(dT)_{16}$, 0.2 mM dNTP and 0.2-0.4 units PRIME RNase inhibitor in a total volume of 5 µl at 70°C for 10 min. A second premix was added to give a total reaction volume of 10 µl. This contained 1x First Strand Buffer [50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3

mM $MgCl_2$], 10 mM DTT and 200 units of Superscript–II™ Reverse Transcriptase (Gibco BRL). Reverse transcription was achieved by incubating the reaction at 37°C for 1 h, this was terminated by incubating at 70°C for 10 min.

PCR reactions using HPRT upper and lower primers (table 2.1) were performed in a total volume of 20 µl containing 2 µl of template cDNA, 1x PCR Buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl], 0.2 mM dNTPs, 5 mM $MgCl_2$, 0.5 µM of each primer and 1 unit of AmpliTaq Gold DNA Polymerase (Perkin Elmer). The reaction was incubated at 94°C for 10 min with 40 cycles of 95°C for 30 sec, 54°C for 45 sec and 72°C for 30 sec with a final extension of 72°C for 5 min. Five microlitre aliquots were analysed by electrophoresis on 1x TAE 1.2% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide.

Oligonucleotide	**Sequence (5'-3')**
HPRT upper PCR primer	TTCCTCCTCAGACCGCTTTT
HPRT lower PCR primer	ACTTTTATGTCCCCCGTTGA
Oligo$(dT)_{16}$	TTTTTTTTTTTTTTTT
SMART II oligonucleotide	AAGCAGTGGTAACAACGCAGA**GTAC**GCGG
cDNA synthesis (CDS) primer	AAGCAGTGGTAACAACGCAGAGTAC$T_{(30)}N_{-1}N$
SMART PCR primer	AAGCAGTGGTAACAACGCAGAGT
Adaptor 1	CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAG**GT** GGCCCGTC**CA**
Adaptor 2R	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAG**GT** GCCGGCTC**CA**
PCR primer 1	CTAATACGACTCACTATAGGGC
Nested PCR primer 1	TCGAGCGGCCGCCCGGGCAG**GT**
Nested PCR primer 2R	AGCGTGGTCGCGGCCGAG**GT**
G3PDH 5' PCR primer	ACCACAGTCCATGCCATCAC
G3PDH 3' PCR primer	TCCACCACCCTGTTGCTGTA

Table 2.1 Oligonucleotides and adapters.

Rsa I restriction endonuclease recognition site used for the digestion of SMART PCR-generated cDNA and subsequently the removal of adapter sequences prior to subtractive hybridisation is shown in bold. Both adaptors contain a region of double stranded DNA towards the 3' end of the top strand and half an *Rsa* I site (bold). PCR primer 1 is identical to the first half of both adaptors; the two nested PCR primers are identical to the second half of the respective adaptor.

SMART PCR cDNA synthesis & PCR-Select cDNA subtraction.

Subtractive hybridisation was performed between poly(A)$^+$ RNA derived from mammary gland tumours from an MMTV-*neu* mouse and an MMTV-*neu/S100A4* mouse. Poly(A)$^+$ RNA derived from the same MMTV- *neu/S100A4* mammary gland tumour and a metastasis, from the same mouse, was also used to perform a second subtractive hybridisation experiment. A pre-amplification step was incorporated to enable SSH to be performed using RNA extracted from small amounts of tissue. All steps required for the preparation of SMART cDNA for subtractive hybridisation were performed for each of the poly(A)$^+$ RNA samples. Subtractive hybridisation was then performed in both forward and reverse orientations for both experiments. Clontech Laboratories supplied all reagents used for SMART PCR cDNA synthesis and PCR-Select cDNA subtractive hybridisation, unless otherwise stated. Procedures were as directed in the manuals supplied with the kits, also unless otherwise stated.

2.24 SMART cDNA synthesis.

First strand cDNA was synthesised by incubating 3-4 µl poly(A)$^+$ RNA, 1 µM cDNA synthesis (CDS) primer and 1 µM SMART II oligonucleotide (table 2.1), in a total volume of 5 µl, at 70°C for 2 min. At room temperature, 2 µl 5x first strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 30 mM $MgCl_2$], 1 µl 20 mM DTT and 200 units of Superscript-II^{TM} Reverse Transcriptase (Gibco BRL) were added. A drop of mineral oil was added and the reaction was incubated at 42°C for 1 h. The first strand synthesis reaction was terminated by adding 90 µl TE buffer [10 mM Tris-HCl (pH7.6), 1 mM EDTA] and heating to 72°C for 7 min. SMART cDNA was stored at -20°C.

2.25 Amplification of SMART cDNA.

Amplification of SMART cDNA was optimised to avoid over-amplification, and therefore possible misrepresentation of cDNAs in the amplified cDNA population, by setting up triplicate reactions, each containing 10 µl diluted cDNA, 1x Advantage 2 PCR buffer, 0.2 mM dNTP, 0.2 µM PCR primer (table 2.1) and 1x Advantage 2 Polymerase mix in a total volume of 100 µl. All three reactions were amplified by incubating at 95°C for 1 min with 15 cycles of 95°C for 15 sec, 65°C for 30 sec and

68°C for 6 min. At this point two reactions were removed and stored at 4°C, 15 µl of the third reaction was removed and the remaining 85 µl was subjected to three additional cycles. Again 15 µl was removed and the remaining 70 µl was subjected to three further cycles. This was repeated to obtain 15 µl aliquots after 15, 18, 21, and 24 cycles. 5 µl of each aliquot was analysed by electrophoresis on a 1x TAE 1.2 % (w/v) agarose gel containing 0.5 µg/ml ethidium bromide.

The optimum number of cycles of amplification was determined to be one less than the number required to reach an amplification plateau, as estimated by agarose gel analysis (see above). Typically 17-20 cycles were required for the 3 tumour mRNA samples. The two reactions previously removed and stored at 4°C were subjected to the additional number of cycles required; 5 µl aliquots were analysed by electrophoresis as before.

2.26 Purification of SMART PCR-generated cDNA.

The two amplification reactions, for each tumour sample, were pooled together and purified by vortex mixing for 1 min with an equal volume (190 µl) of phenol/chloroform/isoamyl alcohol (25:24:1, Sigma) and centrifuging for 10 min at 13,000 rpm in a microcentrifuge. The aqueous (upper) phase, containing DNA, was removed to a clean 1.5 ml microcentrifuge tube and further purified by adding 700 µl n-butanol (Severn Biotech Ltd.), vortex mixing for 1 min and centrifuging for 1 min at 13,000 rpm in a microcentrifuge. The upper (butanol, organic) phase was discarded leaving the lower phase of a volume between 40-70 µl. Butanol-extracted DNA was further purified using a Chroma Spin-1000 column. Columns were prepared by inverting several times to resuspend the gel matrix, eluting column buffer and equilibrated by adding 1.5 ml 1x TNE buffer [10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.1 mM EDTA] and allowing buffer to drain through the column. Sample DNA was applied to the column followed by 25 µl 1x TNE buffer, and once all the buffer had drained through the column a further 150 µl 1x TNE buffer was added and collected. 320 µl of 1x TNE buffer was added and this was collected as the purified cDNA fraction, 75 µl of 1x TNE buffer was added and collected as a second purified cDNA fraction. Aliquots (10 µl) of both purified fractions were

analysed by electrophoresis to monitor yield of SMART amplified cDNA.

2.27 *Rsa* I digestion of purified SMART PCR-generated cDNA.

SMART amplified cDNA was digested to generate shorter, blunt-ended double stranded cDNA fragments. Purified cDNA was mixed with 36 µl 10x *Rsa* I restriction enzyme buffer [100 mM bis Tris propane-HCl (pH 7.0), 100 mM $MgCl_2$, 1 mM DTT] and 15 units *Rsa* I and incubated at 37°C for 3 h. Digestion was analysed by electrophoresis of 10 µl undigested and 10 µl digested cDNA on a 1x TAE 1.2% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide.

2.28 Purification of digested cDNA.

Rsa I digested cDNA was purified by QIAquick PCR purification (Qiagen) as described in section 2.5, yet eluted in 50 µl sterile dH_2O. The eluted cDNA was ethanol precipitated by adding 50 µl 4 M ammonium acetate and 375 µl of 95% (v/v) ethanol, incubating at -20°C for up to 1 h and centrifuging at 13,000 rpm for 30 min in a microcentrifuge. The cDNA pellet was washed in 500 µl 80% ethanol, centrifuged at 13,000 rpm for 10 min in a microcentrifuge, air dried at 37°C for 10 min and resuspended in 6.7 µl 1 x TNE buffer. Purified DNA was analysed by electrophoresis, quantified by spectrophotometry (2.2) and subsequently diluted to 300 ng/µl in 1x TNE buffer.

2.29 PCR-Select cDNA Subtraction.

Rsa I digested, purified cDNA (300 ng/µl) represented the driver cDNA. A portion of each experimental tumour cDNA sample was also further treated to prepare adapter-ligated tester cDNA.

2.30 Ligation of adaptors to tester cDNA.

Three ligation reactions were prepared, two contained 50 ng cDNA, 2 µM of either adaptor 1 or 2R (table 2.1), 1x ligation buffer [50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 2 mM DTT, 0.05 mg/ml BSA] and 400 units T4 DNA ligase in a volume of 10 µl. The third was to prepare a non-subtracted tester control sample, this contained 2 µl of the two adaptor ligation reactions. Ligation was performed by incubating at

16°C overnight. The reaction was terminated by adding 1 µl 0.2 M EDTA/1 mg/ml glycogen and heating to 72°C for 5 min.

The efficiency of the ligation reactions was determined (as per manufacturers instructions) by PCR using PCR primer 1 and G3PDH 3' primer and comparing the intensity of PCR product to those obtained when using G3PDH 3' and 5' PCR primers (table 2.1). The ligation reactions were diluted 1/200 and 1 µl was mixed with 0.4 µM of each PCR primer, 0.2 mM of each dNTP, 1 x Advantage 2 PCR buffer and 1x Advantage 2 cDNA polymerase mix. Reactions were incubated at 75°C for 5 min to extend the single strand adaptors, followed by PCR conditions of 94°C for 30 sec, with 20 cycles of 94°C for 10 sec, 65°C for 30 sec and 68°C for 2.5 min. 5 µl aliquots were analysed by electrophoresis on a 1x TAE 2.0% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide.

2.31 Hybridisation.

Two rounds of hybridisation were performed using *Rsa* I digested driver cDNA derived from one tumour and adaptor-ligated tester cDNA derived from a different tumour. The first hybridisation was performed in 2 separate tubes, each contained 450 ng driver cDNA, 15 ng tester cDNA and 1x hybridisation buffer (20 mM HEPES-HCl (pH 8.3), 50 mM NaCl, 0.2 mM EDTA (pH 8.0), 10% (w/v) PEG 8000) in a volume of 4 µl. The first tube contained tester cDNA ligated with adaptor 1, the second contained tester cDNA ligated with adaptor 2R. Reaction mixtures were heated to 98°C for 1.5 min and then at 68°C for 8 h. The second hybridisation contained 75 ng freshly denatured driver cDNA and 1x hybridisation buffer combined with the two previous hybridisation reactions, samples were incubated at 68°C overnight. Hybridisation reactions were diluted in 200 µl dilution buffer (20 mM HEPES-HCl (pH 8.3), 50 mM NaCl, 0.2 mM EDTA (pH 8.0)) and heated to 68°C for 7 min.

2.32 PCR amplification.

Subtracted cDNA fragments were enriched using 2 rounds of PCR. Control PCR was also performed on non-subtracted tester DNA. 1 µl of diluted subtracted samples, and 1 µl of non-subtracted tester control cDNA diluted 1/1000 were mixed with 1x

Advantage 2 PCR buffer, 0.2 mM dNTP, 0.4 μM PCR primer 1 (table 2.1) and 1x Advantage 2 cDNA polymerase mix. Reactions were incubated at 75°C for 5 min followed by cycling conditions of 94°C for 25 sec and 27 cycles of 94°C for 10 sec, 66°C for 30 sec and 72°C for 1.5 min. Primary PCR reactions were diluted 1/10 and 1 μl was mixed with 1x Advantage 2 PCR buffer, 0.2 mM dNTP, 0.4 μM of both nested PCR primer 1 and 2R (table 2.1) and 1x Advantage 2 cDNA polymerase mix. This was re-amplified for 12 cycles of 94°C for 10 sec, 68°C for 30 sec and 72°C for 1.5 min. Both primary and secondary PCR were analysed by subjecting 8 μl to electrophoresis through a 1x TAE 2.0% (w/v) agarose gel containing 0.5 μg/ml ethidium bromide.

2.33 TA Cloning.

DNA fragments generated by PCR were cloned directly into pCR2.1™ using the Original TA Cloning Kit (Invitrogen). Subtracted cDNA libraries were generated by cloning differentially expressed cDNA fragments enriched by secondary PCR following subtractive hybridisation. Non-subtracted cDNA libraries were also created in this way by cloning SMART amplified cDNA derived from the three tumours that were used for subtractive hybridisation.

1 μl amplified cDNA was mixed with 50 ng linearised pCR2.1 vector, 1x ligation buffer [6 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 5 mM NaCl, 0.1 mg/ml bovine serum albumin, 7 mM β-mercaptoethanol, 0.1 mM ATP, 2 mM dithiothreitol, 1 mM spermidine] and 4.0 Weiss units T4 DNA Ligase (all Invitrogen) in a total volume of 10 μl. Ligation reactions were incubated overnight at 16°C.

2.34 Transformation of competent bacterial cells.

pCR™2.1 containing DNA inserts were transformed into One Shot TOP10F' competent cells (Invitrogen). Positive transformants containing vector and ligated-insert DNA were identified using α-complementation of X-Gal, with LB-agar plates containing 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) and 40 μg/ml 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-Gal). For each experiment, 10 pg pUC18 vector and sterile water controls were also transformed into the competent cells.

A vial of competent cells was thawed on ice for each ligation reaction to be transformed. Two µl β-mercaptoethanol and 2 µl of ligation reaction were mixed, by swirling of the pipette tip, with 50 µl competent cells. The cells were incubated, on ice, for 30 min, heat shocked for 30 sec at 42°C and immediately returned to ice for 2 min. Four hundred and fifty µl SOC medium [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose] or LB medium was added to the cells and incubated, shaking at 225 rpm for 60 min at 37°C in a shaking incubator. Transformed cells were placed on ice, and 50 µl and 200 µl aliquots were spread on the surface of LB-agar plates containing 50 µg/ml ampicillin (Sigma), 0.1 mM X-Gal and 40 µg/ml IPTG. Transformation mixture was allowed to soak into the agar plates for 30 min at 37°C before incubating inverted petri dishes overnight at 37°C. The following day plates were incubated at 4°C for 2-3 h to allow for colour development of blue colonies.

2.35 Construction of non-normalised, non-subtracted cDNA Libraries.

Non-normalised, non-subtracted cDNA libraries representing the tumours utilised in the subtractive hybridisation were constructed by directly cloning SMART PCR-generated cDNA. Amplification of cDNA was performed and optimised as described in section 2.24. The complex cDNA library was ligated into pCR2.1 by TA cloning as described in section 2.32 and transformed into TOP10F' competent *E.coli* (2.33).

Several colonies were picked, grown overnight in 5 ml LB medium containing 50 µg/ml ampicillin and plasmid DNA extracted by Qiagen plasmid DNA Miniprep protocol (2.11). Restriction digestion of plasmid DNA was performed to confirm that cloning of DNA fragments had occurred.

Characterisation of subtracted libraries and subtracted cDNA clones.

Cloned cDNA inserts, amplified by colony-PCR and purified by QIAquick 96 PCR purification (Qiagen), were identified by DNA sequencing using one of the two methods described below. The expression of subtracted cDNA clones was also assessed by dot blotting DNA on to nylon membranes and hybridising with radiolabelled probes derived from tumour RNA. Each membrane was made in triplicate so that three probes derived from the three types of tumour used in

subtractive hybridisation could be hybridised to identical membranes at the same time. Probes have been made by two different means, by incorporating the radioactive label during SMART cDNA amplification and during first-strand cDNA synthesis.

2.36 Colony-PCR.

Colonies containing cloned cDNA molecules, identified by blue/white colony screening, were picked and used to inoculate 200 µl LB-broth containing 50 µg/ml ampicillin set up in 96-well, flat-bottom, tissue culture plates (Nalge Nunc International). Cultures were grown overnight at 37°C in an incubator shaking at 225 rpm. Overnight cultures were diluted 10 fold in sterile distilled water and the remaining culture was stored at –80°C containing 15% (v/v) glycerol. Two µl of the diluted culture was amplified by PCR using the nested PCR primers 1 and 2R (table 2.1). Reactions contained 1x PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl], 0.2 mM dNTP, 2.5 mM $MgCl_2$, 0.4 µM of each primer and 1 unit Amplitaq Gold DNA Polymerase (Perkin Elmer) in a total of 25 µl. The reaction conditions were 94°C for 10 min, 35 cycles of 95°C for 30 sec, 68°C for 1 min, 72°C for 2.5 min followed by 72°C for 5 min.

PCR products were visualised by electrophoresis of a 1 µl aliquot of each reaction on 1x TAE 1.5% (w/v) agarose gels containing 0.5 µg/ml ethidium bromide and purified by QIAquick 96 PCR purification (Qiagen) using a QIAvac 96 vacuum manifold and QIAquick 96 well plates (2.5). The 50 µl eluted DNA was dried by incubating at 37°C for several h, and resuspended in 15-20 µl sterile ddH_2O.

2.37 Cycle sequencing using ABI 377 automated DNA sequencer.

Sequencing reactions (10 µl reaction volume) contained 4 µl (typically 30-90 ng) of Qiagen-PCR purified DNA, either of the two nested PCR primers 1 or 2R (Table 2.1; 0.16 pmol) and 4 µl Terminator Ready Reaction Mix (contains AmpliTaq DNA Polymerase FS, dNTP, A-Dye Terminator labelled with dichloro(R6G), C-Dye Terminator labelled with dichloro(ROX), G-Dye Terminator labelled with dichloro(R110), T-Dye Terminator labelled with dichloro(TAMRA), 2 mM $MgCl_2$ and 20 mM Tris-HCl buffer (pH 9.0), PE Applied Biosystems). Cycle sequencing

was performed by incubation at 96°C for 2 min followed by 25 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. Sequencing extension products were ethanol precipitated to remove excess dye terminators. DNA pellets were resuspended in 2 µl loading buffer containing a 5:1 ratio of deionized formamide and 25 mM EDTA (pH 8.0) with blue dextran (50 mg/ml). Samples were incubated at 95°C for 2 min to denature the DNA and placed on ice, 0.8 µl of each sample was loaded onto a 5% Long Ranger-Gel (AT Biochem) and electrophoresis was performed on an ABI PRISM 377 DNA Sequencer (PE Applied Biosystems).

2.38 Cycle sequencing using MegaBACE 1000 sequencer.

All steps were performed in a 96-well PCR plate (Advanced Biotechnologies). QIAquick PCR purified DNA template, 0.2-2 µg, was mixed with 4 µl DYEnamic ET reagent premix (Amersham Pharmacia Biotech) and 5 µM of either nested PCR primer 1 or 2R (Table 2.1) in a total reaction volume of 10 µl. Cycle sequencing was performed for 25 cycles of 95°C for 20 sec, 50°C for 15 sec and 60°C for 60 sec. Unincorporated dye terminators were removed by ethanol precipitation by adding 0.1 volume of 7.5 M ammonium acetate and 2.5 volumes of 95% ethanol. Precipitated DNA was pelleted by centrifugation in a Sorvall RT 6000D benchtop centrifuge at 2,900 rpm (1,400g) for 45 min at room temperature. The supernatant was removed by a brief centrifugation with the PCR plate inverted on a piece of paper towel. DNA pellets were washed by adding 100 µl 70% ethanol and centrifugation at 2,900 rpm (1,400g) for 15 min at room temperature. DNA pellets were dried by repeating the inverted plate centrifugation and incubating at 37°C for 10 min. DNA was resuspended in 10 µl MegaBACE loading buffer (Amersham Pharmacia Biotech).

2.39 DNA sequence analysis.

DNA sequences were extracted from results of both methods of sequencing and were analysed using a variety of sequence analysis software (PE Applied Biosystems and Amersham Pharmacia Biotech). Sequences were initially screened using Factura software (PE Applied Biosystems) to identify heterozygous base calls where one, or more, peaks were present at a minimum of 75% of the main called peak. Sequence Navigator and Sequence Analysis software were subsequently used to remove poor sequence and occasionally to correct ambiguously called bases. Sequencher (Gene

Codes Corporation) was utilised to screen all sequences for clones exhibiting overlapping regions of identical sequence. Sequence identification was determined using the Basic Local Alignment Search Tool (BLAST; www.ncbi.nlm.nih.gov/blast/; Altschul *et al.*, 1990) for searching Genbank (Benson *et al.*,1999), the EMBL data Library (Stoesser *et al.*, 1998) and the DNA Databank of Japan (Yoshio *et al.*, 1998). Default search parameters were used.

2.40 Preparation of DNA.

Control DNA was used on each Reverse Northern membrane. This included HPRT, actin and p9Ka cDNAs. HPRT 5' cDNA fragment was prepared by PCR (2.23) and Qiagen PCR purification (2.5). Both actin and p9Ka cDNAs were present in pAT153 plasmid vector, cDNAs were isolated from plasmid DNA via *Pst* I digestion and gel purification (2.6) of the excised *Pst* I cDNA fragments. The DNA yield of each cDNA was estimated by gel electrophoresis alongside 1 μg and 0.5 μg of λ DNA digested with *Hind* III.

Cloned cDNA inserts were prepared for immobilising on to Reverse Northern membranes by colony-PCR. Agarose gel electrophoresis was performed for each PCR reaction to monitor the quality and approximate yield of amplified cDNA.

All DNAs were diluted in sterile distilled water to give a volume of 80-85 μl, precisely 25 μl (between 10-100 ng DNA) was carefully aliquoted to three 96-well PCR plates. The excess volume was to ensure that exactly the same volume was aliquoted to each plate, any left over DNA was discarded.

2.41 Preparation of membranes.

DNA was denatured at 95°C for 10 min and immediately cooled in a 96-well Stratacooler (Stratagene). 25 μl 20x SSC [3 M NaCl, 0.3 M Na_3citrate] was added to the DNA, mixed and applied to a Bio-Dot blotting apparatus (Bio-Rad), connected to a vacuum manifold, containing a piece of Hybond-N^+ nylon membrane (approximately 12 cm x 8 cm in size; Amersham Life Sciences) pre-wetted in 10x SSC. DNA was drawn through onto the membrane by applying the vacuum. To ensure all the DNA was applied to the membrane, the wells of the 96-well plate were

washed in 100 µl 10x SSC and the wells of the BioDot apparatus were washed with 200 µl 10x SSC, both were applied to the membrane using the vacuum. The DNA was denatured by removing the membrane from the apparatus and placing it, DNA side up, on filter paper soaked in denaturing solution [1.5 M NaCl, 0.5 M NaOH] for 5 min. The membrane was neutralised by incubating for 1 min on filter paper soaked in neutralising solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA] and then left to air dry for 1 h at room temperature. The DNA was cross-linked to the membrane by subjecting to a UV exposure (wavelength 254 nm) of 70,000 microjoules/cm^2 (optimised pre-set mode for Hybond-N^+ nylon membranes) in a UV Crosslinker (Amersham Life Science). Membranes were wrapped in Saranwrap^{TM} and stored at 4°C.

2.42 First-strand cDNA synthesis probes.

20 µg of total RNA, derived from whole tumour RNA preparations, was incubated with 7.5 µg oligo$(dT)_{16}$ at 70°C for 10 min, at room temperature for 5 min and then on ice for at least 5 min. Reactions contained 20 units of human placental ribonuclease inhibitor (HPRI, Amersham Life Sciences), 1x first strand buffer [50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$], 10 mM DTT, 1 mM each of dATP, dGTP and dTTP, 6 µM dCTP, 5 µl [α^{32}P] dCTP (3000Ci/mmol) and 200 units of Superscript^{TM} Reverse Transcriptase (Gibco BRL). These were incubated at 42°C for 1 h, another 200 units of Superscript^{TM} Reverse Transcriptase was added and incubated for a further 1 h at 42°C. RNA was removed by adding 0.4% (w/v) SDS/20mM EDTA/360 mM NaOH and incubating at 68°C for 30 min and at room temperature for 15 min. Reactions were neutralised by adding 1 M Tris-HCl (pH 7.0)/2N HCl.

2.43 SMART PCR-generated labelled cDNA probes.

Slight modifications to the previously described SMART cDNA amplification protocol (section 2.25) were employed to label SMART cDNA. On each occasion PCR was optimised to prevent over amplification occurring. 2 µl SMART cDNA was amplified in a reaction volume of 20 µl containing 3 µl (α^{32}P) dCTP (3000Ci/mmol), 10 mM each of dATP, dGTP and dTTP and 60 µM dCTP.

Probes were monitored for incorporation efficiencies by TCA precipitation and scintillation counting (section 2.15) and purified using Sephadex G50 spun columns (section 2.16). Purified probes were re-assessed by TCA precipitation and scintillation counting to assess recovery from spun columns and to ensure that the three probes used for hybridisation were of the same, approximate, strength.

2.44 Reverse Northern hybridisation.

Membranes were prehybridised in 25 ml of 5x SSC, 5x Denhardts, 0.5% (w/v) SDS and 100 μg/ml denatured salmon sperm DNA for 2-4 h at 68°C in a rotating hybridisation oven (Hybaid mini hybridisation oven or Techne hybridiser HB-1D). 50x Denhardt's reagent consisted of 1% (w/v) ficoll 400, 1% (w/v) bovine serum albumin and 1% (w/v) polyvinylpyrrolidone. Probes were denatured by boiling for 10 min and quenching on ice for a minimum of 2 min. An equal number of counts (typically $5x10^6$ cpm) for each probe was added to hybridisation buffer [prehybridisation buffer containing 10% (w/v) dextran sulphate] and hybridisation was performed overnight at 68°C. Hybridised membranes were washed in 2 changes of 0.1x SSC/0.1% SDS solution for 30 min and 60 min at 68°C in a shaking water bath. Washed membranes were wrapped in Saranwrap™ and exposed to autoradiography film (Kodak) at -80°C for 16 hr-2 weeks. Images were also acquired using the Bio-Rad GS-363 Molecular Imaging System. Membranes were exposed to Imaging Screen BI (BioRad) for 24 hr-7 days and developed using a GS-363 scanner. Spot intensities were quantified using Molecular Analyst™ software (Bio-Rad).

2.45 Colony lifts and hybridisation.

Approximately 1000 colonies from 3 cDNA libraries (non-normalised, non-subtracted; produced in section 2.35) and the 4 subtracted cDNA libraries (produced in section 2.33), transformed into TOP10F' competent cells were grown on nylon membrane for hybridisation screening experiments. The method used was described by Sambrook et al. (1989). Transformation mixtures were diluted in LB medium and 100 μl was spread directly on to Millipore HA nylon filters pre-soaked on top of freshly made LB-agar plates (containing 50 μg/ml ampicillin, 40 mg/ml X-gal and 0.1 mM IPTG) to give a colony density of approximately 200-250 white colonies/8 cm plate. Inverted plates were incubated at 37°C for approximately 20 h to allow

sufficient numbers of colonies to grow.

Master plates were made by placing membrane filters, colony side down, on to fresh LB agar plates (containing 50 μg/ml ampicillin, 40 mg/ml X-gal and 0.1 mM IPTG) and applying even pressure over the whole of the membrane. Filters were carefully removed and the master plates were incubated at 37°C to allow duplicated colonies to replenish. Plates were sealed with parafilm and stored at 4°C.

Nylon membrane filters containing colonies were treated to fix plasmid DNA to the membrane. Each filter was incubated, colony side up, on a series of pieces of 3MM paper saturated in the following solutions: 10% (w/v) SDS for 3 min, denaturing solution [1.5 M NaCl, 0.5 M NaOH] for 5 min, neutralising solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2)] for 5 min, 20x SSC for 5 min. Membranes were placed colony side up on paper towel between each incubation to remove excess liquid. Membranes were air-dried for 1h before being fixed by subjecting to a UV exposure (wavelength 254 nm) of 70,000 microjoules/cm^2 in a UV Crosslinker (Amersham Life Science). Membranes were stored at 4°C.

25 ng (DNA template) actin, HPRT or p9Ka cDNA fragment was labelled with [α^{32}P] dATP (3000mCi/mmol) using the random primed DNA labelling kit (2.13) with Strip-Ez nucleotide mix (Ambion) incorporated into the reaction. The level of radioactive incorporation was monitored by TCA precipitation (section 2.15) and the probe was purified using a Sephadex G50 purification column (2.16). Colony filters were incubated in the same composition of prehybridisation (25 ml/10 filters) and hybridisation buffers (15 ml/10 filters) as described for the hybridisation of Reverse Northern membranes (section 2.44). Denatured probe was added to the hybridisation buffer and incubated with the colony filters overnight at 68°C in a rotating hybridisation oven. The following day membranes were washed at 68°C in several changes (typically 250 ml) of 0.1xSSC, 0.1% SDS for 3x 45 min each. Hybridised filters were exposed to autoradiographic film overnight at –80°C.

Northern & Virtual Northern Analysis of Subtracted Clones.

The expression of specific subtracted cDNA clones in multiple tumours, cell lines and normal tissues from the transgenic mice was determined by hybridisation to

Northern and Virtual Northern blots. In both cases, gels were visualised prior to blotting on to nylon membrane, and were again visualised post-blotting to assess for loading accuracy and transfer efficiency.

2.46 Northern Blotting of Total RNA.

Equipment for use in Northern blot analysis was treated with 0.1% (v/v) DEPC to inactivate RNase before use. 1% (w/v) formaldehyde agarose gels were prepared using 1x formaldehyde gel running buffer [FGRB: 0.1 M MOPS (pH 7.0), 40 mM sodium acetate, 10 mM EDTA] and 18% (v/v) formaldehyde/NaOH [30 ml formaldehyde/22.5 µl 10 M NaOH]. The gel was subjected to electrophoresis for 5 min, prior to loading RNA samples, in 1x FGRB at 5 V/cm with the buffer being circulated from the cathode to the anode.

10 µg aliquots of total RNA were prepared in a total volume of 20 µl containing 1x FGRB, 17.5% (v/v) formaldehyde and 50% (v/v) deionised formamide. Samples were heat denatured at 65°C for 15 min and immediately cooled on ice. 2.5 µl of formaldehyde gel loading buffer [50% (v/v) glycerol, 1 mM EDTA (pH 8.0), 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 0.25 mg/ml ethidium bromide] was added to the RNA, mixed and subjected to electrophoresis 3-4 V/cm, with buffer circulating as above.

After electrophoresis, the gel was washed in 50 mM NaOH/10 mM NaCl for 45 min at room temperature on a Stovall "belly dancer" shaker. A second wash was performed in 10 mM NaCl/0.1 M Tris-Cl (pH 7.5), also for 45 min. RNA was transferred to Electran nylon blotting membrane (BDH) by capillary action in 10x SSC as described in Sambrook et al. (1989). The membrane was left to air dry for 1 h and the RNA was fixed onto the membrane by exposure to UV light for 2.5 min. Membranes were wrapped in Saranwrap™ and stored at 4°C.

2.47 Virtual Northern blotting of SMART amplified cDNA.

The expression of specific cDNA clones were examined in SMART PCR-generated cDNA (2.25), derived from the same MMTV-*neu* and MMTV-*neu/S100A4* primary tumours and the bitransgenic metastasis to those used for subtractive hybridisation, by blotting amplified SMART cDNA onto a nylon membrane. This was achieved

using the Southern blotting method described by Sambrook et al. (1989) and was instead of performing Northern blotting of RNA from these tumours due to the limited amount of poly(A)$^+$ RNA extracted from tumour sections. Amplified DNA from 15 and 18 cycles of PCR was subjected to electrophoresis, alongside 1 Kb ladder (Advanced Biotech), on a 1x TAE 1.2 % (w/v) agarose gel containing 0.5 μg/ml ethidium bromide. The gel was washed in 0.25 M HCl for 10 min at room temperature. Further washes were performed in denaturing solution [0.5 M NaOH, 1.5 M NaCl] for 2x 15 min each and in neutralising solution [1.5 M NaCl, 0.5 M Tris-Cl (pH 7.5), 1 mM EDTA (pH 8.0)] for 2 x 15 min each. DNA was transferred, overnight, to Hybond-N$^+$ nylon membrane (Amersham) by capillary action in 10x SSC. The membrane was left to air dry, at room temperature, for 1 h. The DNA was cross-linked to the membrane by subjecting to a UV exposure (wavelength 254 nm) of 70,000 microjoules/cm^2 (optimised pre-set mode for Hybond-N$^+$ nylon membranes) in a UV Crosslinker (Amersham Life Science). Membranes were wrapped in Saranwrap^{TM} and stored at 4°C.

Two methods of UV crosslinking of nucleic acids to nylon membranes were used because the experiments were performed in two different laboratories.

2.48 Hybridisation of specific cDNA probes to the RNA and DNA immobilised to the Northern and Virtual Northern membranes.

Both Northern and Virtual Northern membranes were incubated in prehybridisation solution [5x SSPE, 5x Denhardt's reagent, 0.5% (w/v) SDS, 50% (v/v) deionised formamide, 100 μg/ml denatured sonicated salmon sperm DNA] at 42°C for 3-4 h in a rotating hybridisation oven. 20x SSPE consisted of 3 M NaCl, 200 mM sodium dihydrogen orthophosphate hydrate, 20 mM EDTA, pH 7.4. The prehybridisation solution was removed and replaced with hybridisation solution [prehybridisation buffer containing 5% (w/v) dextran sulphate] or UltraHyb^{TM} (Ambion) for 30 min prior to the addition of denatured labelled probe of typical specific activity of 5x10^8-1x10^9 dpm/μg. Hybridisation was performed at 42°C for at least 16 h. After hybridisation, the membranes were washed in pre-heated washing buffers at 68°C in a shaking water bath. Initial wash was performed with two changes of 2x SSC/0.1% (w/v) SDS for 15 min each, this was followed by 1 wash in 0.2x SSC/0.1% (w/v) SDS for 15 min and 2 washes of 0.1x SSC/0.1% (w/v) SDS for 20-30 min. Washed

membranes were wrapped in Saranwrap™ and exposed to autoradiographic film (Kodak).

cDNA expression array hybridisation experiments.

The cDNA expression hybridisation kit contained reagents for SMART cDNA synthesis, amplification and purification, Klenow extension probe labelling and 2 duplicate nylon membrane arrays (Atlas™ Mouse cDNA Expression Array), each containing 588 different mouse cDNAs (see Appendix or www.clontech.com/atlas/genelists/Mbroad.xls). All reagents required for screening of these cDNA expression arrays were supplied by Clontech, unless otherwise stated. Two experiments were performed, firstly comparing the expression profiles of the same MMTV-*neu/S100A4* primary and metastatic tumours that were used in subtractive hybridisation. And secondly, the expression profile of 5 pooled grade 3, infiltrating ductal carcinomas (IDC) from both MMTV-*neu* mice and MMTV-*neu/S100A4* mice were also compared.

2.49 Template probe preparation.

For the first experiment, SMART PCR-generated cDNA was synthesised from poly(A)$^+$ RNA extracted from approximately 100,000 cells derived from the MMTV-*neu/S100A4* primary and metastatic tumours. In the second experiment, 1 µg of mixed total RNA from both MMTV-*neu* mice and MMTV-*neu/S100A4* mice was used to synthesise and amplify SMART cDNA (2.24-25).

The 2 reactions, for each sample, amplified for the required number of cycles, were combined and subjected to spun column purification to remove reaction constituents such as unincorporated nucleotides and PCR primers. 400 µl NT2 buffer was added to each DNA sample, mixed and applied to a Nucleospin filter column placed in a collection tube. Columns were centrifuged at 13,000 rpm in a microcentrifuge for 1 min and the waste solution was discarded. DNA was washed by applying 3 lots of 500 µl NT3 buffer to the filter and centrifuging at 13,000 rpm for 1 min each time. DNA was eluted by applying 50 µl NE elution butter to the filter, incubating for 2 min followed by centrifugation at 13,000 rpm for 1 min. 5 µl purified DNA was analysed by electrophoresis on a 1x TAE 1.2 % (w/v) agarose gel containing 0.5

μg/ml ethidium bromide. DNA yield was determined by measuring optical density at 260 nm (see method 2.2).

2.50 Labelling of DNA.

SMART amplified cDNA was used as template DNA in radioactive labelling reactions using a gene-specific primer mix (CDS primer mix) specific for cDNAs present on the array. Purified DNA (500 ng) and 1 μl CDS primer mix in 34 μl, was denatured by boiling for 8 min and quenching on ice for 2 min. Denatured DNA, incubated at 50°C for 2 min, was labelled by adding 1x Klenow buffer, 1x dNTP mix, 5 μl [α^{32}P] dATP (3000mCi/mmol) and 1 μl (2 units) Klenow enzyme, in a total volume of 50 μl and by incubating at 50°C for 30 min. The labelling reaction was stopped by adding 2 μl 0.5 M EDTA (pH 8.0). Labelled DNA was purified using the Nucleospin columns; the same method as described above was used, only that 350 μl NT2 buffer was mixed with the cDNA prior to addition to the column, 350 μl NT3 buffer was used to wash the DNA/filter and 100 μl NE elution buffer was used to elute the labelled DNA. Probe labelling and purification efficiency was measured by TCA precipitation (2.15)

2.51 Hybridisation.

Atlas array membranes were prepared by soaking briefly in water and then pre-hybridising in 5 ml ExpressHyb (containing heat denatured 50 μl 20x SSC and 50 μl blocking solution) at 68°C for 1 h in a rotating hybridisation oven. Each membrane was hybridised with 5 ml ExpressHyb containing heat denatured 20xSSC (50 μl), blocking solution (50 μl) and approximately $4x10^7$ cpm of probe. Hybridisation was performed overnight at 68°C in a rotating hybridisation oven.

Membranes were washed the following day in two changes of 2x SSC/1% (w/v) SDS for 30 min each followed by three changes of 0.1x SSC/0.5% (w/v) SDS also for 30 min each in a shaking water bath at 68°C. Washed membranes were sealed in Saranwrap™ and exposed to autoradiographic film for an overnight-7 day exposure and also to a phosphorimaging screen for an overnight and a 1 week exposure. Individual spot intensities were analysed using ImageQuaNT® v5.1 software (Molecular Dynamics).

2.52 Removal of hybridised probe from arrays.

Hybridised arrays were stripped, to enable re-probing, by incubating in 2 changes of boiling 0.5% (w/v) SDS solution for 20 min each in a shaking water bath at 80°C. Stripped membranes were wrapped in Saranwrap™, exposed to autoradiograph film to monitor the efficiency of the stripping process and stored at –20°C.

Safety precautions and COSHH assessments.

All experiments were carried under conditions of standard laboratory practice except where COSHH assessment advised otherwise.

Chapter 3

Screening mouse cDNA expression arrays.

The development of primary tumours and the subsequent progression towards a metastatic phenotype is a multistep process that is associated with dynamic changes to the genome and the subsequent acquisition of new cellular properties (Hart *et al.*, 1989; Bishop, 1991; Devilee and Cornelisse, 1994; Hanahan and Weinberg, 2000).

Transfection and subsequent expression of S100A4 in benign rat mammary tumour cells is able to confer *in vivo* metastatic progression in a rat model of breast cancer (Davies *et al.,* 1993). Elevated levels of S100A4 in primary mammary tumours also appear to confer an increased susceptibility to metastasis in MMTV-*neu/S100A4* transgenic mice compared with MMTV-*neu* transgenic mice (Davies *et al.*, 1996). This is demonstrated by the significant incidence of metastatic lung tumours in some MMTV-*neu/S100A4* transgenic mice, but not in MMTV-*neu* transgenic mice. Not all mice expressing S100A4 in their primary tumours, however, develop detectable metastases suggesting S100A4 over expression alone is insufficient for metastasis to occur. The molecular mechanism(s) and the additional genes involved that drive this progression to the metastatic phenotype are unknown, as is the role of S100A4 in this process.

This chapter describes the use of a cDNA expression hybridisation array for the investigation of the expression profiles of 588 different mouse genes in different tumours from the transgenic mice. A list of all these sequences, selected by the manufacturers, can be found in the appendix. The sequences present on the arrays encode proteins involved in many aspects of cellular biology, including intracellular signal transduction, DNA synthesis and repair, DNA binding, protein synthesis, ion transport, apoptosis, cell-cell communication, cytoskeletal structure and cell motility. The functional relevance of all these sequences to the development and progression of metastatic cancer provides a useful means of screening for aberrant expression of additional sequences that may contribute to the metastatic phenotype in the transgenic mice.

Two hybridisation experiments have been performed using these cDNA arrays. The first compared the relative mRNA levels of the 588 cDNAs between an MMTV-*neu/S100A4* primary mammary gland tumour and metastatic lung tumour.

This comparison allows for the identification of gene expression changes that may be directly related to S100A4-mediated metastatic progression. The second experiment was to compare the expression profiles of the same cDNAs in pooled sets of five primary mammary tumours from MMTV-*neu* mice and five primary tumours from MMTV-*neu/S100A4* mice. This was to identify genes demonstrating expression that was consistently higher in multiple S100A4 positive primary tumours relative to S100A4 negative primary tumours, and vice versa. Thus determining how, or if, S100A4 is affecting metastasis by regulating the expression of particular genes. Direct comparison of genes expressed between these tumours of differing metastatic potential, from both experiments, will reveal which of the 588 genes are being expressed in the different tumours and therefore which, if any, may be contributing towards the gain in metastatic phenotype.

3.1. Experiment 1: MMTV-*neu/S100A4* primary tumour versus metastasis.

The two bitransgenic tumours, a primary mammary tumour and a lung metastasis were from the same MMTV-*neu/S100A4* mouse. This mouse developed multiple mammary tumours which were all classified as grade three infiltrating ductal carcinomas (IDC). The primary tumour utilised for these experiments was the largest of these mammary gland carcinomas. However, due to their similar histology it was unknown which, and how many, of these tumours from this mouse metastasised to produce this particular secondary lung tumour. The mouse developed multiple metastases, both microscopic and macroscopic which covered approximately 20% of the lung tissue area (Figure 1.2). The histological characteristics of these two tumours were very similar and reflected the histology of the majority of tumours from bitransgenic mice as described in detail in section 1.5.3.1 (chapter 1; B. Shoker, personal communication). Frozen tumour sections of the two tumours were cut and histologically characterised to verify the histology observed for methacarn-fixed paraffin-embedded sections. Adjacent frozen sections to those used for histological analysis were used as the source of RNA for these gene expression experiments.

3.1.1 RNA extraction & integrity.

Poly(A)$^+$-containing messenger RNA [Poly(A)$^+$-RNA] was extracted from several frozen sections, approximately 300,000 cells, from both the primary tumour and the

metastasis. The integrity of the poly(A)$^+$-RNA was assessed by RT-PCR rather than Northern blot analysis because of the limited amount of RNA that was expected to be obtained from the small number of cells. PCR primers were designed to amplify the 5' region of the HPRT cDNA which meant that Reverse Transcriptase cDNA synthesis, primed by oligo(dT)$_{16}$, required at least 1250 bp of cDNA synthesis for the detection of positive amplification. This assay was previously developed to determine the quality of RNA extracted from archival microdissected frozen tissue, and so was optimised to detect HPRT target cDNA molecules extracted from the equivalent of less than 10 cells. Poly(A)$^+$ mRNA extracted from approximately 100,000 cells from the two different tumours was reverse transcribed into cDNA. Positive amplification of this cDNA was detected for the two tumours at a dilution of equivalent to approximately 20 cells (Figure 3.1A). Amplification was not attempted at higher dilutions of the cDNA. RNA preparations that did not produce positive HPRT amplification did not subsequently produce cDNA amplification by SMART PCR (Figure 3.1). The assay, therefore, proved to be a good indication of poly(A)$^+$ RNA integrity and the ability to synthesise and amplify SMART cDNA from the frozen tumours (see below and Figure 3.1).

3.2 Experiment 2: MMTV-*neu* primary tumours versus MMTV-*neu/S100A4* primary tumours.

Five primary mammary tumours from MMTV-*neu* mice and five from MMTV-*neu/S100A4* mice were selected for this gene expression profiling experiment. All ten tumours were histologically characterised as grade 3 infiltrating ductal carcinomas which were similar to the primary tumour used for the above experiment and representative of the majority of *neu*-induced mammary gland tumours from these strains of transgenic mice (Figure 1.1). The five bitransgenic mice developed multiple metastases in the lungs. Histological analysis of the lung tissue from the MMTV-*neu* mice revealed that 3 of these mice did not develop detectable metastases, and the other 2 mice had a combined total of 3 detectable clumps of *neu* staining tumour cells (<0.2 mm in size) encapsulated in the blood vessels of the lung tissue.

Total RNA extracted from the frozen part of the same tumours used for histological analysis was assessed for quality by electrophoresis on a formaldehyde-containing agarose gel (not shown). The ribosomal RNA was shown to be intact

3.5 cDNA abundance.

Highly expressed sequences were detected after only a short exposure to phosphorimaging screen (Figures 3.4a and 3.5a). In all four tumour samples, expression of approximately 12% of cDNAs screened were detected after an overnight exposure. Exposure to the imaging screen for 7 days was required, in both experiments, to detect expression of the majority of cDNAs, representative of low abundant sequences (Figures 3.4b and 3.5b). Even after this time the expression of some cDNAs was not detected suggesting the technique was not sensitive to detect expression of very low abundant sequences or that these cDNAs were not expressed in these tumours.

Based on the identification of there being 3 distinct abundance classes of mRNAs in HeLa cells (Bishop *et al.,* 1974) the abundance distribution of these sequences was divided into 3 classes representing high, medium and low abundant transcripts (Table 3.1 and Figure 3.6). The high abundance class contained the top 5% of abundant sequences (30 cDNAs), the medium class contained the next 25% of abundant sequences (149 cDNAs) and the remaining 70% of sequences (418 cDNAs) were of low abundance. The division between these classes was somewhat arbitrary yet reflected the distribution of abundance observed in Figure 3.6, where few sequences were expressed to high levels more sequences were expressed to intermediate levels and many cDNAs were expressed to very low or undetectable levels. Those cDNAs where hybridisation signal was not detected were classified as low abundant because it was unknown whether expression of these sequences were not detected due to sensitivity limitations of the method. Expression levels for these sequences was determined as background, a value that was corrected for prior to determining abundance. The proportion of the message profile contributing to each abundance class was expected to reflect the whole mRNA profile of these cells, this distribution was very similar for the four tumour samples (Table 3.1).

The same abundant cDNAs were identified in the four tumour samples (Table 3.2) including glyceraldehyde 3-phosphate dehydrogenase (G3PDH; M32599), β-actin (M12481), ubiquitin (X51703) and the mammary epithelial cell marker cytoskeletal keratin 18 (M11686). These are expected to be naturally abundant in normal mammary tissue (Ronnov-Jessen *et al.,* 1996). Some of the other high abundant sequences in all four samples are of particular interest and may represent sequences up-regulated by either the *neu* transgene or by cellular responses to the

indicating RNA had not degraded and so 1 μg of the combined tumour RNA for each transgenic strain was used to synthesise SMART cDNA.

3.3 SMART cDNA synthesis and amplification.

SMART cDNA was synthesised from approximately 100,000 cells from poly(A)$^+$-RNA extracted from frozen tumour sections and from 1 μg of pooled total RNA. Amplification of this cDNA was carefully optimised to avoid over amplification and therefore misrepresentation of individual cDNA molecules in the amplified cDNA library. This was a critical step in the preparation of cDNA prior to performing differential gene expression analysis.

SMART cDNA amplification was visualised by a smear of cDNA fragments of a range of sizes between 0.2 kb and 5-6 kb. Inability to clearly detect cDNA amplification suggested too few cycles of amplification had occurred (15 cycles in Figure 3.1). Increasing the number of cycles increased the level of amplification until an 'amplification plateau' was reached, after which point high molecular weight DNA was observed on the gel image (19-23 cycles, Figure 3.2). Note the distinct pattern of different sized bands (approximately 2.5 kb, 1.2 kb, 1.0 kb in Figure 3.2), these cDNAs are expected to represent the most abundant mRNAs in each tumour preparation. The same banding pattern was evident in all tumour samples amplified suggesting the same abundant mRNAs are present in the different tumours. The intensity of these fragments becomes less distinct as the cycle number increases beyond the amplification plateau (19-23 cycles, Figure 3.2), this also demonstrates the occurrence of cDNA over-amplification and so this level of amplification was avoided for these experiments.

SMART cDNA synthesised from poly(A)$^+$ RNA required 18 cycles of amplification (not shown), whereas cDNA synthesised from total RNA required 15 cycles of amplification (Figure 3.2). The different number of cycles required reflected the different efficiency of SMART cDNA synthesis and amplification from total RNA or from poly(A)$^+$ RNA.

Dot blot hybridisation was performed to determine whether cDNA representation was maintained following amplification. Aliquots of total RNA (1 μg and 100 ng) and SMART amplified cDNA (50 ng and 5 ng) from the two sets of pooled primary tumours were spotted onto nylon membrane. SMART amplified cDNA derived from poly(A)$^+$ RNA was also applied to the membrane however

insufficient RNA was extracted from these tumour sections to compare this to the amplified population. The membranes were hybridised with radioactively labelled S100A4 (p9Ka), actin and HPRT cDNA probes (Figure 3.3). The results demonstrate that actin and HPRT are both equally represented before and after amplification in the pooled tumour samples. These two cDNAs are also detected to equal levels in the two single bitransgenic tumours. Expression of S100A4 was not detected in the pooled MMTV-*neu* tumour sample, as expected, yet was evident in both the total RNA and SMART cDNA for the pooled MMTV-*neu/S100A4* primary tumour sample. Evidently S100A4 expression increases with metastatic capability, being highest in the metastasis relative to the primary tumours.

3.4 Hybridisation.

For each hybridisation experiment 500 ng SMART PCR-generated cDNA was used as template for a gene-specific primer mix to prime the synthesis of labelled cDNA (as described in 2.50). Probes of an equal activity (cpm) were used for hybridisation to the duplicate arrays and care was taken to minimise variations in experimental conditions.

Each array contained 9 cDNAs, each in duplicate, as independent controls for differences in probe labelling or hybridisation efficiencies. The hybridisation signal intensity of each of these controls were quantified and an average control cDNA hybridisation signal intensities was calculated. For both experiments, this average was identical between the two tumours/sets of pooled tumours being compared suggesting that probe normalisation was not required.

Hybridisation signal intensities for all 588 cDNAs, each in duplicate, were also quantified and an average intensity for each sequence was calculated. This arbitrary spot volume was used as a measure of the relative abundance of each cDNA in the labelled cDNA probes and was thus used to determine which cDNAs were expressed in each tumour/set of pooled tumours, to what extent they were expressed and to ascertain if any cDNAs were differentially expressed between experimental samples.

abnormal cellular phenotype. For example, several members of the cathepsin family of proteolytic enzymes, cathepsin B1 (M14222), D (X53337), H (U06119) and L (X06086) were all abundant in each of the four tumour samples (Table 3.2). Several cDNAs encoding proteins expressed in response to cell survival were also identified in these tumours (Tables 3.2), in particular three glutathione-S-transferases (J03752, D30687, J04696), three heat shock proteins (M38629, M36830, X53584), clusterin (ApoJ; L08235), defender against cell death (DAD1; U83628) and an endoplasmic reticulum stress protein (Erp72; J05186). High expression of these sequences may reflect the mechanisms activated in response to cellular stresses, such as heat shock or oxidative stress, by the growing tumour cell mass to avoid apoptosis-mediated cell death.

3.6 Differentially expressed cDNA sequences.

Comparison of hybridisation signal intensities determined the relative differences in expression for all cDNAs between the different tumour samples in a particular experiment. 597 cDNA expression comparisons (588 cDNAs plus 9 control cDNAs) were made for each exposure time (overnight and one week) in both experiments.

XY scatterplots comparing the cDNA abundance between tumours (Figure 3.7) emphasises the overall similarity of the expression profiles between these tumour samples. The least-squared regression line ($y = ax + b$, where a = slope and b = intercept) was drawn through the data points. In both experiments (Figures 3.7), the gradient of 1.0 indicates that the hybridisation efficiency and probe-labelling was equivalent between probes. The value of b is close to zero suggesting there was little or no difference in background signal. Clustering of data-points close to the regression line (reflected in R^2 values of 0.941 and 0.956) indicates that the relative expression of the majority of cDNAs is very similar in both samples being compared.

Initial criteria for identifying differential expression of a cDNA between two tumour samples was set at a difference in the level of expression of >2 or <0.5 fold (metastasis/primary; *neu/neu-S100A4*). This was selected as the cut off because other cDNA expression analysis experiments using microarrays have also used a similar differential expression ratio of ≥2 or ≥3 as criteria for identifying differentially regulated cDNAs (Schena *et al.*, 1995, DeRisi *et al.*, 1997, Sgroi *et al.*, 1999, Perou *et al.*, 1999, Yang *et al.*, 1999). Only one cDNA was identified as being differentially

expressed by this factor, defender against cell death 1 (DAD1; U83628). This sequence was 2.3-fold more abundant in the metastatic tumour relative to the primary mammary tumour from the bitransgenic mouse (Figure 3.7A and Table 3.3). The remaining 596 cDNAs in this screening experiment and all 597 cDNAs in the second screening experiment were expressed within this 2-fold selection criteria.

The stringency for identifying differentially expressed sequences was reduced because hybridisation signal intensity for several sequences on the arrays appeared to be visibly different between experimental samples (some of these are indicated by arrows on Figures 3.4 and 3.5). Thus, differences in expression ratios of between >1.25 and < 0.8 were applied. Using this reduced selection criteria a total of 35 cDNAs (6% of the 597 sequences screened) were differentially expressed between the metastasis and the primary bitransgenic tumour, 33 of these cDNAs were more highly expressed in the metastasis (Table 3.3). 17 cDNAs (3%) were differentially regulated between the pooled sets of primary tumours from the MMTV-*neu* and MMTV-*neu/S100A4* mice, 7 were more abundant in the combined MMTV-*neu* primary tumours and 10 were more abundant in the MMTV-*neu/S100A4* primary tumours (Table 3.4).

Seven cDNAs were differentially expressed in both experiments (indicated by * in Tables 3.3 and 3.4). Five of the seven cDNAs were more highly expressed in the pooled MMTV-*neu/S100A4* primary tumours relative to the MMTV-*neu* primary tumours, and were also more abundant in the metastasis relative to the MMTV-*neu/S100A4* primary mammary tumour. Thus, suggesting a potential increase in expression of these sequences is associated with an increase in S100A4 expression (Figure 3.3) and in metastatic capability.

3.7 Conclusions to chapter.

The tumours used to screen these cDNA arrays exhibited similar histological characteristics but differences in S100A4 expression were detected and they ultimately demonstrated different metastatic phenotypes. Unlike primary tumours from the MMTV-*neu/S100A4* mice, primary tumours from the MMTV-*neu* mice were unable to complete metastatic progression, as demonstrated by their inability to colonise lung tissue (Davies *et al.,* 1996). The metastasis, exhibiting higher S100A4 expression relative to the other tumours (Figure 3.3), also possessed higher metastatic capability relative to the MMTV-*neu/S100A4* primary mammary tumour.

Differences in gene expression profiles are expected to coincide with the differences in tumour phenotype (Hart *et al.,* 1989; Hanahan and Weinberg, 2000). However, these cDNA array hybridisation experiments were unable to significantly distinguish between the four tumour samples. The expression profiles for each sample were very similar with the same cDNAs being highly expressed and only 1 cDNA (out of 1194 gene expression comparisons) being differentially expressed by > 2-fold; defender against cell death 1 (DAD1; U83628). The significance of this and the potential influences of expression changes that are >1.25 but <2-fold on phenotype are discussed later.

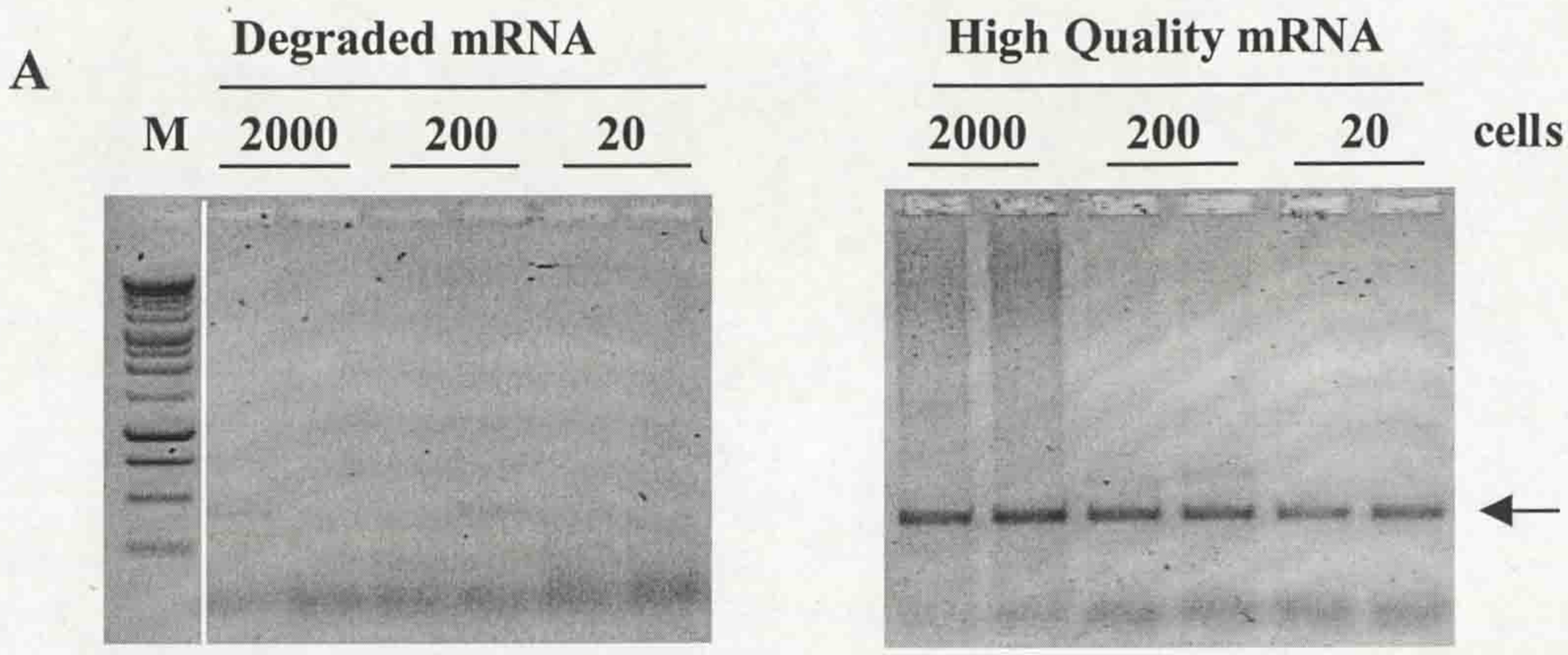


Figure 3.1 Integrity of RNA extracted from archival frozen tumours.

