A study to investigate the expression, structure and function of alternatively spliced progesterone receptor variants in breast cancer cells

A thesis submitted for the degree of Doctor of Philosophy

November 2011

David Michael William Cork

Abstract

The progesterone receptor (PR) gene consists of eight exons and encodes the functionally distinct PR-A and B isoforms. PR alternative splicing events involving deletion of internal exons or intron retention have been reported, potentially generating proteins which lack various internal functional domains or are N-terminally truncated. PR status in breast cancer may be predictive of the efficacy of endocrine therapy and is measured using N-terminally targeted antibodies which detect both PR-A and B.

In this study PR mRNA expression was assessed in breast cancer (MCF-7, T47-D and MDA-MB-231) and non-tumourigenic breast (MCF-10A) cell lines using an RT-PCR based gene walking assay. PR mRNA resulting from alternative splicing was detected in reportedly PR negative MDA-MB-231 cells using this assay. These mRNA could encode the low molecular weight nuclear and cytoplasmic PR proteins which are detected in this cell line using the C-terminal PR antibody C19. Ligand blotting, co-immunoprecipitation and DNA affinity precipitation assays demonstrated the ability of these proteins to bind progesterone, interact with the PR nuclear co-factor PSF, dimerise and bind DNA; thus potentially to function as ligand activated nuclear transcription factors. Validation studies using the gene walking assay demonstrate that alternatively spliced PR mRNA was also present in breast tumours originally characterised using N-terminal antibodies as being both PR+ and PR-. Using a novel non N-terminal PR antibody developed in this study, the nuclear PR status differed from the original status for 2 of 14 tumours examined, and cytoplasmic PR was detected. The results presented in this thesis suggest that PR undergoes extensive alternative splicing, generating potentially functional isoforms, and that expression of variant PR isoforms may affect the PR status of a tumour as determined using antibodies targeting different epitopes.

PR exons 4 and 6 are flanked by consensus 5′ and 3′ splice sites and contain cryptic 3′ splice sites, as well as potential binding sites for a range of RNA binding splicing factors. siRNA knockdown of the individual SR proteins SRSF1, SRSF2, SRSF5 and SRSF6 identified potentially antagonistic roles for SRSF1 and SRSF2 in influencing the inclusion/skipping of PR exons 4 and 6, and also for SRSF6 in regulating exon 6 splicing.

Acknowledgements

The biggest thanks of all must go to Dr Alison Tyson-Capper for supporting and inspiring my work with her endless enthusiasm for science and alternative splicing in particular. Thanks also to Professor Tom Lennard for his help and advice throughout my project, and to the RVI Breast Cancer Appeal for funding this work. A massive debt of gratitude is owed to Mrs Liz Shiells for her patience and technical assistance throughout my lab work, to Dr Hannah Gautrey (who is also owed a massive debt of beer!) for allowing me to use her siRNA knockdowns, and to all other past and present members of the lab, especially Miss Elaine Hong for allowing me to cite her unpublished data. I am also grateful to Mr David Browell for providing access to the breast tumour tissue bank used in this project, and to the Histocytopathology Department at Gateshead Queen Elizabeth Hospital for assistance with immunohistochemistry and breast cancer PR scoring.

Everybody with whom I have shared an office over the last four years has been great in their own way, especially little Kelly, Hedele/Snizz (the office ethnic), Rachel (the most organised person in the world), Julia (the scary Austrian who annexed the office for 3 years), and more recently Stephen (bloody lad!), Chris (who spends half his time in 1945) and Achilleas; you've all contributed to making work fun and have helped me through difficult science times with some random banter. Also, thanks to the Friday pub people; especially Graeme (who I will always owe beer, table chips and cheesy garlic bread), Cat with an O (who drinks more beer/gin than any feline I've ever met and has a freaky blue hand), Joe, Dom, Hannah (again) and David (Fiiive David Swans ♫), as well as to the North Terrace 'Wenches'. These people have all made the time a lot more enjoyable and helped take some of the stress out of doing a PhD.

Finally, a big thank you to my family for all their support (and financial contributions!) especially to my girlfriend Leanne for the massive amount of love and patience she has shown to support me being poor for the huge length of time I have spent writing this thesis (and for her amazing EndNote skills which saved me hours of PubMed searching!).