Agarose gel electrophoresis of RT-PCR (A) and SMART PCR products (B). Failure to detect HPRT PCR product (391 bp, arrow in A) following the RT-PCR RNA integrity assay indicates the presence of degraded or insufficient RNA and the subsequent inability to synthesise and amplify SMART cDNA (B).

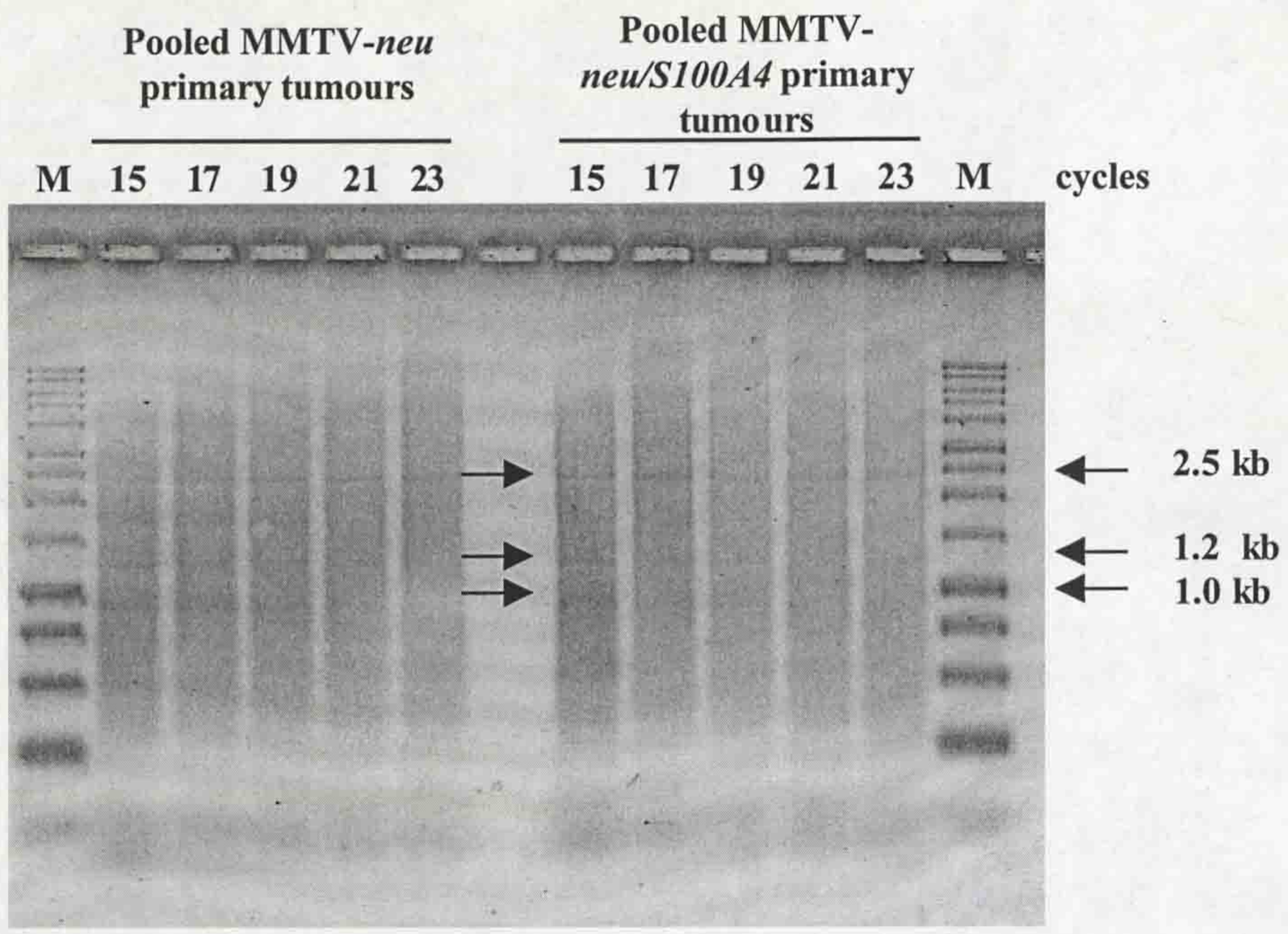


Figure 3.2 SMART-PCR generated cDNA.

SMART-PCR cDNA was used to amplify full length cDNA prior to gene-specific probe labelling and array hybridisation. Amplification was optimised as described in 2.3.25 to avoid using cDNA in differential gene expression experiments that had been over-amplified. The minimum number of cycles of amplification were utilised to produce sufficient cDNA. In this case, 15 cycles of PCR was sufficient to amplify both sets of pooled primary tumours; 18 cycles of amplification was required for the MMTV-*neu/S100A4* primary tumour and metastasis (not shown). Note the distinct pattern of different sized cDNA fragments after 15 cycles (arrows) in both tumour samples, and that these become less distinct with increasing numbers of cycles (19-23).

This agarose gel was stained post-electrophoresis in Sybr Gold (1/10,000 dilution; Molecular probes). M = 1 kb DNA ladder.

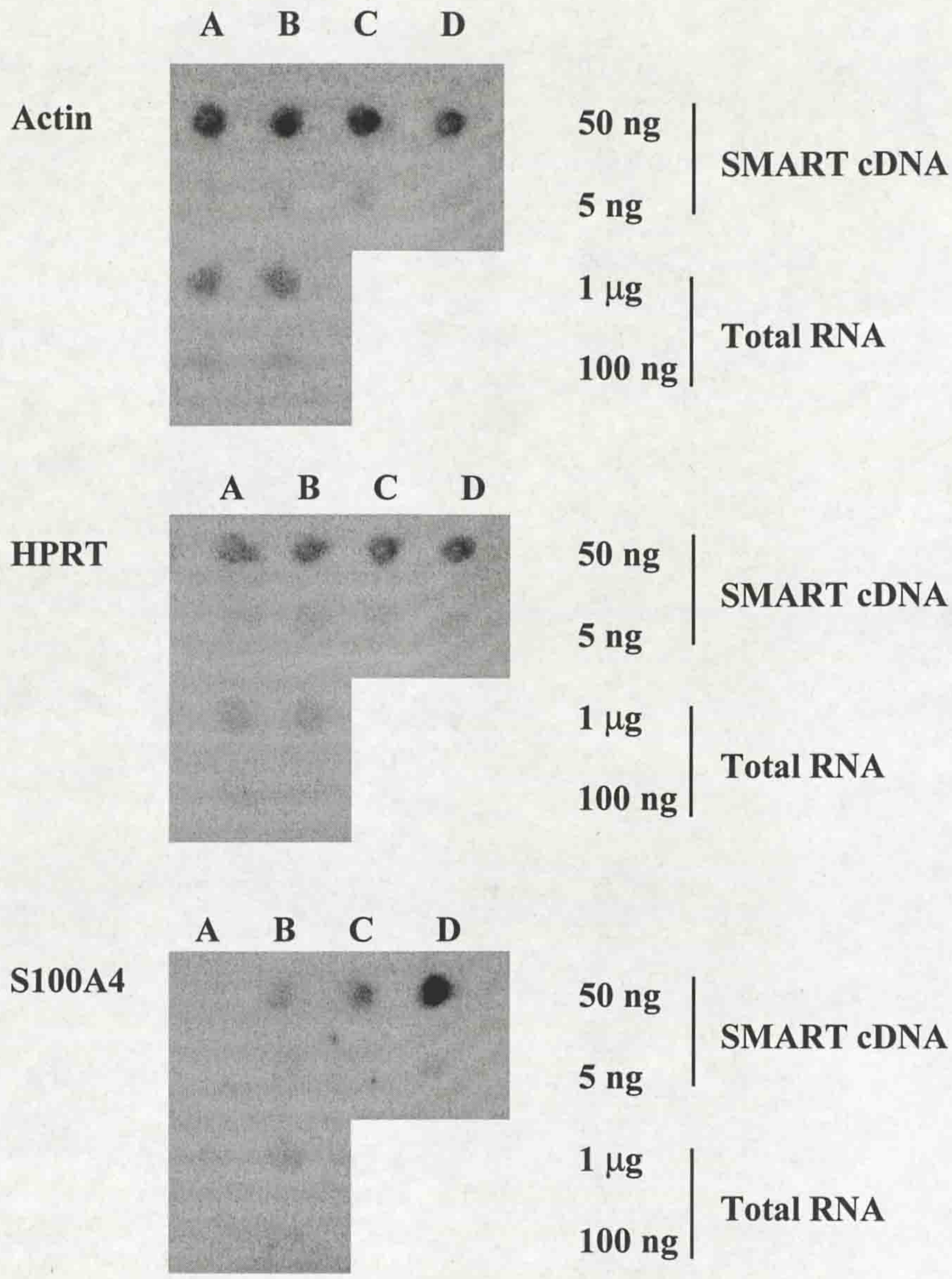


Figure 3.3 Dot blot of total RNA and SMART-PCR cDNA.

Total RNA (1 µg and 100 ng aliquots) and SMART cDNA (50 ng and 5 ng aliquots) from MMTV-*neu* (A) and MMTV-*neu/S100A4* (B) pooled tumour samples were dot blotted onto nylon membrane. Membranes also contained SMART-PCR cDNA derived from the MMTV-*neu/S100A4* primary tumour (C) and metastasis (D); total RNA was not available for these two tumours. Membranes were hybridised with actin, HPRT and S100A4 cDNA probes to determine if the original expression profile was maintained following amplification. Images were developed following 7 day exposure to phosphorimaging screen.

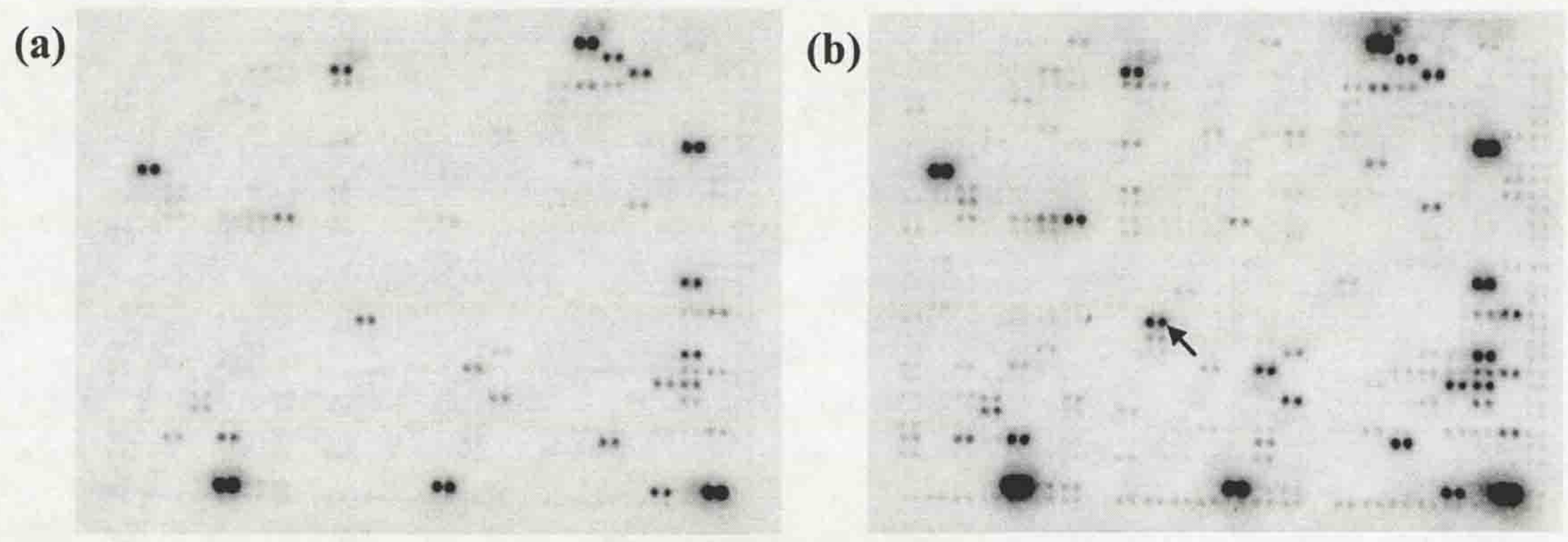
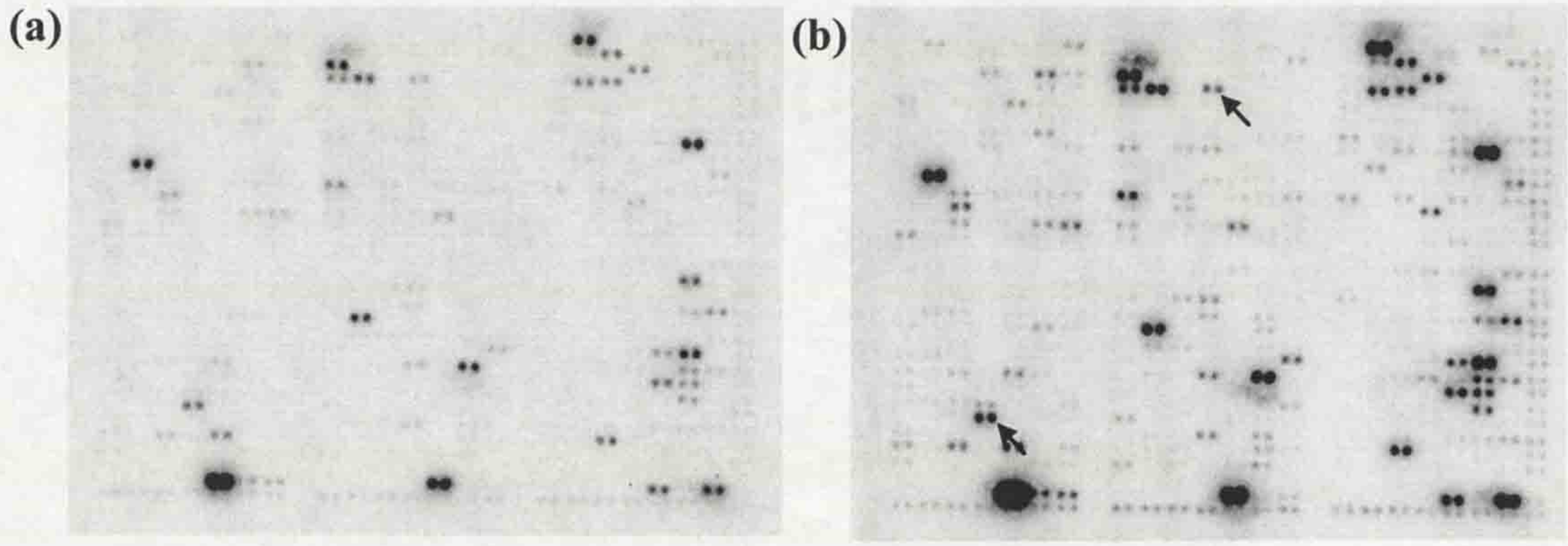


Figure 3.4 cDNA expression array hybridisation.

MMTV-*neu/S100A4* primary mammary tumour versus MMTV-*neu/S100A4* lung metastasis.

SMART PCR-generated cDNA derived from the same two bitransgenic tumours were used as probes to screen the 588 mouse Atlas cDNA arrays. Each array also contains 21 positive and negative control cDNAs to control for differences in probe labelling and hybridisation efficiencies. Hybridised membranes were subjected to (a) an overnight and (b) a week long exposure to a phosphorimaging screen. Genomic DNA spots along the bottom and right hand side were used to aid in orientating the developed image (see b). Some cDNAs that are apparently up-regulated in the primary or in the metastatic tumour are indicated by arrows.

MMTV-*neu* primary mammary tumours



MMTV-*neu/S100A4* primary mammary tumours

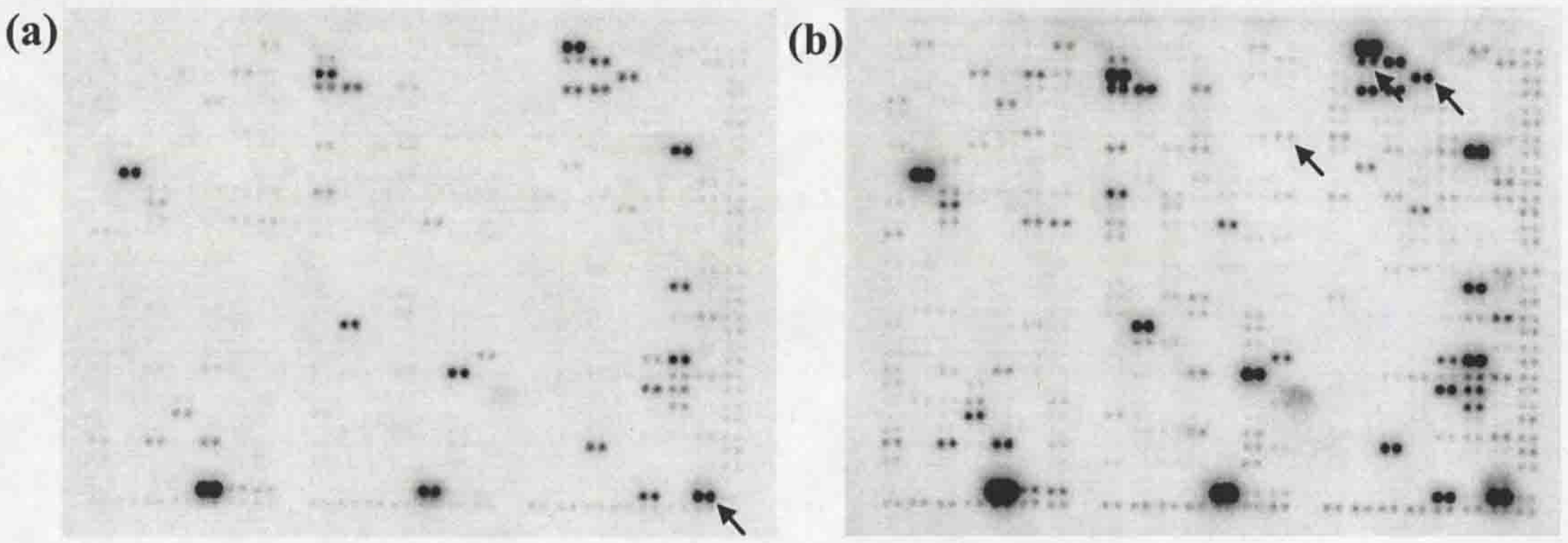


Figure 3.5 cDNA expression array hybridisation.

MMTV-*neu* primary mammary tumours versus MMTV-*neu/S100A4* primary mammary tumours .

SMART PCR-generated cDNA derived from the two sets of pooled mammary tumours were used as probes to screen the 588 mouse Atlas gene expression arrays (Clontech). Each array also contains 21 positive and negative control cDNAs to control for differences in probe labelling and hybridisation efficiencies. Hybridised membranes were subjected to (a) an overnight and (b) a week long exposure to a phosphorimaging screen. Genomic DNA spots along the bottom and right hand side were used to aid in orientating the developed image (see b). Some cDNAs that are differentially regulated in either set of primary tumours are indicated by arrows.

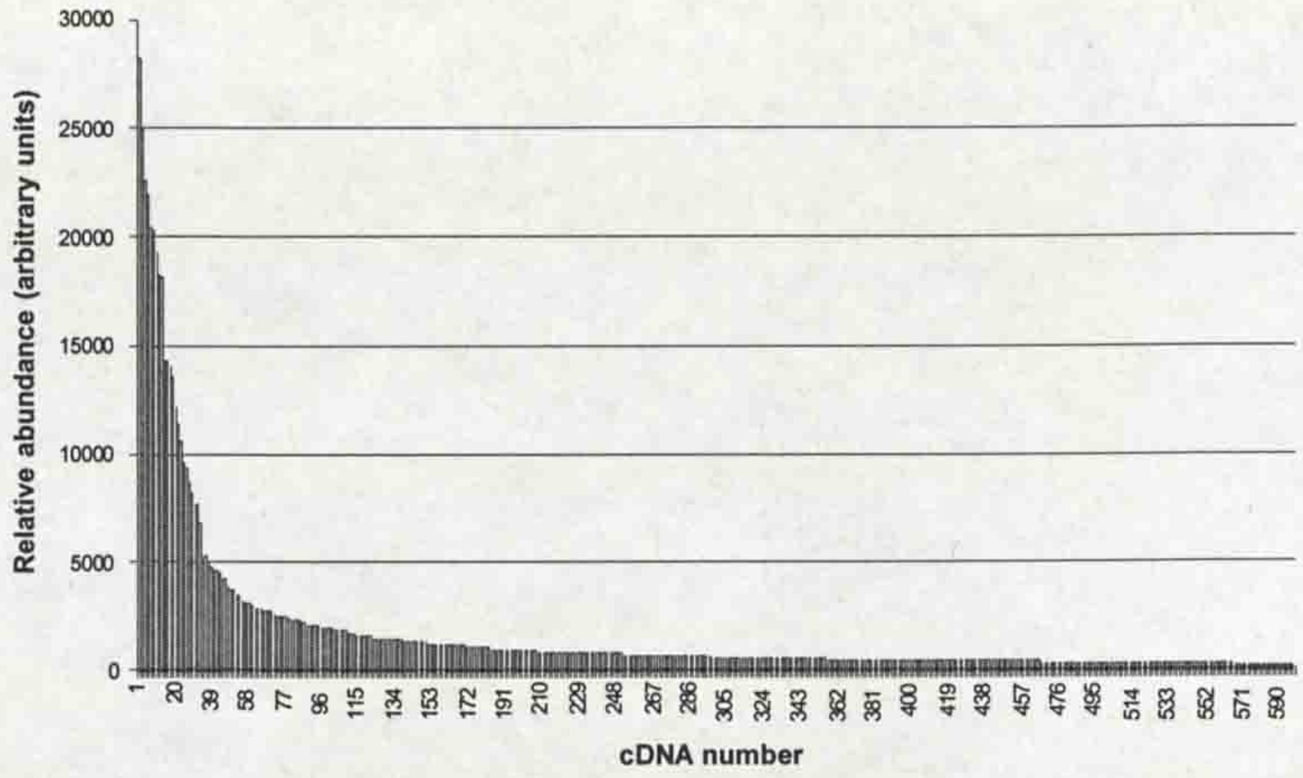


Figure 3.6 Abundance distribution of cDNAs in MMTV-*neu/S100A4* metastasis.

Hybridisation signal intensities for cDNAs indicated the relative abundance of that sequence in the particular cell population. The distribution of high and low abundant sequences, as demonstrated here for the MMTV-*neu/S100A4* metastasis, was essentially the same for all four tumour samples. The majority of cDNAs are representative of low abundant sequences and very few exhibited a high level of expression. Some of the sequences expressed highly in the tumour samples are presented in Table 3.2.

	Proportion of message profile contributing to each abundance class			
Abundance class	**neu primary tumours**	***neu/S100A4* primary tumours**	***neu/S100A4* primary tumour**	***neu/S100A4* metastasis**
High (5%)	**47.5%**	**46.9%**	**50.9%**	**45.2%**
Medium (25%)	**34.7%**	**34.4%**	**31.7%**	**32.7%**
Low (70%)	**17.8%**	**18.7%**	**17.4%**	**22.1%**

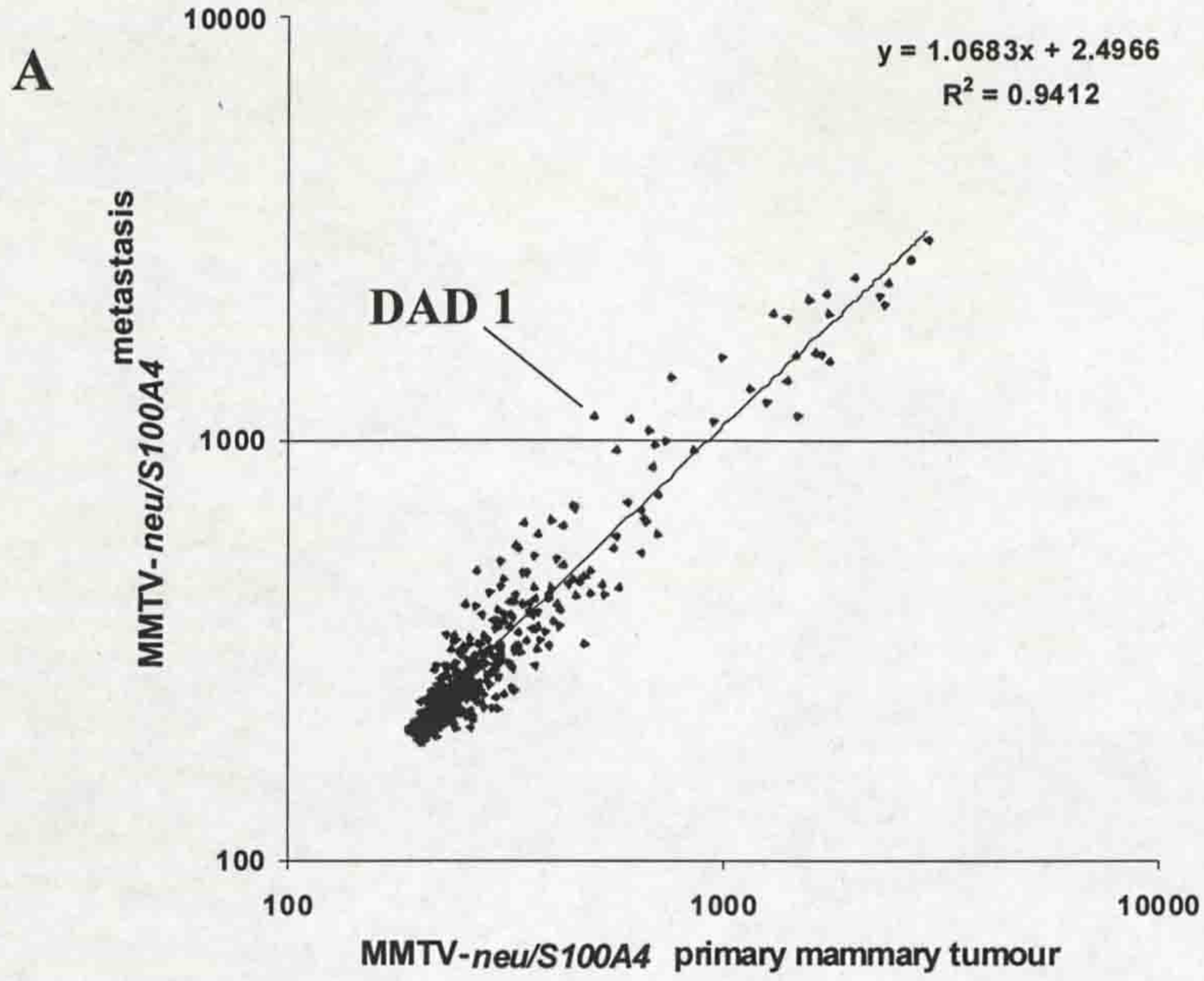
Table 3.1 Abundance distribution of cDNAs in the four tumour samples.

The distribution of high, medium and low abundant sequences observed for the 597 sequences (588 + 9 controls) is expected to be representative of the abundance distribution in the whole mRNA profile. Approximately 5% of cDNAs are high abundant sequences and make up 45-50% of the message profile (total hybridisation signal), 25% of sequences contribute to 32-35% of the total message and are of medium abundance. Low abundant sequences consist of 70% of the different sequences and contribute to the remaining 17-22% of the message profile.

Sequence identification	Accession number	Hybridisation signal MMTV-*neu* pooled	MMTV-*neu/S100A4* pooled	MMTV-*neu/S100A4* primary	MMTV-*neu/S100A4* Metastasis
microsomal glutathione S-transferase (MGST1; GST12)	J03752	26902	28095	29349	30022
HR21spA; DNA double-strand break repair protein	D49429	26184	26258	26866	26922
cathepsin B1 (CTSB)	M14222	24801	24271	19840	24345
84-kDa heat shock protein (HSP84)	M36829	23765	25190	17266	19934
ubiquitin	X51703	23760	23760	23760	23760
glyceraldehyde-3-phosphate dehydrogenase (G3PDH)	M32599	23026	23366	23267	21094
transcription termination factor 1 (TTF1)	X83974	22778	25278	26526	26801
CD14 monocyte differentiation antigen precursor	M34510	21335	23636	11430	13177
40S ribosomal protein S29 (RPS29)	L31609	20937	22871	23760	23760
non-muscle myosin light chain 3 (MLC3NM; MYLN)	U04443	19736	18897	22530	22044
granulocyte-macrophage colony-stimulating factor receptor	M85078	19710	18976	14732	11427
interferon inducible protein 1	U19119	16069	9742	7362	7045
beta-actin	M12481	15820	16653	17411	15382
type I cytoskeletal keratin 18 (KRT1-18; KRT18)	M11686	15792	15917	12459	12356
neuroleukin	M14220	15491	16174	16132	16124
clusterin precursor (CLU); clustrin; apolipoprotein J (APOJ)	L08235	15421	20635	17128	22219
cathepsin D (CTSD)	X53337	14502	11706	9874	15779
HSP86; heat shock 86-kDa protein	M36830	14252	16580	9472	9472
split hand/foot gene	U41626	14218	13783	16655	16061
cathepsin H	U06119	14050	13635	13987	19614
glutathione S-transferase Pi 1 (GSTPIB)	D30687	13684	17127	7538	14106
nucleoside diphosphate kinase B; NM23-M2	X68193	13070	18455	15612	21532
defender against cell death 1 (DAD1)	U83628	12411	17115	5050	11369
heat shock 60-kDa protein1 (HSP60)	X53584	11870	9603	5738	4462
glutathione S-transferase 5 (GST5-5)	J04696	11501	16866	12946	19997
protease nexin 1 (PN-1)	X70296	10643	8781	9395	11054

Table 3.2 Selection of sequences of high abundance in the four tumour samples.

Hybridisation of cDNA expression arrays with probes derived from the four tumour samples revealed a similar pattern of highly expressed sequences. The relative level of expression of some of the most abundant sequences given is in arbitrary units, as determined by hybridisation signal intensities for these cDNAs. The control cDNAs expressed in the four tumour samples demonstrated similar hybridisation signal intensities which negated the need for hybridisation signal normalisation. Thus, the signal intensities given is of raw data.



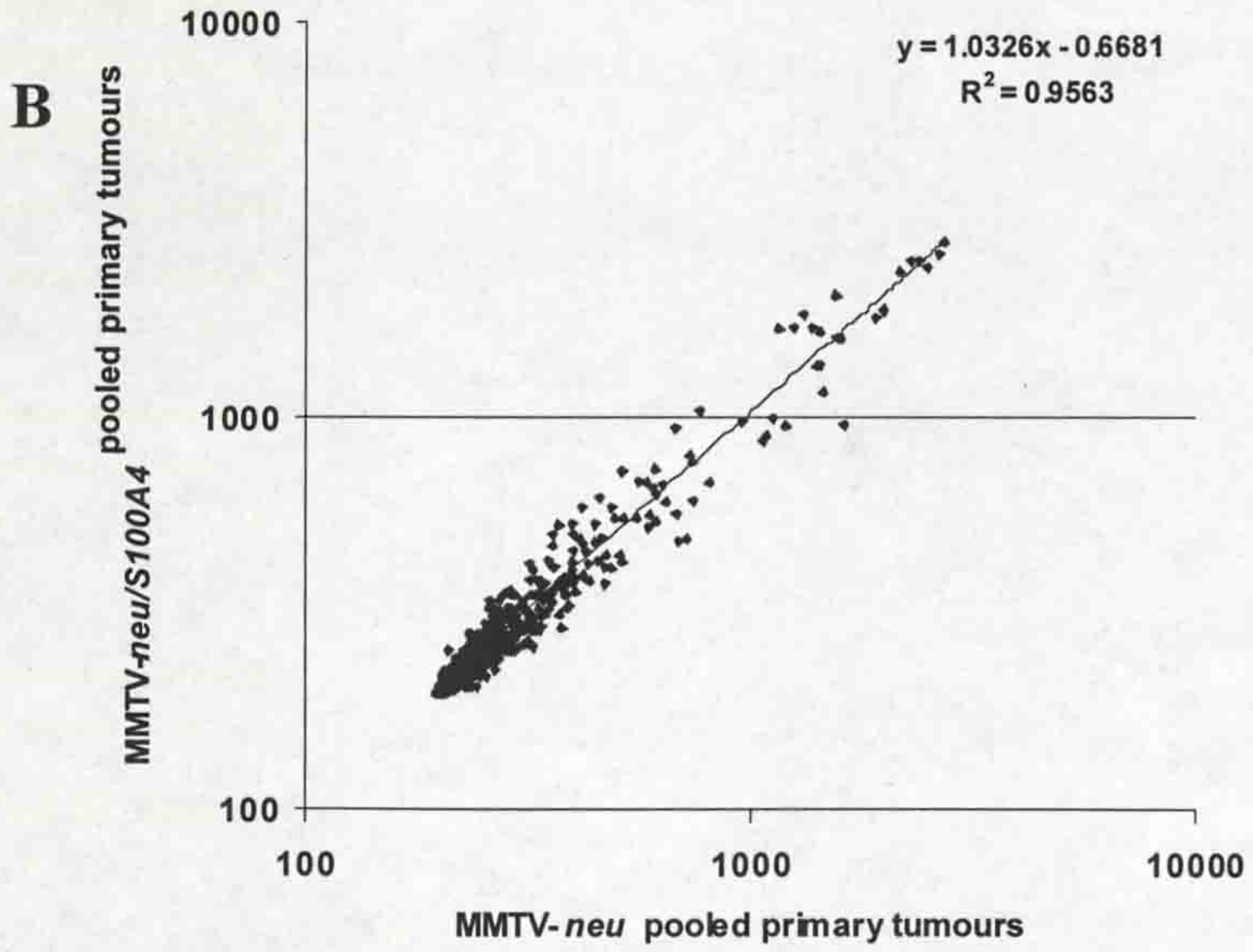


Figure 3.7 XY scatterplots of hybridisation signal intensity for all cDNAs screened in the two experiments.

Hybridisation signal intensities (converted to logarithmic scale) for all cDNAs were compared between the primary mammary tumour and metastasis (A), and the two pooled sets of primary mammary tumours (B). Pair-wise comparison of gene expression profiles between tumours indicate that the relative expression of the majority of these cDNAs is very similar. Only one cDNA was differentially expressed by > 2-fold between compared tumours, that of DAD 1 (indicated in A), which exhibited 2.3-fold higher expression in the metastasis relative to the MMTV-neu/S100A4 primary mammary tumour.

Sequence identification	Grid reference	Accession number	Met/PN Ratio	PN	Met
dystroglycan 1	E6m	U43512	**0.69**		
granulocyte-macrophage colony-stimulating factor receptor	E2e	M85078	**0.72**		
dual-specificity mitogen-activated protein kinase kinase 1 (MAPKK1)	B6a	L02526	**1.25**		
cyclin-dependent kinase 4	A6m	L01640	**1.26**		
Fas I receptor (Apo-1 antigen)	C3f	M83649	**1.26**		
BCL2 binding athanogene 1 (BAG1)	C1e	U17162	**1.27**		
A-Raf proto-oncogene	A3k	M13071	**1.28**		
Stat1; signal transducer and activator of transcription	B4d	U06924	**1.29***		
clusterin precursor (CLU); apolipoprotein J (APOJ)	C3b	L08235	**1.30***		
Stat6; signal transducer and activator of transcription 6	B4g	L47650	**1.31**		
insulin-like growth factor-IA	F3a	X04480	**1.33**		
cyclin-dependent kinase 6 inhibitor (p18-INK6); cyclin-dependent kinase 4 inhibitor C (p18-INK4C)	A7c	U19596	**1.35**		
plasminogen activator inhibitor	F7h	M33960	**1.35**		
cyclin-dependent kinase 7	A7a	U11822	**1.37**		
nucleoside diphosphate kinase B (NDP kinase B); NM23-M2	C4c	X68193	**1.38***		
serine protease inhibitor 2-2 (SPI2-2)	F7j	M64086	**1.39**		
BAX membrane isoform alpha	C1g	L22472	**1.39**		
DNA excision repair protein ERCC1	C6d	X07414	**1.40**		
BAD protein; BCL2 binding component 6 (BBC6)	C1d	L37296	**1.45**		
cathepsin H	F6i	U06119	**1.47**		
tumor necrosis factor 1 (55 kDa)	C5b	X57796	**1.51**		
ALG-2; calcium binding protein required for programmed cell death	C2i	U49112	**1.52**		
glutathione S-transferase 5 (GST5-5)	C2b	J04696	**1.54***		
cathepsin B1 (CTSB)	F6g	M14222	**1.55**		
cathepsin L	F6j	X06086	**1.57**		
cathepsin D (CTSD)	F6h	X53337	**1.60***		
p19ink4; cdk4 and cdk6 inhibitor	A7d	U19597	**1.60**		
PA6 stromal protein; RAG1 gene activator	C6a	X96618	**1.61**		
heat shock 86-kDa protein	B1d	M36830	**1.68**		
interleukin-converting enzyme (ICE)	F7a	L28095	**1.69**		
G2/mitotic-specific cyclin B1 (CCNB1; CYCB1)	A6c	X64713	**1.70**		
Golgi 4-transmembrane spanning transporter	B2d	U34259	**1.82**		
glutathione peroxidase (plasma protein); selenoprotein	C1l	U13705	**1.82**		
glutathione S-transferase Pi 1 (GSTPIB)	C2d	D30687	**1.87***		
defender against cell death 1 (DAD1)	C3d	U83628	**2.26***		

Table 3.3 Differentially expressed cDNAs in MMTV-*neu/S100A4* primary mammary tumour and metastasis.

Quantified hybridisation signal intensities for all cDNAs were compared between the MMTV-*neu/S100A4* (PN) primary mammary tumour and metastatic (Met) lung tumour. Differences in expression ratios (Met/PN) of <0.5 and >2 were used, initially, to select differentially expressed sequences, reduced criteria for selecting differential expression of >1.25 and <0.80 was also applied. * indicates cDNAs differentially expressed in both screening experiments. Hybridisation signals for each cDNA is given for each tumour.

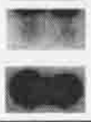
Sequence identification	Grid reference	Accession number	PN/N ratio	N	PN
interferon inducible protein 1	D4k	U19119	**0.61**		
vimentin; intermediate filament protein	F6d	X51438	**0.69**		
Stat1; signal transducer and activator of transcription	B4d	U06924	**0.72***		
phospholipase A2	G6	D78647	**0.75**		
cellular retinoic acid binding protein II (CRABP-II)	D6e	M35523	**0.78**		
cathepsin D (CTSD)	F6h	X53337	**0.80***		
tumor necrosis factor beta (TNF-beta); lymphotoxin-alpha	F4h	M16819	**0.80**		
glutathione S-transferase Pi 1 (GSTPIB)	C2d	D30687	**1.27***		
14-3-3 protein eta	B7g	U57311	**1.27**		
microsomal glutathione S-transferase (MGST1)	C2a	J03752	**1.29**		
TIMP-2 tissue inhibitor of metalloproteinases-2	F7m	X62622	**1.33**		
defender against cell death 1 (DAD1)	C3d	U83628	**1.38***		
nucleoside diphosphate kinase B (NDP kinase B); NM23-M2	C4c	X68193	**1.41***		
structure-specific recognition protein 1 (SSRP1)	B1h	S50213	**1.45**		
clusterin precursor (CLU); apolipoprotein J (APOJ)	C3b	L08235	**1.47***		
glutathione S-transferase 5 (GST5-5)	C2b	J04696	**1.47***		
40S ribosomal protein S29 (RPS29)	G21	L31609	**1.49**		

Table 3.4 Differentially expressed cDNAs in MMTV-*Neu* and MMTV-*neu/S100A4* pooled primary tumour sets.

Quantified hybridisation signal intensities for all cDNAs were compared between MMTV-*neu* (N) and MMTV-*neu/S100A4* (PN) pooled primary mammary tumours. Differences in expression ratios (PN/N) of <0.5 and >2 were used to select differentially expressed sequences, however no sequences demonstrated this level of differential expression. Reduced criteria for selecting differential expression of >1.25 and <0.8 was therefore applied. * indicates cDNAs differentially expressed in both screening experiments. Hybridisation signals for each cDNA is given for each tumour.

Chapter 4.

Production and Sequence Characterisation of Subtracted Libraries.

The previous chapter described the use of cDNA array hybridisation screening for the discovery of genes whose expression was associated, in some way, with S100A4-related metastatic progression. Several interesting sequences were identified that were highly abundant in each of the different tumours analysed which may contribute to the similar histology observed between tumours from the two strains of transgenic mice (MMTV-*neu* and MMTV-*neu/S100A4*). However, the limited nature of these screening experiments meant that the expression of very few cDNAs were analysed and as a consequence very few differentially expressed sequences were identified. It is expected that tumours demonstrating different phenotypic outcomes would exhibit different expression profiles reflecting their ability to metastasise or not. Such differential expression was not detected using the array screening.

Suppression Subtractive Hybridisation (SSH) has been applied to study the gene expression patterns of different tumours more directly. The efficiency of this method for cloning both high and rarely transcribed differentially expressed sequences has been demonstrated (Gurskaya *et al.*, 1996). Messenger RNA abundance between two primary mammary tumours of different metastatic potential and of a primary mammary tumour and a metastasis from the same animal have each been compared by SSH to selectively identify differentially expressed sequences that may influence the S100A4-related metastatic progression that is observed in these transgenic mice. This and subsequent chapters describe the creation and characterisation of subtracted cDNA libraries and the characterisation of candidate differentially expressed mRNAs.

4.1 Selection and histological analysis of tumours

The histology of a panel of primary mammary tumours from both strains of transgenic mouse was assessed to identify suitable candidate tumours for performing subtractive hybridisation experiments. The selection criteria were:

i) transgenic status, MMTV-*neu* and MMTV-*neu/S100A4,*

ii) tumour phenotype, primary mammary disease with or without lung metastasis,

iii) there being frozen samples of the same tumour available for RNA extraction,

iv) histology of the tumours being representative of the majority of mammary tumours from each transgenic type,

v) histological analysis of haematoxylin and eosin stained, frozen sections to verify the same histological characteristics as the methacarn-fixed, paraffin-embedded samples.

Two primary mammary gland tumours were selected for gene expression analysis; one was from an MMTV-*neu* mouse, the other, from an MMTV-*neu/S100A4* mouse. The tumour grade and histology were assessed (B. Shoker, personal communication) based on guidelines of the National Breast Screening Program (NHSBSP, 1997), and involved analysis of the number of tumour cells undergoing mitosis, nuclear pleomorphism and tubule (mammary duct) formation of the epithelial cells. Both the primary tumours selected demonstrated similar histological features and were classified as grade 3 primary infiltrating ductal carcinomas (IDC, Figure 1.1). These were also typical of the majority of tumours from these MMTV-*neu* and MMTV-*neu/S100A4* transgenic mice and similar to human IDC that are often associated with high expression of c-*erb*B-2, as discussed earlier in chapter 1.

One primary mammary tumour was from an MMTV-*neu* mouse which under-went necropsy at 17 months of age, 14 weeks after primary neoplasia was first detected. Two primary mammary tumours had developed, both were grade 3 with the same solid growth pattern marked with necrotic centres and both stained positively for c-*erb*B-2. A single clump of c-*erb*B-2 staining tumour cells, 0.2 mm in size (tumour diameter) was detected encapsulated in a lung blood vessel. That these cells stained for c-*erb*B-2 suggested that they were derived from one of the two primary tumours.

The second primary mammary tumour was from a bitransgenic MMTV-*neu/S100A4* mouse and was the same bitransgenic primary tumour used for cDNA array hybridisation. This mouse was 13.5 months old at necropsy, 18.5 weeks after the identification of palpable mammary tumours. At least 10 mammary gland lesions

developed, each was a grade 3, solid carcinoma with necrotic regions. The primary mammary tumour selected for analysis, tumour A, was the largest identified in this mouse (12 mm), however the sizes of other tumours varied down to as small as 1.5 mm. Multiple metastases developed, both macroscopic (as big as 4 mm) and microscopic (as small as 0.1 mm), which covered approximately 20% of the lung tissue area (see figure 1.2). All the metastases were too small to be graded formally, however they showed the same histopathological characteristics as the primary tumours. A third tumour selected for subtractive hybridisation was a lung metastasis from this same mouse; this tumour was also utilised in cDNA array screening.

We identified frozen sections of the MMTV-*neu* primary tumour that consisted of 90-95% epithelial-derived tumour cells and sections from the MMTV-*neu/S100A4* primary tumour and metastatic tumour that were both 85-90% pure tumour cells. This purity is only slightly lower than that achieved by manual microdissection of frozen tissue sections from the same tumours. Typically 95-100% tumour cells can be isolated from these tissue sections by microdissection, however only small numbers of cells can be extracted at a time. Whole histological sections from the tumours were therefore used as the source of RNA for subtractive hybridisation experiments. Histological sections cut from the two primary tumours contained approximately 50,000 and 100,000 cells respectively. This did not reflect a difference in size of these mammary gland lesions, only a difference in size of the portion of the tumour frozen in liquid nitrogen. The metastatic tumour sections consisted of approximately 20,000 cells as this tumour was smaller than either of the two primary tumours.

4.2 RNA extraction & integrity.

Poly $(A)^+$-containing messenger RNA (mRNA) was extracted from several sections, totalling approximately 240,000-300,000 cells. The integrity of RNA extracted from these sections was analysed by RT-PCR, amplifying the 5' region of the HPRT mRNA following oligo(dT) primed cDNA synthesis, the utility of this assay for determining the quality of RNA extracted was described in chapter 3. Figure 4.1 demonstrates intact RNA was isolated from each of the three frozen tumours.

4.3 Production of subtracted libraries.

All stages of each method were performed, initially, for RNA derived from both primary tumours, creating tester and driver cDNA for each tumour. This enabled both forward and reverse orientated subtractions to be performed between these two tumours. This was repeated using cDNA synthesised from the two bitransgenic tumours, the primary mammary tumour and the lung metastatic tumour. Thus performing two subtractive hybridisations and creating four subtracted libraries (Figure 4.2).

SMART technology was used to prepare a sufficient amount of cDNA from the limited amount of RNA that was available prior to performing subtractive hybridisation. SMART cDNA was synthesised from approximately 100,000 cells derived from each tumour. Amplification of one tenth of this cDNA was carefully optimised to avoid over amplification and therefore misrepresentation of individual cDNA molecules in the amplified cDNA library. This is a critical step in the preparation of cDNA prior to performing suppression subtractive hybridisation (SSH). Between 17 and 20 cycles of amplification were optimal for the three tumour preparations of cDNA (Figure 4.3), producing a visibly amplified smear of DNA fragments of a range of sizes between 0.2kb and 5-6kb. Interestingly, a cDNA fragment 2.5kb in size was the most prominently amplified cDNA fragment, which probably represents the most abundantly expressed message in all three tumours (Figure 4.3). This cDNA molecule is therefore expected to be derived from the mammary epithelial tumour cells and not the surrounding normal tissue, unless the small proportion of normal mammary gland and lung tissue express the same sized cDNA fragment to the same level. The identity of this cDNA was unknown and was not investigated as its abundance appeared to be the same in each sample. This cDNA fragment was used as an internal control for all subsequent SMART cDNA amplifications from these tumours, as over amplification of the cDNA population occurred this 2.5kb fragment became less pronounced (Figure 4.3). Failure to detect this band also suggested sub-optimal PCR or potentially degraded cDNA. Other discrete bands were also faintly visible.

Digestion of amplified cDNA molecules using *Rsa* I reduced the sizes of cDNA molecules from between 0.2kb-6kb to blunt ended fragments, of up to approximately 2 kb in size (Figure 4.4). The distinct 2.5kb band disappeared suggesting successful digestion.

The efficiency of adaptor ligation to both ends of the digested tester cDNA was analysed by PCR, using either two glyceraldehyde 3-phosphate dehydrogenase (G3PDH) gene specific primers, or a single G3PDH primer in combination with PCR primer 1, identical to a sequence present in both adaptors (Table 2.1). Electrophoresis of PCR products after 20 and 25 cycles revealed equal band intensities using both combinations of primers (Figure 4.5). Clontech suggest that sufficient ligation of adaptors has occurred if the band intensity produced following G3PDH-PCR primer 1 amplification is no less than 1/4 of that achieved with the two G3PDH primers. Approximate equal intensity of bands (Figure 4.5) therefore suggested efficient adaptor ligation was achieved.

Two rounds of PCR (primary and secondary) of hybridised cDNA molecules were used to enrich for tester specific cDNA fragments in both subtractions. Amplification was performed in parallel with non-subtracted tester cDNA only. The amplified subtracted molecules were observed as a smear of cDNA fragments ranging in size between approximately 250bp and 2000bp (PN, Met: Figure 4.6), non-subtracted tester cDNA secondary amplification yielded high molecular weight DNA smear (PN-c, Met-c: Figure 4.6). Some discrete bands were visibly differentially amplified in different subtracted or non-subtracted samples (Figure 4.6). This suggests enrichment of potentially differentially expressed cDNA fragments. Note the enrichment for a 1660 bp fragment in PN and a 910 bp fragment in Met and also the loss (removal by subtraction) of a 830 bp and a 430 bp fragment that is observed in both non-subtracted tester controls but not in either of the subtracted samples (figure 4.6).

Secondary PCR products from 12 cycles of amplification were cloned directly into vector pCR2.1^{TM} (InVitrogen) creating the four subtracted cDNA libraries (Figure 4.2). Each library containing cDNA fragments of genes preferentially expressed in each tumour relative to one partnering tumour.

4.4 Characterisation of subtracted libraries.

Characterisation of the four subtracted cDNA libraries, primarily in terms of sequence analysis and expression screening of individual clones, is presented below and in subsequent chapters. These experiments will serve to determine the efficiency of subtractive hybridisation performed, the complexity of subtracted libraries created and the identification of specific differentially regulated sequences.

4.4.1 Size of subtracted libraries.

The four subtracted cDNA libraries were of the same approximate size, containing an average of 4.71×10^5 transformants (range of $3.75 \times 10^5 - 6.45 \times 10^5$) based on colony counts of a proportion of cloned PCR products that were transformed. The Met subtracted cDNA library was larger than the N, P and PN subtracted libraries, although all were of the same order of magnitude (Table 4.1). Colony hybridisation experiments were performed to monitor the success of subtractive hybridisation. Approximately 1000 colonies from each of the four subtracted libraries and from non-normalised and non-subtracted cDNA libraries derived from the same three tumours were screened with a S100A4 (p9Ka) cDNA probe. Six positive colonies were identified, 4 in the Met subtracted library and 2 in the P subtracted library (Table 4.2). Some colony filters from all libraries were also screened using actin and HPRT cDNA probes, a single positive colony was detected for each cDNA, in the N and Met non-subtracted cDNA libraries respectively (Table 4.2).

4.4.2 Colony-PCR of subtracted clones.

A total of 828 white colonies were picked (approximately 200 colonies from each library; Table 4.1), cultured overnight and the cloned inserts were amplified by colony-PCR. Successful amplification of 768 (92.7%) of these cultures was achieved, producing single amplified DNA fragments (see Figure 4.7 as an example of colony-PCR). The remaining 7.3% represented failed PCR or mixed colony populations observed by blank lanes or multiple PCR products respectively. The sizes of cDNA cloned inserts amplified ranged between approximately 200 bp and 2.3 kb which is consistent with the size of cDNA fragments amplified by secondary PCR (Figure 4.6), the average insert size was 790 bp.

4.4.3 Nucleic acid sequence of subtracted clones.

To investigate the identity of each clone that was picked, all 768 single amplified cDNA fragments were sequenced and the DNA sequence was submitted to nucleic acid sequence databases to compare with published sequences for known genes and expressed sequenced tags (ESTs).

Tables A.2 and A.3, in the Appendix, present the sequence identification of all clones from the four subtracted cDNA libraries, and a complete breakdown of database homology search results, for all clones from the four libraries, is presented in Table 4.1. This data includes the number of clones whose DNA sequence showed a high degree (usually greater than 80%) of identity to known genes and ESTs, and also the number which represent potentially novel sequences (those that produced sequence of a high quality yet showed no match to any published sequence in the two databases). In total, 387 (50.4%) clones showed similarity with known genes, 267 (34.8%) clones were similar to EST sequences and 40 (5.2%) clones represented potentially novel sequences. The remaining 74 (9.6%) clones produced poor sequence information which was usually due to poor quality or insufficient template DNA in the sequencing reactions.

Some clones initially produced poor sequence information due to the presence of a poly(dT)$_n$ sequence adjacent to the sequencing primer. Repeat sequencing of these clones from the other direction, with the other sequencing primer, produced readable sequence which often terminated with a poly(dA)$_n$ tail (depending on the insert size). These sequences were derived from the poly(dA)$_n$ tail found at the 3' end of mRNA fragments and therefore indicate the amplification and cloning of 3' end cDNA fragments. 31% of clones contained such sequences (Table 4.1).

Figure 4.8 displays the high quality of sequence that was obtainable from sequencing of the cloned cDNA, three examples are shown, for regions of clones N2E4, P1H12 and P2D6. Database search results are presented for these three clones (Figures 4.9-11) and serve as examples of the identification of a known gene (P1H12, osteopontin), a clone with sequence identity to an EST sequence (N2E4) but not to any sequence in the database of known sequences, and a clone (P2D6) which has no similarity with any sequence in either database as at 02/06/2000.

Some database search results identified clones with homology to several different known genes of apparently similar sequence. Closer examination of the clone sequence, the search results and related literature was often required to further establish the sequence identity of the clone. For example, the sequence of clone Met2F11 possessed 91% sequence similarity with several different known genes. These included pig, soluble angiotensin-binding protein (accession number dbj|D11336), pig endopeptidase 24.16 (dbj|AB000425), rat ribosomal protein L36a

(gb|M19635), human ribosomal protein L44 (ref|NM_001001), pig beta-1, 2-N-acetylglucosaminyltransferase II (emb|Y09537), mouse alpha-galactosidase A (gb|U50715) and rat V1b vasopressin receptor (dbj|AB042197). These sequences have quite diverse and apparently unrelated cellular functions, yet they share sequence identity. The sequence identity of clone M2F11 is likely to be a mouse ribosomal protein, since the region of sequence similarity to the other matches was actually to intron sequences.

Clone P1B9 demonstrated a high degree of sequence identity (93%, 234/240 nucleotides match) to mouse endothelial monocyte activating polypeptide I (gb|U41341) and partial similarity to rabbit calgizzarin (gb|D10586) and human S100C (gb|D49355, which is now called S100A11). These represent species specific versions of the same cDNA.

The sequence of clones P2B10 and P1H12 (see Figures 4.9) have 96-98% (682/693) homology to different named mouse genes, minopontin (emb|X13986), osteopontin (emb|X51834), secreted phosphoprotein 1 (Spp1, ref|NM_009263) and early T-lymphocyte activation 1 protein (emb|X16151). These all appear to represent the same gene, osteopontin, sequenced from the same species by independent groups.

4.4.3.1 Multiple cDNA sequences/clone

The sequencing of two independent clones, by forward and reverse directional sequencing, revealed the presence of two different cDNA sequences. Clone N1A4 contained an MMTV related sequence (gb|AF228550) adjacent to NonO (non-POU domain containing octamer binding protein; gb|S64860) and clone N1A7 contained cDNA fragments of both the transferrin (emb|X77158) and BRP39 (gb|X93035). Presumably these were blunt end ligated during the adaptor ligation, as, in both cases, an *Rsa* I restriction enzyme site separated the adjacent sequences. This *Rsa* I site was utilised to cleave the two fragments in order to assess the expression pattern of each sequence individually.

4.4.3.2 Sequence redundancy of clones in the subtracted cDNA libraries.

Of the 768 clones that were sequenced, 697 clones produced good quality sequence (Figure 4.8) and a total of 546 different sequences were identified. This was based on the number of clones showing sequence similarity to the same known or ESTs or by aligning clones exhibiting overlapping regions of identical sequence

into contigs (Figure 4.12). The majority of these sequences, 499, were represented as singletons in that they were only picked and sequenced once in an individual library. The remaining 47 different sequences were identified multiple times and represented a total of 198 different clones (Table 4.1). Subtracted cDNA libraries P, PN and Met (Figure 4.2) exhibited the same relative level of sequence redundancy, in that 76%, 77% and 82% of clones, respectively, were non-redundant (singleton) sequences. The N subtracted library contained more sequences identified multiple times, and only 55% of clones were of non-redundant sequence.

4.4.3.3 Redundant sequences identified.

Some redundant sequences identified were of interest due to their prevalence in the different libraries (Table 4.3). It is expected that these are either high abundant sequences in the relevant tumours and/or are differentially expressed between these tumours. For example MMTV-related sequences, particularly in the N subtracted cDNA library, and mitochondrial genome sequences were common. A group of sequences previously identified as *neu* and *ras* specific tumour molecular markers (Morrison and Leder, 1994) were identified collectively in the four libraries, BRP39 and an EST sequence similar to human cellular retinol binding protein (CRBP) were specific to the N subtracted library whereas others, transferrin, kappa casein and WDNM 1 were found in multiple libraries. Pulmonary surfactant protein c (PSP-c) is reported to be a lung specific protein and this was reflected by the fact that it was only identified in the metastatic lung tumour derived subtracted library.

4.4.3.4 Non-Redundant sequences identified.

Many non-redundant sequences were identified in the four subtracted cDNA libraries; some possess interesting sequence identification in terms of the subsequent biological/cellular function of the proteins with respect to cancer development and metastatic progression. For example several sequences were identified with known roles in matrix degradation, cell adhesion, cytoskeletal structure, signal transduction and gene expression (see appendix).

4.5 Conclusions to chapter.

SSH has been performed between different tumours from the transgenic mice to identify sequences whose expression relates to metastatic progression associated

with S100A4 expression. The important technical aspects and success of this method in achieving this goal are discussed later (chapter 7).

Many different sequences were identified from these subtracted libraries. Significant differences in the number of times particular sequences were identified between libraries suggest possible differential expression of these sequences. The identification of non-redundant sequences in a library indicates normalisation of sequence abundance has occurred, however it is difficult to select which, if any of these are differentially expressed between the transgenic mouse tumours. This is because clones are expected to be identified which represent background, non-differentially expressed sequences, a feature common to these types of experiments. Thus the expression patterns of cDNA clones must be determined to identify which of these sequences are differentially expressed in these circumstances and which are of sufficient interest to warrant further investigation.

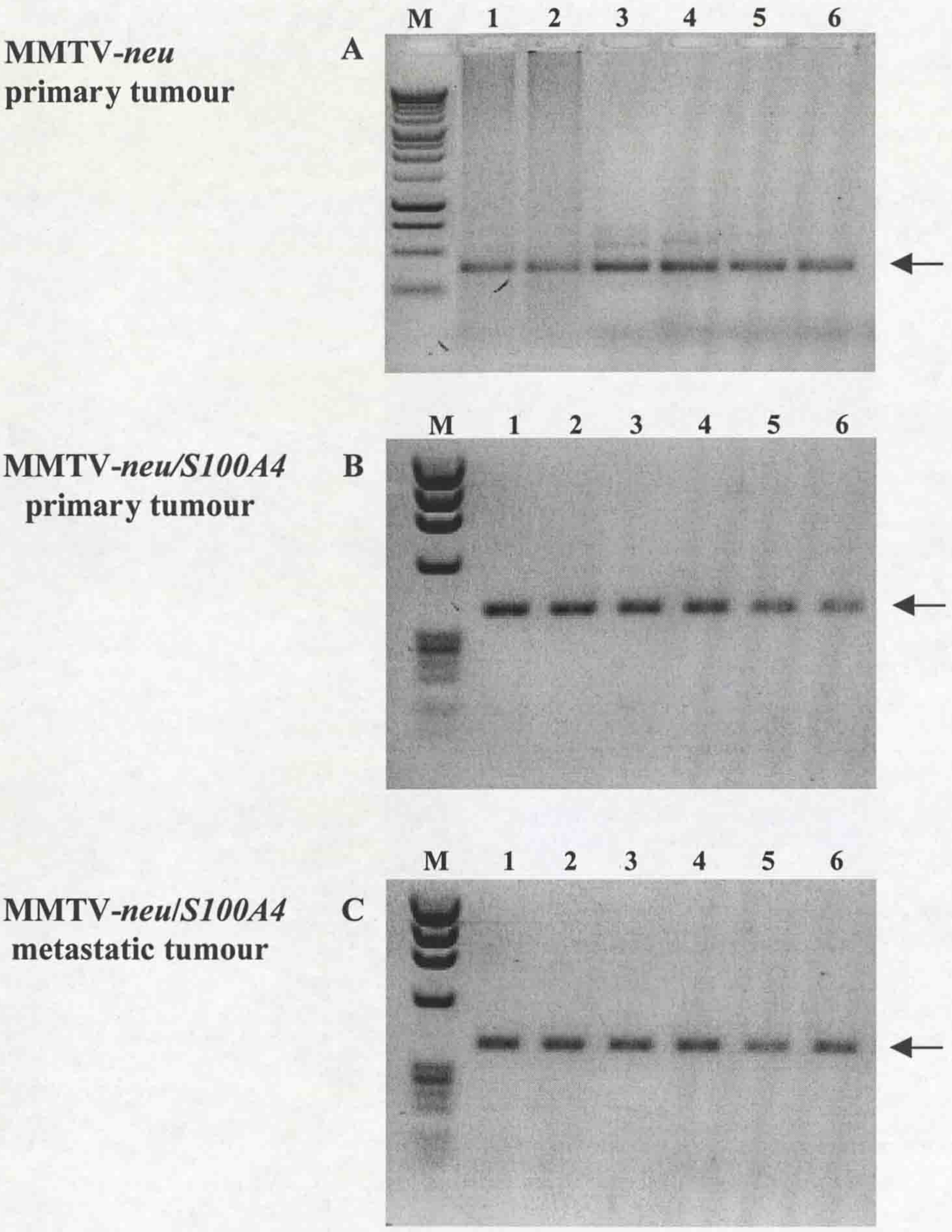


Figure 4.1 HPRT RT-PCR.

The integrity of poly(A)$^+$-containing RNA extracted from frozen tumour sections was analysed by RT-PCR for the 5' region of the *HPRT* gene. Poly(A)$^+$ RNA from an equivalent of approximately 100,000 cells was reverse transcribed into cDNA from each of the three tumours. Serial dilutions of cDNA containing the equivalent to approximately 2,000 cells (lane 1+2), 200 cells (lanes 3+4) and from 20 cells (lanes 5+6) were amplified by PCR. Positive amplification was detected by electrophoresis of PCR reactions on 1.2% agarose gel and was indicated by a single DNA fragment of 392bp (arrowed) when compared to DNA marker (M) such as 1 kb ladder (A) and *Hae* III digested ϕX174 DNA (B & C).

Subtraction 1

MMTV-*neu* mammary tumour	vs	**MMTV-*neu/S100A4* mammary tumour**
Subtracted cDNA library **N**		Subtracted cDNA library **P**

Subtraction 2

MMTV-*neu/S100A4* mammary tumour	vs	**MMTV-*neu/S100A4* metastasis**
Subtracted cDNA library **PN**		Subtracted cDNA library **Met**

Figure 4.2 Summary of the two subtractive hybridisation experiments performed between the three tumours and the four subtracted cDNA libraries created. The four libraries were designated a specific code according to the tumour and transgenic status of the tumour which they represent (**N** = *neu* primary tumour; **P** = *neu* and *S100A4* (*p9Ka*) primary tumour; **PN** = *neu* and *S100A4* (*p9Ka*) primary tumour, yet distinguishable from P; **Met** = metastasis). This code will be subsequently used when referring to a specific subtracted library, and will prefix the name of all subtracted clones discussed as an indication of the library from which they were derived.

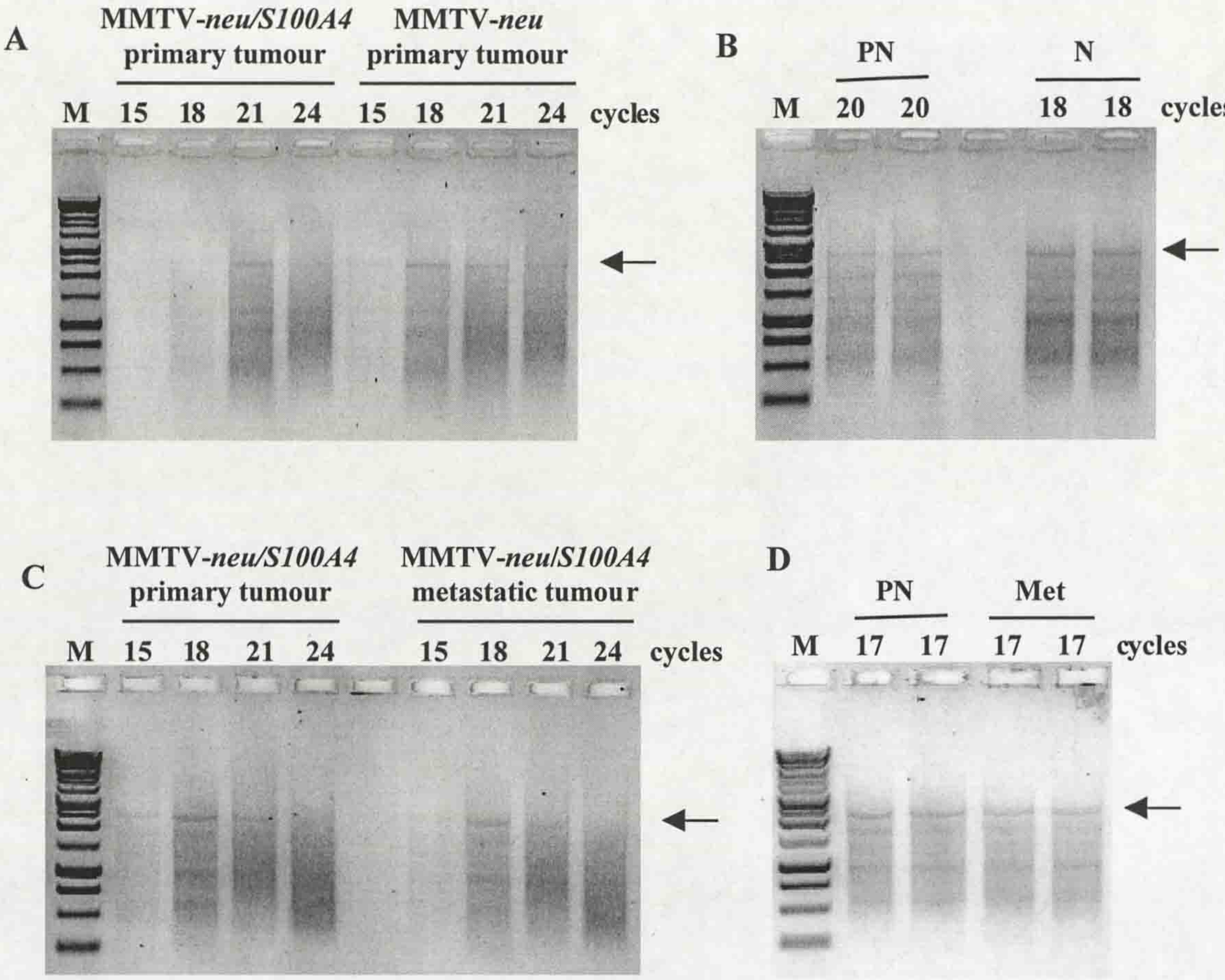


Figure 4.3 SMART PCR cDNA amplification.

SMART cDNA was synthesised from approximately 100,000 cells from each of the three tumours. Independent amplification was performed for the two separate subtractions, the two primary tumours (A & B) and the two bitransgenic tumours (C & D). Amplification of the cDNA was optimised to avoid over-amplification and therefore misrepresentation in the resulting cDNA library. For each cDNA, three reactions were prepared and amplified for 15 cycles, two reactions were stored at 4°C while the third was left in the PCR machine for the remaining 9 cycles. Ten microlitre aliquots were removed after 15, 18, 21 and 24 cycles and 5 µl of each was analysed by electrophoresis on a 1.2% agarose gel containing 0.5 µg/ml ethidium bromide(A & C). The two samples stored at 4°C were returned to the PCR machine for the required additional number of cycles (B & D). Amplified cDNA appears as a smear of cDNA fragments 0.5-6kbp in size. Abundantly expressed cDNAs appear as distinct bands. Over-amplification can be observed after 24 PCR cycles by the reduced intensity of detection of individual bands such as the 2.5kbp fragment (arrow). N = MMTV-neu primary tumour, PN = bitransgenic primary tumour, Met = metastasis.

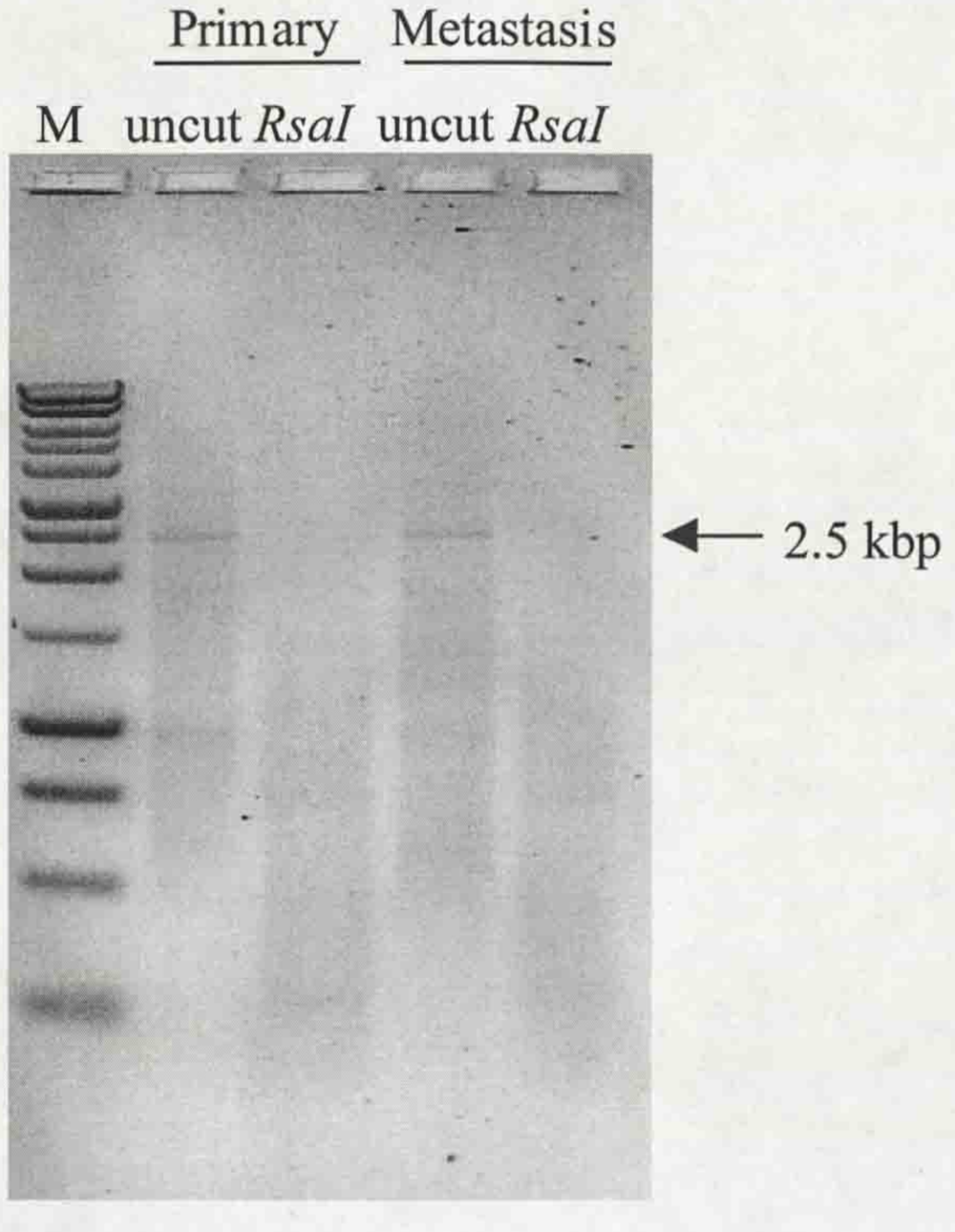


Figure 4.4 *Rsa*I digest of SMART PCR-generated cDNA.

Prior to performing subtractive hybridisation, cDNA from each tumour was digested with *Rsa*I, a restriction enzyme with a 4 bp recognition site. The size distribution of cDNA fragments is reduced from approximately 0.25 kb-5-6 kb to 0.1-2 kb, fragment sizes that have more favourable hybridisation kinetics. The DNA size marker (M) used was 1kbp ladder. The figure displays cDNA before (uncut) and after (*Rsa* I) digestion for the MMTV-*neu/S100A4* primary tumour and metastasis.

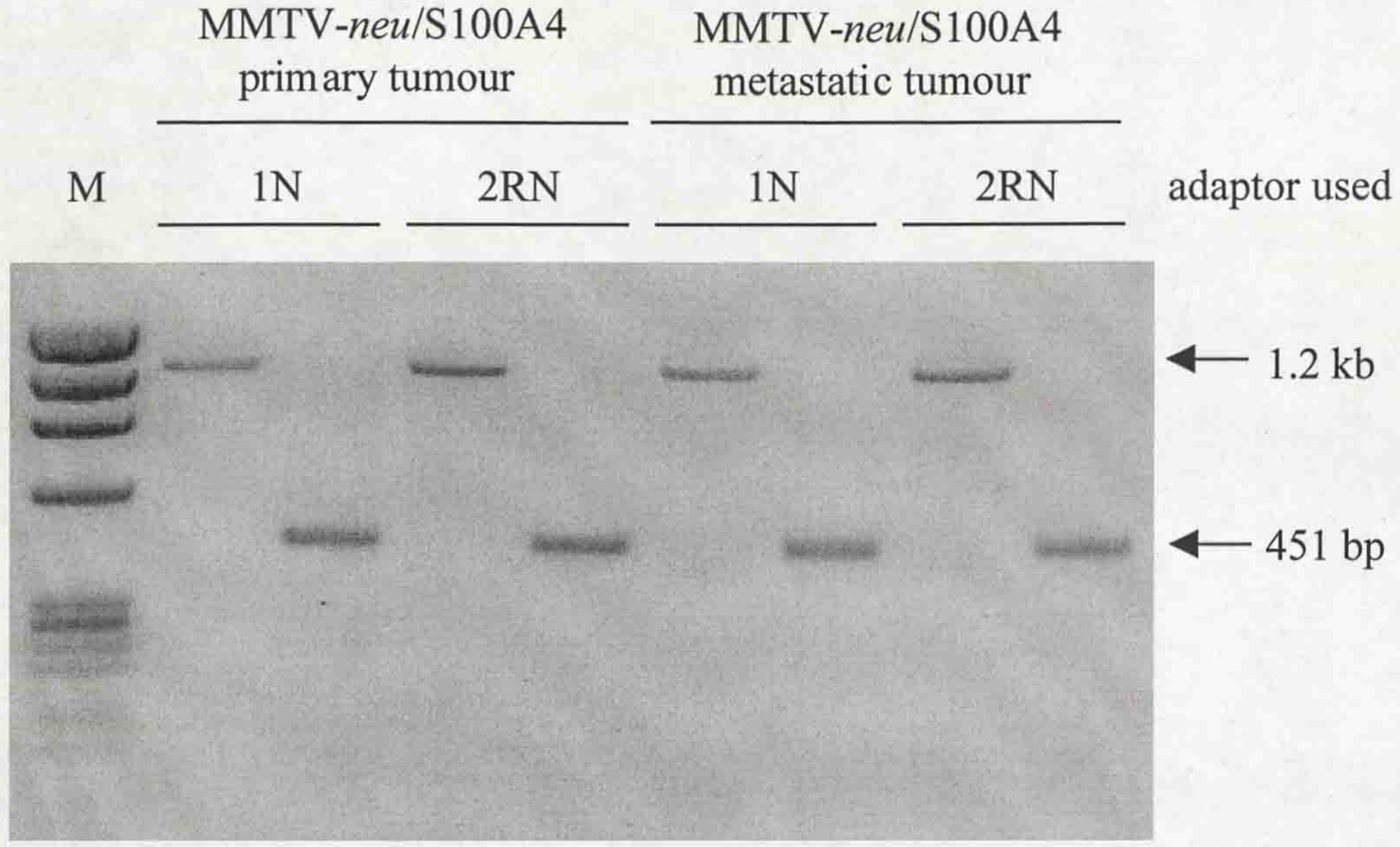


Figure 4.5 PCR to determine efficiency of adaptor ligation.

The efficiency of ligation of adaptors 1 and 2R to *Rsa* I digested tester cDNA was estimated by PCR. cDNA from each tumour sample was used to prepare both tester and driver cDNA so that subtractive hybridisation could be performed in both forward and reverse directions. Tester cDNA for each tumour was divided into two and a different adaptor (1N or 2RN) was ligated to each. Ligation reactions were amplified using PCR primer 1 with a gene specific G3PDH 3' PCR primer (table 2.1) producing a PCR product of 1.2 kb. Amplification using two gene specific G3PDH PCR primers produced a product of 451 bp. An equal band intensity, after 25 cycles, between the 1.2 kb fragment and the 451 bp fragment suggested efficient ligation of adaptors had occurred. Sizes were determined using *Hae* III digested ϕx174 DNA marker (M).

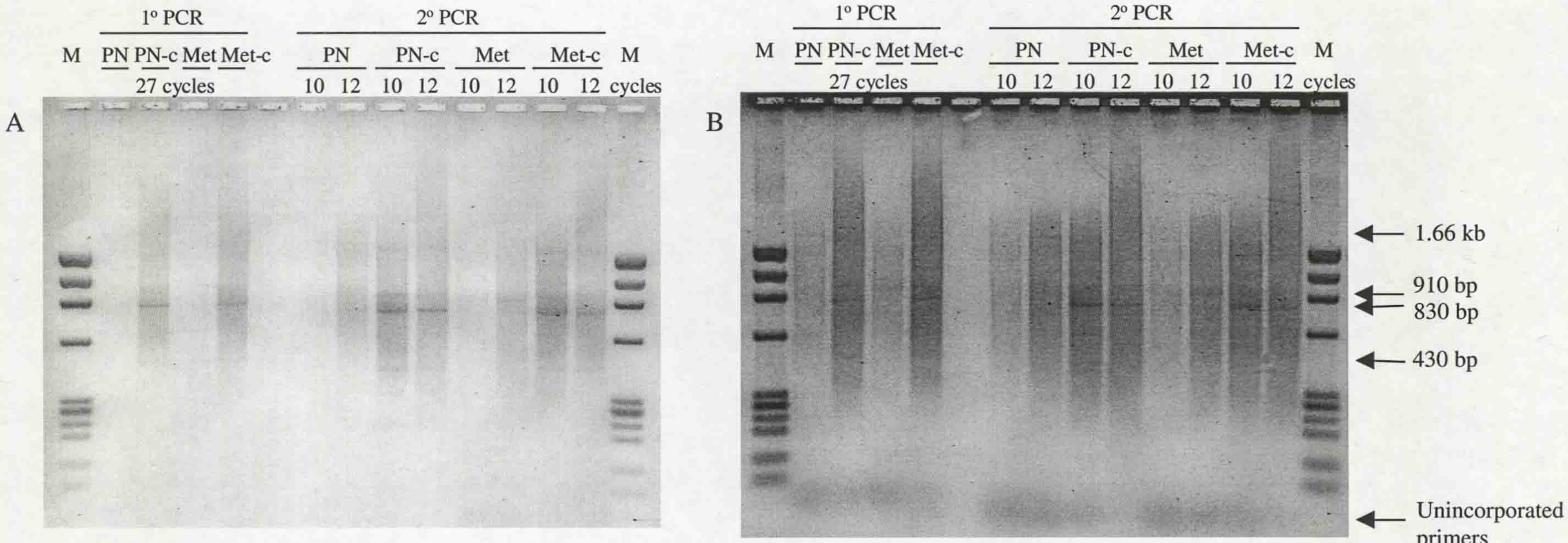


Figure 4.6 Primary & secondary PCR of cDNA.

Selective enrichment of subtracted cDNA fragments was obtained by two rounds (1º and 2º) of PCR amplification. This was performed alongside amplification of control non-subtracted tester cDNA (PN-c, Met-c) from each tumour. 27 cycles of 1º PCR was followed by 10 and 12 cycles of 2º PCR.

Images (A) and (B) are an example of 1º and 2º PCR of subtracted cDNA between MMTV-*neu*/*S100A4* primary tumour and metastasis. Image (A) is of the gel stained with 0.5 μg/ml ethidium bromide, image (B) is the same gel stained post-electrophoresis with Sybr Gold (1/10000 dilution). PN indicates amplification of primary tumour specific cDNA fragments, PN-c indicates amplification of unsubtracted primary tumour tester cDNA, Met indicates amplification of metastatic tumour specific cDNA fragments, Met-c indicates amplification of unsubtracted metastatic tumour tester cDNA. cDNA sizes were estimated by comparison to *Hae* III digested ϕX174 DNA marker (M).

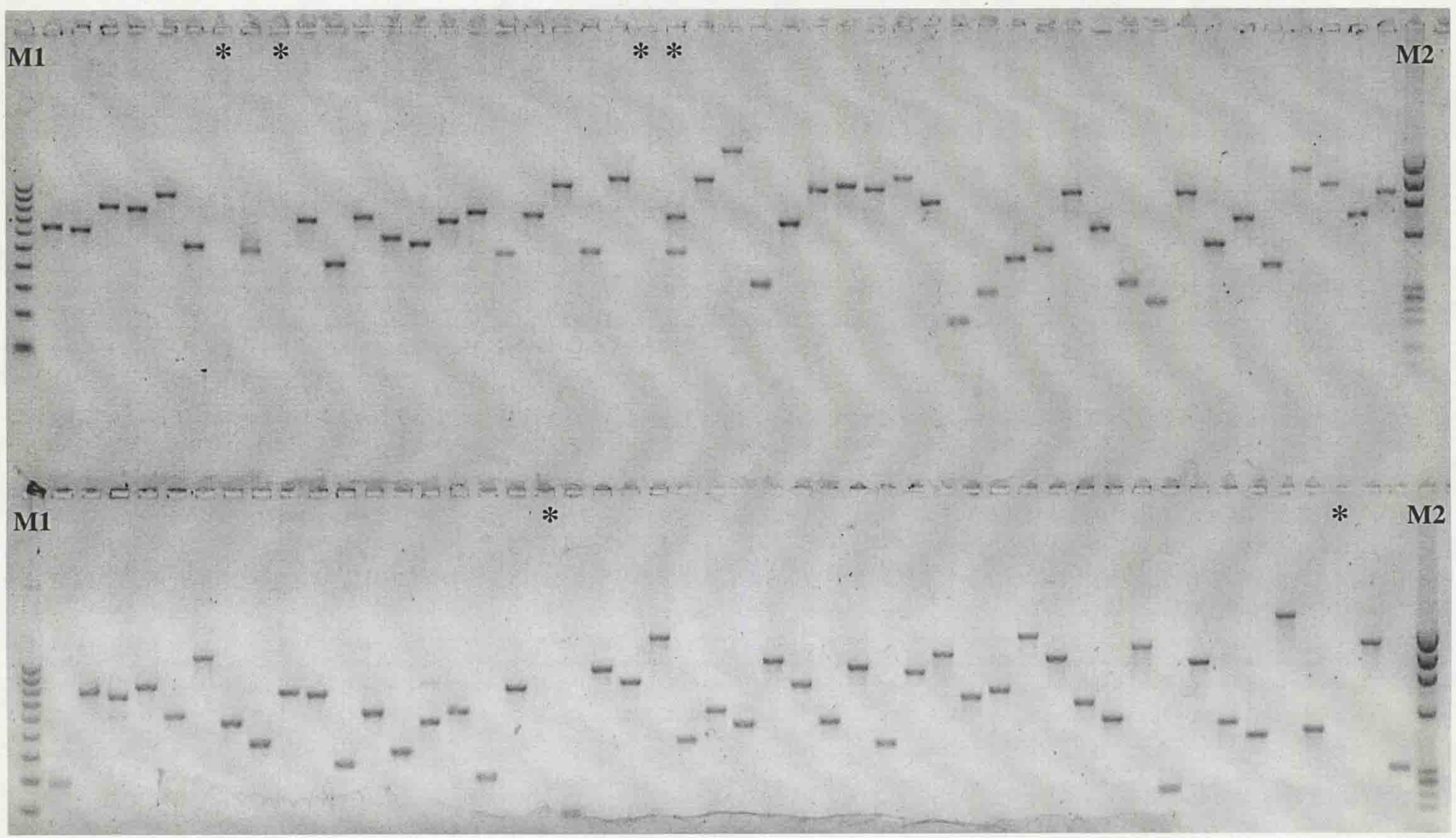


Figure 4.7 Colony-PCR of PN/2 clones.

A total of 828 colonies from 4 subtracted cDNA libraries were picked and cultured overnight in 96-well microtitre plates. Each bacterial culture was diluted 1/10 in sterile water and 2 µl was amplified by colony-PCR. One microlitre aliquots of each PCR were subjected to electrophoresis on a 1x TAE 1.2% agarose gel containing 0.5 µg/ml ethidium bromide, together with DNA markers 100 bp ladder (M1) and *Hae* III digested ϕX174 DNA (M2). 94% of reactions amplified single cDNA inserts. * indicates failed PCR or mixed culture populations (see text).

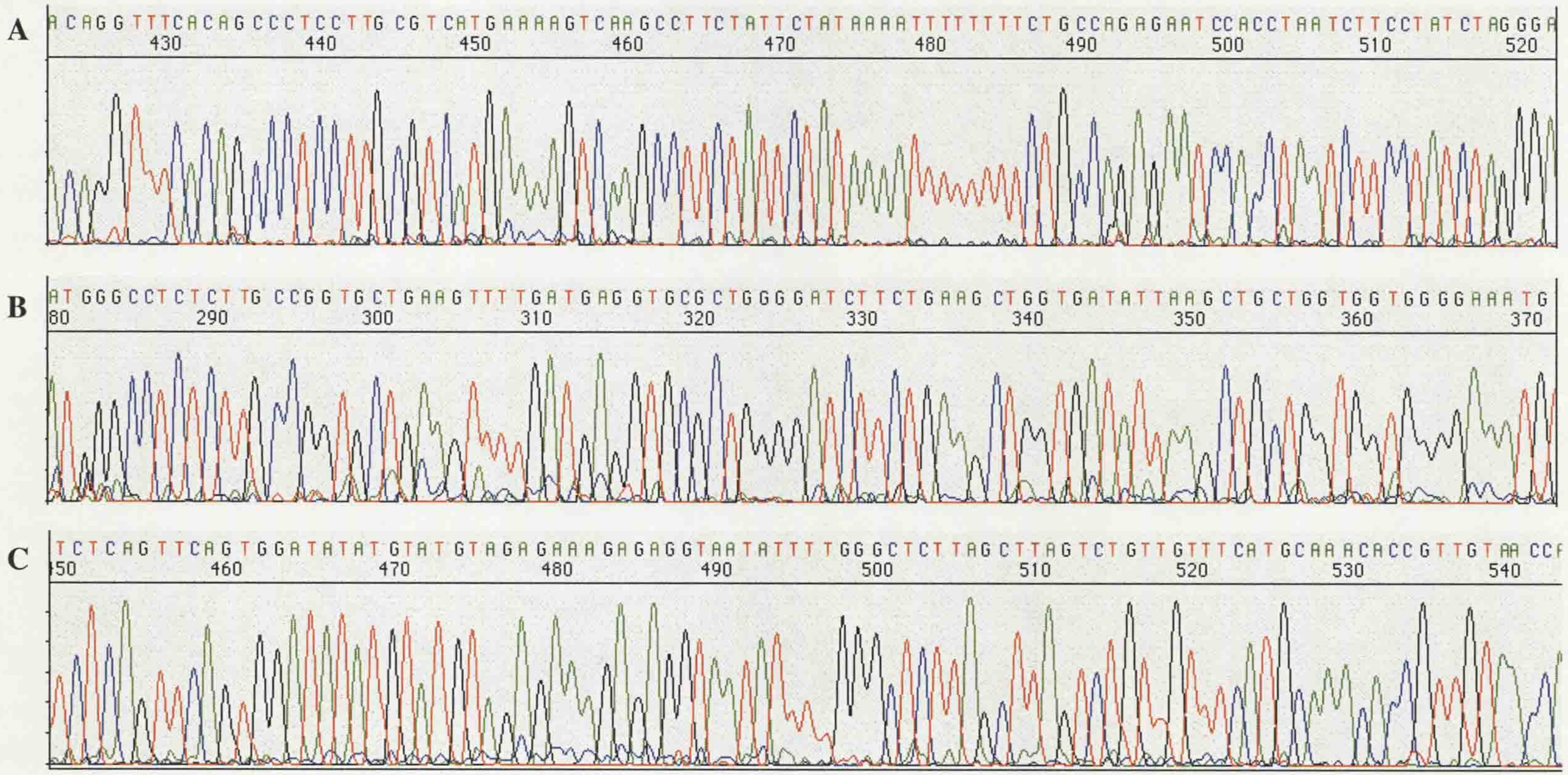


Figure 4.8 DNA Sequence of selected subtracted cDNA clones.

DNA sequence, together with electropherogram data from fluorescent DNA sequencing for 3 clones P2D6 (A), N2E4 (B) and P1H12 (C).

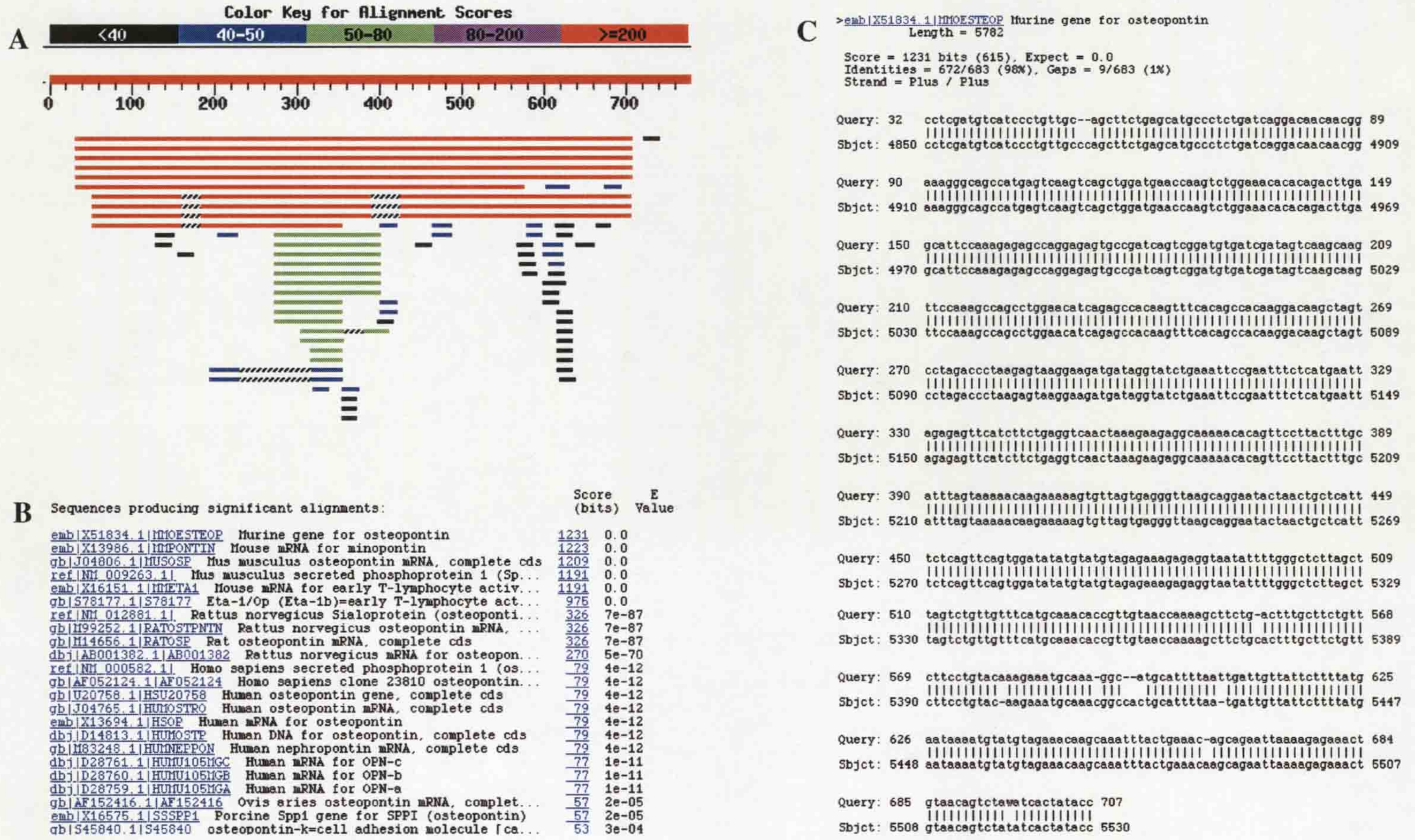
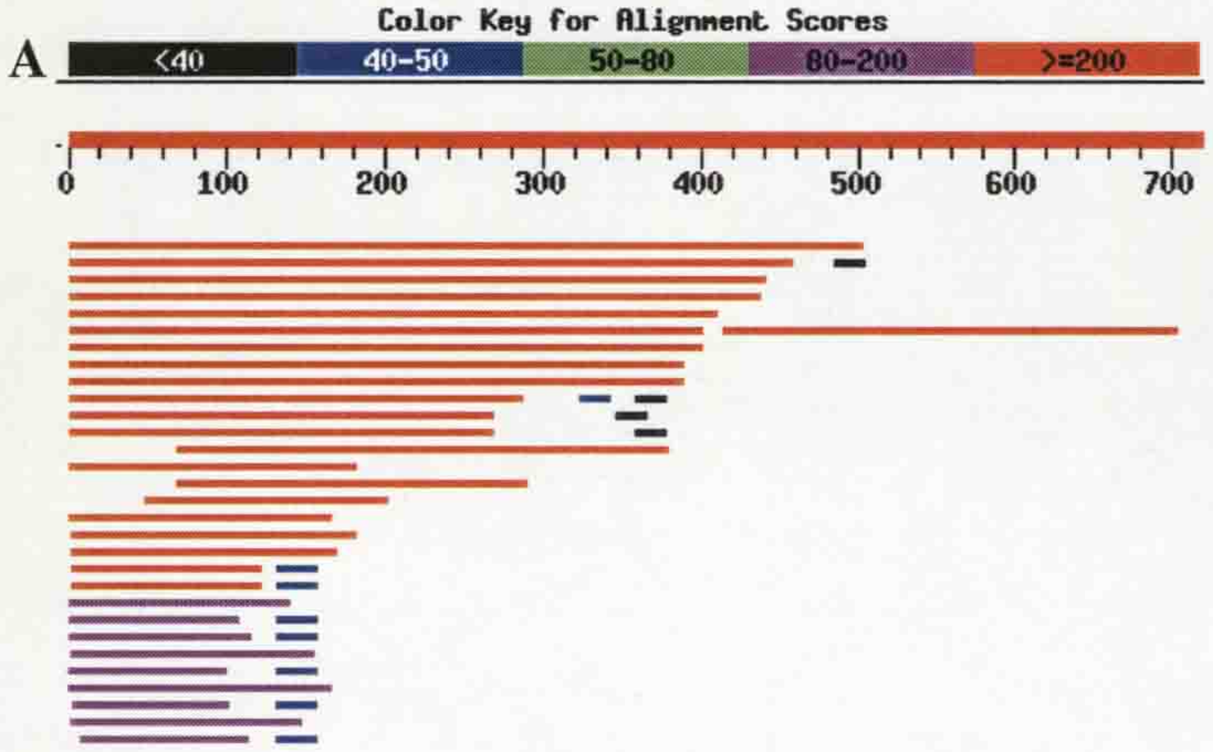


Figure 4.9 Database homology search results for clone P1H12.

Sequence homology searches using the NCBI BLAST search tool (http://www.ncbi.nlm.nih.gov/blast/) revealed clone P1H12 as having significant sequence similarity to mouse osteopontin (gb/J04806), as did clone P2B10. Search results came in the format given in these figures with alignment of the submitted sequence (the graduated band at the top of each figure) with sequences in the database, both in terms of a colour coded alignment score (A), a list of sequences producing the alignment (B) and the extent of homology of these sequences to that submitted (C).

A



B

Sequences producing significant alignments:

		Score (bits)	E Value
gb|AI746416.1|AI746416	ul06g07.x1 Sugano mouse kidney mkia ...	825	0.0
gb|AA498630.1|AA498630	vh31e02.r1 Barstead mouse pooled org...	780	0.0
gb|AI875667.1|AI875667	uk51c01.x1 Sugano mouse kidney mkia ...	768	0.0
gb|AW611445.1|AW611445	un64d08.x1 Sugano mouse kidney mkia ...	736	0.0
gb|AI790412.1|AI790412	uk99h03.x1 Sugano mouse kidney mkia ...	720	0.0
gb|AI042843.1|AI042843	uc84e03.x1 Sugano mouse kidney mkia ...	718	0.0
gb|AI174044.1|AI174044	ud67a10.x1 Sugano mouse liver mlia M...	708	0.0
gb|AA140117.1|AA140117	mq39e06.r1 Barstead MPLRB1 Mus muscu...	708	0.0
gb|AI173482.1|AI173482	ue91f05.x1 Sugano mouse embryo mewa ...	512	e-142
gb|AA277819.1|AA277819	va77d12.r1 Soares mouse NML Mus musc...	480	e-133
gb|AI875491.1|AI875491	uk48h03.x1 Sugano mouse kidney mkia ...	478	e-132
dbj|BB498168.1|BB498168	BB498168 RIKEN full-length enriched...	409	e-112
emb|AL022813.1|AL022813	s8105b52 Beddington mouse dissected...	377	e-102
gb|AI464543.1|AI464543	mv87g09.x1 GuayWoodford Beier mouse ...	309	1e-81
dbj|BB496585.1|BB496585	BB496585 RIKEN full-length enriched...	283	7e-74
gb|AW260387.1|AW260387	um77h09.x1 Sugano mouse liver mlia M...	271	3e-70
gb|AA689940.1|AA689940	vt63e04.r1 Barstead mouse irradiated...	241	3e-61
dbj|BB504637.1|BB504637	BB504637 RIKEN full-length enriched...	233	9e-59
dbj|BB253960.1|BB253960	BB253960 RIKEN full-length enriched...	207	6e-51
gb|AI098056.1|AI098056	ue28a02.x1 Sugano mouse kidney mkia ...	207	6e-51

C

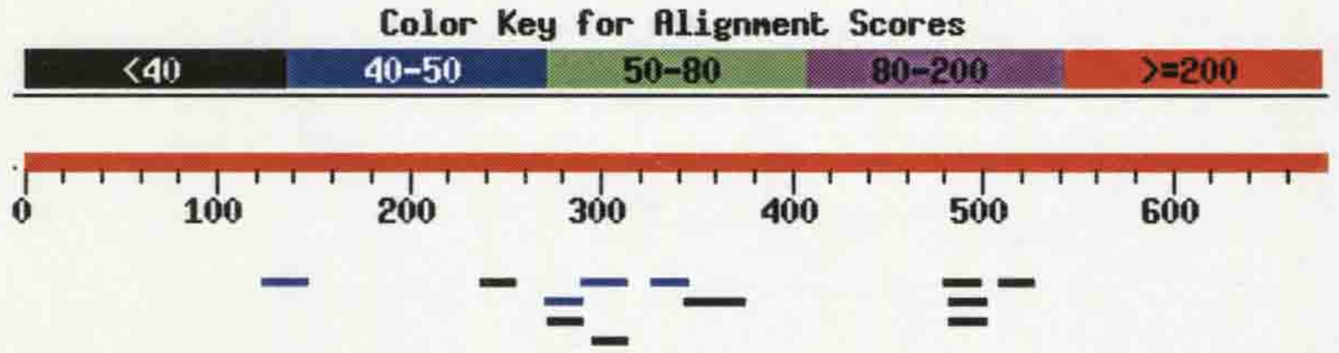
```
>gb|AI790486.1|AI790486 ul01g04.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
          IMAGE:2064822 3'.
         Length = 614

 Score =  923 bits (460), Expect = 0.0
 Identities = 495/504 (98%), Gaps = 4/504 (0%)
 Strand = Plus / Plus

Query: 1   acagacwgacggatagagycacacacaaaatgcacatacaagaactataaactatcaatc 60
           ||||||| |||||||||||| |||||||||| |||||||||||||||||||||||||||||
Sbjct: 113 acagacagacggatagagtcacacacaaa-tgcacatacaagaactataaactatcaatc 171

Query: 61  aaaattatgaatgatgaggattaggcaatttatatattgggagttggaggagggaaagga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 172 aaaattatgaatgatgaggattaggcaatttatatattgggagttggaggagggaaagga 231

Query: 121 aatgaacaggaaggagtggtcagagaaggctcaatgagcaagttctccaacgcatgggga 180
           || |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Sbjct: 232 aa-gaacaggaaggagtggtcagagaaggctcaatgagcaagttctccaa-gcatgggga 289

Query: 181 tgacttagttaaaaggaaatggcttccaggtcagcatatcagcttgaacacaagtaatat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 290 tgacttagttaaaaggaaatggcttccaggtcagcatatcagcttgaacacaagtaatat 349

Query: 241 ggtggamataggaaaggcatgttaagggamacctcacatatgggcctctcttgccggtgc 300
           ||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct: 350 ggtggaaataggaaaggcatgttaagggaaacctcacatatgggcctctcttgccggtgc 409

Query: 301 tgaagttttgatgaggtgcgctggggatcttctgaagctggtgatattaagctgctggtg 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct: 410 tgaagttttgatgaggtgcgctggggatcttctgaagctggtgatattaagctggtggtg 469

Query: 361 gtggggaaatgtctctggtttatga-ccccatctaccacagaagaaatggagcattttaa 419
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct: 470 gtggggaaatgtctctggtttatgacccccatctaccacagaagaaatggagcattttaa 529

Query: 420 gaccgtggaaataaaactgttggacagggaaccagtctccgttctcccacctcaagcaca 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 530 gaccgtggaaataaaactgttggacagggaaccagtctccgttctcccacctcaagcaca 589

Query: 480 atcccaggcagaagttcaccaatt 503
           |||||||||||||||||||||||
Sbjct: 590 atcccaggcagaagttcaccaatt 613
```

Figure 4.10 Database homology search results for clone N2E4.

Sequence homology searches using the NCBI BLAST search tool (http://www.ncbi.nlm.nih.gov/blast/) revealed clone N2E4 as having no significant matches to known sequences, but which demonstrated sequence identity to an EST sequence (gb/AI790486). The same EST also showed sequence similarity to clones N1A6, N2A1 and N3C9. Search results came in the format given in these figures with alignment of the submitted sequence (the graduated band at the top of each figure) with sequences in the database, both in terms of a colour coded alignment score (A), a list of sequences producing the alignment (B) and the extent of homology of these sequences to that submitted (C).

Clone P2D6 submitted to database of known sequences.

A



B

Sequences producing significant alignments:

		Score (bits)	E Value
gb|AF040658.1|CELW07G9	Caenorhabditis elegans cosmid W07G9	43	0.16
gb|AE003623.1|AE003623	Drosophila melanogaster genomic scaf...	41	0.65
gb|AC004817.2|AC004817	Homo sapiens PAC clone RP6-91H8 from...	41	0.65
gb|AF182093.1|AF182093	Scophthalmus maximus microsatellite ...	41	0.65
gb|AC002081.1|AC002081	Homo sapiens BAC clone CTA-331C24 fr...	39	2.6
gb|AC005661.1|AC005661	citb_54_o_2, complete sequence [Homo...	39	2.6
gb|AC004381.1|HUAC004381	Homo sapiens Chromosome 16 BAC clo...	39	2.6
gb|AC007488.15|AC007488	Homo sapiens 3q27 BAC RPCI11-246B7 ...	39	2.6
gb|AC004066.1|AC004066	Homo sapiens chromosome 4 clone B353...	39	2.6
emb|AL135922.4|AL135922	Human DNA sequence from clone RP5-8...	39	2.6
emb|Z80232.1|HS463A9	Human DNA sequence from PAC 463A9, on ...	39	2.6
emb|Z83820.1|HS215K18	Human DNA sequence from PAC 215K18 on...	39	2.6

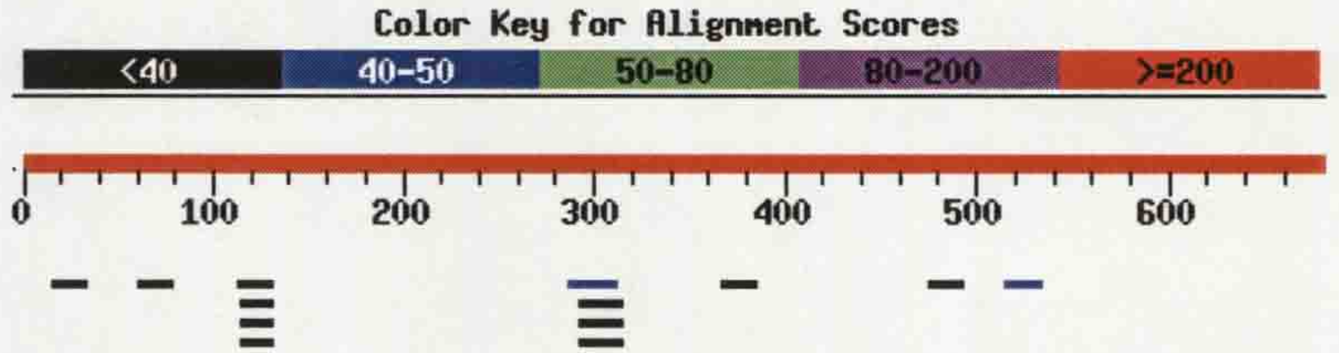
Alignments

C

```
>gb|AF040658.1|CELW07G9 Caenorhabditis elegans cosmid W07G9
          Length = 26767

 Score = 42.9 bits (21), Expect = 0.16
 Identities = 24/25 (96%)
 Strand = Plus / Minus


Query: 290  tgtttgtctaaatatctggttttca 314
            ||||||| ||||||||||||||||
Sbjct: 1053 tgtttgtttaaatatctggttttca 1029
```

Clone P2D6 submitted to database of EST sequences.

A



B

Sequences producing significant alignments:

		Score (bits)	E Value
dbj|AV172813.1|AV172813	AV172813 Mus musculus brain C57BL/6...	43	0.14
dbj|C18586.1|C18586	C18586 Human placenta cDNA (TFujiwara) ...	41	0.55
gb|BE131228.1|BE131228	L48-1185T3 Ice plant Lambda Uni-Zap ...	39	2.3
gb|BE043051.1|BE043051	ho32c07.x1 NCI_CGAP_Lu24 Homo sapien...	37	9.1
gb|AW407454.1|AW407454	UI-HF-BL0-adi-e-08-0-UI.r1 NIH_MGC_3...	37	9.1
gb|AW402835.1|AW402835	UI-HF-BK0-aaq-e-10-0-UI.r1 NIH_MGC_3...	37	9.1
gb|AW297262.1|AW297262	UI-H-BW0-aji-c-11-0-UI.s1 NCI_CGAP_S...	37	9.1
gb|AW268427.1|AW268427	xv34d11.x1 Soares_NFL_T_GBC_S1 Homo ...	37	9.1
gb|AW175483.1|AW175483	fi25f09.x1 Sugano Kawakami zebrafish...	37	9.1
gb|AW170018.1|AW170018	xj35h03.x1 Soares_NFL_T_GBC_S1 Homo ...	37	9.1
gb|AW117059.1|AW117059	fe17f09.x1 Zebrafish WashU MPIMG EST...	37	9.1
gb|AW059125.1|AW059125	fe25f03.x1 Zebrafish WashU MPIMG EST...	37	9.1

C

```
>dbj|AV172813.1|AV172813 AV172813 Mus musculus brain C57BL/6J adult Mus musculus cDNA clone
          3632413018.
          Length = 264

 Score = 42.9 bits (21), Expect = 0.14
 Identities = 24/25 (96%)
 Strand = Plus / Minus


Query: 288 cctgtttgtctaaatatctggtttt 312
           ||||||||| ||||||||||||||
Sbjct: 137 cctgtttgtttaaatatctggtttt 113
```

Figure 4.11 Database homology search results for clone P2D6.

Sequence homology searches using the NCBI BLAST search tool (http://www.ncbi.nlm.nih.gov/blast/) revealed clone P2D6 as having no significant matches to known or EST database sequences. Search results came in the format given in these figures with alignment of the submitted sequence (the graduated band at the top of each figure) with sequences in the database, both in terms of a colour coded alignment score (A), a list of sequences producing the alignment (B) and the extent of homology of these sequences to that submitted (C).

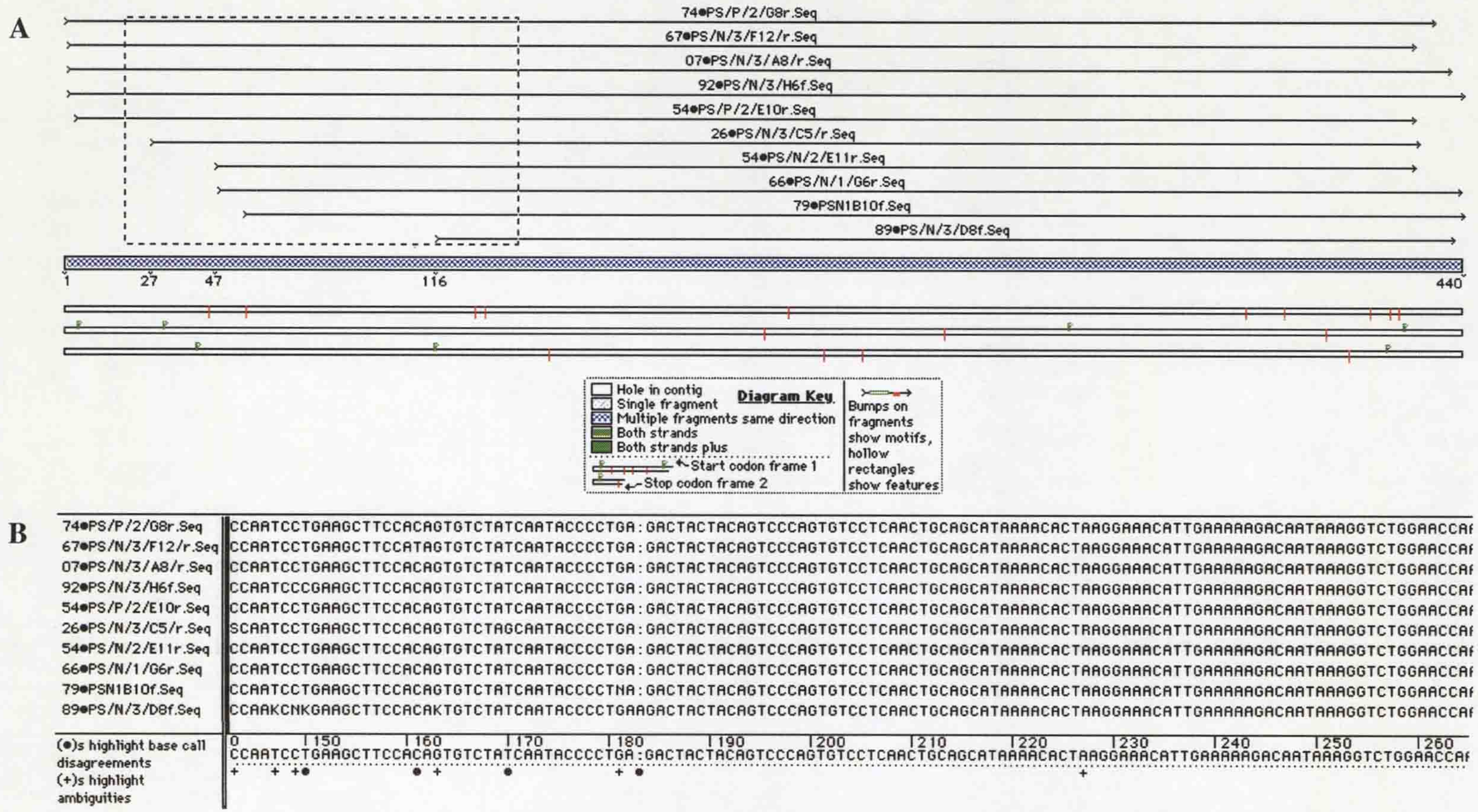


Figure 4.12 Identification of common sequences using Sequencher™ software.

Sequencher™ software (Gene Codes Corporation) has been utilised to identify clones with overlapping regions of identical sequence. Here ten clones from both the N and the P subtracted cDNA libraries were aligned (A) into a contig due to the identical sequence (B) obtained for each clone. These clones had 99% sequence homology to mouse kappa casein (gb|M10114).

Subtracted Library	N	P	PN	Met
Size of library	$4.38x10^5$	$3.75x10^5$	$4.25x10^5$	$6.45x10^5$
Colonies picked	240	204	192	192
PCR product	**218** (91%)	**188** (92%)	**181** (94%)	**181** (94%)
Number 3' fragments	78 (36%)	72 (38%)	38 (21%)	51 (28%)
Known genes	113 (52%)	89 (47.5%)	89 (49.2%)	96 (53%)
ESTs	69 (31.5%)	59 (31.5%)	71 (39.2%)	68 (37.6%)
Unknown genes	16 (7.5%)	5 (2.5%)	11 (6.1%)	8 (4.4%)
No sequence	20 (9%)	35 (18.5%)	10 (5.5%)	9 (5.0%)
Number of common sequences	12	9	16	10
Number of clones representing common sequences	89 (45%)	37 (24%)	40 (23%)	32 (18%)
Number of singletons	109 (55%)	116 (76%)	131 (77%)	143 (82%)

Table 4.1 Characterisation of the four subtracted cDNA libraries.
Subtracted cDNA fragments were cloned into pCR2.1™ (Invitrogen) and transformed into TOP10F' competent *E.coli*. Approximately 0.05% of colonies in each library were studied by colony-PCR, 93% of all colonies contained single amplifiable cDNA inserts. DNA sequence analysis of all clones revealed that 31% of clones represented 3' sequences which contained a poly$(A)_n$ sequence. Nucleotide sequence homology searches using the BLAST search engine was used to determine the identity of each sequenced cDNA clone. The distribution of known, EST and unknown sequences is given for each library. Of the clones producing good sequence, a number from each library had identical sequence or were representative of the same known mRNA. The majority of clones from each of the four libraries, however, were non-redundant sequences (singletons).

	cDNA libraries			Subtracted cDNA libraries			
	N	P/PN	Met	N	P	PN	Met
Number of colonies screened	453	428	579	451	400	432	401
Positive actin colonies	1	0	0	0	0	0	0
Number of colonies screened	621	450	432	456	412	397	502
Positive HPRT colonies	0	0	1	0	0	0	0
Number of colonies screened	1074	1091	1011	907	812	829	903
Positive S100A4 colonies	0	0	0	0	2	0	4

Table 4.2 Colony hybridisation data.
Transformed cDNA libraries (non-subtracted and subtracted) were grown on nylon filters to a density of approximately 200-300 white colonies/8 cm plate. Filters were independently hybridised with actin, HPRT and S100A4 (p9Ka) cDNA probes. The number of positive colonies detected for each probe in each cDNA library is given.