Abbreviations

AF	Transactivation domain
AI	Aromatase inhibitor
AR	Androgen receptor
ASE	Alternative splicing event
ASV	Alternatively spliced variant
BSA	Bovine serum albumin
BPS	Branch point sequence
BUS	B-upstream segment
CDK	Cyclin dependent kinase
CERES	Composite exonic regulatory element of splicing
CFTR	Cystic fibrosis transmembrane conductance regulator
Co-IP	Co-immunoprecipitation
Δ	Exon deletion ASE
DAPA	DNA affinity precipitation assay
DBD	DNA binding domain
DCIS	Ductal carcinoma in situ
DD	Dimerisation domain
EDA	Fibronectin extra domain A
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ER	Oestrogen receptor
ERE	Oestrogen response element
Erk	Extracellular signal-regulated kinase
ERKO	Oestrogen receptor knockout
ESE	Exon splice enhancer
ESS	Exon splice silencer
FFPE	Formalin fixed paraffin embedded
GFP	Green fluorescent protein
GR	Glucocorticoid receptor
HER2	Human epidermal growth factor receptor 2
hnRNP	Heterogenous nuclear ribonucleoprotein
HRP	Horseradish peroxidase
HRT	Hormone replacement therapy
ID	Inhibitory domain
ISE	Intron splice enhancer
ISS	Intron spice silencer
LBD	Ligand binding domain
MAP	Multiple antigenic peptide
MAPK	Mitogen activated protein kinase
MEC	Mammary epithelial cell
NLS	Nuclear localisation sequence
NMD	Nonsense mediated decay
NTC	No template control
NTD	N-terminal domain
PI3K	Phosphatidylinositol 3-kinase
P-POD	Peroxidase conjugated progesterone
PPT	Polypyrimidine tract
PR	Progesterone receptor
PRE	Progesterone response element
PRKO	Progesterone receptor knockout

PSF	Polypyrimidine tract binding protein associated splicing factor
PTC	Premature termination codon
PTEN	Phosphatase and tensin homologue
qRT-PCR	Real-time quantitative RT-PCR
RANKL	Receptor activator of NFκB ligand
RBP	RNA binding protein
RIPA	Radioimmunoprecipitation assay
RRM	RNA recognition motif
RS domain	Arginine/serine rich domain
SAP	Shrimp alkaline phosphatise
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment
SERM	Selective oestrogen receptor modulator
siRNA	Small interfering RNA
snRNP	Small nuclear ribonucleoprotein particle
Sp1	Specificity protein 1
SPRM	Selective progesterone receptor modulator
SRp	Serine/arginine rich protein
SRC	Steroid receptor coactivator
SRSF	Serine/arginine rich splicing factor
Stat	Signal transducer and activator of transcription
TGF	Transforming growth factor
UTR	Untranslated region
U2AF	U2 auxiliary factor
VEGF	Vascular endothelial growth factor

Contents

List of figures

List of tables

1 Introduction

1.1 Progesterone Receptor

1.1.1 Progesterone

Progesterone is a steroid hormone, predominantly released by the corpus luteum in response to luteinising hormone prior to ovulation, in order to maintain the environment necessary for implantation. If fertilisation, and subsequent implantation, does not occur, the regression of the corpus luteum withdraws the progesterone signal. During pregnancy, progesterone is involved in regulating uterine proliferation and suppressing myometrial contractility (Graham and Clarke, 1997). In the endometrium progesterone has a largely anti-proliferative effect, opposing the action of oestrogen. In mammary tissue high levels of serum progesterone correlate with periods of extensive epithelial proliferation and differentiation to mature secretory structures during pregnancy in preparation for lactation (Lydon et al., 1995), whilst also suppressing milk production until parturition (Graham and Clarke, 1997). Progesterone signals through the progesterone receptor (PR) which exists as two nuclear receptor isoforms PR-B and the N-terminally truncated PR-A. PR is a member of the steroid/thyroid hormone receptor family of ligand activated nuclear transcription factors (Evans, 1988). The structure and function of progesterone receptors will be discussed.