Sequence Identification	Identity	Accession number	Subtracted Libraries N	P	PN	Met
Endogenous **mouse mammary tumor virus** Mtv1 complete proviral genome Exogenous mouse mammary tumor virus C3H complete proviral genome Mouse mammary tumor virus LTR DNA	224/229 (97%)	gb|AF228550 gb|AF228552 gb|L37517	28	1	3	0
Mouse **mitochondrion**, complete genome	791/816 (96%)	gb|J01420	11	10	2	5
Mouse **kappa-casein** mRNA	411/414 (99%)	gb|M10114	9	3	0	0
M.musculus mRNA for **BRP39** protein	594/629 (94%)	emb|X93035	7	0	0	0
Mus musculus mammary gland **cDNA** clone similar to felca major allergen I polypeptide chain 2 precursor	436/442 (98%)	gb|AW989042	6	0	0	0
Mus musculus mammary gland **cDNA clone** similar to human cellular retinol binding protein I Mus musculus mammary gland cDNA clone similar to human cellular retinol binding protein II	282/288 (97%) 283/290 (97%)	gb|AA466092 gb|AI037751	7	0	0	0
R.norvegicus mRNA for **transferrin**	350/422 (82%)	emb|X77158	3	0	8	4
Human **elongation factor 1-alpha** (EF 1-alpha)	256/317 (80%)	gb|M29548	0	5	1	1
Mouse **Id2** protein (Id-2) mRNA	174/187 (93%)	gb|M69293	3	3	0	0
M.musculus mRNA for **WDNM1** protein	410/418 (98%)	emb|X93037	0	3	5	1
Sugano mouse kidney mkia Mus musculus **cDNA clone**	497/504 (98%)	gb|AI790486	4	0	0	0
R.norvegicus (Sprague Dawley) **rab28** mRNA for ras-homologous GTPase	265/282 (93%)	emb|X78606	0	4	0	0
Murine gene for **osteopontin**	682/693 (98%)	emb|X51834	0	2	0	0
Mus musculus **pulmonary surfactant protein PSP-C** (SFTP2) gene	135/148 (91%) (67-213) 111/117 (94%) (211-327) 101/109 (92%) (334-441) 103/124 (83%) (447-568)	gb|M38314	0	0	0	8
Mouse **lysozyme M** gene, exon 1 Mouse lysozyme M gene, exon 2 Mouse lysozyme M gene, exon 3	152/164 (92%) (9-167) 154/166 (92%) (167-333) 61/67 (91%) (332-398)	gb|M21047 gb|M21048 gb|M21049	0	0	1	7
Mus musculus **Ca(2+)-sensitive chloride channel 2** (Cacc) mRNA,	433/483 (89%)	gb|AF108501	0	0	3	0
Mus musculus **matrix Gla protein**	341/353 (96%)	gb|S77350	0	0	0	2
Mus musculus **flavo-binding protein** mRNA, complete cds	590/600 (98%)	gb|AF174535	0	0	0	2
Mus musculus **cDNA clone** MNCb-4259	350/367 (95%)	dbj|AU079415	2	0	0	0
Mus musculus **ribosomal protein S3a** (Rps3a), mRNA	527/566 (93%)	ref|NM_016959	0	0	2	0
Mus musculus **Nedd5** mRNA for septin	450/464 (96%)	dbj|D49382	0	0	3	0
Mus musculus **TNFR2-TRAF signalling complex protein** mRNA Mus musculus inhibitor of apoptosis protein 2 mRNA	411/457 (89%) 411/457 (89%)	gb|L49433 gb|U88909	0	0	2	0
Mouse E7.5 Embryonic Portion cDNA Library Mus musculus **cDNA clone**	350/369 (94%)	gb|AW537697	0	0	2	0
Soares mouse embryo NbME13.5 14.5 Mus musculus **cDNA clone** Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone	215/224 (95%) (48-272) 183/200 (91%) (273-472)	gb|AA003680 gb|W91109	0	0	2	0
Mus musculus **Cope1** mRNA for nonclathrin coat protein epsilon-COP Mus musculus cDNA clone similar to Q60445 COATOMER EPSILON SUBUNIT	132/132 (100%) 191/193 (98%)	dbj|AB039837 gb|AI526939	0	0	2	0
Mus musculus **ubiquitin protein ligase Nedd-4** mRNA, complete cds	382/431 (88%)	gb|U96635	0	0	2	0
Mus musculus **PHR1 isoform 1** mRNA, alternatively spliced, complete cds	679/747 (90%)	gb|AF000272	2	0	0	0

Table 4.3 Sequences identified multiple times in subtracted libraries.
List of sequences of cDNA clones identified multiple times by database homology searching and construction of contigs. Database accession numbers, extent of identity of one of the clones for each type of sequence and the number of times each sequence was identified in each subtracted library is given. In the case of PSP-c, the same clone matched several regions of the PSP-c gene sequence, each region relating to consecutive exon sequences. Similarly, a different clone matched 3 database entries for lysozyme M, each entry referring to a different exon sequence. For one of the clones, two ESTs match separate regions of the same clone, both these cDNA clones are given with the region of homology observed. For each, the extent and region of identity of the clone to the database sequence is given.

Chapter 5.

Expression screening of subtracted cDNA clones.

Sequencing several hundred cDNA clones from each library can reveal details concerning the complexity of the subtracted libraries created and the types of genes expressed in the different tumours. A vast amount of data was generated by these sequencing experiments and database homology searches, with the identification of many known genes that can be grouped into numerous different gene or protein families according to their predicted or determined function. Some of these identified include proteins involved in DNA repair, cell cycle control, gene expression, protein synthesis, cytoskeletal proteins, proteases, protease inhibitors and cell surface antigens. Altered expression of genes corresponding to many of these known proteins and, of course, of proteins corresponding to the many ESTs and uncharacterised sequences that were also identified, may have implications in tumour development and/or metastatic progression.

Differential expression of these clones can be speculated upon, particularly for those sequences identified multiple times in specific libraries (Table 4.3; for example BRP39, emb|X93035; EF1A, emb|X13661) and also for those known genes which have tissue specific patterns of gene expression under normal circumstances (PSP-c, gb|M38314). However determination of the expression pattern for each clone is required to confirm if these sequences are in fact differentially expressed between these tumours of different metastatic potential.

5.1 Reverse Northern differential expression screening of subtracted cDNA clones.

Reverse Northern analysis was used to determine the expression pattern of all 768 sequenced subtracted cDNA clones in the two primary tumours and in the metastatic tumour. Eight different 96-well dot blot membranes were produced, in triplicate, each containing cDNA clones from the four subtracted libraries and control cDNAs.

As shown earlier (section 3.5), the expression of high abundant sequences were detected after short exposure (overnight) of hybridised blots to

phosphorimaging screen or autoradiographic film. This was of benefit to accurately detect hybridisation signal with reduced noise, i.e. from neighbouring spotted clones that produced strong hybridisation signal or from cases of high, non-specific background signal. Longer exposures were required for the detection of low abundant cDNA clones which produced a weak hybridisation signal, however these were only of benefit when the images were clean with little background hybridisation signal. Long exposure to autoradiographic film was a problem as strong hybridisation signal was overexposed and so could not be quantified. Overexposure was less of a problem for signal detection using the phosphorimaging screen, which exhibits a linear dynamic range that extends over 5 orders of magnitude. Hybridisation signal obtained from the phosphorimager was therefore used to quantify hybridisation spot intensity, and therefore the relative abundance of each cDNA clone in each tumour.

5.2 Validation of Reverse Northern screening process.

To test the technique was reproducible, hybridisation signal intensities for control cDNAs, included on all membranes, and cDNA clones representing the same contig or known message present on multiple membranes were compared. Also, a selection of differentially expressed clones was re-spotted onto fresh Reverse Northern membranes in duplicate. These clones were re-screened by the same means to confirm that the expression pattern originally observed was reproducible.

5.2.1 Control cDNAs.

Several control cDNAs were used on each membrane to enable immediate determination of differences in probe labelling and hybridisation efficiencies between tumour samples. Actin, HPRT and S100A4 (p9Ka) cDNA were used, in duplicate, and in later experiments subtracted cDNA clones representing PSP-c, lysozyme, kappa casein, transferrin, WDNM 1 and BRP39 were also included (although not in duplicate; figure 5.1). The expression pattern of these clones were determined in early Reverse Northern experiments and their spot intensities in each of the tumours helped determine the relative success and reliability of the hybridisation and quantitation process. Actin, HPRT, kappa casein, transferrin and WDNM 1 were all expressed within the 2-fold cut off between the three different tumours, suggesting similar expression levels for each of these genes in each tumour

sample. BRP39 was more highly expressed in the MMTV-*neu* tumour compared with equally low levels of expression in the two bitransgenic tumours. PSP-c and lysozyme were both abundant in the metastatic tumour sample relative to similarly low detectable expression in the two primary tumours. S100A4 (p9Ka) expression was not detectable in the MMTV-*neu* primary tumour and demonstrated progressively higher abundance from the primary tumour to the metastasis from the MMTV-*neu/S100A4* mouse. This was expected following hybridisation of S100A4 cDNA to the same two bitransgenic tumour cDNAs and representative cDNA (and total RNA) from several pooled MMTV-*neu* primary tumours and several pooled MMTV-*neu/S100A4* primary tumours (Figure 3.3). This pattern of expression was consistent for all these cDNAs in each Reverse Northern experiment. The expression of actin, HPRT and in particular S100A4 and the lung specific sequence PSP-c produced predicted hybridisation signal intensities in each of the three tumours suggesting that the screening process was also able to differentiate between different levels of expression for different sequences between different tumours.

5.2.2 Membrane preparation.

The accuracy of DNA preparation (dilutions and aliquoting) and dot blotting was investigated by comparing hybridisation signals of all DNA samples that were immobilised to the nylon membrane in duplicate. This included all controls and subtracted clones that were re-spotted and re-screened.

The identification of differences in hybridisation signal between duplicate spots, probed with the same tumour sample, revealed a small number of dot blotting errors. In total, 708 cDNA spots were checked, 22 displayed differences between duplicates, indicating a relatively efficient sample and membrane preparation. Thus suggesting that possibly 3% of all the clones analysed may have produced spurious hybridisation signal due to inaccurate preparation of the membranes.

5.2.3 Probe labelling as a source of variability.

Occasionally inconsistent expression patterns were observed for different clones representing the same mRNA molecule that were present on separate membranes. For example, Id2 (gb|M69293) was sequenced 6 times (N=3, P=3; Table 4.3). Some of these clones exhibited differential expression in favour of the MMTV-*neu/S100A4* primary tumour relative to the other two tumours; however, other clones appeared to

be expressed equally in the three tumours. The inconsistency appeared to be due to different probe labelling efficiencies between separate experiments as the differentially expressed Id2 clones were screened in one experiment and independent to the non-differentially expressed Id2 clones. This inconsistency meant that Id2 was not determined to be differentially expressed.

For all redundant sequences (identified multiple times), differential expression was only confirmed if a consistent expression pattern was observed in multiple experiments. Likewise, non-redundant clones (singletons) that were re-screened on separate membranes were only confirmed as being differentially expressed if the same pattern of expression was observed on each occasion. Some non-redundant clones, however, were not re-screened and so their expression pattern was not validated by Reverse Northern analysis.

In total 71 clones were re-screened in additional Reverse Northern experiments (Figure 5.2), 56 of these clones (79%) demonstrated a consistent expression pattern to that observed previously. Of the remaining clones, 8 (11%) did not produce strong enough hybridisation signal too reliably detect expression (suggesting weaker hybridisation probes was used) and 7 (10%) were not differentially expressed.

5.3 Identification of differentially expressed cDNA clones.

Spot intensities for each clone were quantified, normalised for differences in probe strength and compared between the three different tumours. Criteria for selecting differentially expressed clones required a difference in expression ratio between two tumours to be ≥ 2 or ≤ 0.5. As stated previously, other cDNA expression analysis experiments using microarrays or Reverse Northern analysis have used a similar differential expression ratio of ≥ 2 or ≥ 3 as criteria for identifying differentially regulated cDNAs (Schena *et al.*, 1996; DeRisi *et al.*, 1997; Sgroi *et al.*, 1999; Perou *et al.*, 1999; Yang *et al.*, 1999). In total 192 clones were differentially expressed between two of the three tumours by these criteria. This was 25% of the total number of clones screened and was comprised of 125 different sequences (Table 5.1). A comprehensive list of all clones identified as being differentially expressed in this way is presented in Tables 5.2-5.5. Of the differentially expressed sequences, 87 were derived from the N subtracted cDNA library, 29 were from the P library, 27 were from the PN library and 49 were from the Met library.

By utilising a higher cut off of >5 or <0.2-fold difference, 80 differentially expressed clones were identified. 55 of these clones (representing 21 different sequences) were from the N subtracted cDNA library, 5 different sequences were from the PN library and 20 different clones (representing 6 different sequences) were from the Met library (Table 5.1). No cDNA clone from the P subtracted cDNA library exhibited a differential expression ratio of >5-fold. The majority of clones that were differentially expressed by 5-fold or more, exhibited detectable expression in one tumour yet very low or undetectable expression in another tumour.

5.3.1 Differentially expressed clones in the N and P subtracted cDNA libraries.

Tables 5.2 and 5.3 present all the sequences identified as being differentially expressed from the N and P subtracted cDNA libraries. The table provides fold expression ratios of these sequences in the MMTV-*neu* primary tumour and in the metastasis relative to expression in the MMTV-*neu/S100A4* primary tumour. Sequencing information, determined in chapter 4, for these clones is also given in this table, in terms of the identification of each clone and the number of times (redundancy) each sequence was identified in these two libraries.

79 of the 87 differentially expressed cDNA clones from the N subtracted library were expressed more highly in the MMTV-*neu* primary tumour compared to the MMTV-*neu/S100A4* primary tumour, furthermore 69 of these clones also demonstrated a similarly low or lower expression level in the metastatic tumour compared to the MMTV-*neu/S100A4* primary tumour. 5 of the remaining 8 clones were equally expressed in both primary mammary tumours, yet were either expressed higher (2 clones) or lower (3 clones) in the metastasis relative to the MMTV-*neu* tumour. 3 cDNA clones from the N subtracted library were expressed more highly in the bitransgenic primary tumour compared to the MMTV-*neu* primary tumour.

29 cDNA clones from the P subtracted library were differentially expressed by 2 fold or greater between any two of the three tumours, 12 clones were less abundant in the mRNA from the MMTV-*neu* primary tumour relative to the bitransgenic primary tumour. 8 of these clones exhibited expression in the metastatic tumour that was also ≥2-fold higher than in the MMTV-*neu* primary. The remaining 17 clones were either expressed in the wrong orientation (11 clones) or were of the same approximate abundance in both primary tumours yet were either more abundant

(1 clone) or less abundant (5 clones) in the metastatic tumour.

5.3.2 Differentially expressed clones in the PN and Met subtracted cDNA libraries.

Tables 5.4 and 5.5 present all the sequences derived from the PN and Met subtracted cDNA libraries that have been identified as being differentially expressed.

14 of the 27 differentially expressed cDNA clones from the PN subtracted library were less abundant in the metastasis relative to the PN primary. 5 of these 14 clones also demonstrated a similar or higher level of expression in the MMTV-*neu* primary tumour compared to this MMTV-*neu/S100A4* primary tumour. 8 of the remaining 13 clones were equally expressed in both bitransgenic tumours, yet were either expressed higher (5 clones) or lower (3 clones) in the MMTV-*neu* primary tumour. 5 clones from the PN library were expressed more highly in the metastasis compared to the MMTV-*neu/S100A4* primary tumour.

48 Met subtracted cDNA clones were differentially expressed by 2-fold or greater between any two of the three tumours, 25 of these clones were more abundant in the mRNA from the metastasis relative to the MMTV-*neu/S100A4* primary tumour and all of these 25 clones exhibited expression in the MMTV-*neu* primary tumour that was similar or lower than in the MMTV-*neu/S100A4* primary tumour. 6 clones were expressed in the wrong orientation to that expected and the remaining 17 clones were of the same approximate abundance in both bitransgenic tumours yet demonstrated either higher (3 clones) or lower (14 clones) expression in the MMTV-*neu* primary tumour.

5.3.3 Maintenance of expression of subtracted clones with metastatic progression.

If expression was detected to be similar in two tumours and different by > 2-fold in a third tumour (MMTV-*neu* or metastasis) then expression was said to be maintained with metastatic progression. For example, clone N2H5 (Table 5.2) exhibited expression that decreased with metastatic progression (N = 0.97, PN = 1, Met = 0.37). Conversely, clone Met2E9 (Table 5.5) demonstrated maintained expression with metastasis (N = 0.37, PN= 1, Met = 1.22). In total, 118 clones exhibited expression patterns that were said to be maintained in metastatic progression. In some cases differential expression was enhanced towards metastatic

progression, where expression increased or decreased by > 2-fold as S100A4 expression increased (see Figures 3.3 and 5.1) and therefore also did metastatic capability. For example, clone M1C11 (Table 5.5) demonstrated increasing expression as S100A4 expression increased (N= 0.49, PN = 1, Met = 3.57), conversely, clones N2B5 (Table 5.2; N= 9.98, PN = 1, Met = 0.38) and PN1B9 (Table 5.4; N= 4.72, PN = 1, Met = 0.23) exhibited the opposite pattern of expression. A total of 21 clones exhibited such expression in these three tumours.

5.4 Low abundant subtracted cDNA clones.

The expression pattern of 98 clones (12.8% of the total number of clones screened) could not be determined by Reverse Northern analysis because no hybridisation signal or only a very weak signal was detected for these clones in all three tumours. This was most notable for clones from the PN subtracted cDNA library, in which 49 clones (27% of clones from this library) produced a hybridisation signal that was too weak to reliably detect differential expression. This may explain why fewer clones from this library were identified as being differentially expressed, in comparison to other libraries (table 5.1).

Interestingly, 60% of these low expressing clones represent ESTs or previously uncharacterised sequences, suggesting that they may be expressed at similarly low levels in other mouse tissues or in other species. Some of the known sequences for which expression was not detected by Reverse Northern are presented in Table 5.6.

5.5 Reverse Northern expression screening of subtracted cDNA clones in other tumours from the transgenic mice.

Reverse Northern membranes were also probed with cDNA derived from additional tumours of similar transgenic status, histopathology and tumour phenotype to those, which were used to create the subtracted libraries. This was performed to identify clones which demonstrated the same differential expression profile in multiple tumours, therefore to aid in the selection of specific, interesting clones for further characterisation.

In an initial experiment 190 clones were screened for expression levels in two additional primary tumours and an additional metastatic tumour. A further 350 clones were screened for expression in the two primary tumours only. In both

experiments, the probes were prepared by incorporating the radioactive label during first strand cDNA synthesis.

Slightly different expression profiles were observed between matched tumours of the same transgenic type and phenotype (Figure 5.3), suggesting different gene expression profiles may exist between tumours of matched histology, transgenic status and tumour phenotype.

The method of probe generation (SMART PCR or first strand cDNA synthesis) also appeared to contribute to the different patterns of expression seen between matched tumours from the two types of screening experiments. Differences were observed in the hybridisation signal intensity obtained for different cDNA clone fragments that represent the same gene. For example both transferrin and α-enolase were represented on one Reverse Northern membrane by multiple subtracted cDNA clones, these clones had sequence identity to different regions of these two cDNA molecules (Figure 5.4). All fragments, for each cDNA, produced a hybridisation signal of equal intensity when probed with SMART PCR-generated cDNA probe. Yet when these clones were hybridised with the first strand cDNA synthesis probes, fragments closer to the 3' end of the cDNA molecule gave a detectable signal, whereas fragments closer to the 5' end of the cDNA molecule gave no or only a weakly detectable hybridisation signal (Figure 5.4). This suggests that first strand cDNA synthesis was relatively inefficient at generating transcripts beyond the 3' end of the mRNA molecule.

This data also demonstrates the ability of SMART cDNA synthesis to create full length or near full-length cDNA molecules which, during SSH produced subtracted cDNA clones representative of 3', middle and 5' regions of respective cDNAs. Sequence analysis of subtracted clones supported this, where 31% of subtracted clones represent the 3' region of cDNAs as indicated by the identification of poly(A^+) sequences (Table 4.1). Consequently the expression of the remaining 69% of clones, representing sequences that are 5' to this 3' end, may not be detected using the first strand cDNA synthesis-generated Reverse Northern probes.

These additional screening experiments highlighted another feature of expression hybridisation. The membranes previously hybridised with strong SMART cDNA probes that were stripped and re-probed with first-strand cDNA synthesis probes produced a low overall hybridisation signal with a much reduced level of sensitivity (Figure 5.3c and 5.3d). This may have been due to a combination of

factors including inefficient stripping which left less target DNA available for re-hybridisation and the use of a weak probe as the second hybridisation probe. Re-probing was not, however, a problem when a strongly labelled SMART PCR probe was used (Figure 5.3a and 5.3b).

These difficulties may help account for the fact that the expression of a large number of the clones screened (243, 45% of the total number of clones screened) could not be detected, or were only very weakly detected, after one week exposure to phosphorimaging screen in each of the tumours screened. Hence only a few differentially expressed clones were identified, 47 clones (9% of those screened) were differentially expressed by 2-fold or more between two of the three tumours (Table 5.7). Of these clones, 8 (2%), demonstrated a similar differential pattern of expression as observed in tumours used to produce the subtracted cDNA libraries (marked by * in Table 5.7). Suggesting differential expression of these clones was consistent in multiple tumours of matched histology, transgenic status and tumour phenotype.

5.6 Conclusions to chapter.

Instances of variability suggested that the Reverse Northern cDNA expression screening process was not perfect and possible sources for this variability have been discussed here, some aspects of this process will also be discussed in more detail in chapter 7. However, the technique was sufficient to screen a large number of subtracted cDNA clones for expression patterns in multiple tumours from the transgenic mice and to identify a number of differentially expressed sequences. These include a selection of known genes previously reported to be differentially regulated in breast and in other types of cancer and sequences that are of interest regarding metastatic progression due to the proposed function of the identified sequence or related sequence.

For example, elevated osteopontin (clone P2B10, expression was higher in S100A4 expressing tumours) expression has been linked with tumourigenesis in a number of *in vitro* and *in vivo* models of cancer and is over-expressed in human tumours (Brown *et al.*, 1994; Tuck *et al.*, 1999). In particular, the ability of enhanced osteopontin to induce metastasis has been demonstrated in a rodent model of breast cancer (Oates *et al.*, 1996). Doc-1 (clone N3C3, expression decreased with metastasis) is a putative tumour suppresser identified in oral cancer development

(Todd *et al.*, 1995) which exhibits loss of heterozygosity and subsequent reduced expression in malignant oral keratinocytes (Todd *et al.*, 1995; Tsuji *et al.*, 1998). A mouse EST (clone Met2H4) demonstrates highest expression in the metastasis and sequence homology to human Nip1. Nip1 is a protein which interacts with adenovirus E1B 19 kDa protein and Bcl-2 (Boyd *et al.*, 1994; Zhang *et al.*, 1999) to suppress cell death.

This combination of SSH and Reverse Northern screening also identified a number of ESTs and uncharacterised, potentially novel sequences that are differentially regulated in these circumstances. However, the differential expression of specific clones identified by these means must be verified and further characterised to investigate the relative influence of the expression of these sequences in determining tumour phenotype.

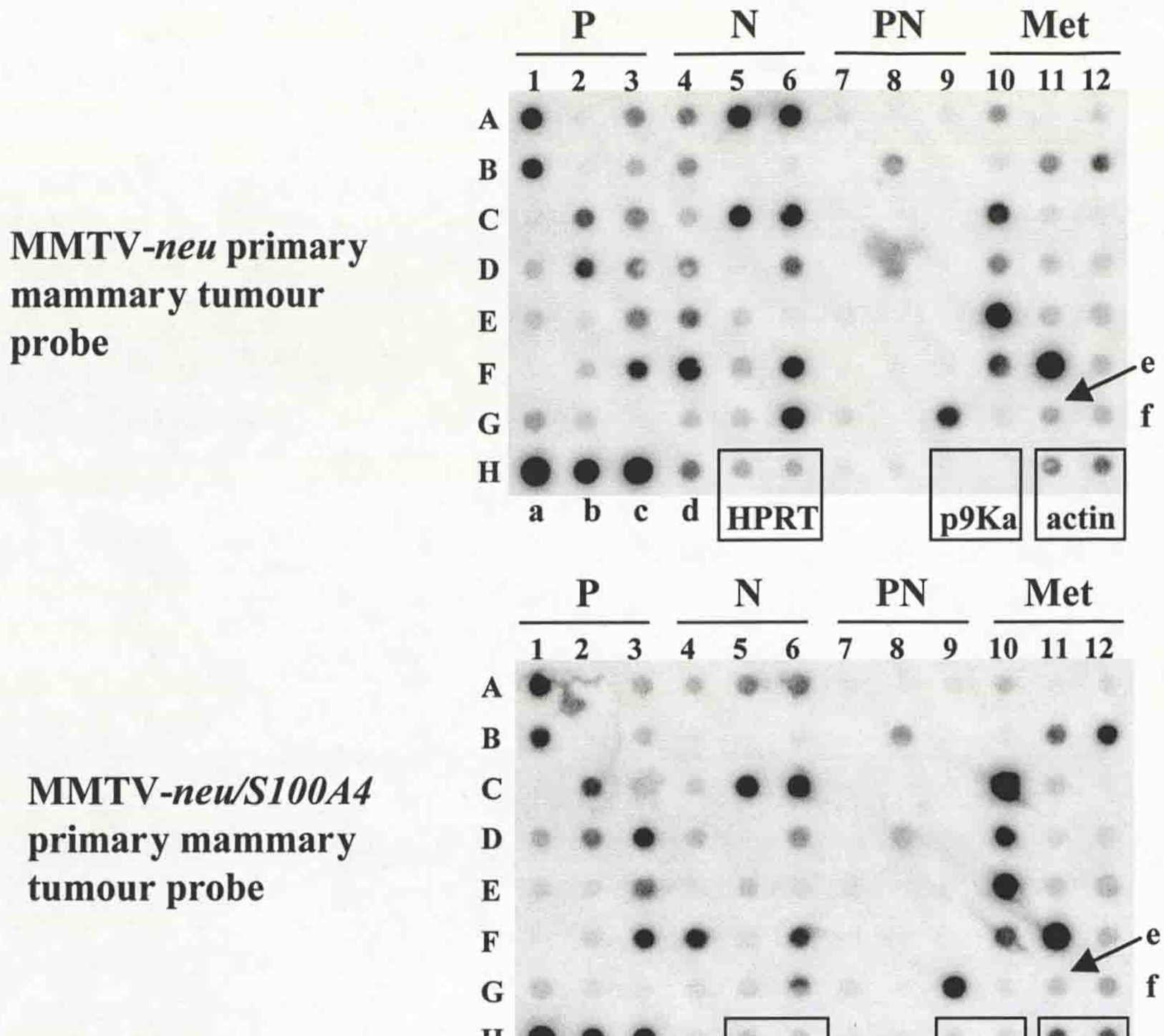


Figure 5.1 Reverse Northern screening of subtracted cDNA clones.

Triplicate sets of dot blot membranes were created containing cDNA clones from each of the four subtracted libraries, plus the cDNA controls (in duplicate) p9Ka (S100A4), actin and HPRT. Additional controls were also included on this membrane for which repeated expression data had already been obtained, these were (a) WDNM 1, (b) transferrin, (c) kappa casein, (d) BRP 39, (e) PSP-c and (f) lysozyme. Each identical membrane was hybridised with a SMART amplified cDNA probe derived from the same tumours used to create the subtracted libraries. Hybridised membranes were exposed to phosphorimaging screen for 1-7 days. The above example is of membrane 8 exposed to a phosphorimaging screen for 6 days.

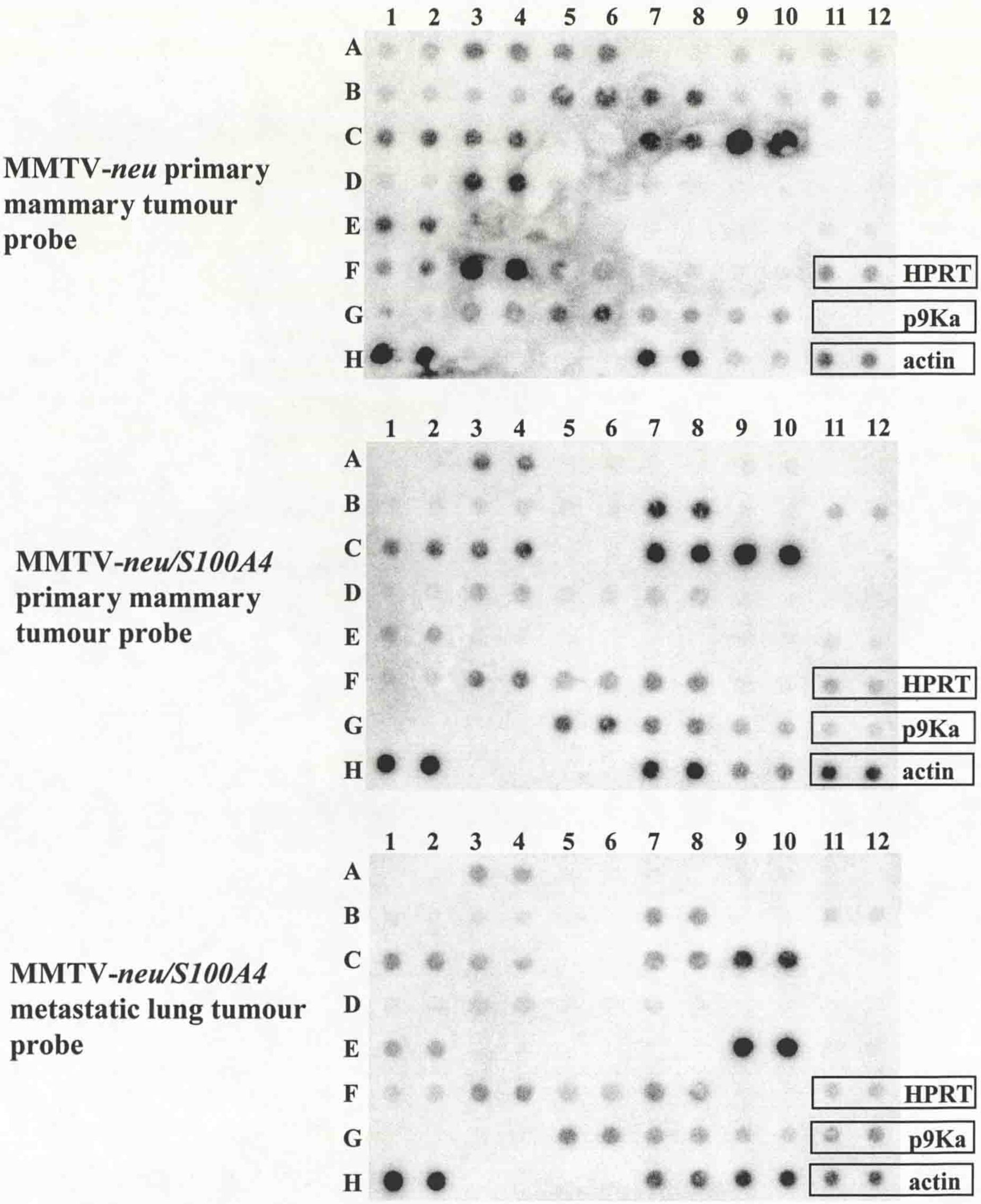


Figure 5.2 Reverse Northern re-screening of selected subtracted cDNA clones.

Some clones were re-spotted, in duplicate, onto fresh membranes and re-screened for expression in the same three tumours. This was to confirm that the differential expression pattern observed in the first screening was reproducible. These membranes also contained three control cDNAs p9Ka (S100A4), HPRT and actin.

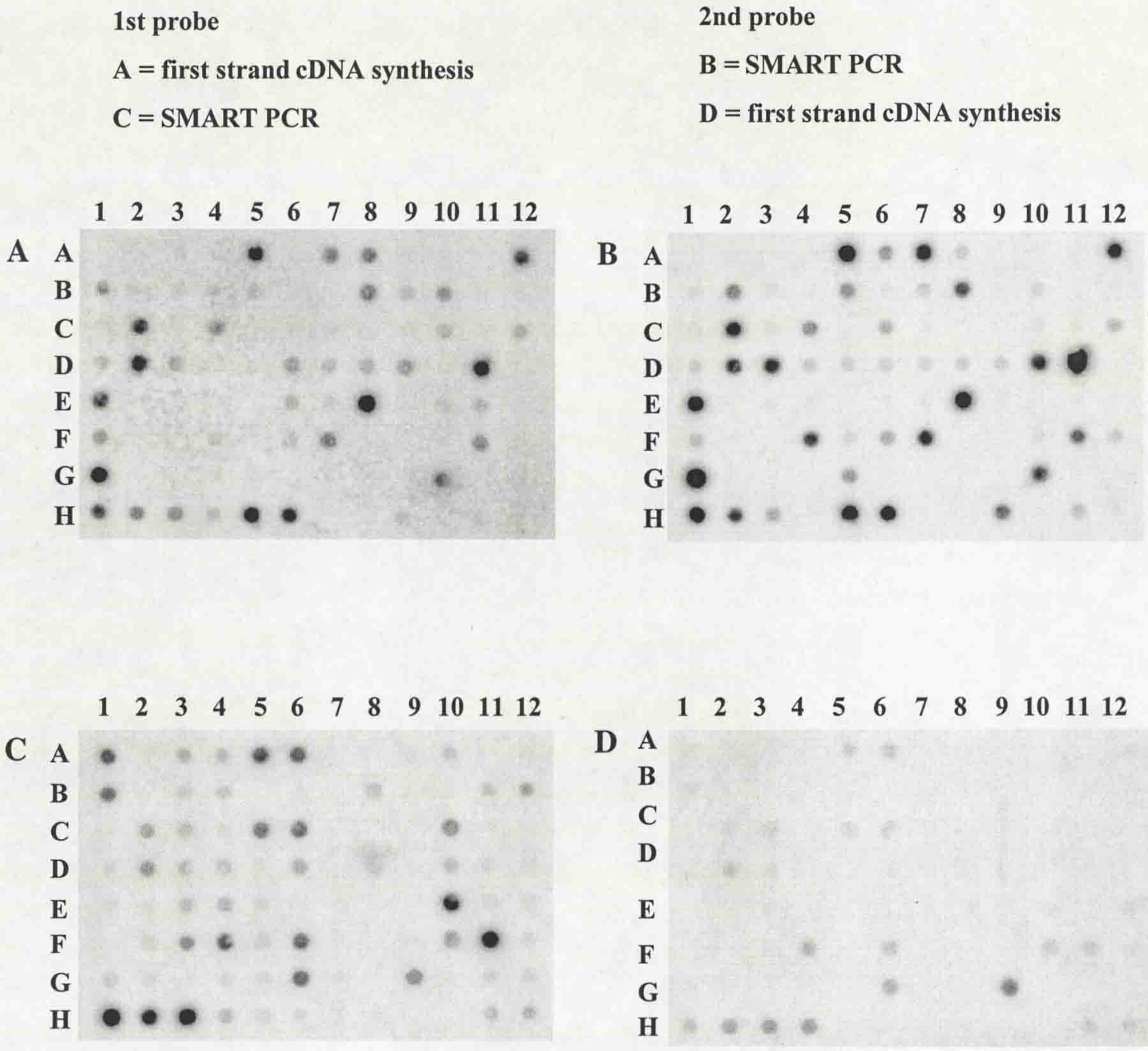


Figure 5.3 First-strand cDNA synthesised probes and SMART PCR-generated cDNA probes.

One membrane was probed firstly with a first-strand cDNA synthesis probe (A), stripped, and re-probed with a SMART amplified probe (B). MMTV-*neu*/*S100A4* primary mammary tumour probes were used for these hybridisations.

A second membrane was probed initially with SMART PCR probes (C) and, following stripping, with first-strand cDNA synthesis probes (D). This membrane was hybridised with MMTV-*neu* primary mammary tumour probes.

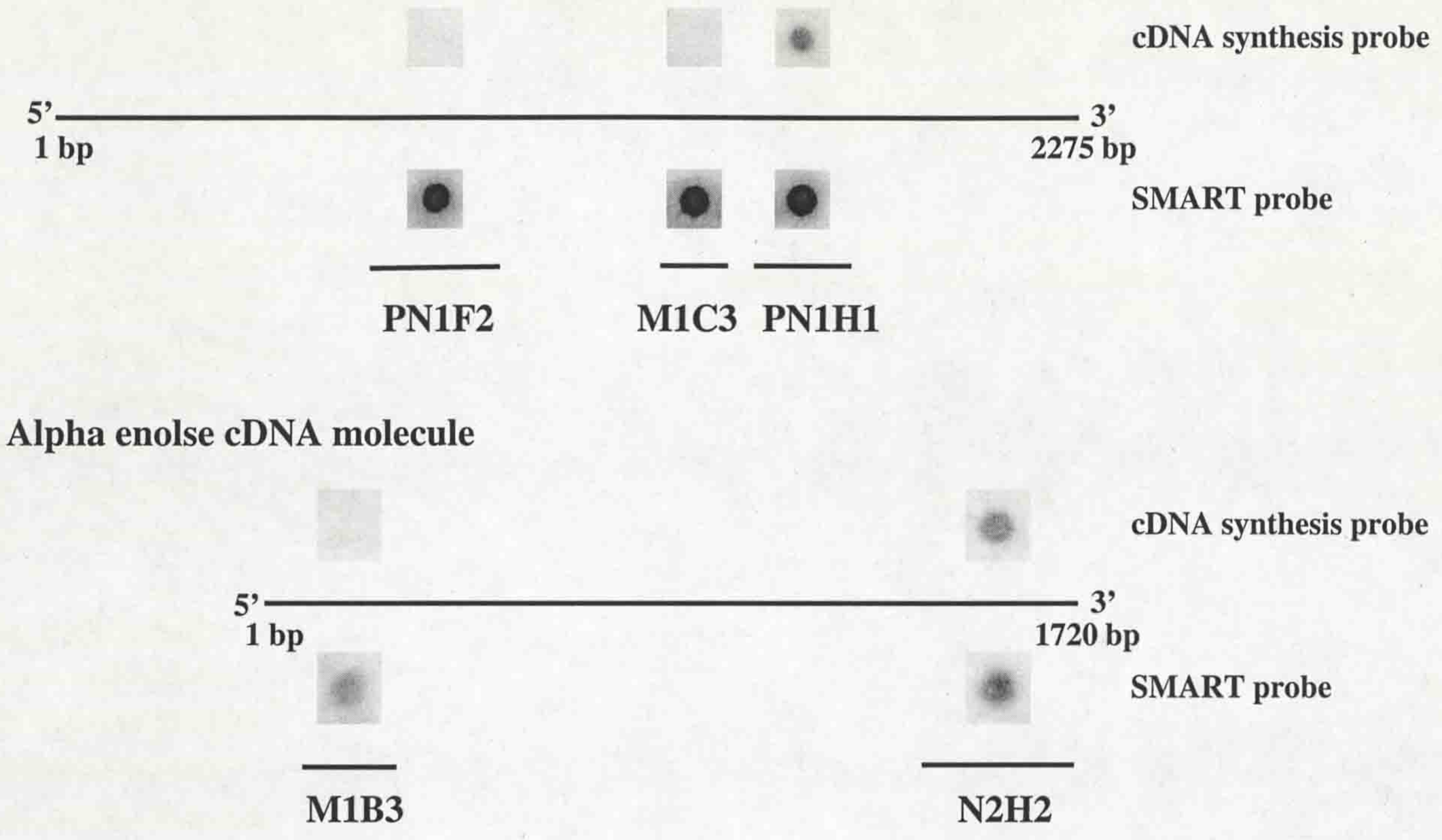


Figure 5.4 First-strand cDNA synthesised probes and SMART PCR-generated cDNA probes.

Diagrammatic representation of the transferrin and alpha enolase cDNA molecules, showing the position of several clones (PN1F2, PN1H1, M1C3, N2H2 and M1B3) whose sequence represents different regions of the two cDNAs.

Spot intensities for each clone following Reverse Northern screening with a cDNA synthesis probe and a SMART amplified probe indicates the ability of the SMART technology to synthesise full length cDNA. Conversely, the limitations, in terms of sensitivity and length of extension obtained from an oligo$(dT)_{16}$ primer (conventional first-strand cDNA synthesis) is demonstrated.

Diagrams are drawn to an approximate scale of 1 cm = 200 bp.

Subtracted Library	N	P	PN	Met
Number of clones screened	218	188	181	181
Number of clones with detectable signal	209	166	132	164
Number different by >2 fold	87 (42%)	29 (17%)	27 (20%)	48 (30%)
Number different by >5 fold	55 (25%)	0 (0%)	5 (3%)	20 (11%)
Number different in wrong orientation	3 (1%)	11 (7%)	5 (4%)	6 (3%)

Table 5.1 Reverse Northern screening of subtracted cDNA clones.
All clones were screened for expression levels in the three different tumours. This table presents the data produced by these experiments including the number of clones from each library differentially expressed (also given as a % of the number of clones with detectable expression) between any two of the three tumours by 2 fold or greater (refer to tables 5.2-5.5 for a comprehensive list of all differentially expressed clones), and by 5 fold or greater. The expression of a number of clones was too weak to detect by these means, particularly from the PN subtracted library. A small number of clones exhibited an expression pattern in the wrong orientation to that expected.

Clone	Sequence identification	Acc. number	Biological function
N2C5	c-kit	Y00864	receptor tyrosine kinase
P2G11	Thrombospondin-4	AF102887	extracellular matrix adhesion protein
PN1A10	MHR23A	X92410	nucloetide excision repair
PN1C5	MutS homolog 2	U21011	DNA mismatch repair
PN2D12	Serine protease inhibitor 3 (Sp13)	U25844	protease inhibitor
PN2F1	Maspin	U54705	protease inhibitor

Table 5.6 Selection of low expressing clones with sequence similarity to known genes of interesting physiological function.

Clone	Sequence Identification	Accession number	Redundancy	Fold expression data		
				N	PN	Met
N3G11	Mouse cDNA (diaphragm)	AA066911	1	**20.38**	1	1.25
N1B9	Mouse cDNA (myotubes)	AA597175	2	**12.17**	1	**2.39**
N2B3	Mouse mitochondrial genome	V00711	2	**11.37**	1	0.75
N2F4	?		1	**11.35**	1	0.83
N2B5	MMTV sequence	AF228552	2	**9.98**	**1**	0.38
N3E4	Mouse cDNA similar to endogenous MMTV (M11024)	AW231061	1	**8.73**	1	0.68
N3B2	MMTV sequence	L37517	3	**7.77**	1	0.68
N2G3	Mouse BRP 39	X93035	8	**7.14**	1	1.00
N/R11	MMTV sequence	L37517	5	**6.46**	1	1.15
N2F5	Mouse cDNA (skin)	AA789934	1	**6.32**	1	0.80
N2C6	Mouse cDNA (mammary gland) similar to cat major allergen I polypeptide chain 2 precursor (P30440)	AW989042	6	**6.98**	1	1.29
N3B4	Mouse apolipoprotein D	L39123	1	**5.64**	1	1.05
N3E1	Mouse cDNA similar to CRBP I	AA466092	7	**5.55**	1	0.56
N3C3	Mouse oral tumour suppresser homolog (Doc-1)	AF011644	1	**5.16**	1	1.10
N3F2	Mouse T cell specific serine protease	M13226	1	**4.64**	1	0.98
N3G10	MMTV sequence	X63024	4	**4.63**	1	0.56
N2B8	Mouse cDNA (neonate skin)	AV237160	1	**3.70**	1	0.66
N1B6	MMTV sequence	M11024	5	**3.50**	**1**	0.42
N1G5	?		1	**3.47**	1	1.21
N1A4	Mouse NonO (non-POU domain containing octamer-binding protein)	S6460	1	**3.46**	1	0.63
N1A6	Mouse cDNA clone (kidney)	AI790486	4	**3.36**	1	1.25
N3A6	Mouse mitochondrial genome/mouse perforin	L07096/X60165	2	**3.21**	1	1.80
N3H5	Mouse mitochondrial genome	V00711	4	**3.20**	1	**3.01**
N1E12	Mouse ribosomal protein L7 (rpL7)	M29016	1	**3.15**	1	1.30
N3F5	Mouse cDNA (kidney) similar to human 40s ribosomal protein S3A	AA208274	1	**2.97**	1	0.59
N2B9	Mouse cDNA clone	BF228009	1	**2.94**	1	0.92
N2D11	Mouse cDNA clone	AA170470	1	**2.93**	1	1.45
N2C11	Mouse cDNA clone (mammary gland)	AI606479	1	**2.71**	1	**5.01**
N/R1	Mouse cDNA clone (mammary gland)	AI152341	1	**2.65**	1	1.30
N1C5	Mouse cDNA clone	AI838686	1	**2.48**	1	**2.07**
N3G4	Mouse cDNA clone (placenta)	AA015548	1	**2.43**	1	0.88
N3E2	Mouse cDNA clone (liver)	AW012440	1	**2.31**	1	1.97
N3H11	Mouse MHC locus classIII region	AF049850	1	**2.25**	1	1.31
N2A3	Mouse cDNA clone (mammary gland)	BF452749	1	**2.22**	1	1.00
N/R10	No sequence		1	**2.18**	1	**2.18**
N2C12	Mouse cDNA clone (mammary gland)	BE625496	1	**2.11**	1	0.91
N3C2	Mouse cDNA clone (2 cell embryo)	AA422633	1	**2.11**	1	1.04
N2D8	Mouse adenine nucleotide translocase (Ant-2)	X70847	1	**2.04**	1	0.64
N3A3	Human splicing factor, arg/ser rich 7 (SFR S7)	L41887	1	**1.60**	1	0.75
N3C12	Rat SERP1	AB018546	1	**1.21**	**1**	0.40
N2H5	Mouse cDNA clone	AW989386	1	**0.97**	**1**	0.37
N2G2	Human basic transcription element binding protein BTEB1 (GC box binding protein)	D31716	1	0.85	1	**3.32**
N3H8	Mouse cDNA clone (kidney)	AI315022	1	0.53	1	**1.19**
N2E5	Human U5 snRNP-specific 200KDa protein	Z70200	1	0.42	**1**	0.69
N3G9	mouse ferritin light chain	J04716	1	0.39	**1**	**2.03**
N1A11	mouse cDNA clone (kidney)	AI647880	1	0.33	**1**	0.30

Table 5.2 Differentially expressed clones from the N subtracted cDNA library.
Subtracted cDNA clones were screened for expression in a MMTV-*neu* primary tumour and in MMTV-*neu/S100A4* primary and metastatic tumours. Signal intensities for each clone were quantified using Molecular Imager software and normalised for differences in probe strength. Fold expression differences for each selected clone were determined and those demonstrating a expression ratio (N/PN or Met/PN) of > 2-fold or < 0.5-fold were selected as being differentially expressed. Expression in the MMTV-*neu* primary tumour and in the MMTV-*neu/S100A4* metastasis are given in respect to expression in the MMTV-*neu/S100A4* primary tumour. Highlighted expression ratios (bold) denotes > 2-fold difference in expression between any two of the three tumours. For redundant clones and for clones analysed on multiple membranes, average fold expression differences were calculated. Sequence identification and associated accession numbers are given, as is the redundancy of each clone in this library. '?' refers to clones whose sequence does not match any sequence in the databases. 'no sequence' refers to clones that gave poor sequence.

Clone	Sequence Identification	Accession number	Redundancy	Fold expression data		
				N	PN	Met
P2E1	Mouse transcytosis associated protein p115	AF096868	1	**2.99**	1	**4.28**
P1H2	Mouse ubiquitin-homolgy domain protein	AF033353	1	**2.73**	1	1.38
P2F1	Rat alkaline phosphodiesterase	D30649	1	**2.62**	1	**3.65**
P2C6	Mouse mitochondrial genome	V00711	2	**2.50**	1	**2.33**
P2H1	Rat rab28 for ras-homolog GTPase	X78606	4	**2.12**	1	**2.48**
P2C12	No sequence		1	**2.05**	1	0.87
P2F9	No sequence		1	**2.03**	1	0.83
P2D5	Rat 14-3-3 protein theta subtype	D17614	1	**1.68**	1	0.72
P2D3	Mouse cDNA clone (macrophage)	AA867412	1	1.59	1	**2.68**
P2B12	Mouse cDNA clone (kidney)	AI648172	1	**1.14**	1	0.53
P1B2	No sequence		1	**1.04**	1	0.51
P2F12	Mouse non muscle myosin alkali light chain	U04443	1	**0.98**	**1**	0.47
P1A2	Mouse mitochondrial genome	J01420	1	**0.84**	**1**	0.24
P1A3	No sequence		1	0.49	**1**	0.52
P1B9	Mouse endothelial monocyte activation polypeptide I	U41341	1	0.49	**1**	**1.18**
P1E8	Mouse carcinoembryonic antigen (Cea10a)	L38422	1	0.34	**1**	0.65
P1B7	Mouse cDNA clone (embryo)	AI842801	1	0.35	**1**	**0.92**
P1D7	Mouse cDNA clone	AI047818	1	0.33	**1**	0.58
P1B8	?		1	0.31	**1**	**0.96**
P2B10	Mouse osteopontin	J04806	2	0.35	**1**	**0.79**
P2D6	?		1	0.30	**1**	**0.68**
P1A9	Mouse cDNA clone (adult male epididymus)	AV381752	1	0.29	**1**	**1.03**
P1F7	Mouse cDNA clone (melanoma)	AA153272	1	0.28	**1**	**0.94**
P2H5	Mouse mitochondrial genome/ATPase subunit 6	AF093677	1	0.27	**1**	0.26

Table 5.3 Differentially expressed clones from P subtracted cDNA library.

Subtracted cDNA clones were screened for expression in a MMTV-*neu* primary tumour and in MMTV-*neu/S100A4* primary and metastatic tumours. Signal intensities for each clone were quantified using Molecular Imager software and normalised for differences in probe strength. Fold expression differences for each selected clone were determined and those demonstrating a expression ratio (N/PN or Met/PN) of > 2-fold or < 0.5-fold were selected as being differentially expressed. Expression in the MMTV-*neu* primary tumour and in the MMTV-*neu/S100A4* metastasis are given in respect to expression in the MMTV-*neu/S100A4* primary tumour. Highlighted expression ratios (bold) denotes > 2-fold difference in expression between any two of the three tumours. For redundant clones and for clones analysed on multiple membranes, average fold expression differences were calculated. Sequence identification and associated accession numbers are given, as is the redundancy of each clone in this library. '?' refers to clones whose sequence does not match any sequence in the databases. 'no sequence' refers to clones that gave poor sequence.

Clone	Sequence Identification	Accession number	Redundancy	Fold expression data		
				N	PN	Met
PN1B9	MMTV sequence	K00556	1	**4.72**	**1**	0.23
PN1D4	Mouse stearoyl-CoA desaturase gene exon 6	M21285	1	0.38	**1**	0.27
PN2A1	Human cDNA KIAA0324	AA347209	1	**0.80**	**1**	0.28
PN1E8	Mouse endogenous ecotopic murine leukemia provirus locus 3 (Emv-3) envelope glycoprotein	L37057	1	0.36	**1**	0.39
PN1F11	MMTV sequence	AF228552	1	**2.28**	**1**	0.40
PN2G6	Mouse protein synthesis elongation factor 1-alpha	M22432	2	0.47	**1**	0.40
PN1E1	Mouse TANK binding kinase TBK1 (tbk1)	AF191839	1	0.26	**1**	0.42
PN1C6	Mouse chloride channel CaCC	AF052746	3	0.34	**1**	0.48
PN1G9	Mouse putative basolateral Na-K-2Cl co-transporter	U13174	1	0.34	**1**	0.49
PN2C8	cDNA clone	AW537697	2	**2.04**	**1**	0.50
PN1F12	Mouse serum deprivation response	S67386	1	**1.75**	1	0.81
PN1E11	MMTV sequence	M151122	1	**1.93**	1	0.82
PN2G7	Mouse protein kinase Chk2	AF086905	1	0.39	**1**	**0.86**
PN1A8	Human cDNA (mammary gland)	BE554003	1	0.44	**1**	**0.89**
PN1H2	Mouse cDNA (mammary gland)	BE197187	1	**2.51**	1	1.10
PN2F3	cDNA clone	AW53118	1	0.52	1	**1.39**
PN1H3	Mouse TNFR2-TRAF signalling complex protein	L49433	2	**2.07**	1	1.82
PN1H10	Mouse apolipoprotein D	L39123	1	**2.25**	1	**2.37**
PN1G2	Mouse cDNA (embryo)	AA003680	1	**3.35**	1	**2.89**
PN2E6	Rat cDNA (ovary)	AI406961	1	0.54	1	**3.01**
PN2E2	Mouse cDNA (stromal cell line) similar to human basic-leucine zipper nuclear factor	BE655833	1	1.12	1	**5.24**
PN2H1	Mouse cDNA (mammary gland)	AW744109	1	0.63	1	**8.04**

Table 5.4 Differentially expressed clones from PN subtracted cDNA library.

Subtracted cDNA clones were screened for expression in a MMTV-*neu* primary tumour and in MMTV-*neu/S100A4* primary and metastatic tumours. Signal intensities for each clone were quantified using Molecular Imager software and normalised for differences in probe strength. Fold expression differences for each selected clone were determined and those demonstrating a expression ratio (N/PN or Met/PN) of > 2-fold or < 0.5-fold were selected as being differentially expressed. Expression in the MMTV-*neu* primary tumour and in the MMTV-*neu/S100A4* metastasis are given in respect to expression in the MMTV-*neu/S100A4* primary tumour. Highlighted expression ratios (bold) denotes > 2-fold difference in expression between any two of the three tumours. For redundant clones and for clones analysed on multiple membranes, average fold expression differences were calculated. Sequence identification and associated accession numbers are given, as is the redundancy of each clone in this library.

Clone	Sequence Identification	Accession number	Redundancy	Fold expression data		
				N	PN	Met
Met1C2	Mouse Flavo-binding protein	AF174535	2	**0.99**	1	0.28
Met2A11	Mouse Calmodulin synthesis (CAM)	M27844	1	0.55	**1**	0.41
Met1B3	Mouse alpha enolase	X52379	1	**1.16**	**1**	0.43
Met2C4	No sequence		1	0.38	**1**	0.43
Met1C1	Mouse cDNA clone (diaphragm)	AA572388	1	**1.44**	**1**	0.49
Met2C10	Human cDNA clone (muscle)	AW176514	1	0.36	**1**	0.71
Met2D4	Mouse cDNA clone similar to human cAMP-dependent protein kinase type II beta regulatory chain	AI049217	1	0.28	**1**	**0.74**
Met1G6	Mouse cDNA clone (mammary gland)	AW988905	1	0.42	**1**	0.76
Met1F4	Mouse mitogen response 96KDa phosphoprotein/Doc-2	U18869	1	**4.55**	1	0.98
Met2H3	Mouse protein kinase inhibitor p58	U28423	1	**2.98**	1	0.98
Met1D1	Mouse cDNA clone (stomach)	BE852308	1	**2.21**	1	1.01
Met2D7	No sequence		1	0.36	**1**	**1.02**
Met1F7	Mouse cDNA (pancreas)	AU051688	1	0.42	**1**	**1.07**
Met1B7	Human SRp20	L10838	1	0.37	**1**	**1.09**
Met2G9	Mouse cDNA clone (embryo) similar to human MHC antigen A-24	AI173231	1	0.38	**1**	**1.09**
Met2E9	Mouse cDNA clone (mammary gland)	198680	1	0.37	**1**	**1.22**
Met1E5	Mouse cDNA clone (embryo)	AF052746	1	0.44	**1**	**1.22**
Met2F9	Mouse cDNA clone similar to acetyl CoA synthetase	AI790449	1	0.36	**1**	**1.25**
Met2B6	No sequence		1	0.40	**1**	**1.28**
Met2D8	Mouse cDNA clone	AA145325	1	0.57	1	**1.32**
Met2H7	Mouse calpactin I light chain (p11)	M16465	1	0.41	**1**	**1.37**
Met2H9	Human melastatin 1 (MLSN1)	AF071787	1	0.52	1	**1.40**
Met2E6	Mouse cDNA (kidney)	AI317707	1	0.60	1	**2.23**
Met1H4	Mouse cDNA clone	AI3190146	1	0.93	1	**2.37**
Met1H2	Mouse IL-8/gro-alpha receptor	L13239	1	0.59	1	**2.85**
Met2F6	Mouse intracellular calcium binding protein (MRP14)	M83219	1	0.87	1	**2.91**
Met2H4	Mouse cDNA clone (embryo) similar to NIP1	AW908668	1	0.86	1	**2.96**
Met2F11	Rat ribosomal protein L36a	M19635	1	0.56	1	**3.05**
Met1D7	Mouse cDNA clone (MG)	AI182473	1	0.63	1	**3.36**
Met2B8	Mouse matrix Gla protein	S77350	2	0.53	1	**3.46**
Met1C11	Mouse lysozyme	M21047	7	0.49	**1**	**3.57**
Met2G5	Mouse integrase interactor 1a protein (INI1A)	AJ011739	1	0.69	1	**3.89**
Met1A10	Mouse pulmonary surfactant protein C	M38314	8	0.88	1	**19.13**

Table 5.5 Differentially expressed clones from Met subtracted cDNA library.
Subtracted cDNA clones were screened for expression in a MMTV-*neu* primary tumour and in MMTV-*neu/S100A4* primary and metastatic tumours. Signal intensities for each clone were quantified using Molecular Imager software and normalised for differences in probe strength. Fold expression differences for each selected clone were determined and those demonstrating a expression ratio (N/PN or Met/PN) of > 2-fold or < 0.5-fold were selected as being differentially expressed. Expression in the MMTV-*neu* primary tumour and in the MMTV-*neu/S100A4* metastasis are given in respect to expression in the MMTV-*neu/S100A4* primary tumour. Highlighted expression ratios (bold) denotes > 2-fold difference in expression between any two of the three tumours. For redundant clones and for clones analysed on multiple membranes, average fold expression differences were calculated. Sequence identification and associated accession numbers are given, as is the redundancy of each clone in this library. '?' refers to clones whose sequence does not match any sequence in the databases. 'no sequence' refers to clones that gave poor sequence.

Clone	Sequence identification	Accession number	Fold expression data N	PN	Met	
PN2G6	Mouse protein synthesis elongation factor 1-alpha	M22432	**4.17**	1	N/D	
PN1G2	Mouse cDNA clone (embryo)	AA003680	1.16	1	**5.06**	
PN1H12	Mouse cDNA clone (mammary gland) similar to mouse ST2L protein	BE851645	0.46	**1**	N/D	
Met1C1	Mouse cDNA clone (mammary gland)	AA572388	**16.67**	1	0.83	
Met2F4	Mouse ribosomal protein S6	Y00348	**3.85**	1	N/D	
Met2B5	Mouse cDNA clone (mammary gland)	AA690429	**2.04**	1	N/D	
Met2D4	Mouse cDNA clone (similar to human cAMP-dep PK type II beta regulatory chain)	AI049217	**2.04**	1	N/D	
Met1F6	Mouse transcobalmin II (Tcn2)	AF090686	**1.52**	1	0.67	
Met2H5	Mouse cDNA clone	AW048554	0.45	**1**	N/D	
Met1H11	Mouse cDNA clone similar to TGR-CL10	AA087124	0.40	**1**	N/D	
N2C4	Mouse BRP39	X93035	**3.85**	**1**	0.42	*
N1A4	MMTV sequence joined to Nono (Pou domain octamer DNA binding protein)	L37517/S6460	**3.45**	1	N/D	*
N2B5	MMTV sequence	AF228550	**3.23**	1	N/D	*
N2G3	Mouse BRP39	X93035	**2.70**	**1**	0.22	*
N2C10	Mouse BRP39	X93035	**2.56**	1	N/D	*
N1A2	Mouse ADP-ribosylation-like factor homolog ARL6	AF031903	**2.33**	1	1.14	
N2B4	Mouse ubiquitin-conjugating enzyme (HR6A)	AF089812	**2.33**	1	0.58	*
N1C2	Mouse cDNA clone similar to human NADH-ubiquinone oxidoreductase subunit B9 homolog		**2.17**	1	N/D	
N1A3	Mouse cDNA clone similar to human 60s ribosomal protein L22		**2.13**	1	1.85	
N2A2	No sequence		**2.13**	1	1.13	
N1A11	Mouse cDNA clone	AI647880	1.72	1	0.60	
N2B3	Mouse mitochondrial genome	V00711	**1.27**	1	0.53	
N2G2	Human basic transcription element binding protein BTEB1	D31716	1.04	1	**3.92**	
N1A6	Mouse cDNA clone (kidney)	AI790486	**0.94**	**1**	0.24	*
N2A8	No sequence		0.50	**1**	N/D	
N2A11	No sequence		0.48	**1**	N/D	
N1C10	Mouse cDNA clone (mammary gland) similar to major allergen 1 polypeptide chain 2 precursor	AW989042	0.40	**1**	N/D	
N3A6	Mouse mitochondrial genome/perofrin	L07096/X60165	0.39	**1**	N/D	
N3G2	Mouse cDNA clone (mammary gland) similar to major allergen 1 polypeptide chain 2 precursor	AW989042	0.38	**1**	N/D	
N2D5	Mouse cDNA clone (mammary gland)	AI607404	0.38	**1**	**1.41**	
N2E11	Mouse kappa casein	M10114	0.36	**1**	N/D	
N2G11	No sequence		0.36	**1**	N/D	
N1A8	Mouse cDNA clone (kidney) similar to human putative thioredoxin like protein	AW108411	0.33	**1**	**1.02**	
P1A7	Mouse Id2	M69293	**2.56**	1	N/D	
P1A8	MMTV sequence	M11024	**2.38**	1	N/D	
P1C6	Mouse cDNA clone (mammary gland) similar to human glutathione s-transferase	AI642156	**1.69**	1	0.69	
P1A6	Mouse cDNA clone (mammary gland)	AA759618	**1.20**	**1**	0.52	
P1D1	No sequence		**0.90**	**1**	0.44	
P1D6	Mouse cDNA clone (embryo) similar to interferon-inducible protein 1-8D	AI931291	**0.83**	**1**	0.28	
P1D4	Human initiation factor 4B	X55733	0.66	**1**	0.48	
P1G1	?		0.66	**1**	0.49	
P/F3	Mouse cDNA clone (mammary gland)		0.50	**1**	N/D	
P1D10	Mouse ubiquitin-like protein	D10918	0.50	**1**	N/D	
P1C11	Mouse cDNA clone (embryo)	AV117154	0.45	**1**	N/D	
P1B9	Mouse endothelial monocyte activation polypeptide	U41341	0.42	**1**	N/D	*
P1A2	Mouse mitochondrial genome	J01420	0.20	**1**	0.36	
P1C3	No sequence		0.18	**1**	0.35	

Table 5.7 Differentially expressed clones identified by screening Reverse Northern membranes with probes from alternative tumours.

First strand cDNA synthesis probes were derived from tumours of matched transgenic status, histology and phenotype to those that were used to create subtracted libraries. The same criteria for identifying differential expression in previous Reverse Northern experiments was applied here, all clones demonstrating > 2-fold or < 0.5-fold difference in expression are presented. Some membranes were not re-screened with probes derived from a second metastasis, thus expression data for these clones in this tumour type was not determined (N/D). Those clones exhibiting a similar differential pattern of expression to that previously identified are indicated (*).

Chapter 6.

Characterisation of differentially expressed candidate cDNA clones.

Suppression Subtractive Hybridisation has been used to create cDNA libraries that are enriched for differentially expressed sequences between transgenic mouse tumours of differing metastatic capability. Expression screening of these sequences, by Reverse Northern hybridisation, determined that 25% of subtracted clones were differentially expressed between these tumours by ≥ 2-fold (Tables 5.1-5.5 and Table 5.7). These included known genes, ESTs and some potentially novel sequences. Data determined thus far, for these clones, is sufficient to associate differential expression with the S100A4-related metastatic progression of these individual tumours. However, further characterisation of these sequences is necessary to, firstly, confirm the differential expression pattern observed in the Reverse Northern hybridisation and, secondly, to determine how the expression of these clones changes relative to S100A4 expression or metastatic capability in other tumours from the transgenic mice.

A panel of candidate cDNA clones, deemed of particular interest has been selected for further characterisation. These sequences were selected based on identity to known, EST or unknown sequences, in combination with a significant pattern of expression with respect to S100A4 expression. The expression of these clones is investigated in normal transgenic mouse tissues, in multiple tumours from MMTV-*neu* and MMTV-*neu/S100A4* mice and in a series of cell lines derived from MMTV-*neu* and MMTV-*neu/S100A4* mammary tumours.

6.1 Confirmation of differential expression.

Virtual Northern analysis was used as a means of confirming the differential expression pattern of clones observed by Reverse Northern hybridisation. Virtual Northern membranes contained amplified SMART cDNA derived from the three tumours that were used to create the subtracted cDNA libraries and to screen Reverse Northern membranes.

6.1.1 Expression patterns of control cDNAs.

The expression pattern of control cDNAs actin, HPRT and S100A4 (p9Ka) were analysed and shown to exhibit similar expression patterns to those observed for the same cDNAs by dot blot hybridisation (Figure 3.3) and by Reverse Northern hybridisation (Figures 5.1 and 5.2). Actin was abundantly detected in all three tumours (Figure 6.1), and although slightly lower expression was observed in the metastasis, expression exhibited a <2-fold difference between these tumours. Likewise, HPRT expression was slightly lower in the metastasis, again by less than 2-fold, relative to expression in the two primaries (Figure 6.1). Relative to actin, HPRT was less abundant in these tumours, exhibiting a weaker signal following overnight exposure to autoradiographic film (Figure 6.1) also consistent to that observed previously. S100A4 transgene expression was progressively higher with increasing metastatic capability, being just detectable in the MMTV-*neu* primary tumour, highest in the metastasis and expressed to an intermediate level in the bitransgenic primary (Figure 6.1). Detection of the same expression patterns for control cDNAs to that previously observed for the same sequences by Reverse Northern hybridisation and dot blot hybridisation suggest that expression patterns detected by Virtual Northern for subtracted cDNAs clones is likely to reflect their abundance in these tumours.

6.1.2 Expression patterns of subtracted cDNA clones.

Subtracted cDNA clones N2G3, N1A6, N2F4, N3C3, P1B9, P2B10, P2D6, PN1B9 and Met1D7 have been screened for expression in these tumours. Reverse Northern analysis demonstrated that all these clones were differentially expressed between these three tumours (Tables 5.2-5.5). Of this panel of 9 clones, N2G3, N1A6 and P1B9 demonstrated a similar pattern of differential expression in additional tumours by first-strand cDNA synthesis probing of Reverse Northern membranes (Table 5.7).

The expression of 7 cDNA clones, observed by Virtual Northern, confirmed the pattern of expression that was previously detected by Reverse Northern hybridisation. Clones N2G3, N1A6, N2F4, N3C3 and PN1B9 demonstrated expression that was higher in the MMTV-*neu* primary tumour relative to the two bitransgenic tumours (Figures 6.1). Clones P2B10 and P2D6 exhibited expression that reflected S100A4 expression, which was higher in the two bitransgenic tumours

relative to the MMTV-*neu* primary tumour (Figure 6.1). The approximate sizes of transcript detected for each cDNA clone are shown alongside the differences in expression identified for these clones in these tumours (Figure 6.1).

Clones N2F4 and P2D6 exhibit sequence that is unique to any sequence in the database, these clones hybridised to cDNA fragments approximately 2.0 kb and 2.1 kb in size respectively (Figure 6.1). Expression of N2F4 was only detected after a 5 day exposure to autoradiographic film suggesting this sequence was of a low abundance in the MMTV-*neu* primary tumour and of an even lower abundance (undetectable) or was not expressed in the bitransgenic tumours. This reflects the significant difference in expression of this sequence observed by Reverse Northern between these tumours (N=11.35, P=1, Met=0.83; Table 5.2). Expression of clone P2D6 was detected in all three tumours by Virtual Northern hybridisation and expression was approximately 10-fold higher in the bitransgenic tumours relative to the MMTV-*neu* tumour (N=0.09, P=1, Met=0.99; Figure 6.1). This was slightly different to that observed by Reverse Northern, where the abundance of P2D6 was slightly lower in the metastasis (by 1.5-fold; Table 5.3) relative to the bitransgenic primary tumour.

Interestingly, cDNA clone PN1B9, which possessed sequence similarity to an MMTV-derived sequence, hybridised to three cDNA transcripts (3.8 kb, 2.1 kb and 1.2 kb) in the MMTV-*neu* primary tumour and to two (3.8 kb and 1.2 kb) transcripts in the MMTV-*neu/S100A4* primary tumour (Figure 6.1); no expression of these fragments was detected in the metastasis. The relative signal intensity for each of these bands was higher in the MMTV-*neu* primary relative to the bitransgenic primary by 11 to 30-fold, which, although reflecting the same expression pattern to that seen previously, is much greater than the 4.7-fold difference in expression between these two tumours determined by Reverse Northern (Table 5.4).

Clone P1B9, previously suggested to be more abundant in the bitransgenic tumours relative to the MMTV-*neu* primary tumour was actually equally expressed in the three tumours suggesting the identification of a false positive clone (not shown). P1B9 was a high abundant sequence in these tumours as determined by the short exposure time to autoradiographic film that was required to detect expression. The expression of a mouse mammary gland cDNA, clone Met1D7, was not detected by Virtual Northern suggesting a low abundant transcript.

6.2 Expression patterns of subtracted cDNA clones in transgenic mice.

Northern hybridisation was used to screen for the expression of this panel of subtracted cDNA clones in normal mouse tissue, in tumours from different MMTV-*neu* mice and MMTV-*neu/S100A4* mice and in cell lines derived from rodent mammary gland carcinomas. A selection of normal mammary gland tissue from both transgenic and non-transgenic mice was screened to determine if the expression of these clones varied or was related to the different physiological states of the normal tissue in these mice. Lung, spleen, brain, kidney and lymph node were also screened to determine expression patterns of selected clones under normal circumstances. Several tumours from the mice were also screened to determine the expression levels of the selected clones in these tumours relative to the (i) normal tissues, (ii) transgenic status and (iii) tumour phenotype of mice bearing these tumours. The selected primary mammary tumours were from MMTV-*neu* and MMTV-*neu/S100A4* mice, they were all grade 3 IDC tumours with similar histological characteristics to the majority of tumours from these mice and to those used previously for gene expression analysis experiments. RNA derived from a single metastasis was also present on Northern membranes.

RNA derived from six cell lines generated from transgenic mouse mammary tumours were also screened. These were generated as part of a separate Ph.D. project by sub-cloning cells derived from an MMTV-*neu* primary mammary tumour (cell lines 4N and 8N) and an MMTV-*neu/S100A4* primary mammary tumour (cell lines PN1 and PN2). Cell lines 4N and 8N were also stably transfected with a genomic DNA fragment containing the rat S100A4 gene to generate the additional cell lines 4NP and 8NP.

6.2.1 Expression patterns of control cDNAs.

The expression pattern of the control cDNAs actin, HPRT and S100A4 were analysed in all tumours, cell lines and normal tissues to provide expression patterns with which to compare the expression seen for subtracted cDNA clones.

Actin was detected in all samples, although to a variable degree (Figure 6.2 and 6.3), which is to be expected in normal tissues of different origins and therefore containing cells of different physiological functions. This can be appreciated in tumours of abnormal cell phenotype, particularly when the expression of S100A4, a potential actin binding protein and regulator of cytoskeletal dynamics (Gibbs *et al.*,

1994; Davies *et al.*, 1993; Watanabe *et al.*, 1993), is being expressed in some tumours and is influencing the metastatic potential of some of these cells.

The expression of HPRT was also analysed in these samples, however expression was not detected in any samples reflecting the relative low abundance of this transcript as determined by the low signal intensity obtained by Reverse Northern (Figure 5.1 and 5.2) and by Virtual Northern (Figure 6.1). Hybridisation with the HPRT cDNA probe was the last hybridisation experiment performed on these Northern membranes suggesting that compromised sensitivity due to repeated probing and stripping may have contributed to this lack of detection.

S100A4 expression was detected with variability in these samples. With reference to previous experiments, expression was high in the normal lung tissue and low in the brain (Davies *et al.*, 1995, 1996) and was also relatively high in all cell lines examined (Figure 6.2 and 6.3). Expression was weakly detected in all MMTV-*neu/S100A4* tumours, in mammary gland and spleen but was not detected in the other normal tissues or in any MMTV-*neu* tumours examined (Figure 6.2 and 6.3). This weak expression of S100A4, however, is not clearly visible in the figures of the scanned autoradiographs.

6.2.2 Expression patterns of subtracted cDNA clones.

The same panel of clones screened by Virtual Northern hybridisation (N2G3, N1A6, N2F4, N3C3, P1B9, P2B10, P2D6, PN1B9 and Met1D7) was also screened by Northern hybridisation to determine the expression in the different samples of RNA.

The expression of clone M1D7 was not detected by Northern hybridisation, suggesting this sequence was either expressed to a level below that of the sensitivity of Northern hybridisation, or that this sequence was not expressed in these tissues or tumour samples.

Expression of clone N3C3 was barely detectable in any RNA sample (Figure 6.2) suggesting a normally very low abundance in these tissues or a very specific tissue distribution. Expression was detected to equal levels in only three tumours, one MMTV-*neu* and two MMTV-*neu/S100A4* primaries (Figure 6.2) after 5 day exposure to autoradiographic film. The size of transcript (approximately 1.0 kb) matches that reported for Doc-1 (AF011644) to which this clone possesses significant sequence identity.

The size of RNA transcript detected using clone N2G3 (BRP39) as a probe was 1.6 kb. Expression was detected to a low level in mammary gland and brain and to a high level in lung. Expression was detected, to a variable level, in all tumours examined. MMTV-*neu* primary tumours exhibited the highest and lowest relative level of expression, which differed by approximately 45-fold (lanes 7 and 8, Figure 6.2).

Clone P1B9 (mouse endothelial monocyte activation polypeptide I, EMAP I) hybridised to a transcript approximately 300 bp in size. Expression was observed in mammary gland, lung, spleen, lymph node and kidney, but was not detected in brain. All tumours expressed this transcript to an equivalent level (Figure 6.2).

The expression pattern of osteopontin in these RNA samples was determined using clone P2B10 as a probe that hybridised to an RNA transcript of 1.4 kb. Expression was particularly high in the majority of mammary gland tissues and in kidney. Three out of 7 MMTV-*neu* primary tumours and 6 out of 8 MMTV-*neu/S100A4* tumours demonstrated variable levels of expression (Figure 6.2 and 6.3). Expression was also high in all the cell lines that were derived from primary tumours of the transgenic mice.

PN1B9 demonstrated very limited expression, being detected in 2 bitransgenic mammary gland tumours (where 2 transcripts were detected of 2.7 kb and 7.5 kb) and in the metastasis (2.7 Kb fragment). The detection of more than one transcript in these tumours reflects that detected by Virtual Northern in the two primary tumours. The sizes are different however and may reflect the expression of different MMTV-transcripts in these tumours.

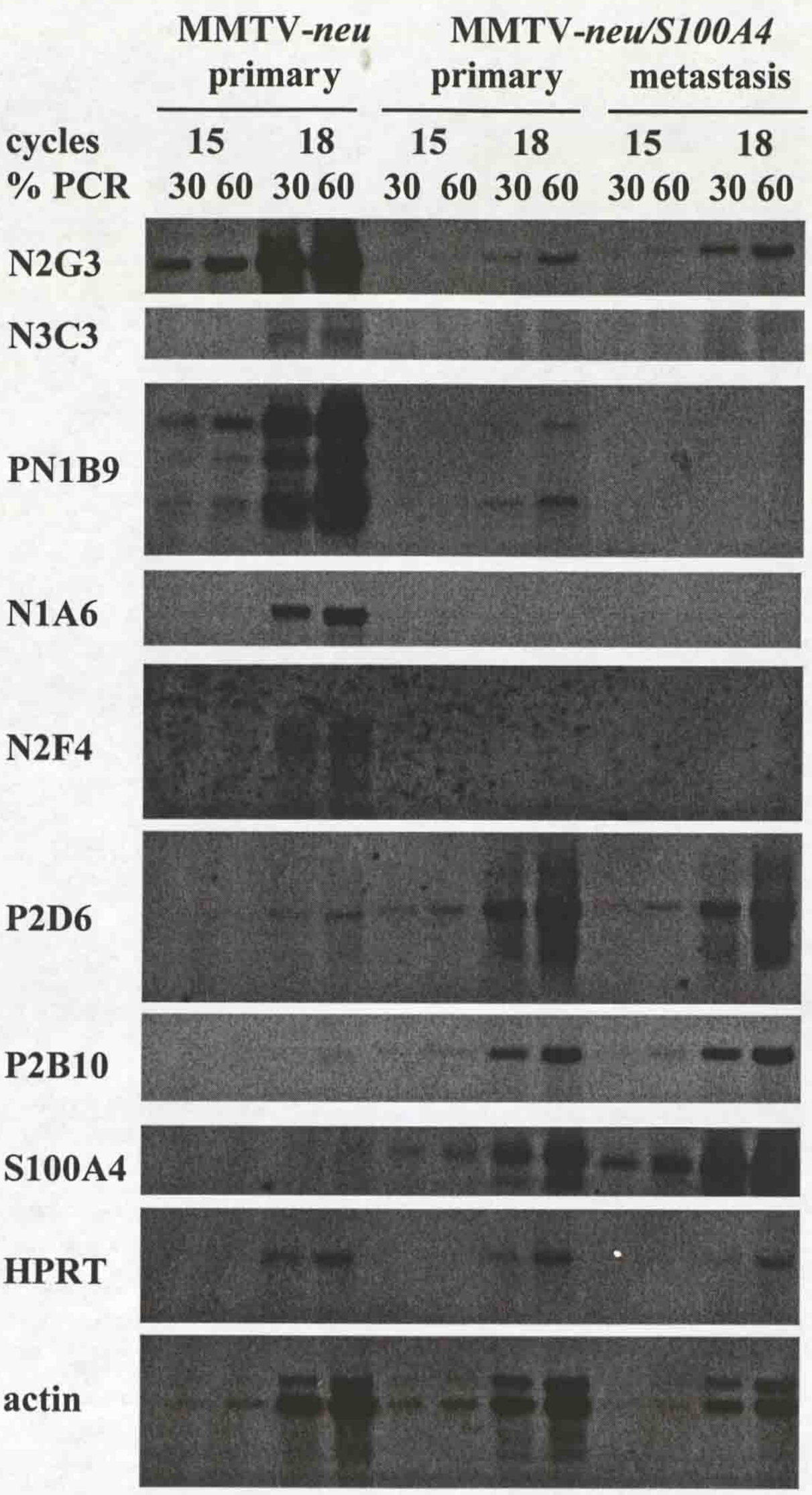
Expression of cDNA clone N1A6 was detected in kidney but in no other normal tissue examined. Many of the tumours demonstrated expression of this sequence yet to a lower abundance than that observed in the kidney (Figure 6.2) and to a variable level between tumours. Two bitransgenic mouse primary tumours exhibited the highest level of expression of this 1.9 kb transcript.

Multiple hybridisation signals were detected by Northern analysis with probes N2F4 and P2D6, some of which may be non-specific (Figures 6.3). Three distinct bands of 2.5 kb, 3.2 kb and >7 kb are observed on different membranes hybridised with N2F4. The smaller 2 bands appear as a doublet of equal abundance in 3 out of 6 MMTV-*neu* primary tumours (lanes 8,10 and 11, Figure 6.3), a fourth tumour demonstrated lower abundance of the larger of the two transcripts (lane 9).

One out of 8 MMTV-*neu/S100A4* tumours (lane 12, figure 6.3) demonstrated expression of this doublet to the same low abundance as the MMTV-*neu* primary tumour in lane 9. Faint detection of this doublet, or one of the two transcripts, was detected in 2 out of 8 MMTV-*neu/S100A4* primary tumours and in the 6 cell lines. Hybridisation with clone P2D6 produced signals reflecting the positions of the 28S and 18S ribosomal bands, also a signal of a very high molecular weight was detected in tumour derived RNA samples only.

6.3 Conclusions to chapter.

The expression of several candidate, differentially expressed cDNA clones were selected for further investigation on the basis of maintained or enhanced expression relative to progressively enhanced S100A4 expression and metastatic capability. Virtual Northern analysis has confirmed the expression pattern observed by Reverse Northern analysis for 7/9 candidate clones. The expression of the remaining clones was proven not to be differentially expressed (clone P1B9) or was not detected by these means (clone M1D7). Fold expression differences between tumours was relatively consistent between methods. Although quantitative comparison was not relevant when clones exhibited detectable expression in one tumour and undetectable expression in another. For example, expression of clones N1A6, N2F4, N3C3 and PN1B9 were not detected in the metastasis and expression of clone P2B10 was not detected in the MMTV-*neu* primary tumour. Northern analysis of the same clones was undertaken and also demonstrated differential expression of these sequences between multiple tumours from the transgenic mice. Differential expression was not confined to tumours of different transgenic status since, for example, the expression of BRP39 was different by approximately 45-fold between two MMTV-*neu* primary tumours of matched grade and histology. Thus it appears that despite all the tumours being initiated by the *neu* transgene and displaying similar histological characteristics, they demonstrate some degree of tumour heterogeneity.



	Transcript size	Fold expression N	PN	Met
N2G3	1.6 kb	16.3	1	1.14
N3C3	1.0 kb	5.7	1	0.97
PN1B9	3.8 kb	30.8	1	ND
	2.1 kb	ND	ND	ND
	1.2 kb	11.0	1	ND
N1A6	1.9 kb	ND	ND	ND
N2F4	2.0 kb	ND	ND	ND
P2D6	2.1 kb	0.09	1	0.99
P2B10	1.4 kb	0.15	1	1.57
S100A4	0.3 kb	0.01	1	7.01
HPRT	1.3 kb	1.13	1	0.77
actin	1.9 kb	0.81	1	0.71

Figure 6.1 Virtual Northern analysis of candidate clones.

The differential expression of candidate subtracted clones was confirmed by hybridisation to Virtual Northern membranes containing SMART PCR-generated cDNA for the three tumours utilised to create subtracted libraries. Membranes were prepared by electrophoresis of one third and two thirds of each SMART PCR reaction amplified for 15 and 18 cycles and transferring to membrane overnight. The size of transcripts detected and the fold expression difference for each tumour, relative to the MMTV-*neu/S100A4* primary tumour is given. ND denotes expression not detected in one or two tumours.

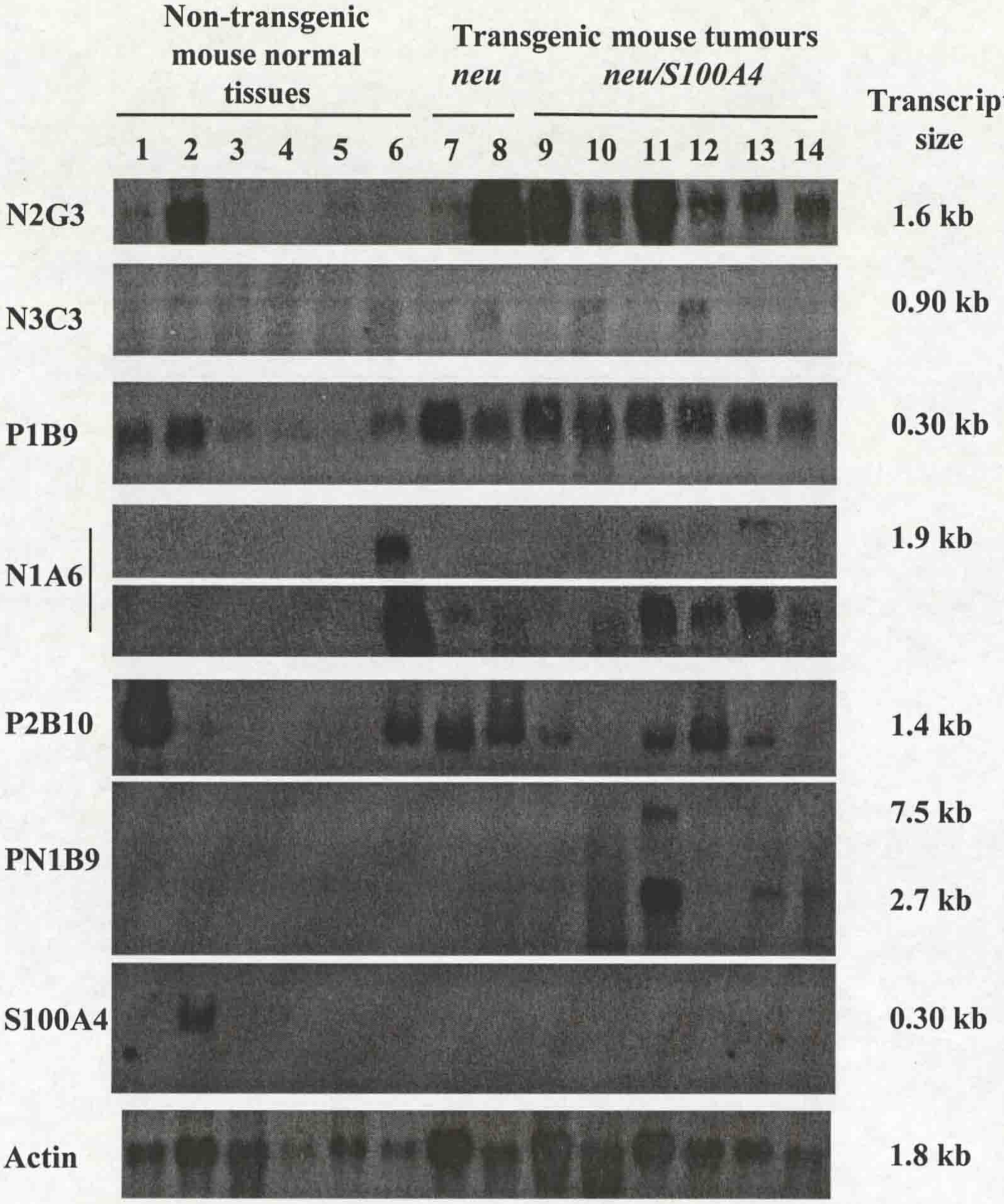


Figure 6.2 Northern hybridisation of selected candidate cDNA clones.

Several cDNA clones shown to be differentially expressed by Reverse Northern and Virtual Northern have been screened for expression in other tissues and tumours from the transgenic mice. Northern membranes contained RNA from the following normal tissues: mammary gland (lane 1), lung (2), spleen (3), lymph node (4), brain (5) and kidney (6). Several tumours from the transgenic mice were also screened, these were from 2 MMTV-*neu* mice (lanes 7 and 8) and from 5 MMTV-*neu/S100A4* mice (lanes 9-13) and a lung metastasis (lane 14). 10 μg total RNA was analysed for all tissues except for the metastasis, where 5 μg total RNA was used. Two exposures are given for N1A6, one following overnight exposure and the second following a 10 day exposure to autoradiographic film.

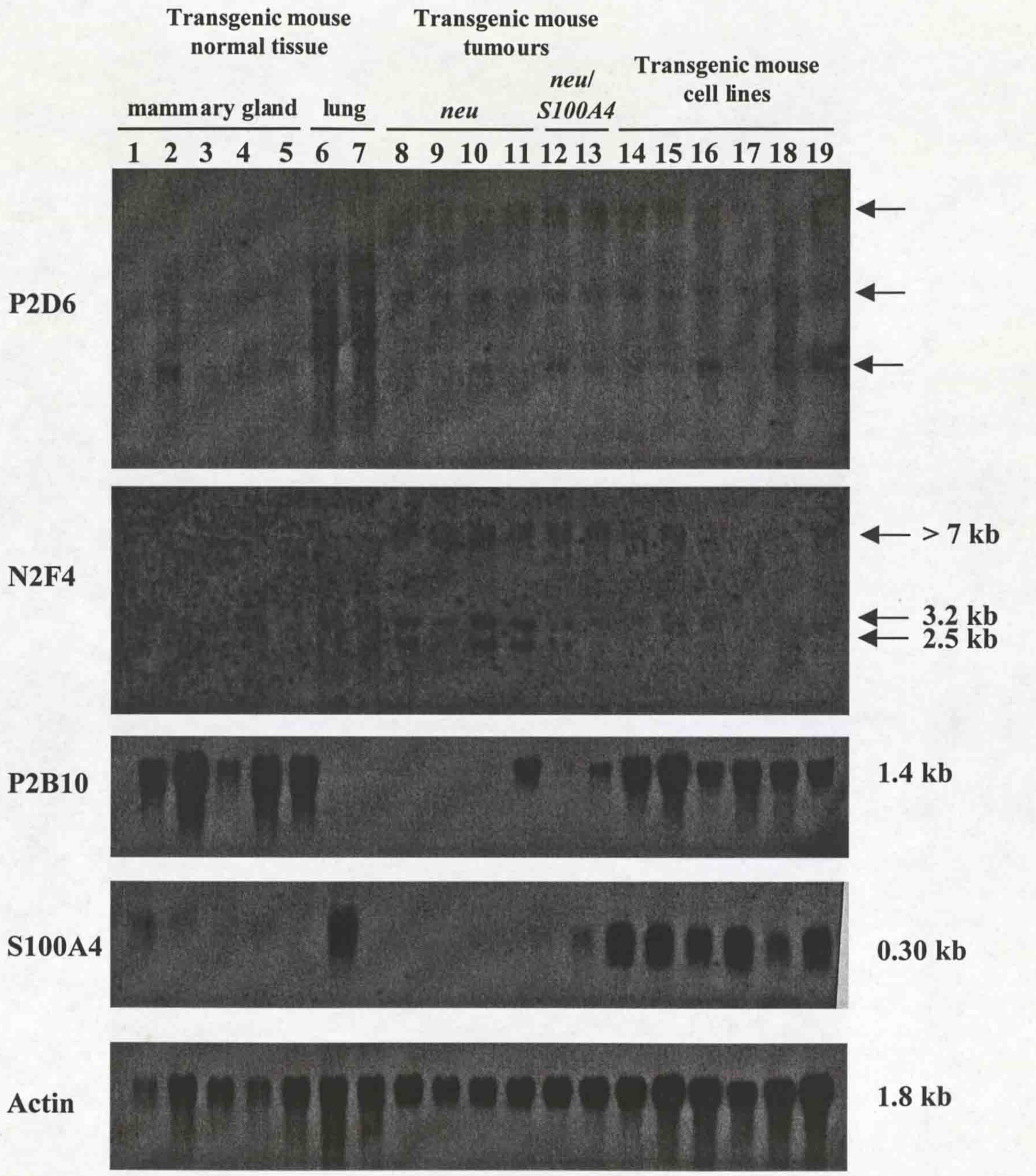


Figure 6.3 Northern hybridisation of selected candidate cDNA clones.

Several cDNA clones shown to be differentially expressed by Reverse Northern and Virtual Northern have been screened for expression in other tissues and tumours from the transgenic mice. Northern membranes contained RNA from the normal mammary gland (lane 1-5) and lung (6,7), from MMTV-*neu* primary mammary tumours (8-11) from MMTV-*neu/S100A4* primary mammary tumours (12, 13) and from cell lines derived from transgenic mouse mammary tumours (lane 14 = cell line PN1, 15 = PN2, 16 = 4N, 17 = 4NP, 18 = 8N and 19 = 8NP; see text).

Chapter 7.

Discussion.

7.1 S100A4 in metastasis

The role of S100A4 in metastasis is largely unknown yet considerable evidence indicates that it has an important part to play in the metastatic progression of some cancers. The transfection of S100A4 into normally benign rodent or human cells is able to induce metastasis (Davies *et al.*, 1993; Grigorian *et al.*, 1996; Lloyd *et al.*, 1998) and the expression of S100A4 in non-metastatic mammary gland tumours of transgenic mice enhances the metastatic capability of these cells (Davies *et al.*, 1996; Ambartsumian *et al.*, 1996). The expression of S100A4 in a series of human tumours has also associated elevated expression of this protein with metastatic spread of breast and colorectal carcinomas (Platt-Higgins *et al.*, 2000; Rudland *et al.*, 2000; Takenaga *et al.*, 1997).

7.2 Identifying gene expression changes co-ordinated with S100A4 expression.

It has been demonstrated that multiple genetic changes are required for the generation of a tumourigenic cell population (Hahn *et al.*, 1999) and it has been speculated that a minimum of 4-7 genetic changes, and the associated acquisition of new cellular properties, are required for the progressive conversion of a normal cell to a metastatic cell population (Renan, 1993; Hanahan and Weinberg, 2000). It is also widely accepted that the complex series of events required for metastatic progression occurs as a consequence of an accumulation of genetic aberrations (Hart *et al.*, 1989; Devilee and Cornelisse, 1994). In line with this, specific genetic anomalies have been identified in cancer development and progression, some of which have been utilised to identify different subtypes of cancer (c-*erb*B-2; Slamon *et al.*, 1987, 1989), or used as markers to indicate the potential of a tumour to spread (loss of E-cadherin; Christofori and Semb, 1999; up-regulation of uPA or cathepsin D; Duffy, 1996). It has also been inferred that gene expression profiles can be utilised to classify tumour types (Perou *et al.*, 1999, 2000; Alizadeh *et al.*, 2000). However, it is unknown if specific expression anomalies co-ordinate with S100A4 to

induce metastasis or whether alternative changes in gene expression occur in different tumours to produce the same effect.

We have used a transgenic model of metastatic breast cancer to investigate gene expression changes that are associated with S100A4-related metastatic progression. These mice, discussed in detail earlier, have been generated to contain two genetic changes, that of the activated *neu* oncogene (Bouchard *et al.*, 1989) and multiple copies of the metastasis-related gene S100A4 (Davies *et al.*, 1995, 1996). Expression of Neu alone is insufficient for tumour development since the occurrence of primary mammary gland tumours is sporadic. Expression of the Neu protein is also insufficient for metastasis to occur, yet elevated expression of both neu and S100A4 in the same cells sporadically enhances the metastatic capability of mammary gland tumours. Thus, it appears that elevated level of both these transgenes alone is insufficient for completion of metastatic progression (Davies *et al.*, 1996). This difference is unlikely to be associated with the level of S100A4 expression since as few as 1% of human breast carcinoma cells staining positive for S100A4 signifies a poor disease outcome in a group of patients with advanced breast cancer (Figure 1.4a; Rudland *et al.*, 2000). It is expected that additional genetic changes have occurred in association with elevated S100A4 expression that enhances the metastatic capability of some MMTV-*neu/S100A4* tumours. The identification of these changes would help to determine the difference between tumour phenotype that is observed in different animals and the identification of altered gene expression patterns that associate with *neu* and *S100A4* to induce metastasis.

7.3 Selection of tumours for gene expression analysis.

Gene expression comparisons were made, by both cDNA array screening and in SSH, between an MMTV-*neu/S100A4* primary tumour and a metastasis from the same bitransgenic mouse as a means of discovering genetic changes directly associated with metastatic progression. In addition, the expression profiles of MMTV-*neu/S100A4* primary mammary tumour(s) were compared with those from MMTV-*neu* primary mammary tumour(s) to identify expression changes that are related to, or associated with elevated S100A4 expression. In cDNA array experiments, several primary mammary tumours from the two strains of mice were pooled so that potential molecular heterogeneity between tumours could be avoided

and that cDNAs consistently differentially expressed in multiple tumours could be identified.

Tissue microdissection has become a major aspect in investigating the molecular pathology of cancer (Sirivatanauksorn *et al.*, 1999) particularly in determining the expression pattern or mutational state of target sequences (Krizman *et al.*, 1996; Weber *et al.*, 1998; Fend *et al.*, 1999; Davies *et al.*, 1999). This technology was available for the generation of pure populations of tumour cells for these gene expression studies, however, this was not applied, in part because of the importance of the reactive stromal tissue that surrounds and interacts with the tumour cells. Complex interactions between the growing tumour and its local microenvironment are likely to be important determining factors in the overall behaviour of a tumour. This is observed by the expression of matrix degrading enzymes by fibroblasts and macrophages at the leading edge of an invading tumour (Carriero *et al.*, 1994; Tetu *et al.*, 1999; Sternlicht *et al.*, 2000), and by the activation, via paracrine signalling from tumour cells, of endothelial cell growth and angiogenesis (Carmeliet, 1999; Kerbel, 2000). Thus, the inclusion of reactive stromal tissue in such gene expression experiments is important and therefore enhances the prospect of discovering important metastasis-related genes that are not necessarily expressed by tumour cells. RNA was therefore extracted from either whole frozen tumour lumps, or from frozen histological tissue sections.

7.4 Technical Issues.

Several new technologies have been utilised during this project. Some of the important technical issues relating to these methods will be discussed prior to demonstrating their effectiveness at identifying differentially expressed genes.

7.4.1 SMART PCR-generated cDNA.

This technology has been utilised in a number of experiments and so its utility in gene expression profiling requires some justification. The inclusion of an amplification step prior to investigating differential gene expression patterns raises the possibility of PCR bias or amplification efficiency. The misrepresentation of transcripts due to over-amplification is an obvious concern for these experiments. Careful optimisation of SMART PCR was employed on each occasion it was utilised to avoid over-amplification of the cDNA, demonstrated in Figures 3.2 and 4.3. Dot

blot hybridisation was also performed to determine if the representation of 3 cDNAs (actin, HPRT and S100A4) was maintained following amplification. The relative abundance of the three cDNAs remained the same pre- and post-amplification in the tumour samples where both SMART cDNA and sufficient total RNA was available (Figure 3.3) suggesting the representation of this limited number of cDNAs was maintained following PCR.

The manufacturers of the technology suggest that the representation of the original message profile is maintained following amplification, providing samples are not over-amplified. They have demonstrated this in dot blot experiments investigating the relative expression of certain cDNAs in multiple samples both pre- and post-amplification (Zhumabayeva *et al.*, 2000), and by studying the representation of 45 specific cDNAs in SMART cDNA, synthesised from 50 ng total RNA, relative to cDNA synthesised by conventional means from 5 μg poly(A)$^+$ RNA. 42 out of the 45 cDNAs demonstrated expression pattern that was maintained post-amplification (Chenchik *et al.*, 1998). The use of this technology in gene expression profiling was also investigated independently of Clontech researchers. Endege *et al.*, (1999) showed that SMART PCR amplified cDNA libraries maintained the representation of the original message profile.

In the present experiments SMART cDNA transcripts were of full length or near to full length which was not achieved by conventional cDNA synthesis in Reverse Northern probe labelling reactions. Figure 5.4 demonstrates the inability of first-strand cDNA to contain the 5' ends of mRNAs for α-enolase and transferrin, whereas SMART cDNA was used to both produce the cDNA prior to SSH and also to detect expression of the subtracted sequences in Reverse Northern hybridisation. Detection of the 5' end of HPRT following SMART cDNA synthesis was also evident by Reverse Northern analysis (Figure 5.1). Virtual Northern analysis of candidate cDNA clones demonstrated the approximately same sized transcripts to those detected for the clones by Northern hybridisation (Figures 6.1-6.3). This indicates the utility of this technology in creating near to full length cDNA and also highlights its possible use in obtaining full length clones of specific cDNAs, for example of previously unknown sequences P2D6 or N2F4.

7.4.2 cDNA array hybridisation for identifying differentially expressed genes.

The availability of mouse Atlas arrays provided an opportunity to screen for potential differential expression of 588 known cDNAs (plus 9 control cDNAs). All the cDNAs possess a functional relevance to cancer development and metastatic progression suggesting that these arrays were ideal for identifying expression changes associated with S100A4-related metastatic progression. These arrays have been extensively used to identify genes related to cancer progression (for example see Smid-Koopman *et al.*, 2000)

The relative level of intensity of hybridisation signal was used as a measure of the relative mRNA abundance in the tumour samples, thus, the pattern of signal intensity was related to the expression profile of cDNAs in the four tumour samples. Comparisons of abundance of all cDNAs between tumours revealed a striking similarity in the expression profiles of all four tumour samples. Only one cDNA (out of a total of 1194 comparisons made) was differentially expressed by > 2-fold, that for 'defender against cell death 1' (DAD 1, U83628), which was 2.3-fold more abundant in the metastasis relative to the primary tumour from the same mouse. The expression of the remaining 596 cDNAs between the bitransgenic primary tumour and metastasis, and all 597 cDNAs between the two pooled sets of primary tumours exhibited expression within this 2-fold difference ratio.

DAD1 was identified as a suppresser of mammalian cell death, in that expression of this cDNA was associated with cells that did not undergo apoptosis (Nakashima *et al.*, 1993; Sugimoto *et al.*, 1995). It was later found that this protein was an integral subunit of the enzyme oligosaccharyltransferase (OST) and that loss of DAD1 from this complex results in defective N-linked glycosylation of proteins and the induction of apoptosis (Kelleher and Gilmore, 1997). This evidence highlights the essential nature of N-linked glycosylation in the survival of mammalian cells, and suggests that loss rather than gain in expression of this gene is likely to have a greater affect on cell phenotype.

The same sequences were found to be highly expressed in the four tumour samples examined including several members of the cathepsin family of matrix degrading enzymes, cathepsin B1 (M14222), D (X53337), H (U06119) and L (X06086). Enhanced expression of these sequences have been previously associated with invasive breast carcinomas (Duffy, 1996; Lah and Kos, 1998), and so their high abundance in each tumour sample may relate to the invasive nature of all of the *neu*-

induced tumours. Several other cDNAs encoding proteins associated with cell survival in response to cellular stress were also highly expressed, particularly glutathione-metabolising proteins (microsomal glutathione S-transferase (J03752), glutathione S-transferase Pi 1 (D30687) and glutathione S-transferase 5 (J04696)), heat shock proteins (M38629, M36830, X53584) and clusterin (ApoJ; L08235). Glutathione S-transferase has previously been identified as being differentially expressed in cancer (Murray *et al.*, 1993; Salinas and Wong, 1999). Elevated expression of clusterin (ApoJ; L08235) protected a human epidermoid cancer cell line from apoptosis during cellular stresses such as heat shock and oxidative stress (Viard *et al.*, 1999) and has recently been identified as over-expressed in human breast carcinoma (Redondo *et al.*, 2000). The transfection and subsequent elevated expression of clusterin also protected L929 cells (a murine fibrosarcoma cell line) against tumour necrosis alpha (TNFα)-mediated cytotoxicity (Humphreys *et al.*, 1997), suggesting clusterin expression plays a protective role in response to cellular stresses and TNFα-induced cell death. High expression of this group of genes, and the distinctly similar expression profile of the majority of cDNAs may reflect the closely matched histology of these, and the majority of neu-induced mammary gland primary tumours and metastases in these transgenic mice. However this also suggests that the expression screening using these arrays was ineffectual at distinguishing between tumours of different metastatic capability.

A subtle difference in the expression patterns of certain cDNAs was observed between different tumours or sets of pooled tumours (Figures 3.4-3.5, arrows; Tables 3.3-3.4). These were identified by utilising less stringent differential expression criteria of >1.25- or < 0.8-fold between tumour samples. The majority of these sequences, 33/52 (63%), exhibited slightly higher expression in the metastasis relative to the primary tumour from the same mouse (Table 3.3), and includes a number of the sequences discussed above.

The validity of using this selection criterion is difficult to predict since the reliability in measuring small changes in expression using cDNA array studies is rarely investigated. Often researchers do not need to question this reliability or variation since sufficient numbers of sequences differentially expressed by >2-, >5- or >10-fold are often identified (Hufton *et al.*, 1999, Sgroi *et al.*, 1999). It must be questioned whether these small differences in expression are related to biological variation between tumour samples or are acquired due to technical variations in the

experiments. If these small differences are related to differences in transcriptional regulation, it must also be considered whether a minimal difference in expression of specific genes or of a co-regulated group of genes is sufficient to affect cell phenotype. This would be difficult to ascertain, yet has been discussed by Hughes *et al.*, (2000) with respect to gene expression profiling of yeast (*S. cerevisiae*). The expression of 6,300 genes under 300 experimental conditions were investigated which required processing of 700 gene chips and analysis of 10 million data points. This is a considerable amount of data, however it enabled groups of co-regulated genes of known and similar function and unknown function, to be reliably clustered on the basis of different expression profiles, including differences in expression of less than 2-fold or less than 1.5-fold. The substantial number of expression profiles generated by these experiments highlighted that small changes in expression of groups of genes were reproducibly identified and were considered to be important in determining cell phenotype. It has also been demonstrated that elevated osteopontin expression by 1.6-fold correlated with induced metastasis in an *in vivo* model of metastasis (Chen *et al.*, 1997). Thus, it is possible that the marginal increases in expression of the group of sequences in the metastasis (Tables 3.3 and 3.4) can contribute towards a gain in metastatic capability.

Nevertheless, the limitations of the technology mean that the expression of a very small fraction of the total number of genes likely to be expressed in a mammalian cell (12,000-33,000 different transcripts; Bishop *et al.*, 1974; Axel *et al.*, 1976) are investigated. Although these cDNAs are potentially very interesting, there is no scope for discovering new genes or screening genes not previously implicated in cancer development or metastatic progression. S100A4, for example, was also not present on these arrays and so it is also likely that the expression of other genes expected to be involved in S100A4-related metastatic progression will be missed using this technology.

7.4.3 Combined SSH and Reverse Northern analysis for identifying differentially expressed genes.

A direct analysis of the gene expression profiles that are specific to these different tumours was undertaken by using SSH to clone and identify potentially differentially regulated sequences. Reverse Northern hybridisation was subsequently

applied to screen large numbers of subtracted cDNA clones to determine differential expression profiles for specific cDNAs.

7.4.3.1 Size of subtracted cDNA libraries.

Few published experiments employing SSH to identify differentially expressed genes have presented the actual size of the subtracted libraries produced. Diatchenko *et al.*, (1999) predict 10^4 clones would be produced by cloning 1 µl of the secondary PCR products, equating to a total of approximate $2.5x10^5$ independent clones, this is a similar number to that produced in the subtractive hybridisation of these transgenic mouse tumours (Table 4.1). Hufton *et al.,* (1999) created a subtracted repertoire of 35,000 clones, which is an order of magnitude lower than the four libraries produced here. The size of the library produced, however, is dependent on several factors, in particular the number of genes that are differentially expressed between the samples being compared, but also the efficiency of the PCR amplification, the subtractive hybridisation and the cloning of subtracted cDNA fragments. The efficiency of subtractive hybridisation is discussed in detail below. cDNA fragments generated by 12 cycles of secondary PCR, used to enrich for subtracted cDNA fragments, were cloned into pCR2.1™ to create the subtracted libraries. The Sybr Gold stained gel in Figure 4.6 however, demonstrates that PCR products generated by 10 cycles of amplification could also have been cloned. This would be expected to produce smaller subtracted libraries due to the smaller number of amplified cDNA fragments generated from the fewer PCR cycles. The cloning efficiency of secondary PCR products was high, as determined by the average 93% of individual colonies picked that contained amplifiable inserts by colony-PCR (Table 4.1).

7.4.3.2 Complexity of subtracted cDNA libraries.

Estimating the overall complexity of each subtracted library created is useful to determine whether the clones that were characterised are representative of the whole library or whether studying additional clones would identify more non-redundant sequences (singletons). Diatchenko *et al.,* (1999) suggested that, in general, this could be achieved by sequencing 500-1000 colonies from a subtracted library, however we were unable to analyse this many clones from each of the four subtracted libraries.

Approximately 200 independent colonies representing an estimated 0.05% of each subtracted library were sequenced and identified by nucleotide database homology searching. In each library some sequences, representing single mRNAs, were identified multiple times, other clones were non-redundant sequences (Table 4.3). In each library the majority of clones studied were non-redundant, for example, 76%, 77% and 82% of clones from the P, PN and Met libraries, respectively, were of sequence only identified once. This distribution of sequence abundance does not reflect the expression profile of these genes in the starting material due to the equalisation of mRNA abundance that occurs during subtractive hybridisation (see below). However it does reflect the complexity of the subtracted libraries created, and infers that these three libraries were of an equally high complexity. The portion of clones studied from these libraries is unlikely to reflect the full diversity of clones present in the whole library. Therefore characterisation of additional individual clones from these libraries would be expected to identify more non-redundant sequences at a regular frequency. A greater number of redundant clones were identified from the N subtracted library, 45% of clones were sequenced multiple times. Additional sequencing of more independent colonies from this library, relative to the other three libraries, would be required to characterise additional non-redundant clones from this less complex library.

7.4.3.3 Colony Hybridisation.

Colony hybridisation experiments made it possible to screen a number of colonies to determine the distribution of individual cDNAs in the four subtracted libraries and also in non-normalised, non-subtracted cDNA libraries. This was of particular importance as the rat transgene S100A4 was not identified in any of the four subtracted libraries. This was despite S100A4 being known to be differentially expressed between the two primary and metastatic tumours, as observed on dot blots of total RNA and SMART cDNA (Figure 3.3) and on Reverse Northern membranes (Figure 5.1). S100A4 expression was highest in the metastatic lung tumour and lowest in the MMTV-*neu* mammary tumour. Successful subtractive hybridisation would therefore see enrichment of S100A4 cDNA fragments in the P and Met subtracted cDNA libraries relative to the N and PN subtracted cDNA libraries respectively. However S100A4 cDNA was not identified in any these libraries by sequence analysis.

Approximately 1000 colonies (an estimated 0.2% of subtracted libraries) from each of the four subtracted cDNA libraries were screened using an S100A4 cDNA probe. At the same time approximately 1000 colonies from non-normalised, non-subtracted cDNA libraries derived from the same three tumours were also screened. Only 6 positive colonies were detected, 2 from the P subtracted library and 4 from the Met subtracted library (Table 4.2). This supports data indicating differential expression of this transgene in the three tumours and suggests that it has been subtracted efficiently. The experiment also suggests that picking a minimum of 500 colonies from the P subtracted library and 250 colonies from the Met subtracted library would be required to detect the S100A4 cDNA, further highlighting the need to screen more colonies in order to characterise a representative portion of each subtracted library. The transgene was not detected in either of the non-normalised, non-subtracted cDNA libraries. Unlike the subtracted cDNA libraries, the non-subtracted cDNA libraries have not been normalised for mRNA abundance, so the number of different cDNA molecules actually screened in these 1000 colonies is expected to be few and representative of the high abundant mRNA molecules such as glutathione S-transferase, cathepsin B1, β-actin, glyceraldehyde 3-phosphate dehydrogenase, ubiquitin and cytoskeletal keratin 18, as determined by Atlas cDNA array hybridisation experiments in chapter 3.

Successful enrichment for S100A4 was further emphasised in respect to the prevalence of actin and HPRT cDNAs in subtracted and non-subtracted libraries, both were only identified once in the non-normalised, non-subtracted libraries only (Table 4.2). Actin and HPRT were shown to be more abundant than S100A4 in the primary tumours and to not be differentially expressed between any of these tumours by dot blot (Figure 3.3) and Reverse Northern (Figure 5.1). Subtraction of these sequences would therefore be expected, and this is consistent with results of colony hybridisation.

7.4.3.4 Normalisation of mRNA abundance.

It was suggested in chapter 3, based on previous reports of distinct abundance classes in mRNA prevalence (Bishop *et al.*, 1974; Axel *et al.*, 1976), that as few as 5% of the different transcripts expressed in these tumours (high abundant sequences) contribute to approximately 50% of the message profile. Conversely the majority of sequences (70% of transcripts) contribute to less than 20% of this profile. Thus,

without equalising mRNA abundance during subtractive hybridisation, subtracted libraries are expected to closely reflect non-subtracted libraries, containing differentially expressed sequences of high and medium abundance (Wan *et al*., 1996; Kuang *et al*., 1998).

A key aspect of the SSH technique, therefore, is the normalisation of mRNA abundance that occurs during hybridisation (Gurskaya *et al*., 1996) which aids in the cloning of rarely transcribed, as well as high abundant differentially expressed sequences. This is particularly important as many regulatory proteins are rare transcripts. A small change (for example 2-fold) in expression of these mRNAs may have profound effects on cellular phenotype. In the present experiments successful normalisation of mRNA abundance has occurred in all four subtractive hybridisation experiments, demonstrated by the distribution of individual sequences and the different hybridisation signal intensities obtained for specific clones in Reverse Northern, Virtual Northern and Northern blot analysis.

The complexity of the four subtracted libraries demonstrates the low redundancy of the libraries, particularly those derived from the two bitransgenic tumours (libraries P, PN and Met). This alone indicates successful normalisation because a non-normalised cDNA library is expected to predominantly contain the highly abundant messages, with the less abundant messages in lower prevalence. The lower complexity of the N subtracted cDNA library does not reflect poor normalisation because many of these redundant sequences were shown to be differentially expressed by Reverse Northern, including BRP39, MMTV derived sequences, cDNA clones similar in sequence to CRBP and major allergen polypeptide and other cDNA clones such as N1A6 (Table 5.2).

Although the amount of cDNA applied to Reverse Northern membranes was not consistently the same for all clones, it was usually more than the 10 ng that is applied to commercial membrane arrays (such as those from Clontech that were utilised in chapter 3), and was generally between 10-100 ng. Therefore the hybridisation signal intensity and the length of exposure time required to detect this signal, is expected to reflect the relative mRNA abundance of each particular sequence in the original tumour sample rather than the amount of DNA immobilised to the membrane. The hybridised Reverse Northern membranes (Figure 5.1) demonstrated a wide range of hybridisation signal intensities. Strong signals were detected after only an overnight exposure to autoradiographic film or

phosphorimaging screen, and represent clones expressed with high abundance. Transferrin, WDNM1, protein synthesis elongation factor 1-α, kappa casein, endothelial monocyte activating polypeptide I (EMAP I) were all highly expressed in these tumour specimens. The expression of clones present at a medium abundance produced a relatively strong hybridisation signal after an exposure time of one week and weak or no hybridisation signal was detected for very low abundantly expressed sequences. This indicates that subtracted cDNA clones represent a mixed population of high, medium and low abundant mRNAs. The same relative abundance demonstrated by Reverse Northern was observed following characterisation of selected clones by Virtual Northern and Northern analysis. Actin, BRP39, EMAP produced very strong hybridisation signals in these experiments, whereas Doc-1, S100A4 and cDNA clone N1A6 required longer exposure times to detect expression. The expression of some clones, for example M1D7, was not detected by these means. Despite the clear differences in abundance shown here, many of these clones were sequenced in a non-redundant manner, indicating normalisation towards cDNAs of lower abundance has occurred.

One major advantage of SSH over other differential screening techniques therefore is the normalisation of mRNA abundance and the cloning and identification of rarely transcribed messages. However the insufficient sensitivity of Reverse Northern and Northern expression screening experiments means that expression of some of these have to be investigated using more sensitive techniques such as quantitative RT-PCR.