1.1.2 PR gene structure

The PR gene consists of eight exons punctuated by introns ranging in length from the relatively short intron 1 (1275bp) to the long intron 2 (34130bp) (Misrahi et al., 1993). The long first exon (2380bp) contains two oestrogen responsive promoter regions which initiate transcription of mRNA encoding the nuclear PR isoforms. Transcription of PR-B is initiated at positions +1 and +15, encoding a 933 amino acid protein. Transcription of the N-terminally truncated PR-A isoform is initiated at positions +737 and +842 encoding a 769 amino acid protein which is truncated of the N-terminal 164 amino acids of PR-B (Kastner et al., 1990). The expression of PR protein is responsive to oestrogen stimulation, however sequencing of the PR gene revealed that the promoter regions do not contain consensus oestrogen response element (ERE) pallindromic sequences but do contain an ERE half site in proximity to potential binding sites for the

transcription factor specificity protein (Sp)1 (Kastner et al., 1990). It was later demonstrated by gel shift assays that the ERE half site in the PR-A promoter is capable of binding oestrogen receptor (ER)α, and oestrogen stimulated transcription of a chloramphenicol acetyltransferase (CAT) reporter gene containing the PR-A promoter ERE half site and Sp1 sites (Petz et al., 2004). The proximity of the Sp1 sites to the ERE half site, and the ability of Sp1 to interact with ERα via the DNA binding domain (DBD) and ligand binding domain (LBD) (Petz et al., 2004) demonstrate a potential mechanism for combinatorial regulation of PR expression, allowing response to external stimuli resulting from changes in the cellular environment. The remainder of the PR gene is made up of six short internal exons of 152bp, 117bp, 306bp, 145bp, 131bp and 158bp. Exon 8, the terminal exon, is 9648bp containing a 153bp coding region and a long 3′ untranslated region (UTR) (Misrahi et al., 1993). The structure of the PR gene can be seen in Figure 1.1.

1.1.3 Nuclear PR isoforms

The full length PR-B and N-terminally truncated PR-A isoforms contain distinct functional domains characteristic of ligand activated nuclear transcription factors (Misrahi et al., 1987) (Figure 1.1). Both isoforms contain a long N-terminal domain encoded by exon 1, containing a transactivation domain (AF1) which is common to both PR-A and PR-B and responsible for mediating interactions with transcriptional co-regulators. The N-terminal 164 amino acids of PR-B, the B-upstream segment (BUS), which are truncated from PR-A form the PR-B specific AF3. The N-terminal region also contains an inhibitory domain (ID), mapping to the first 140 amino acids of PR-A (Giangrande et al., 1997). Exons 2 and 3 each encode one zinc finger of the DBD, responsible for interacting with progesterone response elements (PREs) of target gene promoters. The 5′ part of exon 4 encodes a hinge region responsible for the conformational change necessary to expose the DBD and dimerisation domain (DD) after ligand binding. Exon 4 also encodes a nuclear location sequence (NLS) and the 3′ part of exon 4, along with the rest of exons 5 to 8, encode the LBD, DD and a further transactivation domain (AF2) which becomes functional following dimerisation (Misrahi et al., 1993; Kastner et al., 1990; Misrahi et al., 1987). At least thirteen phosphorylation sites exist within PR, six in the PR-B specific BUS, and seven in regions common to both PR-A and PR-B. Some of these are basally phosphorylated, exhibiting increased phosphorylation upon ligand binding and others are phosphorylated

only after ligand binding (Ward and Weigel, 2009). The functional implications of PR phosphorylation will be discussed. The domain structure of PR-A and PR-B are shown in Figure 1.1.