7.4.4 Validation of Reverse Northern hybridisation.

The reliability of this technique was analysed to determine how suitable the technique was for distinguishing between differences in sequence abundance of specific mRNAs between different tumours. For example, hybridisation signal intensities for control cDNAs, of redundant clones on multiple membranes and the re-screening of redundant and non-redundant clones in multiple experiments gave relatively consistent data. Although, this was not always reproducible, as demonstrated for clones with sequence similarity to Id2 (section 5.2.3) and for the subsequent screening of clone P1B9 (EMAP I, section 6.1.2) by Virtual Northern analysis (which discounted this clone from being differentially expressed). Virtual Northern analysis did, confirm the differential expression of 7 out of 9 candidate

cDNAs, indicating the relative efficiency of Reverse Northern analysis at identifying differential expression of subtracted clones.

7.4.4.1 Data interpretation.

Careful data interpretation was required to confirm that the Reverse Northern quantified data obtained was correct and reflected the expression pattern observed on the autoradiographs and phosphorimages. Inconsistencies between the images and the data were occasionally identified; this appeared to be due to high background signal intensity enhancing the relative expression level of individual clones. This was particularly evident for clones which produced low or undetectable hybridisation signal or for clones which were adjacent to highly expressed clones producing a strong hybridisation signal that interfered with neighbouring spot intensities. As a consequence, falsely called differential expression of some clones and missed differential expression of other clones was encountered, yet this was accounted for by detailed examination of the hybridised membranes and the quantified data.

7.4.4.2 Cross-hybridisation.

Expression hybridisation experiments such as the Reverse Northern technique or the use of microarrays can suffer from cross-hybridisation problems due to sequences that are common to gene families or functionally related sequences. This can lead to false hybridisation signals and therefore the identification or loss of identification of differentially expressed sequences. Manufacturers of arrays attempt to overcome this problem with careful design and selection of the cDNA fragment regions that are immobilised to the solid support (membrane or glass slide), often employing the 3' untranslated region (UTR) sequences and avoiding repeat sequences and poly(A)$^+$ tails. The cDNA expression array hybridisation experiments developed by Clontech, and described in chapter 3, enhance the specificity of hybridisation signal by reducing the complexity of the labelled probe, this also serves to increase the sensitivity of the experiment. Reduced probe complexity is achieved by utilising a gene specific primer mix for probe labelling, rather than global SMART cDNA amplification or oligo(dT) cDNA synthesis as used in Reverse Northern experiments and in routine array hybridisation experiments. This produces probes that are complementary only to target sequences on the membrane, thus reducing background and cross hybridisation. These technical modifications were

sought to minimise cross-hybridisation and therefore ensuring hybridisation signals are specific to the cDNA sequences in question. Such luxuries were not available in our Reverse Northern experiments due, primarily, to the random selection of colonies from the transformed subtracted cDNA libraries. Approximately 30% of these clones contain poly(A)$^+$ sequences, some also have regions of repeat motifs suggesting that cross-hybridisation is a possibility in some screening experiments.

High stringency hybridisation and washing conditions were applied to limit the hybridisation of cDNAs that do not match the full length of the immobilised target DNA sequences. This seemed to be relatively effective as many cDNAs did not produce a hybridisation signal, suggesting, firstly, that they were low abundantly expressed sequences, and secondly that non-specific hybridisation interactions had not occurred for these clones. In fact, for those clones where expression was detected, a range of hybridisation signal intensities was observed (strong, medium and weak signals) reflecting the relative different abundances of the different messages. This suggests, not only that probe over-amplification has not occurred, which would have resulted in a general, low level of hybridisation signal intensity for many medium, low and non-detectable clones, but also that the high stringency of hybridisation and washing has limited non-specific hybridisation interactions.

7.5 Effective cDNA subtraction.

Comparison of hybridisation signal intensities for specific clones screened in different tumours was used to determine which of the subtracted cDNA clones were differentially expressed. The numbers of differentially expressed sequences between these tumours indicate the success of these subtractive hybridisation experiments. A total of 192 subtracted clones (25% of the total number of clones with detectable expression) were identified which exhibited differential expression between these transgenic mouse tumours by ≥ 2-fold, 80 of these clones exhibit ≥ 5-fold differences in expression. Screening of 540 of the same subtracted clones for expression in different tumours of matched histology from the transgenic mice identified 47 differentially expressed sequences.

Comparison can be made between the success demonstrated in the present experiments relative to those observed for other systems using the same technology. However it must be stated, again, that the number of differentially expressed cDNA clones identified is dependent on the similarity in expression profiles between

experimental samples and so this comparison can only be utilised to further emphasise the efficiency of the procedure. SSH was used for analysing differences in the expression profiles of estrogen receptor (ER) positive, MCF7, and ER negative, MDA-MB-231, breast cancer cell lines. Characterisation of 42 subtracted clones by Northern hybridisation revealed 29 (69%) were differentially expressed (Kuang *et al.*, 1998). Subsequently, 332 clones generated in the same MCF7 subtracted library were spotted onto a microarray and screened for expression in the same two cell lines. With a differential expression ratio of 3, 76 (23%) were considered to be differentially expressed. The same clones were also screened for expression in two other breast carcinoma cell lines: ER positive T47D and ER negative HBL-100. Only 22 clones (7%) were differentially expressed, by the same factor of 3-fold, in both ER positive cell lines relative to the ER negative cell lines (Yang *et al.*, 1999). Hufton *et al.*, (1999) described the use of SSH and Reverse Northern analysis to screen a subtracted library created to contain genes over expressed in human colorectal cancer (CRC) compared to normal colonic mucosa. Of the 400 clones screened 45 (11%) were detectably over expressed in the tumour compared to the normal tissue (Hufton *et al.*, 1999). Similarly, 12.5% of subtracted clones (625 of 5000 clones analysed) were identified as being differentially expressed by Reverse Northern of cDNAs from a rat pancreatic metastatic adenocarcinoma cell line, Bsp73-ASML, subtracted from its non-metastatic counterpart, Bsp73-1AS (von Stein *et al.*, 1997). The numbers of differentially expressed clones identified in these latter experiments is somewhat low considering the nature of the tissue used (cell lines) and that only 26% of the genes over-expressed in colorectal cancer cell lines were also over-expressed in primary tumour tissue (Zhang *et al.*, 1997).

The individual subtracted libraries exhibited different proportions of differentially expressed sequences. The N subtracted cDNA library contained the greatest enrichment for differentially expressed sequences, where 87 clones (42%) were differentially regulated by ≥ 2-fold. This was relative to the other three subtracted libraries which, by the same criteria identified 29 clones (17%), 27 clones (20%) and 49 clones (30%) in the P, PN and Met subtracted libraries respectively. Enrichment for differentially expressed sequences in all four libraries is comparable to the examples described from the literature. This also demonstrates significant enrichment for differentially expressed sequences relative to Atlas array experiments which demonstrated distinct similarities in the expression profiles between the same

MMTV-*neu/S100A4* primary tumour and metastasis, and also between multiple mammary gland tumours from MMTV-*neu* and MMTV-*neu/S100A4* mice.

The different levels of enrichment for differentially expressed sequences exhibited by different libraries is likely to reflect biological differences between tumours rather than technical variations between experiments. The N subtracted library has been shown to be less complex than the other three libraries, yet this was not due to poor normalisation as many of the redundant sequences were differentially expressed. A high level of normalisation was achieved in the P, PN and Met libraries, as demonstrated by the high library complexity and the mixed range of abundance exhibited by P, PN and Met cDNA clones. The PN library in particular demonstrated high level of normalisation due to the high proportion of low abundant sequences identified (27%; Table 5.1 and as shown clearly in Figure 5.1).

It appears, therefore, that many technical aspects employed for generating the subtracted libraries and screening the subtracted clones was relatively efficient and comparable between libraries. Thus, the differences observed between the N subtracted library and the P, PN and Met subtracted libraries are expected to reflect differences between the tumours themselves. The two primary tumours were of a carefully matched grade and histology yet differed in S100A4 expression and metastatic capability. On face value, therefore, elevated S100A4 expression in the bitransgenic primary tumour appears to correlate with the down-regulated expression of many genes, more so in fact than the up-regulation of genes, as demonstrated by the large number of differentially expressed sequences (by $\geq$ 2-fold and by $\geq$ 5-fold) in the N subtracted library and the few identified in the P library. The similarities in gene expression profiles between the MMTV-*neu/S100A4* primary tumour and metastasis is evident by the few numbers of differentially expressed sequences in the PN and Met libraries, and also by considering the expression of the majority of differentially expressed N and P clones exhibiting similar expression patterns in the two MMTV-*neu/S100A4* tumours (Tables 5.2 and 5.3). This suggests that the two bitransgenic tumours are more closely matched, in terms of their gene expression profiles, than either of these tumours are to the MMTV-*neu* tumour. The reasons for this can be speculated on, not least that these two tumours are from the same animal suggesting the distinct possibility that the metastasis is a clonal expansion of cells derived from the primary tumour. Alternatively, elevated expression of S100A4 in both tumours may be associated with the same changes in gene expression observed

between these two tumours and the MMTV-*neu* tumour. Determining the identity of such changes in expression that are consistently related to S100A4 expression will aid in the discovery of sequences whose expression is directly related to the gain or suppression of metastatic capability in these mice.

7.6 Characterised subtracted cDNA clones.

Sequence analysis and expression screening of subtracted cDNA clones has identified many interesting sequences, some of these have only been characterised by DNA sequencing because Reverse Northern hybridisation failed to detect their expression. Other clones have been shown to exhibit differential expression in the three tumours utilised to create the subtracted libraries and in other tumours from the same strains of transgenic mice.

7.6.1 Rarely transcribed cDNA clones.

Many of the sequences of low abundance (not detectable by Reverse Northern analysis) were uncharacterised ESTs or novel sequences suggesting that they may be expressed at similarly low levels in other mouse tissues or in other species. Table 5.6 highlights some low abundant sequences of particular interest to the development and progression of cancer. Namely maspin that was originally identified by subtractive hybridisation to be down regulated in mammary carcinoma versus normal breast epithelial cells (Sager *et al*., 1997). A member of the serpin family of protease inhibitors, maspin expression limits tumour cell invasion and metastasis and has been linked with improved patient survival in breast, prostrate and oral squamous cell carcinomas (Seftor *et al*., 1998; Xia *et al*., 2000). The identification of these sequences does not necessarily indicate differential expression. Detection of expression of maspin, and other interesting sequences by more sensitive techniques such as quantitative RT-PCR may indicate differential expression in these tumours.

7.6.2 Differentially expressed cDNA clones.

Many differentially expressed sequences were identified by Reverse Northern hybridisation (Tables 5.2-5.5) including a number of known sequences with previous associations with cancer development and/or progression, and known sequences that have not previously been linked with this disease. A number of ESTs and potentially novel sequences were also identified as being differentially expressed in this system.

For example, clone Met2F6 was expressed nearly 3-fold higher in the metastasis relative to both primary tumours (Table 5.5). This clone possessed sequence similarity to mouse MRP14, also known as S100A9, a member of the same family of calcium binding proteins as S100A4. Like S100A4, this protein is normally abundantly expressed in immune cells of a normally motile behaviour. There is reported evidence that S100A9 expression in neutrophils enhances metastasis of co-injected mammary adenocarcinoma cells in a syngeneic rat model of metastasis (McGary *et al.*, 1997), and also that S100A9 expression mediates transendothelial migration of monocytes (Keukhoff *et al.*, 1999). Thus, expression of S100A9 in its normal cell type may contribute to metastasis in MMTV-*neu/S100A4* transgenic mice, or abnormal expression in epithelial tumour cells may also enhance metastatic capability by mediating the motility (possibly in association with S100A4) or intra- and extravasation properties of tumour cells.

The mouse mitochondrial genome (Bibb *et al.*, 1981; accession number gb/J01420) consists of 16kb of sequence, every part of which is utilised for the production of transfer RNA (tRNA) molecules, 12S and 16S ribosomal RNA molecules and several proteins. Between the four libraries, a total of 38 clones represented sequences derived from this genome. These included clones with sequence similarity to cytochrome oxidase subunit I (N=5, P=5, PN=0, Met=2), subunit II (N=3), subunit III (N=1, P=3, Met=1) and to ATPase 6 (N=5, P=2, PN=1, Met=2). It is possible that changes in expression of mitochondrial genes play a part in tumour progression as mitochondrial mutations have been detected in some tumours (Polyak *et al.*, 1998; Fliss *et al.*, 2000). The expression of several mitochondrial genes involved in oxidative phosphorylation have also been shown to be increased in cells transformed by viral or cellular oncogenes (Glaichenhaus *et al.*, 1986; Torroni *et al.*, 1990). Sequences corresponding to different mitochondrial genes were differentially expressed in these tumours. For example, clone N2B3 represents two subtracted clones with homology to unknown 'protein 4', expression of these were significantly higher in the MMTV-*neu* primary tumour relative to the two bitransgenic tumours (Table 5.2).

Eight clones from the Met subtracted cDNA library possessed sequence identity to PSP-c. This is reported to be a lung specific protein and so it is perhaps not surprising that it was not sequenced from the other libraries. Expression of these clones were significantly higher in the lung metastatic tumour sample relative to the

two primary tumour samples (Table 5.5). It is expected that this sequence was derived from the normal lung tissue surrounding the metastatic tumour rather than the mammary gland derived tumour cells, and represents the only known lung tissue specific sequence identified in the Met library. This indicates the success of the subtraction procedure in enriching for a gene expressed in a small proportion of the cells per frozen section.

To determine whether the expression of subtracted cDNA clones were directly associated to S100A4-related metastatic capability and to exclude the possibility that expression occurred irrespective of metastatic propensity, several Reverse Northern membranes were re-screened with additional tumours from the transgenic mice. A group of 8 clones demonstrated comparable expression patterns in these tumours (Table 5.7). These were BRP39 (3 clones), MMTV sequences (3 clones) EMAP I (clone P1B9) and a mouse kidney cDNA (clone N1A6). The expression of these, and a further 5 clones exhibiting significant sequence identity and expression associated to S100A4 expression were further analysed in multiple tumours by Northern hybridisation. With the exception of clone M1D7, all candidates exhibited detectable expression in some or all of these tumours, often this was considerably different between tumours. However this was not confined to differences between tumours of different transgenic type suggesting a molecular heterogeneity exists, contrary to that observed by array hybridisation, between tumours of matched grade, histology and presence or absence of S100A4 expression.

Many sequences, in particular clones from the N subtracted cDNA library, demonstrated sequence identity to the mouse mammary tumour virus genome. The majority matched numerous different database entries based on MMTV sequences, however to facilitate the identification of the specific coding sequence of each clone, the region of homology to three commonly matched entries (Table 4.3) was mapped out. Individual clones, or contigs of clones, possessed similarity to all regions of the MMTV genome, including the gag, pro-pol polyprotein transcript, pol, env and sag genes. This indicates that these sequences are distinct from the MMTV-LTR sequence involved in regulating the expression of the *neu* transgene, which does not contain these additional gene sequences. Thus suggesting the presence of an endogenous MMTV integrated into the mouse genome, or infection with an exogenous MMTV, which is transmitted from mother to pup in the milk during suckling. Some clones, however, demonstrated homology to the LTR sequence that

could be transgene derived. The 3' LTR of both endogenous and exogenous MMTVs contain an open reading frame (ORF) encoding the sag, superantigen, protein (Choi *et al.*, 1991; Acha-orbea and MacDonald 1995), which may explain why expression of this regulatory sequence was detected in these tumours. MMTV-like ENV gene sequences have been discovered, by PCR analysis and hybridisation studies, to be expressed in 37% of human breast tumours, the same sequences have not been detected in normal breast tissue (Wang *et al.*, 1995; Etkind *et al.*, 2000).

Reverse Northern experiments demonstrated that 18 out of 32 MMTV-derived sequences were differentially expressed. All were more abundant in the MMTV-*neu* primary tumour, and tended to display decreasing expression as S100A4 expression increased and as metastatic capability increased (Tables 5.2 and 5.4). This may be an important correlation, in that S100A4 expressing cells that appear to have a greater tendency to metastasise, may not express these MMTV sequences.

Virtual Northern hybridisation confirmed this particular pattern of expression for clone PN1B9 that was not expressed in the metastasis (Figure 6.1). Three cDNA fragments of different sizes were identified (Figure 6.1). These are expected to represent different MMTV derived sequences or different spliced variants of the same sequence. They are unlikely to be derived from the neu transgene regulatory sequences because this would have also been detected in the metastasis. The MMTV genome consists of a number of genes that are sometimes co-transcribed as a single transcript, for example, gag, pol, pro genes can be transcribed individually or as gag-pol, gag-pol-pro polyproteins, which is expected to have occurred to produce these three distinct bands. Northern analysis also demonstrated evidence of two transcript being expressed in other tumours. This also demonstrated, however that this MMTV sequence is expressed in MMTV-*neu/S100A4* tumours, including in the metastasis (Figure 6.2).

A group of commonly identified sequences have previously been associated with primary mammary tumours developed by the expression of the *neu* and/or *ras* oncogenes in transgenic mice (Morrison and Leder, 1994). These sequences tended not to be expressed in primary mammary tumours induced by *myc* and *int*-2 oncogenes, reflecting possible differences in intracellular signalling pathways occurring following the activation of the different oncogenes (Morrison and Leder, 1994). Several of these markers appear to be expressed in one or more of these three MMTV-*neu* and MMTV-*neu/S100A4* transgenic tumours, as expected due to *neu*

activated tumour initiation. BRP39, kappa casein, transferrin, WDNM1, and several mouse ESTs that show sequence similarity to human cellular retinol binding proteins (CRBP I and II) were identified (Table 4.3). Sequence redundancy of some of these clones, in particular BRP39 and CRBP, indicated potential differential expression in these tumours. This was confirmed by Reverse Northern hybridisation (Table 5.2) and for BRP39 (clone N2G3) also by Virtual Northern hybridisation (Figure 6.1). Expression was considerably higher in the MMTV-*neu* tumour relative to the two bitransgenic tumours, which expressed equally low levels of this transcript. The expression of this cDNA was also variable between other RNA samples analysed by Northern hybridisation (Figure 6.2). Expression was particularly high in the lung and in three of the primary tumours, was relatively low in the mammary gland, brain, the other four primary tumours and in the metastasis. No expression was detected in the spleen, lymph node or kidney. The most significant expression data obtained for this clone is the 45-fold difference in expression between the two MMTV-*neu* primary tumours analysed on this Northern membrane (Lanes 7 and 8, Figure 6.2). This data suggests BRP39 expression is potentially up-regulated in the some of the neu-induced tumours, as suggested by (Morrison and Leder, 1994), however expression does not appear to be related to the ability of cells to metastasise.

The expression of osteopontin has previously been shown to enhance metastatic phenotype on transfection into normally benign rat mammary cells (Oates *et al.*, 1996), this protein is also over-expressed in human tumours (Brown *et al.*, 1994; Tuck *et al.*, 1999). The expression of osteopontin in this system was enhanced as S100A4 expression increased as determined by Reverse Northern (Table 5.3) and by Virtual Northern (Figure 6.1). High expression was also detected in each of the normal mammary gland tissues examined and also in the cell lines derived from the transgenic mouse primary tumours (Figures 6.2 and 6.3). Kidney also exhibited relatively high expression consistent with that observed in the rat (Oates, 1995). The tumours examined demonstrated variable expression patterns that did not appear to correlate with transgenic status or tumour phenotype of the different mice. For example, expression was high in 3/6 MMTV-*neu* primaries and 2/7 MMTV-*neu/S100A4* primaries, and lower, but detectable in the other 3 MMTV-*neu* primaries and 4/7 MMTV-*neu/S100A4* primaries, expression was not however, detected in the metastasis.

Four cDNA clones, N1A6, N2A1, N2E4 and N3C9 possessed sequence similarity to the same mouse ESTs (Figure 4.10), that were predominantly derived from sequencing of mouse kidney cDNA libraries. Differential expression was identified by Reverse Northern (Table 5.2) and by Virtual Northern (Figure 6.1), a pattern of expression that was higher in the MMTV-*neu* tumour relative to the two bitransgenic tumours. Northern analysis confirmed the expression that was expected to be identified in the kidney tissue (Figure 6.2), the other normal tissues did not express this transcript to detectable levels. Expression was also detected in the majority of tumours, however, only after long exposure to autoradiographic film (Figure 6.2). The size of the EST transcript detected was 1.9 kb.

Two cDNA clones, N2F4 and P2D6 (Figure 4.11) possessed sequence unrelated to any in the databases that were searched, suggesting that potentially novel cDNA sequences were identified. cDNA clone P2D6 was more highly expressed in the bitransgenic tumours relative to the MMTV-*neu* primary tumour, although there is some element of disagreement between Reverse Northern (Table 5.3) and Virtual Northern (Figure 6.1) as to whether expression was higher in the metastasis or not. Northern analysis does not provide data to solve this anomaly, nor does it provide sufficiently clear data to determine the expression of this sequence in these additional RNA samples. The transcript size of this unknown sequence was 2.1 kb, as determined by Virtual Northern hybridisation. Transcripts detected by Northern include a very high molecular weight band only identified in transgenic mouse tumour derived RNA samples, i.e. this was not detected in normal tissues. A set of two bands detected in all samples on the second membrane (Figure 6.3) are the same size as 28S and 18S ribosomal RNAs, suggesting that these are either non-specific hybridisation signals, or that this clone has ribosomal sequence or ribosomal-like sequence.

The expression of clone N2F4 correlated inversely with S100A4 expression as determined by Reverse Northern (Table 5.2) and Virtual Northern (Figure 6.1). Northern analysis demonstrated that this clone hybridised to multiple RNA transcripts, some of which may be non-specific (Figure 6.3). The expression levels of these transcripts detected were quite low, determined by the exposure times required to detect expression, suggesting a potentially low abundantly expressed sequences. Three distinct bands of the same size are observed in different RNA samples. The smaller transcripts (3.2 kb and 2.5 kb) appear as a doublet and are specifically

detected in 3 out of 6 MMTV-*neu* primary tumours (lanes 8, 10 and 11, Figure 6.3), and to a lower abundance (by approximately 3-fold) in a fourth tumour (lane 9, Figure 6.3). Expression of these transcripts were detected in MMTV-*neu/S100A4* tumours and cell lines but to a lesser extent to that observed in MMTV-*neu* tumours suggesting expression is possibly influenced or down-regulated by increases in S100A4 expression.

Using cDNAs as probes on Northern blots we were able to demonstrate some degree of tumour heterogeneity. Not all tumours of the same transgene genotype expressed these cDNAs to the same extent. This phenomenon was not previously detected when screening the Atlas arrays with individual tumours and pooled tumours, where similar expression profile were obtained for MMTV-*neu* primary tumours, MMTV-*neu/S100A4* primary tumours and a metastasis. Yet this clearly is evident by Northern blot analysis and is occasionally marked by vast differences in expression between tumours of the same transgenic type and histology, for example BRP 39 in MMTV-*neu* tumours (Figure 6.2).

7.7 Further Work.

7.7.1 Identifying additional metastasis-related genes.

Hierarchical clustering of gene expression profiles of multiple tumours has demonstrated the utility of expression profiling to classify tumours (Perou *et al.*, 1999, 2000; Alizadeh *et al.*, 2000). The effective combination of SSH and microarray technology has also been previously reported (Von Stein *et al.*, 1997; Yang *et al.*, 1999). Thus cDNA microarray technology provides the opportunity to perform further expression characterisation of a larger repertoire of subtracted cDNA clones in multiple tumours from the transgenic mice. Whilst a proportion of these may be redundant clones that were identified by previous screening, such as BRP39, MMTV-sequences and PSP-c, screening arrayed subtracted clones may lead to the identification of more sequences which exhibit expression associated with metastatic progression. Repeated hybridisation experiments of multiple tumours from the transgenic mice would be possible to generate gene expression profiles of these subtracted clones. Cluster analysis of expression patterns to identify co-ordinated expression profiles of groups of genes could aid the discovery of gene sequences whose expression is consistently related to S100A4 expression and therefore also to metastatic progression in multiple tumours in these transgenic mice.

7.7.2 Further characterisation of specific candidate metastasis-related genes.

Of all candidate cDNA clones screened, the expression of clone N2F4 appears to be the most interesting, not least because it is of novel sequence and the fact that it demonstrates decreased expression associated with increased S100A4 expression. Considerable further characterisation of this clone will determine more about whether the expression of this sequence is associated with suppressing metastatic progression. Complete characterisation of the full length cDNA sequence, available through screening non-normalised, non-subtracted libraries or 5' and 3' rapid amplification of cDNA ends (RACE) of SMART PCR-generated cDNA, may aid in the identification of a possible function for this gene, by related sequence similarity to other genes. Expression screening in tumours by *in situ* hybridisation and functional studies involving the expression of the protein in normally non-metastatic cell lines (e.g. Rama 37 cells, Dunnington *et al.*, 1983) and in metastatic cell lines (e.g. Rama 800 cells, Dunnington, 1984) provide an opportunity to study the effect of elevated expression on various aspects of the metastatic process, such as in *in vitro* assays for cell motility and invasion and more importantly in *in vivo* metastasis assays. Conversely, it is possible to investigate the consequences of reducing the expression of the protein on metastatic capability by transfection of a ribozyme or antisense molecule directed to the mRNA transcript. The effectiveness of this has been reported in targeting the expression of S100A4 (Maelandesmo *et al.*, 1996; Bjornland *et al.*, 1999).

Further characterisation of known differentially expressed genes where antibodies are available for the translated protein, by immunohistochemical staining of tumour sections, will provide important information concerning the specific cell type expressing the particular candidate. In addition, this may indicate whether the expressed protein is, for example, secreted like matrix degrading enzymes (Duffy, 1996) or is concentrated to specific cells around the leading edge of an invading tumour, as observed for S100A4 in these transgenic mouse tumours (Davies *et al.*, 1996) and in human cancers (Takenaga *et al.*, 1997). Such a staining pattern may indicate a potential role in the metastatic spread of these tumours.

The significance of a change in expression of a particular gene lies in its ability to affect cellular phenotype. Thus the functional role of the related protein, and not simply the expression pattern of the RNA and protein, is the determining factor as to whether expression affects metastatic progression or cellular phenotype.

Expression studies involving the transfection of the full length cDNA or genomic DNA fragment containing the gene of interest into *in vitro* rodent model systems such as the Rama series of mammary tumour cell lines, has provided an opportunity to observe the metastasis inducing properties of S100A4 and osteopontin in breast cancer (Davies *et al*., 1993; Lloyd *et al*., 1998; Oates *et al*., 1996). Such studies are vital in ascertaining the functional significance of elevated expression of candidate sequences on metastatic progression.

The utility of the transgenic mice described here in providing a model of human breast cancer, in particular of subtypes of the disease related to elevated expression of c-*erb*B-2 and *S100A4* was discussed previously. The fully *in vivo* model has provided the opportunity to discover gene expression changes that co-ordinate with these two genes to induce the metastatic phenotype. Therefore the identification of genetic anomalies occurring in these transgenic mouse tumours may also be occurring in human breast cancers, especially those associated with the elevated expression of c-*erb*B-2 and/or *S100A4*. The identification and cloning of human homologues to identified metastasis-related candidate genes and expression screening of such genes, and their related proteins in human breast tumours is therefore important in determining the utility of such genes as markers and possible targets for the therapy of human breast cancer metastasis.

References.

Acha-Orbea H, MacDonald HR. Superantigens of mouse mammary tumour virus. Ann. Rev. Immunol. 1995; 13: 459-86.

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 1991 Jun 21;252(5013):1651-6.

Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, Lukyanov SA, Sverdlov ED, Berg DE. PCR-based subtractive hybridization and differences in gene content among strains of Helicobacter pylori. Proc Natl Acad Sci U S A. 1998 Oct 27;95(22):13108-13.

Albertazzi E, Cajone F, Leone BE, Naguib RN, Lakshmi MS, Sherbet GV. Expression of metastasis-associated genes h-mts1 (S100A4) and nm23 in carcinoma of breast is related to disease progression. DNA Cell Biol. 1998 Apr;17(4):335-42.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000 Feb 3;403(6769):503-11.

Ambartsumian NS, Grigorian MS, Larsen IF, Karlstrom O, Sidenius N, Rygaard J, Georgiev G, Lukanidin E. Metastasis of mammary carcinomas in GRS/A hybrid mice transgenic for the mts1 gene. Oncogene. 1996 Oct 17;13(8):1621-30.

Anandappa SY, Winstanley JH, Leinster S, Green B, Rudland PS, Barraclough R. Comparative expression of fibroblast growth factor mRNAs in benign and malignant breast disease. Br J Cancer. 1994 Apr;69(4):772-6.

Andrechek ER, Hardy WR, Siegel PM, Rudnicki MA, Cardiff RD, Muller WJ. Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. Proc Natl Acad Sci U S A. 2000 Mar 28;97(7):3444-9.

Axel R, Feigelson P, Schutz G. Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. Cell. 1976 Feb;7(2):247-54.

Bargmann CI, Weinberg RA. Oncogenic activation of the neu-encoded receptor protein by point mutation and deletion. EMBO J. 1988 Jul;7(7):2043-52.

Barraclough R, Dawson KJ, Rudland PS. Control of protein synthesis in cuboidal rat mammary epithelial cells in culture. Changes in gene expression accompany the formation of elongated cells. Eur J Biochem. 1982 Dec 15;129(2):335-41.

Barraclough R, Dawson KJ, Rudland PS. Elongated cells derived from rat mammary cuboidal epithelial cell lines resemble cultured mesenchymal cells in their pattern of protein synthesis. Biochem Biophys Res Commun. 1984 Apr 30;120(2):351-8.

Barraclough R, Savin J, Dube SK, Rudland PS. Molecular cloning and sequence of the gene for p9Ka. A cultured myoepithelial cell protein with strong homology to S-100, a calcium-binding protein. J Mol Biol. 1987 Nov 5;198(1):13-20.

Barraclough R. Calcium-binding protein S100A4 in health and disease. Biochim Biophys Acta. 1998 Dec 10;1448(2):190-9.

Beckmann MW, Niederacher D, Schnurch HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. J Mol Med. 1997 Jun;75(6):429-39.

Bennett DC, Peachey LA, Durbin H, Rudland PS. A possible mammary stem cell line. Cell. 1978 Sep;15(1):283-98.

Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BF, Rapp BA, Wheeler DL. GenBank. Nucleic Acids Res. 1999 Jan 1;27(1):12-7.

Beral V, Banks E, Reeves G, Appleby P. Use of HRT and the subsequent risk of cancer. J Epidemiol Biostat. 1999;4(3):191-210; discussion 210-5.

Beral V, Hermon C, Reeves G, Peto R. Sudden fall in breast cancer death rates in England and Wales. Lancet. 1995 Jun 24;345(8965):1642-3.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA. Sequence and gene organization of mouse mitochondrial DNA. Cell. 1981 Oct;26(2 Pt 2):167-80.

Bishop JO, Morton JG, Rosbash M, Richardson M. Three abundance classes in HeLa cell messenger RNA. Nature. 1974 Jul 19;250(463):199-204.

Bishop MJ. Molecular themes in oncogenesis. Cell. 1991 Jan, 64, 235-248.

Bjornland K, Winberg JO, Odegaard OT, Hovig E, Loennechen T, Aasen AO, Fodstad O, Maelandsmo GM. S100A4 involvement in metastasis: deregulation of matrix metalloproteinases and tissue inhibitors of matrix metalloproteinases in osteosarcoma cells transfected with an anti-S100A4 ribozyme. Cancer Res. 1999 Sep 15;59(18):4702-8.

Bouchard L, Lamarre L, Tremblay PJ, Jolicoeur P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-neu oncogene. Cell. 1989 Jun 16;57(6):931-6.

Boyd JM, Malstrom S, Subramanian T, Venkatesh LK, Schaeper U, Elangovan B, D'Sa-Eipper C, Chinnadurai G. Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set of cellular proteins. Cell. 1994 Oct 21;79(2):341-51.

Brown LF, Papadopoulos-Sergiou A, Berse B, Manseau EJ, Tognazzi K, Perruzzi CA, Dvorak HF, Senger DR. Osteopontin expression and distribution in human carcinomas. Am J Pathol. 1994 Sep;145(3):610-23.

Carmeliet P. Developmental biology. Controlling the cellular brakes. Nature. 1999 Oct 14;401(6754):657-8.

Carriero MV, Franco P, Del Vecchio S, Massa O, Botti G, D'Aiuto G, Stoppelli MP, Salvatore M. Tissue distribution of soluble and receptor-bound urokinase in human breast cancer using a panel of monoclonal antibodies. Cancer Res. 1994 Oct 15;54(20):5445-54.

Chambers AF, Naumov GN, Vantyghem SA, Tuck AB. Molecular biology of breast cancer metastasis. Clinical implications of experimental studies on metastatic inefficiency. Breast Cancer Res. 2000, 2: in press.

Chekmareva MA, Kadkhodaian MM, Hollowell CM, Kim H, Yoshida BA, Luu HH, Stadler WM, Rinker-Schaeffer CW. Chromosome 17-mediated dormancy of AT6.1 prostate cancer micrometastases. Cancer Res. 1998 Nov 1;58(21):4963-9.

Chen H, Ke Y, Oates AJ, Barraclough R, Rudland PS. Isolation of and effector for metastasis-inducing DNAs from a human metastatic carcinoma cell line. Oncogene. 1997 Apr 3;14(13):1581-8.

Chenchik A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD. Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. Biotechniques. 1996 Sep;21(3):526-34.

Chenchik A, Zhu YY, Diatchenko L, Li R, Hill J and Siebert PD. Gene Cloning and Analysis by RT-PCR, Eds. Siebert P. & Larrick J. (Biotechniques Books, MA), 1998; pp 213-239.

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry. 1979 Nov 27;18(24):5294-9.

Choi Y, Kappler JW, Marrack P. A superantigen encoded in the open reading frame of the 3' long terminal repeat of mouse mammary tumour virus. Nature. 1991 Mar 21;350(6315):203-7.

Christofori G, Semb H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. Trends Biochem Sci. 1999 Feb;24(2):73-6.

Chu KC, Tarone RE, Kessler LG, Ries LA, Hankey BF, Miller BA, Edwards BK. Recent trends in U.S. breast cancer incidence, survival, and mortality rates. J Natl Cancer Inst. 1996 Nov 6;88(21):1571-9.

Davies BR, Davies MP, Gibbs FE, Barraclough R, Rudland PS. Induction of the metastatic phenotype by transfection of a benign rat mammary epithelial cell line with the gene for p9Ka, a rat calcium-binding protein, but not with the oncogene EJ-ras-1. Oncogene. 1993 Apr;8(4):999-1008.

Davies M, Harris S, Rudland P, Barraclough R. Expression of the rat, S-100-related, calcium-binding protein gene, p9Ka, in transgenic mice demonstrates different patterns of expression between these two species. DNA Cell Biol. 1995 Oct;14(10):825-32.

Davies MP, Rudland PS, Robertson L, Parry EW, Jolicoeur P, Barraclough R. Expression of the calcium-binding protein S100A4 (p9Ka) in MMTV-neu transgenic mice induces metastasis of mammary tumours. Oncogene. 1996 Oct 17;13(8):1631-7.

Davies MP, Gibbs FE, Halliwell N, Joyce KA, Roebuck MM, Rossi ML, Salisbury J, Sibson DR, Tacconi L, Walker C. Mutation in the PTEN/MMAC1 gene in archival low grade and high grade gliomas. Br J Cancer. 1999 Mar;79(9-10):1542-8.

De Vouge MW, Mukherjee BB. Transformation of normal rat kidney cells by v-K-ras enhances expression of transin 2 and an S-100-related calcium-binding protein. Oncogene. 1992 Jan;7(1):109-19.

Debouck C. Differential display or differential dismay? Curr. Opinion in Biotech. 1995. 6:597-599

DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 1997 Oct 24;278(5338):680-6.

Devilee P, Cornelisse CJ. Somatic genetic changes in human breast cancer. Biochim Biophys Acta. 1994 Dec 30;1198(2-3):113-30

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A. 1996 Jun 11;93(12):6025-30.

Diatchenko L, Lukyanov S, Lau YF, Siebert PD. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol. 1999;303:349-80.

Duffy MJ. Proteases as prognostic markers in cancer. Clin Cancer Res. 1996 Apr;2(4):613-8.

Dunnington DJ, Monaghan P, Hughes CM, Rudland PS. Phenotypic instability of rat mammary tumour epithelial cells. J Natl Cancer Inst 1983; 71:1227-1240.

Dunnington DJ, Kim U, Hughes CM, Monaghan P, Rudland PS. Lack of production of myoepithelial variants by cloned epithelial cell lines derived from the TMT-081 metastasizing rat mammary tumour. Cancer Res. 1984 Nov;44(11):5338-46.

Dunnington DJ. The development and study of single cell-cloned metastasizing mammary tumour cell systems in the rat. Ph. D. thesis, University of London, 1984.

Ebralidze A, Tulchinsky E, Grigorian M, Afanasyeva A, Senin V, Revazova E, Lukanidin E. Isolation and characterization of a gene specifically expressed in different metastatic cells and whose deduced gene product has a high degree of homology to a Ca2+-binding protein family. Genes Dev. 1989 Jul;3(7):1086-93.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14863-8.

Endege WO, Steinmann KE, Boardman LA, Thibodeau SN, Schlegel R. Representative cDNA libraries and their utility in gene expression profiling. Biotechniques. 1999 Mar;26(3):542-8, 550.

Engelkamp D, Schafer BW, Erne P, Heizmann CW. S100 alpha, CAPL, and CACY: molecular cloning and expression analysis of three calcium-binding proteins from human heart. Biochemistry. 1992 Oct 27;31(42):10258-64.

Etkind P, Du J, Khan A, Pillitteri J, Wiernik PH. Mouse mammary tumor virus-like ENV gene sequences in human breast tumors and in a lymphoma of a breast cancer patient. Clin Cancer Res. 2000 Apr;6(4):1273-8.

Farina KL, Wyckoff JB, Rivera J, Lee H, Segall JE, Condeelis JS, Jones JG. Cell motility of tumor cells visualized in living intact primary tumors using green fluorescent protein. Cancer Res. 1998 Jun 15;58(12):2528-32.

Fearon ER. Genetic alterations underlying colorectal tumorigenesis. Cancer Surv. 1992;12:119-36.

Fend F, Quintanilla-Martinez L, Kumar S, Beaty MW, Blum L, Sorbara L, Jaffe ES, Raffeld M. Composite low grade B-cell lymphomas with two immunophenotypically distinct cell populations are true biclonal lymphomas. A molecular analysis using laser capture microdissection. Am J Pathol. 1999 Jun;154(6):1857-66.

Feuer EJ, Wun LM, DEVCAN: Probability of Developing or Dying of cancer. Version 4.0 Bethesda MD: National Cancer Institute, 1999.

Fidler IJ, Gersten DM, Hart IR. The biology of cancer invasion and metastasis. Adv Cancer Res. 1978;28:149-250.

Fidler IJ, Radinsky R. Genetic control of cancer metastasis. J Natl Cancer Inst. 1990 Feb 7;82(3):166-8.

Fidler IJ. Cancer metastasis. Br Med Bull. 1991 Jan;47(1):157-77.

Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, Jen J, Sidransky D. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science. 2000 Mar 17;287(5460):2017-9.

Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med. 1971 Nov 18;285(21):1182-6.

Folkman J. Tumor angiogenesis. Adv Cancer Res. 1985;43:175-203.

Ford HL, Zain SB. Interaction of metastasis associated Mts1 protein with non-muscle myosin. Oncogene. 1995 Apr 20;10(8):1597-605.

Ford HL, Silver DL, Kachar B, Sellers JR, Zain SB. Effect of Mts1 on the structure and activity of nonmuscle myosin II. Biochemistry. 1997 Dec 23;36(51):16321-7.

Garne JP, Aspegren K, Balldin G, Ranstam J. Increasing incidence of and declining mortality from breast carcinoma. Trends in Malmo, Sweden, 1961-1992. Cancer. 1997 Jan 1;79(1):69-74.

Gibbs FE, Wilkinson MC, Rudland PS, Barraclough R. Interactions in vitro of p9Ka, the rat S-100-related, metastasis-inducing, calcium-binding protein. J Biol Chem. 1994 Jul 22;269(29):18992-9.

Gibbs FE, Barraclough R, Platt-Higgins A, Rudland PS, Wilkinson MC, Parry EW. Immunocytochemical distribution of the calcium-binding protein p9Ka in normal rat tissues: variation in the cellular location in different tissues. J Histochem Cytochem. 1995 Feb;43(2):169-80.

Glaichenhaus N, Leopold P, Cuzin F. Increased levels of mitochondrial gene expression in rat fibroblast cells immortalized or transformed by viral and cellular oncogenes. EMBO J. 1986 Jun;5(6):1261-5

Goldberg SF, Harms JF, Quon K, Welch DR. Metastasis-suppressed C8161 melanoma cells arrest in lung but fail to proliferate. Clin Exp Metastasis. 1999;17(7):601-7.

Gong JL, McCarthy KM, Telford J, Tamatani T, Miyasaka M, Schneeberger EE. Intraepithelial airway dendritic cells: a distinct subset of pulmonary dendritic cells obtained by microdissection. J Exp Med. 1992 Mar 1; 175(3):797-807.

Goto K, Endo H, Fujiyoshi T. Cloning of the sequences expressed abundantly in established cell lines: identification of a cDNA clone highly homologous to S-100, a calcium binding protein. J Biochem (Tokyo). 1988 Jan;103(1):48-53.

Grigorian M, Ambartsumian N, Lykkesfeldt AE, Bastholm L, Elling F, Georgiev G, Lukanidin E. Effect of mts1 (S100A4) expression on the progression of human breast cancer cells. Int J Cancer. 1996 Sep 17;67(6):831-41.

Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov KA, Vagner LL, Ermolaeva OD, Lukyanov SA, Sverdlov ED. Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Anal Biochem. 1996 Aug 15;240(1):90-7.

Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. Proc Natl Acad Sci U S A. 1992 Nov 15;89(22):10578-82.

Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. Nature. 1999 Jul 29;400(6743):464-8.

Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan 7;100(1):57-70.

Harris JR, Lippman ME, Veronesi U, Willett W. Breast cancer (1).N Engl J Med. 1992 Jul 30;327(5):319-28.

Hart IR, Goode NT, Wilson RE. Molecular aspects of the metastatic cascade. Biochim Biophys Acta. 1989 Jul 28;989(1):65-84.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature. 1984 Mar 8-14;308(5955):149-53.

Hennessy C, Henry JA, May FE, Westley BR, Angus B, Lennard TW. Expression of the antimetastatic gene nm23 in human breast cancer: an association with good prognosis. J Natl Cancer Inst. 1991 Feb 20;83(4):281-5.

Hermon C, Beral V. Br J Cancer 1996 Apr;73(7):955-60 Breast cancer mortality rates are levelling off or beginning to decline in many western countries: analysis of time trends, age-cohort and age-period models of breast cancer mortality in 20 countries.

Hufton SE, Moerkerk PT, Brandwijk R, de Bruine AP, Arends JW, Hoogenboom HR. A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization. FEBS Lett. 1999 Dec 10;463(1-2):77-82.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH. Functional discovery via a compendium of expression profiles. Cell. 2000 Jul 7;102(1):109-26.

Humphreys D, Hochgrebe TT, Easterbrook-Smith SB, Tenniswood MP, Wilson MR. Effects of clusterin overexpression on TNFalpha- and TGFbeta-mediated death of L929 cells. Biochemistry. 1997 Dec 9;36(49):15233-43.

Ilg EC, Schafer BW, Heizmann CW. Expression pattern of S100 calcium-binding proteins in human tumors. Int J Cancer. 1996 Nov 4;68(3):325-32.

Jackson-Grusby LL, Swiergiel J, Linzer DI. A growth-related mRNA in cultured mouse cells encodes a placental calcium binding protein. Nucleic Acids Res. 1987 Aug 25;15(16):6677-90.

Jin H, Cheng X, Diatchenko L, Siebert PD, Huang CC. Differential screening of a subtracted cDNA library: a method to search for genes preferentially expressed in multiple tissues. Biotechniques. 1997 Dec;23(6):1084-6.

Jolicoeur P, Bouchard L, Guimond A, Ste-Marie M, Hanna Z, Dievart A. Use of mouse mammary tumour virus (MMTV)/*neu* transgenic mice to identify genes collaborating with the c-*erb*B-2 oncogene in mammary tumour development. Biochem Soc Symp. 1998;63:159-65.

Kelleher DJ, Gilmore R. DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase. Proc Natl Acad Sci U S A. 1997 May 13;94(10):4994-9.

Kerbel RS. Tumor angiogenesis: past, present and the near future. Carcinogenesis. 2000 Mar;21(3):505-15.

Kerkhoff C, Klempt M, Kaever V, Sorg C. The two calcium-binding proteins, S100A8 and S100A9, are involved in the metabolism of arachidonic acid in human neutrophils. J Biol Chem. 1999 Nov 12;274(46):32672-9.

Kirschmann DA, Seftor EA, Nieva DR, Mariano EA, Hendrix MJ. Differentially expressed genes associated with the metastatic phenotype in breast cancer. Breast Cancer Res Treat. 1999 May;55(2):127-36.

Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971 Apr;68(4):820-3.

Kriajevska MV, Cardenas MN, Grigorian MS, Ambartsumian NS, Georgiev GP, Lukanidin EM. Non-muscle myosin heavy chain as a possible target for protein encoded by metastasis-related mts-1 gene. J Biol Chem. 1994 Aug 5;269(31):19679-82.

Kriajevska M, Tarabykina S, Bronstein I, Maitland N, Lomonosov M, Hansen K, Georgiev G, Lukanidin E. Metastasis-associated Mts1 (S100A4) protein modulates protein kinase C phosphorylation of the heavy chain of nonmuscle myosin. J Biol Chem. 1998 Apr 17;273(16):9852-6.

Krizman DB, Chuaqui RF, Meltzer PS, Trent JM, Duray PH, Linehan WM, Liotta LA, Emmert-Buck MR. Construction of a representative cDNA library from prostatic intraepithelial neoplasia. Cancer Res. 1996 Dec 1;56(23):5380-3.

Kuang WW, Thompson DA, Hoch RV, Weigel RJ. Differential screening and suppression subtractive hybridization identified genes differentially expressed in an estrogen receptor-positive breast carcinoma cell line. Nucleic Acids Res 1998 Feb 15;26(4):1116-23

Kurt RA, Urba WJ, Schoof DD. Isolation of genes overexpressed in freshly isolated breast cancer specimens. Breast Cancer Res Treat. 2000 Jan;59(1):41-8.

Lah TT, Kos J. Cysteine proteinases in cancer progression and their clinical relevance for prognosis. Biol Chem. 1998 Feb;379(2):125-30.

Lakshmi MS, Parker C, Sherbet GV. Metastasis associated MTS1 and NM23 genes affect tubulin polymerisation in B16 melanomas: a possible mechanism of their regulation of metastatic behaviour of tumours. Anticancer Res. 1993 Mar Apr;13(2):299-303.

Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL, Riggins GJ. A public database for gene expression in human cancers. Cancer Res. 1999 Nov 1;59(21):5403-7.

Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature. 1983 Aug 18-24;304(5927):596-602.

Liang P, Averboukh L, Keyomarsi K, Sager R, Pardee AB. Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. Cancer Res. 1992 Dec 15;52(24):6966-8.

Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. 1992 Aug 14;257(5072):967-71.

Linzer DI, Nathans D. Growth-related changes in specific mRNAs of cultured mouse cells. Proc Natl Acad Sci U S A. 1983 Jul;80(14):4271-5.

Lloyd BH, Platt-Higgins A, Rudland PS, Barraclough R. Human S100A4 (p9Ka) induces the metastatic phenotype upon benign tumour cells. Oncogene. 1998 Jul 30;17(4):465-73. 18:

Lochter A, Galosy S, Muschler J, Freedman N, Werb Z, Bissell MJ. Matrix metalloproteinase stromelysin-1 triggers a cascade of molecular alterations that leads to stable epithelial-to-mesenchymal conversion and a premalignant phenotype in mammary epithelial cells. J Cell Biol. 1997 Dec 29;139(7):1861-72.

Luqmani YA, Smith J, Coombes RC. Polymerase chain reaction-aided analysis of gene expression in frozen tissue sections. Anal Biochem. 1992 Feb 1;200(2):291-5.

Luqmani YA, Lymboura M. Subtraction hybridization cloning of RNA amplified from different cell populations microdissected from cryostat tissue sections. Anal Biochem. 1994 Oct;222(1):102-9.

Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. J Natl Cancer Inst. 1995 Nov 15;87(22):1681-5.

Madigan MP, Troisi R, Potischman N, Dorgan JF, Brinton LA, Hoover RN. Serum hormone levels in relation to reproductive and lifestyle factors in postmenopausal women (United States).Cancer Causes Control. 1998 Mar;9(2):199-207.

Maelandsmo GM, Hovig E, Skrede M, Engebraaten O, Florenes VA, Myklebost O, Grigorian M, Lukanidin E, Scanlon KJ, Fodstad O. Reversal of the in vivo metastatic phenotype of human tumor cells by an anti-CAPL (mts1) ribozyme. Cancer Res. 1996 Dec 1;56(23):5490-8.

Mandinova A, Atar D, Schafer BW, Spiess M, Aebi U, Heizmann CW. Distinct subcellular localization of calcium binding S100 proteins in human smooth muscle cells and their relocation in response to rises in intracellular calcium. J Cell Sci. 1998 Jul 30;111 (Pt 14):2043-54.

Martin KJ, Pardee AB. Principles of differential display. Methods Enzymol. 1999;303:234-58.

Masiakowski P, Shooter EM. Nerve growth factor induces the genes for two proteins related to a family of calcium-binding proteins in PC12 cells. Proc Natl Acad Sci U S A. 1988 Feb;85(4):1277-81.

Matz MV, Lukyanov SA. Different strategies of differential display: areas of application. Nucleic Acids Res. 1998 Dec 15;26(24):5537-43.

McGary CT, Pan YC, Michel H, Guntrum WD, Neri A, Welch DR. Elevated expression of the neutrophil calcium-binding protein, MRP-14, in metastasis-enhancing neutrophils. Anticancer Res. 1997 Jan-Feb;17(1A):1-6.

Mester J, Wagenaar E, Sluyser M, Nusse R.Activation of int-1 and int-2 mammary oncogenes in hormone-dependent and -independent mammary tumors of GR mice. J Virol. 1987 Apr;61(4):1073-8.

Morrison BW, Leder P. neu and ras initiate murine mammary tumors that share genetic markers generally absent in c-myc and int-2-initiated tumors. Oncogene. 1994 Dec;9(12):3417-26.

Muller WJ, Sinn E, Pattengale PK, Wallace R, Leder P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. Cell. 1988 Jul 1;54(1):105-15.

Murray GI, Weaver RJ, Paterson PJ, Ewen SW, Melvin WT, Burke MD. Expression of xenobiotic metabolising enzymes in breast cancer. J Pathol 1993 Mar;169(3):347-53.

Nacht M, Ferguson AT, Zhang W, Petroziello JM, Cook BP, Gao YH, Maguire S, Riley D, Coppola G, Landes GM, Madden SL, Sukumar S. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res. 1999 Nov 1;59(21):5464-70.

Nakamura TM, Cech TR. Reversing time: origin of telomerase. Cell. 1998 Mar 6;92(5):587-90.

Nakamura Y, White R, Smits AM, Bos JL Genetic alterations during colorectal-tumor development. N Engl J Med. 1988 Sep 1;319(9):525-32.

Nakashima T, Sekiguchi T, Kuraoka A, Fukushima K, Shibata Y, Komiyama S, Nishimoto T. Molecular cloning of a human cDNA encoding a novel protein, DAD1, whose defect causes apoptotic cell death in hamster BHK21 cells. Mol Cell Biol. 1993 Oct;13(10):6367-74.

Nikitenko LL, Lloyd BH, Rudland PS, Fear S, Barraclough R. Localisation by in situ hybridisation of S100A4 (p9Ka) mRNA in primary human breast tumour specimens. Int J Cancer. 2000 Apr 15;86(2):219-28.

Oates A.J. (1995), Ph.D. Thesis: 'The identification of metastasis-related gene products in a rodent mammary tumour model.' Liverpool University.

Oates AJ, Barraclough R, Rudland PS. The identification of osteopontin as a metastasis-related gene product in a rodent mammary tumour model. Oncogene. 1996 Jul 4;13(1):97-104.

Oates AJ, Barraclough R, Rudland PS. The role of osteopontin in tumorigenesis and metastasis. Invasion Metastasis. 1997;17(1):1-15.

Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, MatsubaraK. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat Genet. 1992 Nov;2(3):173-9.

Ormerod EJ, Rudland PS. Mammary gland morphogenesis in vitro: formation of branched tubules in collagen gels by a cloned rat mammary cell line. Dev Biol. 1982 Jun;91(2):360-75.

Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. Oncogene. 1996 Jul 4;13(1):63-72.

Pedrocchi M, Schafer BW, Durussel I, Cox JA, Heizmann CW. Purification and characterization of the recombinant human calcium-binding S100 proteins CAPL and CACY. Biochemistry. 1994 May 31;33(21):6732-8.

Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A. 1999 Aug 3;96(16):9212-7.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747-52.

Peto R, Boreham J, Clarke M, Davies C, Beral V. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. Lancet. 2000 May 20;355(9217):1822.

Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. Int. J. Cancer, 83, 18-29 (1999).

Platt-Higgins AM, Renshaw CA, West CR, Winstanley JH, De Silva Rudland S, Barraclough R, Rudland PS. Comparison of the metastasis-inducing protein S100A4 (p9ka) with other prognostic markers in human breast cancer. Int J Cancer. 2000 Mar 20;89(2):198-208.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet. 1999 Sep;23(1):41-6.

Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B. Somatic mutations of the mitochondrial genome in human colorectal tumours. Nat Genet. 1998 Nov;20(3):291-3.

Potts BC, Smith J, Akke M, Macke TJ, Okazaki K, Hidaka H, Case DA, Chazin WJ. The structure of calcyclin reveals a novel homodimeric fold for S100 Ca(2+)-binding proteins. Nat Struct Biol. 1995 Sep;2(9):790-6.

Press MF, Pike MC, Chazin VR, Hung G, Udove JA, Markowicz M, Danyluk J, Godolphin W, Sliwkowski M, Akita R, Paterson MC, Slamon DJ. Her-2/neu expression in node-negative breast cancer: direct tissue quantitation by computerized image analysis and association of overexpression with increased risk of recurrent disease. Cancer Res. 1993 Oct 15;53(20):4960-70.

Rassoulzadegan M, Cowie A, Carr A, Glaichenhaus N, Kamen R, Cuzin F. The roles of individual polyoma virus early proteins in oncogenic transformation. Nature. 1982 Dec 23;300(5894):713-8.

Redondo M, Villar E, Torres-Munoz J, Tellez T, Morell M, Petito CK. Overexpression of clusterin in human breast carcinoma. Am J Pathol. 2000 Aug;157(2):393-9.

Renan MJ. How many mutations are required for tumorigenesis? Implications from human cancer data. Mol Carcinog. 1993;7(3):139-46.

Revillion F, Bonneterre J, Peyrat JP. ERBB2 oncogene in human breast cancer and its clinical significance. Eur J Cancer. 1998 May;34(6):791-808.

Ronnov-Jessen L, Petersen OW, Bissel MJ. Cellular changes involved in conversion of normal to malignant breast: Importance of stromal reaction. Physio Reviews 1996 Jan; 76(1): 69-125.

Ross SR, Solter D. Glucocorticoid regulation of mouse mammary tumor virus sequences in transgenic mice. Proc Natl Acad Sci U S A. 1985 Sep;82(17):5880-4.

Rudland PS, Platt-Higgins A, Renshaw C, West CR, Winstanley JH, Robertson L, Barraclough R. Prognostic significance of the metastasis-inducing protein S100A4 (p9Ka) in human breast cancer. Cancer Res. 2000 Mar 15;60(6):1595-603.

Ruley HE. Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. Nature. 1983 Aug 18-24;304(5927):602-6.

Sager R, Sheng S, Pemberton P, Hendrix MJ. Maspin. A tumor suppressing serpin. Adv Exp Med Biol. 1997;425:77-88.

Salinas AE, Wong MG. Glutathione S-transferases--a review. Curr Med Chem. 1999 Apr;6(4):279-309.

Sambrook J. Fritsch EF and Maniatis T. Molecular cloning, a laboratory manual, Cold Spring Harbour NY, Cold Spring Habour Laboratory Press 1989.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995 Oct 20;270(5235):467-70.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci U S A. 1996 Oct 1;93(20):10614-9.

Scheurle D, DeYoung MP, Binninger DM, Page H, Jahanzeb M, Narayanan R. Cancer gene discovery using digital differential display. Cancer Res. 2000 Aug 1;60(15):4037-43.

Seftor RE, Seftor EA, Sheng S, Pemberton PA, Sager R, Hendrix MJ. maspin suppresses the invasive phenotype of human breast carcinoma. Cancer Res. 1998 Dec 15;58(24):5681-5.

Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR Jr, Elkahloun AG. In vivo gene expression profile analysis of human breast cancer progression. Cancer Res. 1999 Nov 15;59(22):5656-61.

Shak S. Overview of the trastuzumab (Herceptin) anti-HER2 monoclonal antibody clinical program in HER2-overexpressing metastatic breast cancer. Herceptin Multinational Investigator Study Group. Semin Oncol. 1999 Aug;26(4 Suppl 12):71-7.

Shattuck-Eidens D, Oliphant A, McClure M, McBride C, Gupte J, Rubano T, Pruss D, Tavtigian SV, Teng DH, Adey N, Staebell M, Gumpper K, Lundstrom R, Hulick M, Kelly M, Holmen J, Lingenfelter B, Manley S, Fujimura F, Luce M, Ward B, Cannon-Albright L, Steele L, Offit K, Thomas A, Goldgar DE. BRCA1 sequence analysis in women at high risk for susceptibility mutations. Risk factor analysis and implications for genetic testing. JAMA. 1997 Oct 15;278(15):1242-50.

Sherbert GV, Lakshmi MS. S100A4 (mts1) calcium binding protein in cancer growth invasion and metastasis. Anticancer Res. 1998. 18: 2415-2422.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 1995 Mar 25;23(6):1087-8.

Sirivatanauksorn Y, Drury R, Crnogorac-Jurcevic T, Sirivatanauksorn V, Lemoine NR. Laser-assisted microdissection: applications in molecular pathology. J Pathol. 1999 Oct;189(2):150-4.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987 Jan 9;235(4785):177-82.

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989 May 12;244(4905):707-12.

Smid-Koopman E, Blok LJ, Chadha-Ajwani S, Helmerhorst TJ, Brinkmann AO, Huikeshoven FJ. Gene expression profiles of human endometrial cancer samples using a cDNA-expression array technique: assessment of an analysis method. Br J Cancer. 2000 Jul;83(2):246-51.

Sternlicht MD, Bissell MJ, Werb Z. The matrix metalloproteinase stromelysin-1 acts as a natural mammary tumor promoter. Oncogene. 2000 Feb 21;19(8):1102-13.

Stoesser G, Moseley MA, Sleep J, McGowran M, Garcia-Pastor M, Sterk P. The EMBL nucleotide sequence database. Nucleic Acids Res. 1998 Jan 1;26(1):8-15.

Strausberg RL, Dahl CA, Klausner RD. New opportunities for uncovering the molecular basis of cancer. Nat Genet. 1997 Apr;15 Spec No:415-6.

Strutz F, Okada H, Lo CW, Danoff T, Carone RL, Tomaszewski JE, Neilson EG. Identification and characterization of a fibroblast marker: FSP1. J Cell Biol. 1995 Jul;130(2):393-405.

Sugimoto A, Hozak RR, Nakashima T, Nishimoto T, Rothman JH. dad-1, an endogenous programmed cell death suppressor in Caenorhabditis elegans and vertebrates. EMBO J. 1995 Sep 15;14(18):4434-41.

Takenaga K, Nakamura Y, Sakiyama S, Hasegawa Y, Sato K, Endo H. Binding of pEL98 protein, an S100-related calcium-binding protein, to non-muscle tropomyosin. J Cell Biol. 1994 Mar;124(5):757-68.

Takenaga K, Nakamura Y, Sakiyama S. Cellular localisation of pEL98 protein, an S100-related calcium binding protein in fibroblasts and its tissue distribution analysed by monoclonal antibodies. Cell Struct. Funct. 1994 19: 133-141.

Takenaga K, Nakamura Y, Sakiyama S. Expression of a calcium binding protein pEL98 (mts1) during differentiation of human promyelocytic leukemia HL-60 cells. Biochem Biophys Res Commun. 1994 Jul 15;202(1):94-101.

Takenaga K, Nakanishi H, Wada K, Suzuki M, Matsuzaki O, Matsuura A, Endo H. Increased expression of S100A4, a metastasis-associated gene, in human colorectal adenocarcinomas. Clin Cancer Res. 1997 Dec;3(12 Pt 1):2309-16.

Tarin D, Price JE, Kettlewell MG, Souter RG, Vass AC, Crossley B. Mechanisms of human tumour metastasis studied in patients with peritoneovenous shunts. Cancer Res. 1984; 44: 3584-3592.

Tarin D, Vass AC, Kettlewell MG, Price JE. Absence of metastatic sequelae during long-term treatment of malignant ascites by peritoneo-venous shunting. A clinico-pathological report. Invasion Metastasis 1984, 4: 1-12.

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Goldgar DE. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nat Genet. 1996 Mar;12(3):333-7.

Tetu B, Brisson J, Lapointe H, Wang CS, Bernard P, Blanchette C. Cathepsin D expression by cancer and stromal cells in breast cancer: an immunohistochemical study of 1348 cases. Breast Cancer Res Treat. 1999 May;55(2):137-47.

Thompson DA, Weigel RJ. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. Biochem Biophys Res Commun. 1998 Oct 9;251(1):111-6.Todd R, McBride J, Tsuji T, Donoff RB, Nagai M, Chou MY, Chiang T, Wong DT. Deleted in oral cancer-1 (doc-1), a novel oral tumour suppressor gene. FASEB J. 1995 Oct;9(13):1362-70.

Torroni A, Stepien G, Hodge JA, Wallace DC. Neoplastic transformation is associated with co-ordinate induction of nuclear and cytoplasmic oxidative phosphorylation genes. J Biol Chem. 1990 Nov 25;265(33):20589-93.

Tsuji T, Duh FM, Latif F, Popescu NC, Zimonjic DB, McBride J, Matsuo K, Ohyama H, Todd R, Nagata E, Terakado N, Sasaki A, Matsumura T, Lerman MI, Wong DT. Cloning, mapping, expression, function, and mutation analyses of the human ortholog of the hamster putative tumour suppressor gene Doc-1. J Biol Chem. 1998 Mar 20;273(12):6704-9.

Tuck AB, Arsenault DM, O'Malley FP, Hota C, Ling MC, Wilson SM, Chambers AF. Osteopontin induces increased invasiveness and plasminogen activator expression of human mammary epithelial cells. Oncogene. 1999 Jul 22;18(29):4237 46.

Van der Valk MA 1981 In: mammary tumours in the mouse. (eds) J Hilgers and M Sluyser pp. 46-115.

van de Vijver MJ, Peterse JL, Mooi WJ, Wisman P, Lomans J, Dalesio O, Nusse R. N Engl J Med 1988 Nov 10;319(19):1239-45. Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science. 1995 Oct 20;270(5235):484-7.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW. Characterization of the yeast transcriptome. Cell. 1997 Jan 24;88(2):243-51.

Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM, Cook BP, Dufault MR, Ferguson AT, Gao Y, He TC, Hermeking H, Hiraldo SK, Hwang PM, Lopez MA, Luderer HF, Mathews B, Petroziello JM, Polyak K, Zawel L, Vogelstein B, Kinzler KW. Analysis of human transcriptomes. Nat Genet. 1999 Dec;23(4):387-8.

Viard I, Wehrli P, Jornot L, Bullani R, Vechietti JL, Schifferli JA, Tschopp J, French LE. C'lusterin gene expression mediates resistance to apoptotic cell death induced by heat shock and oxidative stress. J Invest Dermatol. 1999 Mar;112(3):290-6.

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. N Engl J Med 1988 Sep 1;319(9):525-32 Genetic alterations during colorectal-tumour development.

von Stein OD, Thies WG, Hofmann M. A high throughput screening for rarely transcribed differentially expressed genes. Nucleic Acids Res. 1997 Jul 1;25(13):2598-602.

Wan JS, Sharp SJ, Poirier GM, Wagaman PC, Chambers J, Pyati J, Hom YL, Galindo JE, Huvar A, Peterson PA, Jackson MR, Erlander MG. Cloning differentially expressed mRNAs. Nat Biotechnol. 1996 Dec;14(13):1685-91.

Wang Y, Holland JF, Bleiweiss IJ, Melana S, Liu X, Pelisson I, Cantarella A, Stellrecht K, Mani S, Pogo BG. Detection of mammary tumor virus env gene-like sequences in human breast cancer. Cancer Res. 1995 Nov 15;55(22):5173-9.

Wang G, Rudland PS, White MR, Barraclough R. Interaction in vivo and in vitro of the metastasis-inducing S100 protein, S100A4 (p9Ka) with S100A1. J Biol Chem. 2000 Apr 14;275(15):11141-6.

Warburton MJ, Ferns SA, Hughes CM, Rudland PS. Characterization of rat mammary cell types in primary culture: lectins and antisera to basement membrane and intermediate filament proteins as indicators of cellular heterogeneity. J Cell Sci. 1985 Nov;79:287-304.

Watanabe Y, Kobayashi R, Ishikawa T, Hidaka H. Isolation and characterization of a calcium-binding protein derived from mRNA termed p9Ka, pEL-98, 18A2, or 42A by the newly synthesized vasorelaxant W-66 affinity chromatography.Arch Biochem Biophys. 1992 Feb 1;292(2):563-9.

Watanabe Y, Usada N, Minami H, Morita T, Tsugane S, Ishikawa R, Kohama K, Tomida Y, Hidaka H. Calvasculin, as a factor affecting the microfilament assemblies in rat fibroblasts transfected by src gene. FEBS Lett. 1993 Jun 7;324(1):51-5.

Weber RG, Scheer M, Born IA, Joos S, Cobbers JM, Hofele C, Reifenberger G, Zoller JE, Lichter P. Recurrent chromosomal imbalances detected in biopsy material from oral premalignant and malignant lesions by combined tissue microdissection, universal DNA amplification, and comparative genomic hybridization. Am J Pathol. 1998 Jul;153(1):295-303.

Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis--correlation in invasive breast carcinoma. N Engl J Med. 1991 Jan 3;324(1):1-8.

Weinberg RA. Tumor suppressor genes. Science. 1991 Nov 22;254(5035):1138-46.

Weiss L. Metastatic inefficiency. Adv Cancer Res. 1990;54:159-211.

Winstanley J, Cooke T, Murray GD, Platt-Higgins A, George WD, Holt S, Myskov M, Spedding A, Barraclough BR, Rudland PS. The long term prognostic significance of c-erbB-2 in primary breast cancer. Br J Cancer. 1991 Mar;63(3):447-50.

Xia W, Lau YK, Hu MC, Li L, Johnston DA, Sheng Sj, El-Naggar A, Hung MC. High tumoral maspin expression is associated with improved survival of patients with oral squamous cell carcinoma. Oncogene. 2000 May 11;19(20):2398-403.

Yang GP, Ross DT, Kuang WW, Brown PO, Weigel RJ. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucleic Acids Res. 1999 Mar 15;27(6):1517-23.

Yoshio T, Fukami-Kobayashi K, Miyazaki S, Sugawara H and Gojobori T. Nucleic Acids Res. 1998; 26:8-15.

Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene expression profiles in normal and cancer cells. Science. 1997 May 23;276(5316):1268-72.

Zhang H, Heim J, Meyhack B. Novel BNIP1 variants and their interaction with BCL2 family members. FEBS Lett. 1999 Apr 1;448(1):23-7.

Zhumabbayeva B, Diatchenko L, Chenchik A, Siebert PD. Clontechniques July 2000:22-23.

Zuber J, Tchernitsa OI, Hinzmann B, Schmitz AC, Grips M, Hellriegel M, Sers C, Rosenthal A, Schafer R. A genome-wide survey of RAS transformation targets. Nat Genet. 2000 Feb;24(2):144-52.

Appendix.

Table A.1 Comprehensive list of all cDNAs present on the Atlas Mouse cDNA Expression Array (http://www.clontech.com/atlas/genelists/Mbroad/), stating sequence identification, Genbank accession number and grid reference position on the membrane.

Gene Name	Accession #	Coordinate
adenomatous polyposis coli protein (APC)	M88127	A1a
BRCA1; breast/ovarian cancer susceptibility locus 1 product	U31625	A1b
breast cancer type 2 susceptibility protein (BRCA2)	U65594	A1c
DCC; netrin receptor; immunoglobulin gene superfamily member	X85788	A1d
EB1 APC-binding protein	U51196	A1e
ezrin; villin 2; NF-2 (merlin) related filament/plasma membrane associated protein	X60671	A1f
Madr1; mSmad1; others against dpp protein (Mad) murine homolog; TGF-beta signaling protein-1 (bsp-1)	U58992	A1g
Mdm2; p53-regulating protein	X58876	A1h
NF2; merlin (moesin-ezrin-radixin-like protein); shwannomin; murine neurofibromatosis type 2 susceptibility protein	L27105	A1i
p107; RBL1; retinoblastoma gene product-related protein p107 (cell cycle regulator)	U27177	A1j
p130; retinoblastoma gene product-related protein Rb2/p130 (cell cycle regulator)	U36799	A1k
p53; tumor suppressor; DNA-binding protein	K01700	A1l
retinoblastoma-associated protein 1 (RB1); pp105	M26391	A1m
tumor susceptibility protein 101 (TSG101)	U52945	A1n
maspin precursor; protease inhbitor 5 (PI5)	U54705	A2a
VHL; Von Hippel-Lindau tumor suppressor protein	U12570	A2b
WT1; Wilms tumor protein; tumor suppressor	M55512	A2c
ZO-1; tight junction protein; partially homologous to a dlg-A tumor suppressor in Drosophila	D14340	A2d
A-myb proto-oncogene; myb-related protein A	X82327	A2e
B-myb proto-oncogene; myb-related protein B	X70472	A2f
thyroid hormone receptor alpha 1 (THRA); MR1A1; c-erbA alpha	X51983	A2g
c-fos proto-oncogene; p55	V00727	A2h
transcription factor AP-1; c-jun proto-oncogene; AH119	J04115	A2i
transcription termination factor 1 (TTF1)	X83974	A2j
c-myb proto-oncogene protein	M16449	A2k
c-myc proto-oncogene	X01023	A2l
c-rel proto-oncogene	X15842	A2m
Ear-2; v-erbA related proto-oncogene	X76654	A2n
Elk-1 ets-related proto-oncogene	X87257	A3a
Fli-1 ets-related proto-oncogene	X59421	A3b
fos-B	X14897	A3c
fos-related antigen 2 (FRA2); fos-L2	X83971	A3d
Gli oncogene; zinc finger transcription factor	S65038	A3e
junB	J03236	A3f
junD	J05205	A3g
L-myc proto-oncogene protein	X13945	A3h
Net; ets related transcription factor; activated by Ras	Z32815	A3i
N-myc proto-oncogene protein	X03919	A3j
A-Raf proto-oncogene	M13071	A3k
B-raf proto-oncogene	M64429	A3l
Cot proto-oncogene	D13759	A3m
casein kinase II alpha 1 related sequence 4 (CSNK2A1-RS4)	U51866	A3n
pim1 proto-oncogene	M13945	A4a
c-Fms proto-oncogene; macrophage colony stimulating factor 1 (CSF-1) receptor	X68932	A4b
mast/stem cell growth factor receptor (SCFR); c-kit proto-oncogene	Y00864	A4c
Met proto-oncogene	Y00671	A4d
platelet-derived growth factor receptor alpha precursor (PDGFR-alpha)	M84607	A4e
ret proto-oncogene precursor; c-ret	X67812	A4f
ski proto-oncogene	U14173	A4g
tyro3 precursor; rse; dtk; TK19-2	U18342	A4h
Tie-2 proto-oncogene	S67051	A4i
vascular endothelial growth factor receptor 1 (VEGFR1); fms-related tyrosine kinase 1 (Flt1)	L07297	A4j
abl proto-oncogene	L10656	A4k
c-Fes proto-oncogene	X12616	A4l
c-Fgr proto-oncogene	X52191	A4m
c-Src proto-oncogene	M17031	A4n
lymphocyte-specific tyrosine-protein kinase LCK	M12056	A5a

c-Cbl proto-oncogene (adaptor protein)	X57111	A5b
H-ras proto-oncogene; transforming protein p21	Z50013	A5c
Lfc proto-oncogene	U28495	A5d
N-ras proto-oncogene; transforming G-protein	X13664	A5e
Shc transforming adaptor protein; Src homology 2 (SH2) protein; SHB-related	U15784	A5f
macrophage colony stimulating factor 1 precursor (CSF1; MCSF; CSFM)	X05010	A5g
Int-3 proto-oncogene; NOTCH family member; NOTCH4	M80456	A5h
preproglucagon	Z46845	A5i
beta-protachykinin a	D17584	A5j
c-Mpl; thrombopoietin receptor; hematopoietic growth factor receptor superfamily member	Z22649	A5k
Mas proto-oncogene (G-protein coupled receptor)	X67735	A5l
insulin-like growth factor binding protein 2 precursor (IGF-binding protein 2)	X81580	A5m
Tiam-1 invasion inducing protein; GDP-GTP exchanger-related	U05245	A5n
cyclin A (G2/M-specific)	Z26580	A6a
G2/mitotic-specific cyclin A1 (CCNA1)	X84311	A6b
G2/mitotic-specific cyclin B1 (CCNB1; CYCB1); CCN2	X64713	A6c
G2/M-specific cyclin B2 (CCNB2; CYCB2)	X66032	A6d
cyclin C (G1-specific)	U62638	A6e
cyclin D1 (G1/S-specific)	S78355	A6f
cyclin D2 (G1/S-specific)	M83749	A6g
cyclin D3 (G1/S-specific)	U43844	A6h
cyclin E (G1/S-specific)	X75888	A6i
cyclin F (S/G2/M-specific)	Z47766	A6j
G2/M-specific cyclin G (CCNG)	Z37110	A6k
cyclin G2 (G2/M-specific)	U95826	A6l
Cdk4; cyclin-dependent kinase 4	L01640	A6m
Cdk5; cyclin-dependent kinase 5	D29678	A6n
Cdk7; MO15; cyclin-dependent kinase 7 (homolog of Xenopus MO15 cdk-activating kinase)	U11822	A7a
p58/GTA; galactosyltransferase associated protein kinase (cdc2-related protein kinase)	M58633	A7b
cyclin-dependent kinase 6 inhibitor (p18-INK6); cyclin-dependent kinase 4 inhibitor C (p18-INK4C); CDKN2C	U19596	A7c
p19ink4; cdk4 and cdk6 inhibitor	U19597	A7d
cyclin-dependent kinase inhibitor 1 (CDKN1A); melanoma differentiation-associated protein; WAF1	U09507	A7e
p27kip1; G1 cyclin-Cdk protein kinase inhibitor; p21-related	U10440	A7f
p57kip2; cdk-inhibitor kip2 (cyclin-dependent kinase inhibitor 1B); member of the p21CIP1 Cdk inhibitor family	U20553	A7g
Wee1/p87; cdc2 tyrosine 15-kinase	D30743	A7h
CDC25MM; guanine nucleotide releasing protein (GNRP; RASGRF1)	X59868	A7i
Cdc25a; cdc25M1; MPI1 (M-phase inducer phosphatase 1)	U27323	A7j
Cdc25b; cdc25M2; MPI2 (M-phase inducer phosphatase 2)	S93521	A7k
myeloblastin; trypsin-chymotrypsin related serine protease	U43525	A7l
prothymosin alpha	X56135	A7m
Tob antiproliferative factor; interacts with p185erbB2	D78382	A7n
HSP27; heat shock 27-kDa protein 1	U03560	B1a
heat shock 60-kDa protein1 (HSP60); chaperonin; GroEL homolog	X53584	B1b
84-kDa heat shock protein (HSP84); HSP 90-beta; tumor-specific transplantation 84-kDa antigen (TSTA)	M36829	B1c
HSP86; heat shock 86-kDa protein	M36830	B1d
MTJ1; DNAJ-like heat-shock protein from mouse tumor	L16953	B1e
Osp94 osmotic stress protein; APG-1; hsp70-related	D49482	B1f
MDR1; P-glycoprotein; multidrug resistance protein; efflux pump	M14757	B1g
structure-specific recognition protein 1 (SSRP1); recombination signal sequence recognition protein	S50213	B1h
MmMre11a putative endo/exonuclease	U58987	B1i
C3H cytochrome P450; Cyp1b1	X78445	B1j
ERp72 endoplasmic reticulum stress protein; protein disulfide isomerase-related protein	J05186	B1k
etoposide induced p53 responsive (EI24) mRNA	U41751	B1l
78-kDa glucose regulated protein (GRP78)	D78645	B1m
oxidative stress-induced protein mRNA	U40930	B1n
P-1-450; dioxin-inducible cytochrome P450	M10021	B2a
ATP-binding casette 8; ABC8; homolog of Drosophila white	U34920	B2b
CCHB3; calcium channel (voltage-gated; dihydropyridine-sensitive; L-type)	U20372	B2c
Golgi 4-transmembrane spanning transporter; MTP	U34259	B2d
erythrocyte/brain glucose transporter 1 (GLUT1; GT1); SLC2A1	M23384	B2e
voltage-gated sodium channel	L36179	B2f
B7-2; T-lymphocyte activation antigen CD86; CD28 antigen ligand 2;	L25606	B2g
BST-1; lymphocyte differentiation antigen CD38	D31788	B2h
C-C chemokine receptor type 1 (C-C CKR-1; CCR-1); macrophage inflammatory protein-1 alpha receptor (MIP-1alpha-R); RANTES-R	U29678	B2i
Cf2r; coagulation factor II (thrombin) receptor	L03529	B2j
Eph3 (Nuk) tyrosine-protein kinase receptor	L25890	B2k
Etk1 (Mek4; HEK) tyrosine-protein kinase receptor HEK	M68513	B2l
Frizzled-3; Drosophila tissue polarity gene frizzled homolog 3; dishevelled	U43205	B2m

receptor		
Hek2 murine homolog; Mdk5 mouse developmental kinase; Eph-related tyrosine-protein kinase receptor	Z49086	B2n
Htk; Mdk2 mouse developmental kinase; Eph -related tyrosine-protein kinase receptor	Z49085	B3a
IFNgR2; interferon-gamma receptor second (beta) chain; interferon gamma receptor accessory factor-1 (AF-1)	S69336	B3b
interleukin-6 receptor beta chain; membrane glycoprotein gp130	M83336	B3c
LCR-1; CXCR-4; CXC (SDF-1) chemokine receptor 4; HIV coreceptor (fusin)	D87747	B3d
LFA1-alpha; integrin alpha L; leukocyte adhesion glycoprotein LFA-1 alpha chain; antigen CD11A (p180)	M60778	B3e
prostaglandin E2 receptor EP4 subtype	D13458	B3f
Tie-1 tyrosine-protein kinase receptor	X80764	B3g
transferrin receptor protein (p90; CD71)	X57349	B3h
uPAR1; urokinase plasminogen activator surface receptor (CD87)	X62700	B3i
VEGFR2; KDR/flk1 vascular endothelial growth factor tyrosine kinase receptor	X70842	B3j
retinoic acid receptor beta-2 (beta2-RAR)	S56660	B3k
activating transcription factor 2 (ATF2); cAMP response element DNA-binding protein 1 (CREBP1)	S76657	B3l
I-kappa B alpha subunit (IKB alpha)	U36277	B3m
I-kB (I-kappa B) beta	U19799	B3n
NF-kB p65; NF-kappa-B transcription factor p65 subunit; rel-related polypeptide	M61909	B4a
Pml; murine homolog of the leukemia-associated PML gene	U33626	B4b
RXR-beta cis-11-retinoic acid receptor	X66224	B4c
Stat1; signal transducer and activator of transcription	U06924	B4d
signal transducer & activator of transcription 3 (STAT3); acute phase response factor (APRF)	U06922	B4e
Stat5a; mammary gland factor	Z48538	B4f
Stat6; signal transducer and activator of transcription 6; IL-4 Stat; STA6	L47650	B4g
TRAF family member-associated NF-kappaB activator (TANK)	U51907	B4h
transcription factor A10	L21027	B4i
transcription factor TF II D	D01034	B4j
tristetraproline (TTP); tissue plasminogen activator-induced sequence 11 (TPA-induced sequence 11; TIS11)	M57422	B4k
Cas; Crk-associated substrate; focal adhesion kinase substrate	U48853	B4l
Crk adaptor protein	S72408	B4m
Csk; c-Src-kinase and negative regulator	U05247	B4n
Fyn proto-oncogene; Src family member	U70324	B5a
Hck tyrosine-protein kinase	Y00487	B5b
SLAP; src-like adapter protein; Eck receptor tyrosine kinase-associated	U29056	B5c
Syk tyrosine-protein kinase (activated p21cdc42Hs kinase (ack))	U25685	B5d
non-receptor type 11 protein tyrosine phosphatase (PTPN11)	D84372	B5e
calcium/calmodulin-dependent protein kinase IV catalytic subunit (CAM kinase-GR)	X58995	B5f
cAMP-dependent protein kinase type I-beta regulatory chain	M20473	B5g
extracellular signal-regulated kinase 1 (ERK1); p44-MAPK; ERT2	M61177	B5h
58-kDa inhibitor of RNA-activated protein kinase	U28423	B5i
Jak3 tyrosine-protein kinase; Janus kinase 3	L33768	B5j
Jnk stress-activated protein kinase (SAPK)	L35236	B5k
LIM domain kinase 1 (LIMK1); KIZ-1	U15159	B5l
MAPK; MAP kinase; p38	U10871	B5m
MAPKAPK-2; MAP kinase-activated protein kinase; MAPKAP kinase 2	X76850	B5n
dual-specificity mitogen-activated protein kinase kinase 1 (MAP kinase kinase 1); erk activator kinase 1 (MEK1)	L02526	B6a
MAP kinase kinase 3 (dual specificity; MAPKK3; MKK3; MEK3)	U43187	B6b
MAPKK4; MAP kinase kinase 4; Jnk activating kinase 1	U18310	B6c
MAPKK6; MAP kinase kinase 6 (dual specificity) (MKK6)	X97052	B6d
PKC-alpha; protein kinase C alpha type	M25811	B6e
PKC-beta; protein kinase C beta-II type	X53532	B6f
PKC-delta; protein kinase C delta type	M69042	B6g
PKC-theta; protein kinase C theta type	D11091	B6h
Rsk; ribosomal protein S6 kinase	M28489	B6i
PI3-K p110; phosphatidylinositol 3-kinase catalytic subunit	U03279	B6j
phosphatidylinositol 3-kinase regulatory alpha subunit (PI3-kinase P85-alpha)	M60651	B6k
PLC beta; phospholipase C beta 3	U43144	B6l
PLC gamma; phospholipase C gamma	X95346	B6m
G13; G-alpha-13 guanine nucleotide regulatory protein	M63660	B6n
Gem induced immediate early protein; Ras family member	U10551	B7a
rab2 ras-related protein	X95403	B7b
Rac1 murine homolog	X57277	B7c
R-ras protein (closely related to ras proto-oncogenes)	M21019	B7d
transducin beta-2 subunit	U34960	B7e
Vav; GDP-GTP exchange factor; proto-oncogene	X64361	B7f
14-3-3 protein eta	U57311	B7g
cortactin; protein tyrosine kinase substrate	U03184	B7h
Dvl2; dishevelled-2 tissue polarity protein	U24160	B7i
GapIII; GTPase-activating protein	U20238	B7j

interferon regulatory factor 1 (IRF1)	M21065	B7k
PTPRG; protein-tyrosine phosphatase gamma	L09562	B7l
WBP6; pSK-SRPK1; WW domain binding protein 6 serine kinase for SR splicing factors	U92456	B7m
zyxin (ZYX)	X99063	B7n
caspase-11; ICH-3 cysteine protease; upstream regulator of ICE	U59463	C1a
Caspase-3; Nedd2 cysteine protease; positive regulator of programmed cell death ICH-1 homolog	D28492	C1b
caspase-7; Lice2; ICE-LAP3 cysteine protease	U67321	C1c
BAD protein; BCL2 binding component 6 (BBC6)	L37296	C1d
BCL2 binding athanogene 1 (BAG1)	U17162	C1e
bcl-2 homologous antagonist/killer (BAK1)	Y13231	C1f
BAX membrane isoform alpha	L22472	C1g
B-cell lymphoma protein 2 (BCL2)	M16506	C1h
B-cell lymphoma protein W (BCLW); BCL2L2	U59746	C1i
bcl-x; BCL2L1	L35049	C1j
BH3 interacting domain death agonist (BID)	U75506	C1k
glutathione peroxidase (plasma protein); selenoprotein	U13705	C1l
glutathione reductase	X76341	C1m
glutathione S-transferase A	J03958	C1n
microsomal glutathione S-transferase (MGST1; GST12)	J03752	C2a
glutathione S-transferase 5 (GST5-5); GST mu (GSTM2)	J04696	C2b
gluthathione S-transferase theta 1 (GST theta 1; GSTT1)	X98055	C2c
glutathione S-transferase Pi 1 (GSTPIB); GST YF-YF	D30687	C2d
A20 zinc finger protein; apoptosis inhibitor	U19463	C2e
adenosine A1M receptor	U05671	C2f
adenosine A2M2 receptor	U05672	C2g
adenosine A3 receptor	L20331	C2h
ALG-2; calcium binding protein required for programmed cell death	U49112	C2i
Blk; B lymphocyte kinase; Src family member	M30903	C2j
rac alpha serine/threonine kinase (RAC-PK-alpha); C-akt proto-oncogene	M94335	C2k
CD27; lymphocyte-specific NGF receptor family member	L24495	C2l
CD 30L receptor; lymphocyte activation antigen CD 30; Ki-1 antigen	U25416	C2m
CD40L; CD40 ligand	X65453	C2n
Chop10; murine homolog of Gadd153; growth arrest and DNA-damage-inducible protein	X67083	C3a
clusterin precursor (CLU); clustrin; apolipoprotein J (APOJ)	L08235	C3b
CRAF1; TNF receptor (CD40 receptor) associated factor; TRAF-related	U21050	C3c
defender against cell death 1 (DAD1)	U83628	C3d
fas-associated factor 1 (FAF1)	U39643	C3e
Fas I receptor; Fas antigen (Apo-1 antigen)	M83649	C3f
Fasl; Fas antigen ligand; generalized lymphoproliferation disease gene (gld)	U06948	C3g
FLICE-like inhibitory protein long form (FLIP-L)	U97076	C3h
fms-related tyrosine kinase 3 Flt3/Flk2 ligand	U04807	C3i
growth arrest and DNA-damage-inducible protein 45 (GADD45)	L28177	C3j
insulin-like growth factor receptor II (IGFR II); cation-independent mannose-6-P receptor	U04710	C3k
insulin-like growth factor I receptor alpha subunit (IGF-I-R alpha)	U00182	C3l
inducible nitric oxide synthase (iNOS)	M87039	C3m
interleukin-1 receptor	M20658	C3n
NADPH-cytochrome P450 reductase	D17571	C4a
activator of apoptosis harakiri (HRK); neuronal death protein 5 (DP5)	D83698	C4b
nucleoside diphosphate kinase B (NDP kinase B; NDK B); NM23-M2	X68193	C4c
nur77 early response protein; nuclear hormone receptor (HMR); N10	J04113	C4d
nuclear protein p55CDC	U05341	C4e
programmed cell death 1 protein precursor (PDCD-1; PD-1)	X67914	C4f
protein tyrosine phosphatase	D83966	C4g
presenilin 2 (PSEN2; PSNL2; PS2); ALG3; Alzheimer disease 4 homolog (AD4H)	U57324	C4h
relaxin	Z27088	C4i
receptor interacting protein (RIP; RINP)	U25995	C4j
Sik; Src-related intestinal kinase	U16805	C4k
serine proteinase inhibitor 3 (SPI3)	U25844	C4l
STAM; signal transducing adaptor molecule	U43900	C4m
stromelysin 3; matrix metalloproteinase 11 (MMP11)	Z12604	C4n
T-cell death-associated protein (TDAG51)	U44088	C5a
TNF 55; tumor necrosis factor 1 (55 kDa)	X57796	C5b
TRAIL; TNF-related apoptosis inducing ligand; Apo-2 ligand	U37522	C5c
tumor necrosis factor receptor 1 (TNFR-1)	M59378	C5d
activator-1 140-kDa subunit; replication factor C140 kDa; DNA-binding protein PO-GA	X72711	C5e
DNA-(apurinic/apyrimidinic) lyase; AP endonuclease 1 (APEX nuclease; APEX)	U12273	C5f
Atm; ataxia telangiectasia murine homolog	U43678	C5g
ATP-dependent DNA helicase II 70-kDa subunit; lupus Ku autoantigen protein p70	M38700	C5h
ATP-dependent DNA helicase II 86-kDa subunit; thyroid Ku (p70/p80) autoantigen p86 subunit; p86 Ku	X66323	C5i

DNA ligase I; polydeoxyribonucleotide synthase (ATP) (DNL1) (LIG1)	U04674	C5j
DNA ligase III; polydeoxyribonucleotide synthase (ATP) (DNL3)	U66058	C5k
DNA polymerase alpha catalytic subunit	D17384	C5l
DNA topoisomerase I (Top I)	D10061	C5m
DNA topoisomerase II (Top II)	D12513	C5n
PA6 stromal protein; RAG1 gene activator	X96618	C6a
DNA polymerase delta catalytic subunit (POLD1)	Z21848	C6b
DNAse I	U00478	C6c
DNA excision repair protein ERCC1	X07414	C6d
xeroderma pigmentosum group B complementing protein (XPB); DNA excision repair protein ERCC3	S71186	C6e
xeroderma pigmentosum group G complementing protein (XPG); DNA excision repair protein ERCC5	D16306	C6f
GTBP; G/T-mismatch binding protein; MSH6	U42190	C6g
HR21spA; protein involved in DNA double-strand break repair; PW29; calcium-binding protein	D49429	C6h
RAD23 UV excision repair protein homolog A (MHR23A)	X92410	C6i
MHR23B; Rad23 UV excision repair protein homolog	X92411	C6j
MLH1 DNA mismatch repair protein; MutL homolog	U59883	C6k
meiotic recombination protein DMC1/LIM15 homolog	D64107	C6l
MmRad51; yeast DNA repair protein Rad51 and E coli RecA homolog	D13473	C6m
MmRad52; yeast DNA repair protein Rad52 homolog	Z32767	C6n
MSH2 DNA mismatch repair protein; MutS homolog 2	U21011	C7a
PCNA; proliferating cell nuclear antigen; processivity factor; cyclin	X53068	C7b
photolyase/blue-light receptor homolog	AB000777	C7c
PMS2 DNA mismatch repair protein; yeast PMS1 homolog 2	U28724	C7d
pur-alpha transcriptional activator; sequence-specific ssDNA-binding protein	U02098	C7e
Rad50; DNA repair protein	U66887	C7f
RAG-1; V(D)J recombination activating protein	M29475	C7g
RAG-2; V(D)J recombination activating protein	M64796	C7h
ShcC adaptor; Shc-related; brain-specific	U46854	C7i
translin; recombination hotspot binding protein	X81464	C7j
ubiquitin-conjugating enzyme; yeast Rad6 homolog; murine HR6B	X96859	C7k
Ung1; uracil-DNA glycosylase	X99018	C7l
XPAC; xeroderma pigmentosum group A correcting protein	X74351	C7m
XRCC1 DNA-repair protein (affecting ligation)	U02887	C7n
ablphilin-1 (abi-1); similar to HOXD3	U17698	D1a
activating transcription factor 4 (mATF4)	M94087	D1b
delta-like protein precursor (DLK); preadipocyte factor 1 (PREF1); adipocyte differentiation inhibitor protein	L12721	D1c
AT motif-binding factor (ATBF1)	D26046	D1d
basic domain/leucine zipper transcription factor	L36435	D1e
brain factor 1 (Hfhbf1)	U36760	D1f
brain specific transcription factor NURR-1	S53744	D1g
Brn-3.2 POU transcription factor	S68377	D1h
butyrate response factor 1	M58566	D1i
CACCC Box- binding protein BKLF	U36340	D1j
CCAAT- binding transcription factor (C/ EBP)	X61800	D1k
caudal type homeobox 1 (Cdx1)	M37163	D1l
caudal type homeobox 2 (Cdx2)	S74520	D1m
Sim transcription factor	U42554	D1n
early B cell factor (EBF)	L12147	D2a
engrailed homeobox protein (EN1; MOEN1)	L12703	D2b
engrailed protein (En-2) homolog	L12705	D2c
erythroid transcription factor NF-E2	U01036	D2d
DNA-binding protein SATB1	U05252	D2e
DNA-binding protein SMBP2	L10075	D2f
transcription factor E2F dimerization partner 1 (TFDP1); DRTF1 polypeptide 1	X72310	D2g
E2F-5 transcription factor	X86925	D2h
early growth response protein 1 (EGR1); KROX-24 protein; ZIF/268	M20157	D2i
ets-related transcription factor; E74-like factor 1 (ELF1)	U19617	D2j
epidermal growth factor receptor kinase substrate EPS8	L21671	D2k
ERA-1 protein (ERA-1-993)	M22115	D2l
Erf; Ets-related transcription factor	U58533	D2m
erythroid kruppel-like transcription factor	M97200	D2n
Ets-related protein PEA 3	X63190	D3a
C-ets2	J04103	D3b
Ets-related protein Sap 1A	Z36885	D3c
GA-binding protein beta 2 subunit (GABP-beta 2 subunit; GABBP2)	M74517	D3d
GATA binding transcription factor (GATA-4)	M98339	D3e
trans-acting T-cell-specific transcription factor GATA3	X55123	D3f
Gbx 2	L39770	D3g
glial cells missing gene homolog (mGCM1)	U59876	D3h
gut-enriched Kruppel-like factor (GKLF); epithelial zinc finger protein (EZF)	U20344	D3i
heat shock transcription factor 2 (HSF 2)	X61754	D3j

forkhead-related transcription factor 1 (FREAC1); hepatocyte nuclear factor 3;	L35949	D3k
HMG-box transcription factor from testis (MusSox17)	D49474	D3l
HMG-14 non histone chromosomal protein	X53476	D3m
homeobox protein 1.1 (Hox-1.1)	M17192	D3n
homeobox protein 2.1 (Hox-2.1)	M26283	D4a
homeobox protein 2.4 (Hox-2.4)	X13721	D4b
homeobox protein 2.5 (Hox-2.5)	M34857	D4c
homeobox protein 3.1 (Hox-3.1)	X07439	D4d
homeobox protein 4.2 (Hox-4.2)	J03770	D4e
homeobox protein 7.1 (Hox-7.1)	X14759	D4f
homeobox protein 8 (Hox-8)	X59252	D4g
homeobox protein HOXD-3	X73573	D4h
ikaros DNA binding protein	L03547	D4i
Sp4 zinc finger transcription factor	U62522	D4j
interferon inducible protein 1	U19119	D4k
interferon regulatory factor 2 (IRF 2)	J03168	D4l
lung Kruppel-like factor (LKLF)	U25096	D4m
Lbx 1 transcription factor	X90829	D4n
Mph-1 nuclear transcriptional repressor for hox genes	U63386	D5a
MRE-binding transcription factor	X71327	D5b
myocyte nuclear factor (MNF)	L26507	D5c
myogenic factor 5	X56182	D5d
atonal protein homolog 2 (ATOH2; ATH2); helix-loop-helix protein MATH2	U29086	D5e
NF-1B protein (transcription factor)	D90176	D5f
nuclear factor kappaB p105 subunit (NF-kappaB p105; NFKB1)	M57999	D5g
nuclear factor related to P45 NF-E2	U20532	D5h
nuclear hormone receptor ROR-alpha-1	U53228	D5i
nucleobindin	M96823	D5j
octamer binding transcription factor (Oct 3)	M34381	D5k
PAX-8 (paired box protein PAX 8)	X57487	D5l
split hand/foot gene	U41626	D5m
SRY-box containing gene 3 (Sox3)	X94125	D5n
paired box protein PAX5; B-cell-specific transcription factor; BSAP	M97013	D6a
PAX-6 (paired box protein)	X63963	D6b
POU domain (class 2) associated factor 1	U43788	D6c
PSD-95/SAP90A	D50621	D6d
cellular retinoic acid binding protein II (CRABP-II)	M35523	D6e
retinoic acid receptor gamma (RXR gamma; RXRG)	M84819	D6f
retinoid X receptor interacting protein (RIP 15)	U09419	D6g
T-lymphocyte activated protein	M31042	D6h
transcription factor 1 for heat shock gene	X61753	D6i
transcription factor BARX1; homeodian transcription factor	Y07960	D6j
transcription factor C 1	U53925	D6k
transcription factor CTCF (11 zinc fingers)	U51037	D6l
transcription factor LIM-1	Z27410	D6m
cAMP-dependent transcription factor 3 (ATF3); transcription factor LRG - 21	U19118	D6n
transcription factor NFAT 1 isoform alpha	U02079	D7a
SRY-box containing gene 4	X70298	D7b
transcription factor RelB	M83380	D7c
transcription factor S -II; transcription elongation factor	D00926	D7d
transcription factor SEF2	X91753	D7e
transcription factor SP1P; POUdomain transcription factor	X56959	D7f
U2 small nuclear ribonucleoprotein auxiliary factor 35-kDa subunit-related protein 1 (U2AF1-RS1); SP2	D17407	D7g
transcription factor UBF	X60831	D7h
transcriptional enhancer factor 1 (TEF-1)	S74227	D7i
YB1 DNA binding protein	X57621	D7j
YY1 (UCRBP) transcriptional factor	L13968	D7k
zinc finger Kruppel type Zfp 92	U47104	D7l
zinc finger transcription factor RU49	U41671	D7m
zinc finger X-chromosomal protein (ZFX)	M32309	D7n
activin type I receptor	Z31663	E1a
orphan receptor	U11688	E1b
bone morphogenetic protein receptor IA; BMP2/BMP4 receptor; serine/threonine-protein kinase receptor R5	D16250	E1c
C-C chemokine receptor; monocyte chemoattractant protein 1 receptor (MCP-1RA)	U56819	E1d
CD 4 receptor (T cell activation antigen)	X04836	E1e
CD 40L receptor (TNF receptor family)	M83312	E1f
C5A receptor	L05630	E1g
corticotropin releasing factor receptor	X72305	E1h
endothelin b receptor (Ednrb)	U32329	E1i
granulocyte colony - stimulating factor receptor precursor (GCSF receptor)	M58288	E1j
monotype chemoattractant protein 3	S71251	E1k

D-factor/LIF receptor	D26177	E1l
ERBB-2 receptor; c-neu; HER2 protein tyrosine kinase	L47239	E1m
ERBB-3 receptor	L47240	E1n
erythropoietin receptor precursor (EPOR)	J04843	E2a
fibroblast growth factor receptor 4	X59927	E2b
fibroblast growth factor receptor basic (b FGF-R)	M28998	E2c
G-protein-coupled receptor	D17292	E2d
granulocyte-macrophage colony-stimulating factor receptor	M85078	E2e
growth factor receptor	M98547	E2f
lymphotoxin receptor (TNFR family)	U29173	E2g
macrophage mannose receptor	Z11974	E2h
pre-platelet-derived growth factor receptor	X04367	E2i
snoN; ski-related oncogene	U36203	E2j
transforming growth factor beta receptor 1 (TGF-beta receptor 1; TGFR1)	D25540	E2k
interferon alpha-beta receptor	M89641	E2l
interferon-gamma receptor	M28233	E2m
interleukin-1 receptor type II	X59769	E2n
interleukin-10 receptor	L12120	E3a
somatostatin receptor 2	M81832	E3b
interleukin-2 receptor gamma chain	L20048	E3c
interleukin-3 receptor	M29855	E3d
interleukin-4 receptor (membrane-bound form)	M27959	E3e
interleukin-5 receptor alpha subunit precursor (IL-5R alpha; IL5RA)	D90205	E3f
interleukin-7 receptor alpha (IL-7 receptor alpha; IL-7R alpha)	M29697	E3g
interleukin-8 receptor	D17630	E3h
interleukin-9 receptor	M84746	E3i
androgen receptor	X53779	E3j
calcitonin receptor 1b	U18542	E3k
estrogen receptor	M38651	E3l
glucocorticoid receptor form A	X13358	E3m
growth hormone receptor	M33324	E3n
insulin receptor	J05149	E4a
insulin receptor substrate-1 (IRS-1)	L24563	E4b
prolactin receptor PRLR2	M22959	E4c
low-density lipoprotein receptor precursor (LDL receptor; LDLR)	Z19521	E4d
5-hydroxytryptamine receptor; serotonin receptor type 2 (5HT2)	S49542	E4e
5-hydroxytryptamine receptor 1B receptor (5HT1B; HTR1B); serotonin receptor	Z11597	E4f
5-hydroxytryptamine (serotonin) receptor 1c	X72230	E4g
5-hydroxytryptamine (serotonin) receptor 1e beta	Z14224	E4h
5-hydroxytryptamine (serotonin) receptor 2c	Z15119	E4i
5-hydroxytryptamine (serotonin) receptor 3	X72395	E4j
5-hydroxytryptamine (serotonin) receptor 7	Z23107	E4k
acetylcholine receptor delta submit	K02582	E4l
adrenergic receptor beta 1	L10084	E4m
cannabinoid receptor 1 (brain)	U17985	E4n
macrophage cannabinoid receptor 2 (CB2)	U21681	E5a
dopamine receptor 4	U19880	E5b
G-protein coupled receptor	U46923	E5c
gamma-aminobutyric-acid receptor alpha-1 subunit precursor (GABA-A receptor alpha-1 subunit)	M86566	E5d
sodium- & chloride-dependent GABA transporter 1 (GABT1; GAT1); SLC6A1	M92378	E5e
GABA-A transporter 3	L04663	E5f
sodium- & chloride-dependent GABA transporter 3 (GABT3; GAT3)	L04662	E5g
glutamate receptor; ionotropic AMPA 1	X57497	E5h
glutamate receptor; ionotropic NMDA2A (epsilon 1)	D10217	E5i
glutamate receptor; ionotropic NMDA2B (epsilon 2)	D10651	E5j
nicotinic acetylcholine receptor	M14537	E5k
P-selectin glycoprotein ligand 1 precursor (PSGL1; SELPLG; SELP1)	X91144	E5l
alpha 2 catenin (CTNNA2; CATNA2); alpha catenin-related protein; alphan catenin	D25281	E5m
CD18 antigen beta subunit; leukocyte adhesion LFA-1	X14951	E5n
CD2 antigen	M18934	E6a
T-cell-specific surface glycoprotein CD28 precursor	M34563	E6b
CD3 antigen delta polypeptide	M33158	E6c
CD31; platelet endothelial cell adhesion molecule 1	L06039	E6d
CD44 antigen precursor; HUTCH I; extracellular matrix receptor III ; gp90 lymphocyte homing/adhesion receptor;	M27129	E6e
receptor-type protein tyrosine phosphatase (PTPRCAP); CD45-associated protein (CD45-AP)	U03856	E6f
CD7 antigen	D10329	E6g
CD14 monocyte differentiation antigen precursor; LPS receptor (LPSR)	M34510	E6h
CD22 antigen	L16928	E6i
cell surface glycoprotein MAC-1 alpha subunit	X07640	E6j
CTLA-4 (immunoglobin superfamily member)	X05719	E6k
desmocollin 2	L33779	E6l
dystroglycan 1	U43512	E6m

glutamate receptor channel subunit gamma	X04648	E6n
integrin alpha 2 (CD49b)	X75427	E7a
integrin alpha 4	X53176	E7b
integrin alpha 5 (CD51)	U14135	E7c
integrin alpha 6	X69902	E7d
integrin alpha 7	L23423	E7e
dipeptidyl peptidase IV (DPPIV; DPP4); thymocyte-activating molecule (THAM)	X58384	E7f
integrin beta	Y00769	E7g
integrin beta 7	M95633	E7h
intercellular adhesion molecule 1 precursor (ICAM1); MALA2	X52264	E7i
lamimin receptor 1	J02870	E7j
neuronal-cadherin (N-cadherin)	M31131	E7k
neuronal cell surface protein F3	X14943	E7l
vascular cell adhesion protein 1	M84487	E7m
integrin alpha 3 precursor (ITGA3); galactoprotein B3 (GAPB3)	D13867	E7n
basic fibroblast growth factor (b- FGF)	D12482	F1a
bone morphogenetic protein 1 precursor (BMP1)	L24755	F1b
bone morphogenetic protein 2 (BMP-2) (TGF-beta family)	L25602	F1c
bone morphogenetic protein 4 precursor (BMP4); BMP2B; DVR4	X56848	F1d
bone morphogenetic protein 7 (BMP-7); osteogenic protein 1	X56906	F1e
bone morphogenetic protein 8a (BMP-8a) (TGF-beta family)	M97017	F1f
Cek 5 receptor protein tyrosine kinase ligand	U12983	F1g
Cek 7 receptor protein tyrosine kinase ligand	U14752	F1h
endothelial ligand for L-selectin (GLYCAM 1)	M93428	F1i
epidermal growth factor (EGF)	J00380	F1j
fibroblast growth factor 9	D38258	F1k
follistatin precursor (FST); activin-binding protein	Z29532	F1l
gamma interferon-induced monokine precursor (MIG); M119	M34815	F1m
glial cell line-derived neurotrophic factor	D49921	F1n
granulocyte colony- stimulating factor (G-CSF)	M13926	F2a
growth/ diffferentiation factor 1 (GDF-1) (TGF- beta family)	M62301	F2b
growth/ diffferentiation factor 2 (GDF-2)	X77113	F2c
heparin-binding EGF-like growth factor (diphtheria toxin receptor)	L07264	F2d
hepatocyte growth factor ; hepapoitein	X72307	F2e
hepatoma transmembrane kinase ligand	L38847	F2f
inhibin alpha subunit precursor (INHA)	X69618	F2g
inhibin beta A subunit precursor (INHBA); activin beta A subunit	X69619	F2h
insulin-like growth factor binding protein -6 (IGFBP 6)	X81584	F2i
insulin-like growth factor binding protein-1 (IGFBP-1)	X81579	F2j
insulin-like growth factor binding protein 3 (IGF-binding protein 3)	X81581	F2k
insulin-like growth factor binding protein 4 precursor (IGF-binding protein 4)	X81582	F2l
insulin-like growth factor binding protein 5 precursor (IGF-binding protein 5)	X81583	F2m
insulin-like growth factor-2 (somatomedin A)	M14951	F2n
insulin-like growth factor-IA	X04480	F3a
keratinocyte growth factor FGF-7	Z22703	F3b
K-fibroblast growth factor	M30642	F3c
leukemia inhibitory factor (LIF); cholinergic differentiation factor	X06381	F3d
macrophage inflamatory protein (MIP)	X12531	F3e
macrophage inflamatory protein 1 beta (Act 2)	M35590	F3f
macrophage inflamatory protein 2 alpha (MIP2-alpha)	X53798	F3g
Mad related protein 2 (MADR2)	U60530	F3h
mast cell factor	U44725	F3i
MAX dimerization protein (MAD)	X83106	F3j
nerve growth factor alpha (alpha-NGF)	M11434	F3k
nerve growth factor beta precursor (beta-NGF; NGFB)	K01759	F3l
neuroleukin	M14220	F3m
oncostatin M (OSM)	D31942	F3n
placental ribonuclease inhibitor; angiogenin	U22516	F4a
platelet- derived growth factor (A chain) (PDGF- A)	M29464	F4b
prepro-endothelin-3	U32330	F4c
thrombomodulin	X14432	F4d
thrombopoietin precursor (THPO); megakaryocyte colony stimulating factor (MCSF)	L34169	F4e
transforming growth factor beta 1 (TGF-beta 1; TGFB1)	M13177	F4f
transforming growth factor beta 2 precursor (TGF-beta 2; TGFB2)	X57413	F4g
tumor necrosis factor beta (TNF-beta); lymphotoxin-alpha	M16819	F4h
uromodulin	L33406	F4i
vascular endothelial growth factor precursor (VEGF); vascular permeability factor (VPF)	M95200	F4j
interleukin 1 beta	M15131	F4k
interleukin 10	M37897	F4l
interleukin 11 (IL-11)	U03421	F4m
interleukin 12 (p40) beta chain	M86671	F4n
interleukin 15	U14332	F5a

interleukin 4	M25892	F5b
interleukin 6 (B cell differentiation factor)	X51975	F5c
interleukin 7	X07962	F5d
cardiac myosin heavy subunit alpha isoform (MYH6; MYHCA)	M76601	F5e
CamK II; Ca2+/calmodulin-dependent protein kinase II (beta subunit)	X63615	F5f
CDC42 GTP-binding protein; G25K	U37720	F5g
cytoskeletal epidermal keratin (14 human)	M13806	F5h
type I cytoskeletal keratin 18 (KRT1-18; KRT18); cytokeratin 18; cytokeratin endo B	M11686	F5i
cytoskeletal epidermal keratin (19 human)	M28698	F5j
epidermal keratin (1 human)	M10937	F5k
myosin light subunit 1 atrial/fetal isoform (MLC1A; MLC1EMB)	M19436	F5l
kinesin family protein KIF1A	D29951	F5m
neuronal kinesin heavy chain (NKHC); KIF5C	X61435	F5n
kinesin like protein KIF 3B	D26077	F6a
non-muscle myosin light chain 3 (MLC3NM; MYLN): MYL6	U04443	F6b
Rab-3b ras-related protein	Y14019	F6c
vimentin; intermediate filament protein	X51438	F6d
unconventional myosin VI	U49739	F6e
angiotensin-converting enzyme (ACE) (clone ACE.5.)	J04946	F6f
cathepsin B1 (CTSB)	M14222	F6g
cathepsin D (CTSD)	X53337	F6h
cathepsin H	U06119	F6i
cathepsin L	X06086	F6j
72-kDa type IV collagenase type; 72-kDa gelatinase; gelatinase A; matrix metalloproteinase 2 (MMP2)	M84324	F6k
cytotoxic cell protease 2 (B10)	X12822	F6l
granzyme B(G,H) precursor (GZMB); cytotoxic cell protease 1 (CCP1); CTLA1; fragmentin 2	M12302	F6m
gelatinase B	X72795	F6n
interleukin-converting enzyme (ICE)	L28095	F7a
mast cell protease (MMCP) - 4	M55617	F7b
membrane-type matrix matalloproteinase	X83536	F7c
protease nexin 1 (PN-1)	X70296	F7d
tissue plasminogen activator precursor (T-plasminogen activator; PLAT; TPA)	J03520	F7e
urokinase type plasminogen activator	X02389	F7f
alpha-1 protease inhibitor 2	M75716	F7g
plasminogen activator inhibitor	M33960	F7h
plasminogen activator inhibitor-2	X16490	F7i
serine protease inhibitor 2-2 (SPI2-2); SPI2 proteinase inhibitor; SPI2/EB4	M64086	F7j
serine protease inhibitor 2.4 (SPI2.4)	X69832	F7k
serine protease inhibitor homolog J6	J05609	F7l
TIMP-2 tissue inhibitor of metalloproteinases-2	X62622	F7m
tissue inhibitor of metalloproteinases 3 (TIMP3); SUN	L19622	F7n
ubiquitin	X51703	G5
phospholipase A2	D78647	G6
hypoxantine-guanine phosphoribosyltransferase (HPRT)	J00423	G7
glyceraldehyde-3-phosphate dehydrogenase (G3PDH; GADPH)	M32599	G12
myosin I	L00923	G13
murine ornithine decarboxylase	M10624	G14
beta-actin	M12481	G19
Ca2+ binding protein (CAB45)	U45977	G20
40S ribosomal protein S29 (RPS29)	L31609	G21

Table A.2 Sequence identity of N and P subtracted cDNA clones.
To investigate the identity of subtracted clones, cDNA inserts were sequenced and the DNA sequence was submitted to nucleic acid sequence databases to compare with published sequences for known genes and expressed sequenced tags (ESTs). Sequence identity and database accession numbers of the closest match(es) are given. X denotes colony PCR either failed or amplified multiple bands thus no cDNA fragment was sequenced with this clone name; ? denotes cDNA clone with no sequence similarity to any database entry; blank sequence identity indicates poor sequence was obtained for this clone.

Clone	Sequence Identity	Accession Number
N/R1	Mouse cDNA (mammary gland/macrophage/Rat embryo)	gb/AA717648
	Rat ribosomal associated membrane protein RAMP4	emb/AJ238236
N/R2	Mouse cDNA (mammary gland/embryo/uterus)	gb/AI152341
N/R3		
N/R4		
N/R5	MMTV gag gene	
N/R6	X	
N/R7	X	
N/R8	Mouse mitochondrial genome/ATPase subunit 6	AF093677
N/R9	Mouse cDNA (mammary gland) similar to Mouse (P30440) major allergen I polypeptide chain 2 precursor	gb/AI157611
N/R10		
N/R11		
N/R12	?	
N/1/A2	Mouse ADP-ribosylation-like factor homolog ARL6	AF031903
N/1/A3	MousecDNA similar to Human 60s ribosomal protein L22	
N/1/A4	MMTV sequence fusion with Mouse NonO (non-POU domain containing octamer-binding protein)	S6460
N/1/A5	Human CGI-101 protein	gb/AF151859
N/1/A6	Mouse cDNA (kidney, also liver/embryo similar to Mouse Tie 2, X71426)	gb/AI790486
N/1/A7	Mouse BRP39/Transferrin fusion!!	
N/1/A8	Mouse cDNA (mammary gland)	gb/AI504734
	Mouse cDNA (kidney) similar to Human putative thioredoxin like protein (AJ010841)	gb/AW108411
N/1/A9	Mouse sterol carrier protein-2 (SCP-2)	M62361
N/1/A10	Mouse mitochondrial genome	
N/1/A11	Mouse cDNA	
N/1/A12	Mouse cDNA (pooled organs/kidney)	gb/AA498450
N/1/B2	Mouse kappa casein	M10114
N/1/B3	Mouse cDNA (also embryo/lymph node/mammary gland/irradiated colon)	gb/AI841052
N/1/B5	?	
N/1/B6	MMTV sequence	
N/1/B7	?	
N/1/B8	Mouse ATP-dep metalloprotease FtsH1	AF090436
N/1/B9	Mouse cDNA (myotubes/brain/embryonic stem cell)	
	Rat Na+,K+-ATPase alpha isoform catalytic subunit	M14512
N/1/B10	Mouse kappa casein	gb/M10114
N/1/B11	Mouse L6 antigen	L15443
N/1/B12	Bos taurus/Human NAD+-dep Isocitrate DH	U07681/U07681
N/1/C2	Mouse cDNA similar to Human NADH-ubiquinone oxidoreductase subunit B9 homolog (AF070653)	
N/1/C5	Mouse cDNA	gb/AI838686
N/1/C10	Mouse cDNA (mammary gland)	
N/1/C11	Mouse cDNA (mammary gland) similar to Human cellular retinol binding protein I (P50120)	gb/AI037751
N/1/D2		
N/1/D6	Mouse cDNA (placenta/mammary gland/T cell)	gb/AA024124
N/1/D7	?	
N/1/D11	MMTV env gene	
N/1/D12	Rat outer mitochondrial genome membrane receptor rTOM20	gb/U21871
	Mouse tie=receptor tyrosine kinase	gb/S79346
N/1/E3	Mouse cDNA (small intcDNAine)	dbj/AV060648
	Mouse IgG receptor	gb/M63159
N/1/E4	Mouse cDNA (unfertilised egg)	dbj/AU024548
	Mouse MHC class I region	gb/AF111102
N/1/E5	Mouse mitochondrial genome	
N/1/E12	Mouse ribosomal protein L7 (rpL7)	M29016
N/1/G5	?	
N/1/G6	Mouse kappa casein	M10114
N/2/A1	Mouse cDNA (kidney also embryo/liver/irradiated colon)	AI790486

Clone	Sequence Identity	Accession Number
P/F1	Human cDNA (fetus)	gb/AA453209
P/F2	Human cDNA	gb/AW118171
P/F3	Mouse cDNA (mammary gland, also irradiated colon/lymph node/embryo)	gb/AA763394
P/F4	Human elongation factor 1-alpha (EF1A)	M29548
P/F5	Mouse mitochondrial genomechondrial related protein (L23mrp)	gb/U84902
P/F6	Rat collagen XII alpha I (Col12a1)	gb/U57362
P/F7	Mouse cDNA (mammary gland)	gb/AA544714
P/F8	Rat cDNA (embryo)	gb/AI008022
P/F9	X	
P/F10	X	
P/F11	Mouse cDNA (kidney/irradiated colon)	gb/AI573844
P/F12	Mouse cDNA (embryo)	gb/W83532
P/1/A1	Mouse cDNA (also skin)	gb/AI836511
P/1/A2		
P/1/A3		
P/1/A4	X	
P/1/A5		
P/1/A6	Mouse cDNA (mammary gland/melanoma/diaphragm/Rat ovary)	gb/AA759618
P/1/A7	Mouse Id2	
P/1/A8	MMTV LTR	gb/L37517
P/1/A9	Mouse cDNA (adult male epididymus)	djb/AU381752
P/1/A10	Human cDNA (KIAA0244/fetal liver/retina/glioblastoma)	emb/Z45576
P/1/A11	Mouse C1 inhibitor	gb/AF010254
P/1/A12		
P/1/B1	Rat rab28 for ras-homolog GTPase	X78606
P/1/B2		
P/1/B3	Human elongation factor 1-alpha (EF1A)	M29548
P/1/B4	Mouse glucocorticoid-attenuated response gene	U43086
P/1/B5	Mouse cDNA (also mammary gland/liver/T cell/blastocyst/skin)	gb/AW212041
P/1/B6	Mouse cDNA (mammary gland/brain/uterus/liver/embryonic region)	gb/AI642215
P/1/B7	Mouse cDNA (embryo/uterus/lymph node)	gb/AI842801
P/1/B8	?	
P/1/B9	Mouse endothelial monocyte activating polypeptide I	gb/U41341
	Mouse cDNA (mammary gland) similar to rabbit (P59543) calgizzarin	gb/AI462387
P/1/B10	X	
P/1/B11	X	
P/1/B12	Mouse Son of Sevenless 2	Z11664
P/1/C1	Mouse ecotropic viral intetration site 2 (Evi2) ORF	gb/M34897
P/1/C2		
P/1/C3		
P/1/C4	Mouse cDNA (myotubes/T cell/embryo/pooled organs/Rat placenta)	gb/AA637743
P/1/C5		
P/1/C6	Mouse cDNA (adult testis/lymph node)	dbj/AU280758
	Mouse cDNA (mammary gland) similar to Human (J03746) Glutathione S- transferase, microsomal	gb/AI642156
P/1/C7		
P/1/C8	Mouse cDNA (mammary gland/myotubes/embryonic region)	gb/AI117805
P/1/C9	Mouse cDNA (mammary gland/embryo/Rat spleen)	gb/AI786489
P/1/C10	Mouse cDNA similar to *B.subtilis* YQJC like protein	
P/1/C11	Mouse cDNA (embryo/mammary gland/blastocyst/diaphragm)	dbj/AV117154
P/1/C12	Mouse cDNA (also embryo)	gb/AW121086
P/1/D1		
P/1/D2		
P/1/D3	Mouse gp49B gene	gb/U05265
P/1/D4	Human initiation factor 4B	emb/X55733
P/1/D5	Human elongation factor 1-alpha (EF1A)	M29548
P/1/D6	Mouse cDNA (embryo/mammary gland/myotubes/lung) similar to Human (x57351) interferon-inducible protein 1-8D	gb/AI931291

N/2/A2		
N/2/A3	Mouse cDNA (mammary gland)	
N/2/A4	Mouse cDNA (lymph node/proximal colon/mammary gland/myotubes)	gb/AA389974
N/2/A5	X	
N/2/A6	?	
N/2/A7	Human CGI-55 protein	gb/AF151813
N/2/A8		
N/2/A9	Mouse mitochondrial genome/Rat receptor for hyaluranan-mediated motility	U87983
N/2/A10	Mouse Id2	M69293
N/2/A11		
N/2/A12	X	
N/2/B1	X	
N/2/B2	Rat adenosine kinase	U90340
N/2/B3	Mouse mitochondrial genome	V00711
N/2/B4	Mouse ubiquitin-conjugating enz (HR6A)	gb/AF089812
N/2/B5	MMTV superantigen/MTV-1 gene	
N/2/B6	?	
N/2/B7		
N/2/B8	Mouse cDNA (neonat skin/stomach/fertilised egg/Rat spleen)	dbj/AV237160
N/2/B9	Mouse cDNA (uterus/fertilised egg/embryo)	
N/2/B10	Mouse D6MM5e & DOK protein genes (also Rat & Human ribosomal protein L34 & cat c-kit)	AF084363
N/2/B11	Mouse ribosomal protein S3	X76772
N/2/B12	Homosapiens mRNA	ref/NM014751
N/2/C1	X	
N/2/C2	X	
N/2/C3	Mouse cDNA (embryo/stromal cell line/mammary gland/heart) similar to Rat bithoraxoid-like protein(AF073839)	
N/2/C4	Mouse BRP 39	X93035
N/2/C5	Mouse c-kit	Y60864
N/2/C6	Mouse cDNA (mammary gland/skin) similar to cat major allergen I polypep chain 2 precursor (P30440)	
N/2/C7	Mouse cDNA similar to Human (J05096) Na+/K+-transporting ATPase a-1 chain	gb/W41563
	Rat Na+,K+-ATPase a isoform catalytic subunit	gb/M14512
N/2/C8	Mouse cDNA similar to CRBP I	
N/2/C9		
N/2/C10	Mouse BRP 39	X93035
N/2/C11	Mouse cDNA (mammary gland/embryo/T cell/heart/Rat lung)	gb/AI606479
N/2/C12	Mouse cDNA (embryo/mammary gland/lymph node)	
N/2/D1	X	
N/2/D2	MMTV-LTR	gb/L37517
N/2/D3	Mouse calcium channel protein 3 (Clcn3)	X78874
N/2/D4	X	
N/2/D5	Mouse cDNA (mammary gland/skin/embryo)	gb/AI607404
N/2/D6	X	
N/2/D7		
N/2/D8	Mouse adenine nucleotide translocase (Ant-2)	X70847
N/2/D9		
N/2/D10		
N/2/D11	Mouse cDNA (lymph node/embryo/irradiated colon)	
N/2/D12	X	
N/2/E1		
N/2/E2	Mouse 26s ptoteasome non-ATPase subunit	Y13071
N/2/E3	Mouse mitochondrial genome	V00711
N/2/E4	?	
N/2/E5	Human U5 snRNP-specific 200KDa protein	emb/Z70200
N/2/E6	MMTV gag gene	
N/2/E7	Mouse cDNA (embryo/mammary gland) similar to Human cellular retinol binding protein I	
N/2/E8	Mouse mitochondrial genome/ATPase subunit 6	AF093677
N/2/E9	X	
N/2/E10	Mouse cDNA (mammary gland) similar to cat major allergen I polypep. P30440	
N/2/E11	Mouse kappa casein	M10114
N/2/E12	?	
N/2/F1		
N/2/F2	Mouse cellular nucleic acid binding protein	gb/U20326
N/2/F3	?	
N/2/F4		
N/2/F5	Mouse cDNA (skin)	gb/AA789934
N/2/F6		
N/2/F7		
N/2/F8	Mouse cDNA (lung/uterus/mammary gland)	gb/AI552627

P/1/D7		
P/1/D8	X	
P/1/D9	Human cDNA (also fetal skin) similar to Alu repetitive element	gb/N28595
P/1/D10	Mouse ubiquitin-like protein	gb/D10918
P/1/D11	Mouse cotz mRNA for CCT(chaperonin containing TCP-1)	Z31557
P/1/D12	Human putative spliceosome associated protein	gb/AF081788
	Human putative spliceosome associated protein amplified in breast cancer (DAM 1)	ref/NM_005872
P/1/E1		
P/1/E2		
P/1/E3	Mouse cytochrome c oxidase subunit V1a	gb/U08440
P/1/E4		
P/1/E5	Mouse cDNA (myotubes/blastocyst/kidney/embryonic carcinoma)	gb/AA682146
	Mouse MHC class I region	gb/AF111102
P/1/E6		
P/1/E7	Mouse mitochondrial genome	
P/1/E8	Mouse carcinoembryonic antigen (Cea10a)	L38422
P/1/E9	Mouse transcription factor Sp3	gb/AF062567
	Human SPR-2 mRNA for GT box bindinf protien	emb/X68560
P/1/E10	Mouse lymphocyte common antigen (Ly5)	M92933
P/1/E11	Mouse cDNA (pooled organs)	gb/AA272209
P/1/E12	Mouse cDNA	gb/AI450991
P/1/F1	Human cDNA (ovarian cancer/colon)	gb/AA174024
	Human cDNA (ovary tumour) similar to Human (Q07157) tight junction protein	gb/AA477247
P/1/F2	Rat clathrin heavy chain	gb/J03583
	Mouse cDNA (mammary gland/uterus/embryo/heart)	gb/AI646705
P/1/F3	Human CYR61	Y11307
P/1/F4		
P/1/F5	Human beta-2 microglobulin	emb/V00567
P/1/F6	Mouse cDNA (liver/mammary gland) similar to insulin-induced response protein CL-2 (gb/AA462626	gb/AI196009
	Human insulin-induced protein 2	gb/AF125392
P/1/F7	Mouse cDNA (melanoma)	gb/AA153272
P/1/F8		
P/1/F9	Human cDNA	AF086032
P/1/F10	X	
P/1/F11	X	
P/1/F12	Mouse transcription factor NF1-B protein	dbjD90173
P/1/G1	?	
P/1/G2	Mouse cDNA (also blastocyst/kidney/skin/hippocampus)	gb/AW208784
P/1/G3	Mouse reticulocalbin exon 4	dbj/D13003
	Mouse reticulocalbin exon 2 & 3	dbj/D43954
P/1/G4		
P/1/G5	?	
P/1/G6	Mouse alpha tropomyosin slow	gb/U04541
P/1/G7	Mouse ribosomal protein L37a	X73331
P/1/G8	Mouse cDNA (embryo/blastocyst/skin/uterus/irradiated colon)	gb/AA050816
P/1/G9	synthetic Shiva 1a precursor gene interleukin 2 enhabcer/ plasmid vectors/T2 antigen=truncated large Tantigen	L78438/S79053
P/1/G10	Mouse mitochondrial genome	J01420
P/1/G11	Mouse tetraspan TM4SF (Tspan-6)	AF053454
P/1/G12	Mouse mitochondrial genome	V00711
P/1/H1	Human hH3.3B gene for histone H3.3	Z48950
P/1/H2	Mouse ubiquitin-homolgy domain protein	AF033353
P/1/H3	Mouse MHC locus class III regions	gb/AF109905
	Mouse cDNA (lymph node/embryo/mammary gland) similar to putative P64 CLCP (chloride channel) protein	gb/AI450157
P/1/H4		
P/1/H5	Mouse mitochondrial genome/ATPase subunit 6	AF093677
P/1/H6	Mouse 21KDa polypep under translational control (p21?)	X06407
P/1/H7	X	
P/1/H8	Human elongation factor 1-alpha (EF1A)	M29548
P/1/H9		
P/1/H10	Mouse cDNA (uterus)	gb/AI155274
P/1/H11		
P/1/H12		
P/2/A1	Human ribosmal protein S24	M31520
	Human cDNA (prostate) similar to Human (M31520) ribosmal protein S24	gb/AI879766
P/2/A2	Mouse Syntaxin B-interactin protein pallid	gb/AF079530
	Human nucleolar fibrillar centre protein ASE-I	U86751
P/2/A3	Mouse WDNM 1 protein	emb/X93037

N/2/F9	Mouse cDNA (embryo)	gb/AA739089
N/2/F10	Human glycoprotein 6-alpha-L-fucosyltransferase transcript A1	Y17978
N/2/F11		
N/2/F12	Rat ubiquitin & ribosomal protein S27a	X81839
N/2/G1	Mouse cDNA (mammary gland)	AA789587
N/2/G2	Human basic transcription element binding protein BTEB1 (GC box binding protein)	D31716
N/2/G3	Mouse BRP 39	X93035
N/2/G4		
N/2/G5	Mouse cDNA (mammary gland/embryo) similar to Human CRBP I	
N/2/G6	Mouse t-cell specific serine protease (CCP1)	M12302
N/2/G7	Mouse cDNA (embryonic regoin)	gb/AA114673
	Rat lamina associated polypeptide	gb/U18314
N/2/G8	MMTV LTR	
N/2/G9	Mouse BRP 39	X93035
N/2/G10	Mouse 49Kda zinc finger protein	AB013357
N/2/G11		
N/2/G12	X	
N/2/H1	X	
N/2/H2	Mouse a-enolase	X52379
N/2/H3	Mouse intestinal tyrosine kinase	Z48757
N/2/H4	Human cDNA	AI002913
N/2/H5	Mouse cDNA	
N/2/H6	Mouse mitochondrial genome	V00711
N/2/H7		
N/2/H8	Mouse cDNA similar to CRBP I	
N/2/H9	Mouse cDNA (blastocyst/embryo/pancreas/Rat ovary/Human pregnant uterus)	dbj/C78988
N/2/H10	Mouse cDNA similar to GTP-binding protein homolgue (mSARA) L20294	
N/2/H11	X	
N/2/H12	MMTV LTR/superantigen	gb/L37517/U40459
N/3/A1	Rat lamina associated polypep. 2(LAP2)	RN18314
N/3/A2	Mouse mitochondrial genome	V00711
N/3/A3	Human splicing factor, arg/ser rich 7 (SFR S7)	L41887
N/3/A4	X	
N/3/A5	Mouse mitochondrial genome/ATPase subunit 6	AF093677
N/3/A6	Mouse mitochondrial genome/Mouse perforin	L07096/X60165
N/3/A7	Mouse PHR1 isoform 3 (Phr1	gb/AF100613
N/3/A8	Mouse kappa casein	M10114
N/3/A9	Mouse cDNA (proximal colon/embryo/blastocyst/Rat embryo)	gb/AA529921
N/3/A10	MMTV protease gene	L01464
N/3/A11	Mouse mitochondrial genome/Mouse perforin	L07096/X60165
N/3/A12	X	
N/3/B1	Mouse cDNA	
N/3/B2	MMTV proviral gene for retroviridae protein	gb/J02272/L37517
N/3/B3	?	
N/3/B4	Mouse apolipoprotein D	L39123
N/3/B5	X	
N/3/B6	Mouse cDNA (skin/mammary gland/lymph node/embryo)	gb/AA755017
N/3/B7	Mouse cDNA (embryonic region/kidney/endothelial cell)	gb/AF116849
N/3/B8	Mouse cDNA (kidney/embryo/neuron/fertilised egg/irradiated colon)	gb/AW049406
N/3/B9	MMTV env gene & right LTR	gb/M11024
N/3/B10	Mouse alpha-1,3-galactosyltransferase	M85153
N/3/B11	Mouse mitochondrial genome	
N/3/B12	Mouse phosphatidylethanolamine binding protien	U43206
N/3/C1	MMTV LTR	gb/L37517
N/3/C2	Mouse cDNA (2 cell embryo/emnryonic carcinoma)	gb/AA422633
	Human F-box protein (Fbx5)	gb/AF129535
N/3/C3	Mouse oral tumour surpressor homolog (Doc-1)	AF011644
N/3/C4	?	
N/3/C5	Mouse kappa casein	M10114
N/3/C6	MMTV LTR	gb/L37517
N/3/C7	Mouse Id2 protien	gb/M69293
N/3/C8	MMTV LTR/proviral gene for retroviridae protein	gb/L37517/J02272
N/3/C9	Mouse cDNA (kidnek/liver/embryo/pooled organs)	gb/AI875667
N/3/C10	endogenous MMTV LTR	V01179
N/3/C11	MMTV gag gene 3', pol gene 5'	gb/M30519
N/3/C12	Mouse cDNA (liver/embryo/mammary gland/kidney/lymph node)	gb/AI182360
	Rat SERP1	dbj/AB018546
P/2/A4	Mouse mitochondrial genome/Rat cytochrom oxidase subunit I ser-tRNA	J01420/
P/2/A5	Mouse cDNA (lymph node/uterus/liver/embryo)	gb/AI450449
P/2/A6	X	
P/2/A7	Human cytochrome bc-1 complex core protein II	J04973
P/2/A8	X	
P/2/A9	Mouse cDNA (also mammary gland/embryo) similar to Human (M87546) Red cell Acid phosphatase I isoform S	gb/AA276106
P/2/A10	Human MBNL protein	Y13829
P/2/A11	Human cDNA similar to C.elegans RNA binding protein U14946	emb/AL009266
P/2/A12	Mouse cDNA (embryo/lymph node)	gb/AA939440
P/2/B1	Mouse cDNA similar to CE00043 glycine induced repressor	
P/2/B2	X	
P/2/B3	Mouse heparin sulphate 6-sulfotransferase I	dbj/AB024566
P/2/B4	X	
P/2/B5	Mouse calumenin	U81829
P/2/B6	Rat RNA poly II-like protein	Z71925
P/2/B7	Human elongation factor 1-alpha (EF1A)	M29548
P/2/B8	Human cDNA (colon) similar to contains Alu repetitive element	gb/AA099500
P/2/B9	Mouse cDNA (mammary gland/myotubes/embryo/prox.colon)	gb/AI527628
	Cavia porcellus metalloprotease inhibitor TIMP-2	gb/AF127803
P/2/B10	Mouse osteopontin	J04806
P/2/B11	?	
P/2/B12	Mouse cDNA (kidney/embryo/Rat kidney)	gb/AI648172
P/2/C1	Mouse mitochondrial genome/Rat cytochrom oxidase subunit I ser-tRNA	J01420/
P/2/C2	Human chromosome X sequence	gb/AC002907
P/2/C3	Mouse peptidylglycine alpha-amidating monoxygenase (PAM)	gb/U79523
P/2/C4	Human WD repeat domain 3 (WDR3)	gb/AF083217
P/2/C5	Mouse mitochondrial genome	V00711
P/2/C6	Mouse mitochondrial genome	V00711
P/2/C7	Mouse mitochondrial genome	V00711
	Mouse cDNA (liver) similar to Human (L08441) cyt.C oxidase polypep.III	
P/2/C8		
P/2/C9	Mouse cDNA	
	Rat cytochrome P450CMF1b	M25143
P/2/C10	Mouse cDNA (dissected endoderm)	emb/AL022790
P/2/C11		
P/2/C12		
P/2/D1	Mouse mitochondrial gene for subunit I of cytochrome C oxidase	emb/X57780
P/2/D2	Mouse cDNA (macrophage/lymph node/olfactory brain/Rat lung)	gb/AA867412
P/2/D3	Mouse cDNA (macrophage)	gb/AA867412
P/2/D4	Mouse cDNA (also irradiated colon/mammary gland/embryo/heart)	gb/AA289301
P/2/D5	Rat 14-3-3 protein theta subtype	D17614
P/2/D6	?	
P/2/D7	Mouse calcium binding protein (pp52)	gb/M89956
P/2/D8	Mouse PW29 (also structural gene encoding B-1,4-galactosyl transferase-relating protein E11759)	D49429
P/2/D9	Mouse WW domain binding protein 5	U92454
P/2/D10		
P/2/D11		
P/2/D12	Rat rab28 for ras-homolog GTPase	emb/X78606
P/2/E1	Mouse transcytosis associated protein p115	
P/2/E2	Human signal recognition particle subunit 9 (SRP9)	gb/U20998
	Human cDNA (also Human testis/uterus/ovarian cancer)	gb/AI302715
P/2/E3	Mouse mitochondrial genome/Rat cytochrom oxidase subunit I ser-tRNA	J01420/
P/2/E4	Mouse cDNA similar to Human (gb/X51760) zinc finger binding protein ZFP-36	gb/AA655697
P/2/E5		
P/2/E6	Mouse Id2 protien	M69293
P/2/E7	Mouse WDNM 1 protein	emb/X93037
P/2/E8	Mouse WDNM 1 protein	emb/X93037
P/2/E9	Mouse kappa casein (same clone as P/2/E10?)	gb/M10114
P/2/E10	Mouse kappa casein	gb/M10114
P/2/E11		
P/2/E12	X	
P/2/F1	Rat alkaline phosphodiesterase	D30649
P/2/F2	Mouse cDNA (lymph node/embryonic region/uterus/Human fetal heart) similar to CEO3122	gb/AA288972
P/2/F3	Mouse cDNA (embryo)	
P/2/F4	Mouse cDNA (also embryo/mammary gland/urigenital ridge/Human)	gb/L26714
	Mouse cDNA (testis) similar to ring canal protein Q04652	
P/2/F5	Mouse Id2 protien	M69293
P/2/F6		

	Rat ribosome associated membrane protein RAMP4	emb/AJ238236
N/3/D1	Mouse cDNA (brain)	dbj/AU036006
N/3/D2	Mouse cDNA (skin)	gb/AA799241
N/3/D3	Mouse PHR1 isoform 3 (Phr1)	gb/AF100613
N/3/D4	Mouse cDNA (uterus/pooled organs/irradiated colon)	gb/AA183700
N/3/D5	Mouse cDNA (mammary gland)	gb/AI462418
N/3/D6	Mouse BRP39	X93035
N/3/D7	?	
N/3/D8	Mouse kappa casein	M10114
N/3/D9	Mouse chaperonin 10	U09659
N/3/D10	Mouse cDNA (kidney/mammary gland) similar to Human (Y00764) ubiquinol-cytochrome C reductase	gb/AW106975
N/3/D11		
N/3/D12	Mouse cDNA (blastocyst) and Mouse cDNA similar to mouse (gb/M89641) interferon a/b receptor (IFNAR)	dbj/C76356
	Mouse cDNA similar to Mouse (gb/M89641) interferon a/b receptor (IFNAR)	
N/3/E1	Mouse cDNA similar to CRBP I	
N/3/E2	Mouse cDNA (liver/testis/mammary gland/lung/melanoma)	gb/AW012440
N/3/E3	MMTV env gene	
N/3/E4	Mouse cDNA similar to endogenous MMTV (M11024)	gb/AW231061
N/3/E5	Mouse cDNA (liver/testis/mammary gland/lung/melanoma)	gb/AW012440
N/3/E6	MMTV proviral DNA (Mouse mRNA for dexamethasone induced product)	(D44443)
N/3/E7	MMTV proviral DNA for gag-protease-pol polyprotein & env protein	dbj/D16249
N/3/E8	Mouse BRP39	X93035
N/3/E9	Rat transferrin	D38380
N/3/E10	Mouse Transferrin	gb/M23014
N/3/E11	?	
N/3/E12	Mouse mitochondrial genome/ATPase subunit 6	AF093677
N/3/F1	X	
N/3/F2	Mouse T cell specific serine protease	M13226
N/3/F3	Mouse cDNA (also neonate head/lung/fertilised egg/brain/uterus/embryo)	gb/AW228945
N/3/F4	Mouse cDNA similar to CRBP I	
N/3/F5	Mouse cDNA (lung/embryo/brain)	gb/AA220058
	Human claudin-10 (CLDN10)	U89916
N/3/F6	MMTV MTV-1/MTV-6 gene/Mouse DNA for VSAG1 viral superantigen	X63024
N/3/F7	Mouse 49KDa zinc finger protein	AB013357
N/3/F8	Mouse cDNA (mammary gland/skin) similar to cat major allergen I polypep chain 2 precursor (P30440)	
N/3/F9	X	
N/3/F10	?	
N/3/F11	Mouse cDNA similar to bovine NADH-ubiquinone oxidoreductase complex (X63223)	
N/3/F12	Mouse kappa casein	M10114
N/3/G1	endogenous MMTV LTR	V01179
N/3/G2	Mouse cDNA (mammary gland)	
N/3/G3	Mouse BRP 39	X93035
N/3/G4	Mouse cDNA (placenta/embryo)	gb/AA015548
N/3/G5	Mouse mitochondrial genome/ATPase subunit 6	AF093677
N/3/G6	MMTV gag gene	
N/3/G7	?	
N/3/G8	MMTV MTV-1/MTV-6 gene/Mouse DNA for VSAG1 viral superantigen	X63024
N/3/G9	Mouse ferritin light chain	J04716
N/3/G10	MMTV MTV-1/MTV-6 gene/Mouse DNA for VSAG1 viral superantigen	X63024
N/3/G11	Mouse cDNA (diaphragm/mammary gland)	gb/AA066911
N/3/G12	X	
N/3/H1	endogenous MMTV LTR	V01179
N/3/H2	Mouse M2 type pyruvate kinase	X97047
N/3/H3	Mouse cDNA (skin)	
N/3/H4	?	
N/3/H5	?	
N/3/H6	Mouse kappa casein	M10114
N/3/H7	Mouse AP-2.2 gene	X94694
N/3/H8	Mouse cDNA (kidney/liver/mammary gland/embryo/fertilised egg)	gb/AI315022
N/3/H9	Mouse fetuin-like protein IRL685	emb/AJ242927
N/3/H10	Mouse Transposon Etn.	emb/Y17106
N/3/H11	Mouse MHC locus classIII region	AF049850
N/3/H12	?	

P/2/F7	Mouse cDNA (mammary gland/embryo)	gb/AI527530
P/2/F8	Mouse cDNA (mammary gland/embryo)	gb/AI527530
P/2/F9		
P/2/F10	Human telomeric DNA sequence	emb/Z96177
P/2/F11	X	
P/2/F12	Mouse non muscle myosin alkali light chain	U04443
P/2/G1	Mouse Lcn=lumican	S79461
P/2/G2	Mouse cDNA (also mammary gland/embryonic stem cell/skin)	gb/AA867784
	Mouse cDNA similar to Human (gb/X64229) DEK protein	gb/AA839661
	Human DEK (putative oncogene)	emb/X64229
P/2/G3	Mouse cDNA (also ovary/Rat ovary) similar to Mouse (P21073) protein A55	gb/AI386424
	Human cDNA similar to Actin binding protein MAYVEN (Q95198)	gb/AW270940
P/2/G4	Mouse Chromosome BAC	gb/AL003601
P/2/G5	oryctolagus mRNA	db/AB040931
P/2/G6	Human DNA	emb/AL078461
P/2/G7	Human cDNA (also fetus/pineal gland/parathyroid tumour/ovarian cancer)	gb/AF10431
P/2/G8	Mouse kappa casein	M10114
P/2/G9	Mouse kappa casein Mrna	ref/NM007786
P/2/G10	Mouse cDNA (adult male testis)	dbj/AU260557
P/2/G11	Mouse thrombospondin-4	X89963
P/2/G12	Mouse WDNM 1 protein	emb/X93037
P/2/H1	Rat rab28 for ras-homolog GTPase	emb/X78606
P/2/H2	Mouse COL3A1 gene for collagen alpha 1	X52046
P/2/H3	Mouse mitochondrial genome	V00711
P/2/H4		
P/2/H5	Mouse mitochondrial genome/ATPase subunit 6	
P/2/H6		
P/2/H7	Human putative translation initiation factor A121/Suil	gb/AF100737
	Mouse Sui1 homolog	gb/AF129888
P/2/H8	X	
P/2/H9	Human cDNA (also bone marrow stroma/fetus)	gb/AF075010
	Mouse cDNA (also lymph node)	gb/AI605037
P/2/H10	Human WRN gene	gb/AF181896
P/2/H11	Rat rab28 for ras-homolog GTPase	emb/X78606
P/2/H12	Mouse cDNA (mammary gland/stromal cell line/uterus/	gb/AA763406

Table A.3 Sequence identity of PN and Met subtracted cDNA clones.

To investigate the identity of subtracted clones, cDNA inserts were sequenced and the DNA sequence was submitted to nucleic acid sequence databases to compare with published sequences for known genes and expressed sequenced tags (ESTs). Sequence identity and database accession numbers of the closest match(es) are given. X denotes colony PCR either failed or amplified multiple bands thus no cDNA fragment was sequenced with this clone name; ? denotes cDNA clone with no sequence similarity to any database entry; blank sequence identity indicates poor sequence was obtained for this clone.

Clone	Sequence Identity	Accession Number
PN/1/A1	Human origin recognition complex subunit 3 (ORC3)	AF125507
PN/1/A2	Mouse cDNA (lung/irradiated colon) similar to Mouse CD9 antigen	MMAA89222/AF054839
	Human tetraspan TM4SF(TSPAN-2)	
PN/1/A3	Mouse cDNA	
PN/1/A4	X	
PN/1/A5		
PN/1/A6	Mouse cDNA (mammary gland)	
PN/1/A7		
PN/1/A8	Human cDNA/Mouse cDNA (mammary gland)	
PN/1/A9	Mouse cDNA (kidney/liver) similar to C.griseus epsilon-COP (emb/Z32554, 61/73)	emb Z32554
PN/1/A10	Mouse MHR23A	X92410
PN/1/A11	Mouse SYNCRIP (Human NSI-associated protein NSAP-1)	dbj/AB035725.1
PN/1/A12	?	
PN/1/B1		
PN/1/B2		
PN/1/B3	Mouse osteoblast specific factor (OSF-2)	D13664
PN/1/B4	Mouse ribophorin	D31717
PN/1/B5	Rat ribosomal protein S3a	X75161
	Rat & Human putative v-fos transformation effector protein(Fte-1)	M84711
PN/1/B6	Mouse chloride channel CaCC	AF052746
PN/1/B7	X	
PN/1/B8	Mouse cDNA (mammary gland)	
PN/1/B9	MMTV superantigen	K00556
PN/1/B10	Rat collagen XII alpha1	U57362
PN/1/B11	Mouse cDNA neonate skin (also embryo/mammary gland)	dbj/AU23524.1
PN/1/B12	Mouse hsp86	J04633
PN/1/C1	Human KIAA0066	D31886
PN/1/C2	Mouse cDNA embryonic stem cell (also mammary gland/blastocyst)	gb/AI047112
PN/1/C3	Mouse cDNA (skin)	
PN/1/C4	X	
PN/1/C5	Mouse MutS homolog 2 (mMSH2)	U21011
PN/1/C6	?	
PN/1/C7	?	
PN/1/C8	Mouse WDNM 1	X93037
PN/1/C9	Mouse cDNA (embryo/skin/mammary gland) similar to Human G1/S-specific cyclin D2	
PN/1/C10	Mouse Nedd5 mRNA for DIFF 6 or CDC3,10,11,12-like protein with GTPase motif	D49382
PN/1/C11	X	
PN/1/C12	Mouse cDNA (embryo ectoplacental core)	
PN/1/D1	Mouse cDNA Kidney similar to methionyl aminopeptidase homolog	gb/AI956684
PN/1/D2	Mouse cDNA	M14512
	Rat Na+,K+-ATPase alpha isoform catalytic subunit	
PN/1/D3	Mouse cDNA similar to Human SH3 domain binding glutamic acid-rich-like protein (SH3BGR, AF042081)	MMAA71105
PN/1/D4	Mouse stearoyl-CoA desaturase gene exon 6	gb/M21285
PN/1/D5	Mouse ubiquitin-protein ligase Nedd4	U96635
PN/1/D6	Mouse chloride channel CaCC	AF052746
PN/1/D7	?	
PN/1/D8	Mouse MHC region containing Q region of class1	AF111103
PN/1/D9	Rat ribosomal protein L24/ Mouse cDNA (kidney) similar to Mouse MC3nmRNA for proteasome	X78443
PN/1/D10	Mouse cDNA (kidney/stromal cell line) similar to Mouse 60s ribosomal protein L9 (U17332)	
PN/1/D11	Mouse cDNA (mammary gland) similar to oxysterol-binding protein	
PN/1/D12	Mouse cDNA (mammary gland/embryo/lymph node/melanoma)	gb/AI843802.1
PN/1/E1		
PN/1/E2	?	
PN/1/E3	Mouse syntenin	AF077527
PN/1/E4	Mouse cDNA (irradiated colon/mammary gland)	
PN/1/E5	X	
PN/1/E6	?	

Clone	Sequence Identity	Accession Number
Met/1/A1s	Mouse cDNA (mammary gland/placenta/embryo/lymph node) (2 bands small & large)	
Met/1/A1l	Mouse Pulmonary Surfactant Protein C (2 bands small & large)	M38314
Met/1/A2	Mouse mitochondrial genome (ATPase subunit 6)	
Met/1/A3	Mouse cDNA (mammary gland)	gb/AI461793
Met/1/A4		
Met/1/A5	Mouse Pulmonary Surfactant Protein C	M38314
Met/1/A6	Mouse acidic ribosomal phosphoprotein PO	X15267
Met/1/A7	Rat cDNA (mixed tissue library)	gb/AI639161.1
Met/1/A8	Mouse cDNA (lymph node)	gb/AA273952
	Mouse cDNA (kidney)	gb/AI390782
Met/1/A9	Mouse cDNA (liver/lymph node/ovarian cancer)	gb/AA267621
	Mouse cDNA (stromal cell line)	gb/AA870215
Met/1/A10	Mouse Pulmonary Surfactant Protein C	M38314
Met/1/A11	Mouse chimeric 16s ribosomal RNA/mitochondrial gene for nuclear protein	gb/AF089815.1
Met/1/A12	Human CHD2	AF006514
	Mouse cDNA similar to Mouse KYBP mRNA	
Met/1/B1	Mouse cDNA (embryo/mammary gland/kidney/irradiated colon)	gb/AW107939.1
Met/1/B2	Mouse cDNA (embryo/mammary gland/blastocyst/melanoma)	dbj/AV299783
	Human chromobox homolog 3 CBX3 (Drosophila HP1 gamma)	dbj/AB030905
Met/1/B3	Mouse alpha enolase	X52379
Met/1/B4	X	
Met/1/B5	Mouse cDNA (mammary gland/lymph node)	gb/AW909306
Met/1/B6	Human Selenprotein	AF051894
	Mouse cDNA (heart/mammary gland)	
Met/1/B7	Human SRp20	L10838
Met/1/B8	Human DRES 24	U69560
Met/1/B9	Mouse ribosomal protein S14	Y08307
Met/1/B10	Mouse rac1 gene	emb/X57277
	Mouse cDNA (mammary gland) similar to Mouse rac1 gene	gb/AI646062
Met/1/B11	Mouse MHC class1 B(2) microglobulin	M84367
Met/1/B12	Rat mRNA fragment with B2 repetitive sequence	X05700
Met/1/C1	Mouse cDNA (diaphragm/mammary gland) similar to LPE5P	gb/AA572388
	Human VPS28 protein	gb/AF182844
Met/1/C2	Mouse cDNA	
Met/1/C3	Rat Transferrin / Mouse transferrin	dbj/D38380 / gb/M23015
	Mouse cDNA (embryo/mammary gland/lung/liver) similar to Human serotransferrin precursor	gb/AW211839
Met/1/C4	Mouse lysozyme M/P	M21048/X51547
Met/1/C5	Mouse cDNA (8 cell embryo/mammary gland)	dbj/AU019844
Met/1/C6	Mouse cDNA (Melanoma)	gb/AI527560
Met/1/C7	Mouse MHC Class 1 H2-D transplantation antigen gene	L29190
Met/1/C8	Rat Zis (Zinc finger splicing protein)	AF013967
Met/1/C9	Mouse cDNA (blastocyst)	emb/X67813
Met/1/C10	Human Hepatocyte GF activator inhibitor	AB000095
Met/1/C11	Mouse Lysozyme P	X51547
Met/1/C12	?	
Met/1/D1	Mouse cDNA (stomach/fertilised egg)	gb/BE852308
Met/1/D2	Rat ribosomal protein S9	X66370
Met/1/D3	Mouse cDNA	gb/AI836596
Met/1/D4	Rat Transferrin	D38380
Met/1/D5	Mouse WDNM1	gb/BE554305
Met/1/D6	Rat cDNA	gb/AI69440
	Mouse cDNA (placenta) similar to Human M69013 guanine nucleotide binding protein G	
Met/1/D7	Mouse cDNA (mammary gland)	gb/AI182473
Met/1/D8	?	
Met/1/D9		
Met/1/D10	Mouse transcription factor PBX3a	AF020199
Met/1/D11	Mouse cyclophilin	M60456

PN/1/E7	Mouse NMRI fibroglycan (syndecan-2)	U00674
PN/1/E8	Mouse endogenous ecotopic murine leukemia provirus locus 3 (Emv-3) envelope glycoprotein	L37057
PN/1/E9	Mouse cDNA (mammary gland) similar to Human ubiquinol-cytochrome c reductase complex ubiquinone binding	
PN/1/E10	Mouse cDNA similar to Human serotransferrin precursor (gb/S95936)	
PN/1/E11	MMTV proviral DNA for gag-protease-pol polyprotein & env protein	M151122
PN/1/E12	Mouse H3.3A variant histone	emb/X91866
PN/1/F1	Rat transferrin/Mouse transferrin/Mouse cDNA (lung) similar to Human serotransferrin	X77158/M23015
PN/1/F2	Rat transferrin	D38380
PN/1/F3	Mouse alpha 4 protein	L31652
PN/1/F4	Mouse mitochondrial genome (ATPase subunit 6)	
PN/1/F5	Mouse Wdr1 protein	AF020055
PN/1/F6	Mouse integrin alpha 8	gb/AF041409
PN/1/F7	Mouse cDNA (embryo)	gb AA003680
PN/1/F8	Mouse cDNA (mammary gland)	gb/AI020629
PN/1/F9		
PN/1/F10		
PN/1/F11	MMTV complete proviral genome/Mouse cDNA similar to G61634 polypep PR77	
PN/1/F12	Mouse serum deprivation response	S67386
PN/1/G1	Mouse capping protein alpha 2 subunit	U16741
PN/1/G2	Mouse cDNA (embryo)	gb AA003680
PN/1/G3	Mouse cDNA (diaphragm/embryo)	gb/AA003012
PN/1/G4	Human cDNA (neuron)	gb/AA161185
	Mouse cDNA (pancreas) similar to Human zinc finger protein (Q06730)	gb/AW23047.1
PN/1/G5	Mouse MSEC66	D12713
PN/1/G6	Mouse cDNA (kidney)	gb/AI315937
PN/1/G7	Mouse cDNA (kidney/embryo/mammary gland/irradiated colon etc.)	gb/AI875649
PN/1/G8		
PN/1/G9	Mouse putative basolateral Na-K-2Cl cotransporter	U13174
PN/1/G10		
PN/1/G11	Mouse cDNA (embryo) similar to Mouse P25233 NECDIN	
	O. cuniculus corneal endothelial cDNA (C83576)	
PN/1/G12	Mouse paneth cell enhabced expression PCEE	U37351
PN/1/H1	Rat transferrin/Mouse transferrin/Mouse cDNA (lung) similar to Human serotransferrin	X77158/M23015
PN/1/H2	Mouse cDNA (mammary gland/unfertilised egg)	
PN/1/H3	Mouse TNFR2-TRAF signalling complex protein/inhibitor of apoptosis protein 2	L49433/U88909
PN/1/H4	Mouse xanthine dehydrogenase	X62932
PN/1/H5	Mouse UDP-GalNAc	U70538
PN/1/H6	Mouse Sec61 protein complex gamma subunit	U11027
PN/1/H7	?	
PN/1/H8	Mouse lysozyme M	M21050
PN/1/H9	Mouse transferrin/Mouse cDNA (mammary gland/lung/liver) similar to Human serotransferrin	J03299/M23014
PN/1/H10	Mouse apolipoprotein D	L39123
PN/1/H11	Mouse cDNA	AW048554.1
PN/1/H12	Mouse cDNA (mammary gland/kidney/embryo) similar to Mouse ST2L protein (D13695)	
PN/2/A1	Human cDNA KIAA0324	
PN/2/A2	Mouse ubiquitin-protein ligase Nedd4	U96635
PN/2/A3		
PN/2/A4	Mouse cDNA (mammary gland)	
PN/2/A5	MousecDNA (thymus)	
	Mouse cDNA (mammary gland) similar to X51760 Human zinc finger protein ZFP-36	
PN/2/A6	Rat tricarboxlate carrier	S70011
	Mouse cDNA (liver/embryo/mammary gland/skin) similar to Rat tricarboxlate carrier	
PN/2/A7	Rat brain myosin II isoform	Z32519
	Mouse cDNA similar to Human myosin regulatory light chain 2 (X54394)	gb/AW228807
PN/2/A8	Rat phospholipase D	
	Mouse cDNA (melanoma)	
PN/2/A9	Human tetratricopeptide repeatdomain 2	
	Mouse mDj11	dbj/AB028861
	Mouse cDNA (blastocyst) similar to phosphoprotein phosphatase (G567040)	
PN/2/A10	X	
PN/2/A11	Mouse NIX mRNA nuclear gene encoding mitochondrial protein also Mouse cDNA similar to NIP3	AF067395
PN/2/A12	Mouse cDNA (diaphragm)	
PN/2/B1	Rat transferrin	D38380
PN/2/B2	Mouse cDNA (liver/mammary gland) similar to B-galactosidase a-peptide	
PN/2/B3	Mouse mitochondrial genome-> 3x tRNA's	V00665
PN/2/B4	Mouse WDNM1	
PN/2/B5	Mouse translational controlled 40KDa polypep. P40 / 37KDa oncofetal antigen (also see J02870/D25224/J03799)	X06406/AF140348

Met/1/D12	Mouse eIF-4AII pro syn initiation factor	X12507
Met/1/E1	Human G protein gamma-II subunit	U313841
Met/1/E2	Mouse cDNA (melanoma)	gb/AI614971
Met/1/E3	Mouse hnRNP H	L22009
Met/1/E4	Mouse fetal globin inducing factor	AF028722
Met/1/E5	Mouse cDNA embryo	gb/AA409021
Met/1/E6	?	
Met/1/E7	Rat ATP synthase subunit d	D13120
	Mouse cDNA similar to above	
Met/1/E8	Rat 26s proteasome	AJ006340
Met/1/E9	Mouse cDNA (embryonic region)	
Met/1/E10	Mouse cDNA (mammary gland/blastocyst/stromal cell line	gb/AA574803
Met/1/E11	Mouse cDNA irradiated colon	
Met/1/E12	Rat Shal-related potassium channel kv4.3	U42975
	Mouse cDNA (melanoma)	
Met/1/F1	Mouse COP 9 complex subunit 4	gb/AF071314
Met/1/F2	Mouse COP 9 complex subunit 4	AF071314
Met/1/F3	repeat sequence	gb/AF263366
Met/1/F4	Mouse mitochondrialgen response 96KDa phosphoprotein	U18869
	Rat Doc-2 isoform	U95178
	Human mitochondrialgen response phosphoprotein Doc-2	U53446
Met/1/F5	Mouse DJ-1(Human DJ-1 D61380)	AB015652
Met/1/F6	Mouse Transcobalmin II (Tcn2)	AF090686
Met/1/F7	Mouse cDNA (pancreas)	dbj/AU051688
Met/1/F8	Lysozyme M	M21050
Met/1/F9	Lysozyme M	M21050
Met/1/F10	Mouse cyclin T1	AF109179
Met/1/F11	Mouse flavo-binding protein	gb/AF174535
Met/1/F12	Mouse cDNA (also mammary gland)	gb/AI604169
Met/1/G1	Mouse cDNA (mammary gland)	gb/AI464992
Met/1/G2	?	
Met/1/G3	Human CCR4-associated factor 1 (POP2)	gb/AF180476
	Mouse cDNA similar to mCAF1 protein (Mouse CCR4 associated factor 1)	
Met/1/G4	Mouse mitochondrial genome (ATPase subunit 6)	
Met/1/G5	Mouse Pulmonary Surfactant Protein C	M38314
Met/1/G6	Mouse cDNA (mammary gland)	
Met/1/G7	T1 antigen=truncated large T antigen	gb/U75992
Met/1/G8	Rat cDNArogen-responsive uterine	U53184
	Mouse cDNA similar to TNF-a inducible responsive element	
Met/1/G9	Mouse epididymal secretory protein	AB021289
Met/1/G10	Mouse cdc 42	L78075
Met/1/G11	Mouse cDNA similar to Human lactoylglutathione lyase	
Met/1/G12	Mouse cDNA similar to phosphatidyl serine decarboxylase proenzyme	
Met/1/H1	Mouse cDNA	gb/AF209905
Met/1/H2	Mouse IL-8/gro-alpha receptor	L13239
Met/1/H3	?	
Met/1/H4	Mouse cDNA	
Met/1/H5	Mouse mitochondrial genome	
Met/1/H6	Rat mitochondrial. succinyl-CoA synthetase alpha subunit	gb/AF144101
Met/1/H7	Mouse sex determination protein homolog female gene	AF064447
Met/1/H8	Mouse cDNA (myotubes)	gb/AI643683
Met/1/H9		
Met/1/H10	Mouse cDNA	
Met/1/H11	Mouse cDNA similar to TGR-CL10	
Met/1/H12	Human mitochondrial genome	
Met/2/A1	Mouse cDNA similar to Human G a interacting protein GAIP	
Met/2/A2	Mouse cDNA	
Met/2/A3	X	
Met/2/A4	Mouse protein kinase Bub1	AF002823
Met/2/A5	Mouse cDNA (embryonic stem cell)	
Met/2/A6	Mouse ceruloplasmin	U49430 MMU49430
Met/2/A7	Mouse oxysterol-binding protein	AB017026
Met/2/A8	X	
Met/2/A9	Mouse skeletal LIM protein (FHL1)	U77039 MMU77039
Met/2/A10	Mouse GalNAc a-2,6 siayltransferase III	Y11342.1 MMY11342
Met/2/A11	Mouse Calmodulin synthesis (CAM)	M27844 MUSCAM
Met/2/A12	X	

PN/2/B6	Mouse cDNA (mammary gland)	
PN/2/B7	Mouse cDNA (blastocyst/diaphragm/embryonic region)	
PN/2/B8	Mouse cDNA (mammary gland/embryo)	
PN/2/B9	X	
PN/2/B10	Rat microtubule-associated protein 2 (MAP 2)	U30938
	Mouse cDNA (mammary gland) similar to MAP-2	
PN/2/B11	Human spicesomal protein SAP 155 mRNA	AF054284
	Mouse cDNA (embryonic region/mammary gland/macrophage)	
PN/2/B12	Mouse cDNA (embryo)	
PN/2/C1	?	
PN/2/C2	Mouse TNFR2-TRAF signalling complex protein	L49433
	Mouse inhibitor of apoptosis protein 2	U88909
PN/2/C3	Mouse MO25 gene	S51858
PN/2/C4	X	
PN/2/C5	Mouse embryo	
PN/2/C6	Mouse cDNA (T cell)	
PN/2/C7	Mouse cDNA (T cell)	gb/AA546226
PN/2/C8	Mouse cDNA (kidney/mammary gland/embryo/heart etc)	
PN/2/C9	Rat ribosomal protein S3a	X75161
	Rat & Human putative v-fos transformation effector protein(Fte-1)	M84711
PN/2/C10	Mouse cDNA (macrophage)	
PN/2/C11	Mouse cDNA (mammary gland) similar to Rat amidophosphororibosyltransferase precursor	
PN/2/C12	Mouse cDNA (ln/myotubes/kidney/macrophage)	
PN/2/D1	X	
PN/2/D2	Rat liver cytochrome c oxidase subunit VIc (COX-VIc)	M27466
PN/2/D3		
PN/2/D4	Rat Transferrin	
	Mouse cDNA similar to transfferin	
PN/2/D5	Mouse transferin last exon	M30820
PN/2/D6	Mouse kappa casein	
PN/2/D7	Human cDNA KIAA0433	
PN/2/D8	Mous peroxisomal PTS2 receptor	U69171
PN/2/D9	Mouse MHC class1 heavy chain precursor/H2-K gene	U47328/S00400
PN/2/D10	Mouse WDNM 1	X93037
PN/2/D11	Rat syntaxin 5 (Mouse cDNA similar to syntaxin)	L20822
PN/2/D12	Mouse serine proteinase inhibitor (SP13)	U25844
PN/2/E1	X	
PN/2/E2	Mouse cDNA (stromal cell line) similar to Human basic-leucine zipper nuclear factor (JEM1, NM 003666.1)	
PN/2/E3	MMTV env gene region	
PN/2/E4	?	
PN/2/E5	Mouse protein phoshpatase type 1 (dis 2m2)	M27073
PN/2/E6	?	
PN/2/E7	Mouse cDNA	
PN/2/E8	Mouse cDNA (kidney) similar to Human ras-related protein Rab-11(X56740)	
PN/2/E9	Rat cDNA	
PN/2/E10	MousecDNA (melanoma)	
PN/2/E11	Mouse cDNA (embryo/mammary gland)	
PN/2/E12	Mouse breast hsc73	U27129
PN/2/F1	Mouse tumour surpressor maspin	U54705
PN/2/F2	Mouse androgen-related gene RP2	K03414
PN/2/F3	Mouse cDNA (embryo, also mammary gland/blastocyst)	gb/AW53118
PN/2/F4	Mouse cDNA (urigenital ridge NMUB)	
PN/2/F5	Mouse cDNA (embryo/mammary gland/myotubes/uterus/hypothalamus)	
PN/2/F6	?	
PN/2/F7	Rat Sec61 homologue	M96630
PN/2/F8	?	
PN/2/F9	Rat transferrin	D38380
PN/2/F10	Mouse Nedd5 mRNA for DIFF 6 or CDC3,10,11,12-like protein with GTPase motif	D49382
PN/2/F11	Mouse cDNA (kidney) similar to Rat mitochondrial dicarboylate carrier (AJ223355)	
PN/2/F12	Human non-lens beta gamma-crystallin like protein	U83115
PN/2/G1	Mouse cDNA (mammary gland/blastocyst/skin/heart)	
PN/2/G2	Mouse Nedd5 mRNA for DIFF 6 or CDC3,10,11,12-like protein with GTPase motif	D49382
PN/2/G3	Mouse cDNA (mammary gland also kidney/testis/embryo/urigenital ridge)	
PN/2/G4	?	
PN/2/G5	?	
PN/2/G6	Mouse protein synthesis elongation factor 1-alpha	M22432
PN/2/G7	repeat sequence	
PN/2/G8	Mouse alcohol dehydrogenase-B2 (Adh-2)	M84147

Met/2/B1	Mouse mesoderm/mesenchyme forkhead 1 (Mf1)	AF045017
Met/2/B2	X	
Met/2/B3	Rat TM-4 gene -> fibroblast tropomyosin 4	Y00169 RNTM4
	Mouse cDNA (myotubes)	
Met/2/B4	Crictulus griseus UDP-glucose pyrophosphorylase	
Met/2/B5	Mouse cDNA (mammary gland/liver/embryo/irradiated colon)	
Met/2/B6		
Met/2/B7	Mouse cDNA	
Met/2/B8	?	
Met/2/B9	X	
Met/2/B10	?	
Met/2/B11	Mouse ribosomal protein S7	AF043285
Met/2/B12	Mouse cDNA (mammary gland/blastocyst/etc) similar to NADH-ubiquinone oxidoreductase MLRQ Subunit	
Met/2/C1	X	
Met/2/C2	Mouse phosphatidylethanolamine binding protien	U43206 MMU43206
Met/2/C3	Human DEAD/H box polypeptide 10 (RNA helicase)	U28042
	Mouse cDNA (lymph node) similar to Human DEAD/H box polypeptide 10 (RNA helicase)	
Met/2/C4		
Met/2/C5		
Met/2/C6	Mouse cDNA (also tongue/Rat embryo)	gb/AI552167
Met/2/C7	X	
Met/2/C8	Mouse repeat sequence	
Met/2/C9	Mouse mitochondrial genome/ Mouse cDNA similar to NADH-ubiquinone oxidoreductase chain 5	V00711
Met/2/C10	Human cDNA (muscle)	gb/AW176514
Met/2/C11	Mouse cDNA (mammary gland)	
Met/2/C12	Rat transferrin	X77158 RNTF
	Mouse cDNA (mammary gland/liver/lung/embryo) similar to Human serotransferrin precursor	
Met/2/D1	Mouse cDNA (ln/mammary gland/blastocyst/liver) similar to Human N-terminal acetyltransferase complex ARD1 subunit homolog	
Met/2/D2	X	
Met/2/D3	Mouse ubiquitin fusion-degradation 1 like protein (ufdl1)	U64445
Met/2/D4	Mouse cDNA	
	Mouse cDNA similar to Human cAMP-dep PK type II beta reulatory chain	
Met/2/D5	Mouse pulmonary surfactant protein C (Rat hydrophobic surfactant associated protein C)	
Met/2/D6	X	
Met/2/D7	?	
Met/2/D8	Mouse cDNA	gb/AA145325
Met/2/D9	Human RasGAP-related protein (IQGAP2)	U51903
Met/2/D10	Human CGI-43 protein	AF151801
Met/2/D11	Mouse lysozyme M/P	M21048/X51547
Met/2/D12	Mouse cDNA (embryo/mammary gland/macrophage) similar to Human interferon-inducible protein I-8D (X57351)	gb/AA980815
Met/2/E1	Mouse cDNA (mammary gland)	
Met/2/E2	Mouse cDNA	
Met/2/E3	Mouse ubiquitin/60s ribosomal fusion	AF118402
Met/2/E4	Mouse C57BL/6J Sec61 protein complex gamma subunit	U11027
Met/2/E5	Mouse cDNA (pooled organs) & Human CGI-35 protein	Human AF132969
Met/2/E6	MousecDNA (kidney also heart, blastocyst,mammary gland,lung etc)	
Met/2/E7	Mouse cDNA (embryo)	
Met/2/E8	Mouse protein synthesis elongation factor 1-alpha	M22432
Met/2/E9	Mouse cDNA (mammary gland)	
Met/2/E10	Mouse matrix Gla protein	D00613
Met/2/E11	Mouse repeat sequence	
Met/2/E12	Mouse cDNA (mammary gland) similar to Human glia maturation factor B	
Met/2/F1	Mouse galactose-1-phosphate uridyl tarnsferase (GALT)	M96265
Met/2/F2	Mouse G3PDH	
Met/2/F3	Human BAC/Mouse cDNA	
Met/2/F4	Mouse ribosomal protein S6/a globin transcription factor CP	Y00348/M76987
Met/2/F5	Mouse ribosomal protein L41	
Met/2/F6	Mouse intracellular calcium binding protein (MRP14)/Human S100A9 (CalgranulinB)	M83219
Met/2/F7	Mouse cDNA	
Met/2/F8	Mouse mitochondrial genome	
Met/2/F9	Mouse repeat sequence (MHC classIII)	
	Mouse cDNA similar to acetyl CoA synthetase	
Met/2/F10	Rat XY40/p47	Y10769/AB002086
Met/2/F11		
Met/2/F12	Mouse cDNA similar to uridine kinase	
Met/2/G1	Mouse translation repressor NAT1 (Mouse translation initiation factor Eif4g2)	U76112
Met/2/G2	?	
Met/2/G3	?	

PN/2/G9	Mouse cDNA (mammary gland) similar to Human destrin (actin depolymerisation factor) & Mouse cofiln (D00472)	
	Mouse cDNA (mammary gland) similar to Mouse cofiln (D00472)	
PN/2/G10	Mouse cDNA (urigenital ridge NMUB/liver/mammary gland/embryo/HCC)	
PN/2/G11	Mouse arginase II	
PN/2/G12		
PN/2/H1	Mouse cDNA (embryo/mammary gland)	
	Human MAGE tumour antigen D1 (MAGE-D1, AF124440.1)	
PN/2/H2	Mouse adenylsuccinate synthetase	L24554
PN/2/H3	Mouse cDNA (kidney)	
PN/2/H4	Mouse cDNA (pooled organs)	
	Rat mRNA for unknown protein (PIPPin)	X89962
PN/2/H5	Mouse EGP314 precursor	M76124
PN/2/H6	Mouse cDNA	gb/AW122449
PN/2/H7	Mouse liver arginase	U51805
PN/2/H8	Mouse cDNA (kidney/liver) similar to C.griseus epsilon-COP	emb Z32554
PN/2/H9	Mouse steroyl-CoA desaturase (SCD2)	M26270
PN/2/H10		
PN/2/H11	Mouse glycoprotein-associated amino acid transporter y+LAT1a	AJ012754
PN/2/H12	Mouse cDNA	

Met/2/G4	Mouse lysozyme M exon 4	
Met/2/G5	Mouse integrase interactor 1a protein (INI1A)	AJ011739
Met/2/G6	X	
Met/2/G7	Mouse cDNA (embryo)	
Met/2/G8	Human cDNA KIAA0841 protein	dbj/AB020648
	Mouse cDNA (embryo)	gb/AA052556
Met/2/G9	Mouse cDNA (embryo/mammary gland/T cell/kidney) similar to Human MHC antigen A-24 & Mouse MHC class I Qa-Tla	
Met/2/G10	Mouse pulmonary surfactant protein C	gb/M38314
Met/2/G11	Mouse cDNA	
Met/2/G12	Mouse cDNA similar to Mouse T-cell specific tcn factor (X61385)	
Met/2/H1	Mouse cDNA (blastocyst/myotubes/embryo/irradiated colon)	
Met/2/H2		
Met/2/H3	Mouse protein kinase inhibitor p58	U28423
Met/2/H4	Mouse cDNA (embryo/mammary gland) similar to NIP1	
	Human BCL2/adenovirus E1B 19KD interacting protein 1 & Human Nip1 isoform S3/S2/S1	
Met/2/H5	Mouse cDNA	AW048554.1
Met/2/H6	Mouse cDNA (mammary gland/embryo)	
	Human cDNA similar to contains element MER22 repetitive element	
Met/2/H7	Mouse calpactin I light chain (p11)	M16465
Met/2/H8	Rat transferrin	
Met/2/H9	?	
Met/2/H10	Mouse cDNA (myotubes)	
Met/2/H11	?	
Met/2/H12	Mouse cDNA similar to Mouse D12644 KIF2 protein	gb/AI639935