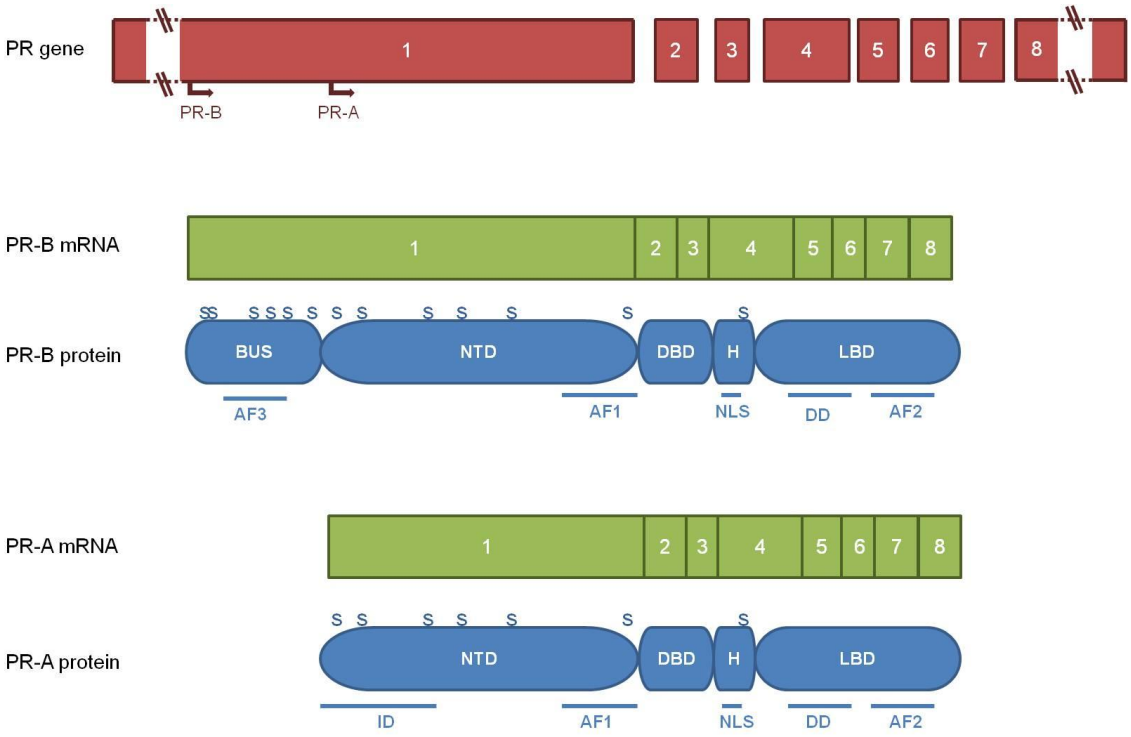


Figure 1.1: Structure of the PR gene and nuclear PR-A and PR-B isoforms.

Diagramatic representation of the eight exons of the PR gene, depicted relative to the length in bases. Exon 1 contains a long 5′ UTR and exon 8 a long 3′ UTR indicated by the broken lines. The eight exons are separated by introns (not depicted by length). The sites of transcriptional initiation resultig from promoter B (PR-B) and promoter A (PR-A) are indicated by the arrows. The mRNA for each isoform is shown with the structure of the encoded proteins beneath. Structural domains are depicted relative to their length in amino acids and position within the protein sequence. BUS= B-upstream segment, NTD= N-terminal domain, DBD= DNA binding domain, H= Hinge, LBD= Ligand binding domain, ID= Inhibitory domain, AF= Transactivation domain, NLS= Nuclear location signal, DD= Dimerisation domain, S= Serine.

1.1.4 Mechanism of PR action

In the absence of ligand PR exists in heteromeric complexes with chaperone proteins including the heat shock proteins (Hsp)90, Hsp70 and Hsp40 (Cheung and Smith, 2000). Protein complexes containing Hsp90 are required to establish the mature hormone binding conformation of PR (Smith, 1993). The chaperone complex also appears to provide protection from proteosomal degradation of non ligand bound PR, since blocking complex assembly increases the rate of PR degradation (Lange et al., 2000). In living cells, non-ligand bound PR localises predominantly in the nucleus but is

undergoing constant nucleocytoplasmic shuttling in response to the NLS (Wan et al., 2001). PR-LBD constructs locate predominantly in the cytoplasm (Kakar et al., 2006), presumably since they lack the hinge region constitutive NLS. Cytoplasmic localisation of non-ligand bound PR-LBD is mediated by interaction with Hsp90, however Hsp90 is also required for nuclear retention of ligand bound PR-LBD, suggesting a role for chaperones in regulating transport of PR across the nuclear membrane (Kakar et al., 2006).

Ligand binding causes specific conformational changes and dissociation from chaperones, exposing the DD allowing the formation of PR dimers (Tsai and O'Malley, 1994). The importance of the hinge region in facilitating this altered conformation and thus facilitating dimerisation was indicated by the inability of C-terminal PR constructs containing the DD, but lacking the N-terminal and hinge regions, to form dimers detectable by co-immunoprecipitation with full length PR, pull down with tagged immobilised peptide constructs or by ligand binding assays (Tetel et al., 1997).

Ligand activated, dimerised PR then translocates to the nucleus, largely in response to a constitutive NLS located within the hinge region (Kastner et al., 1990). Mutation of this sequence leads to cytoplasmic accumulation of PR and delayed translocation in response to progesterone (Kakar et al., 2006). Present within the NLS sequence of the hinge region is an acetylation domain and ligand induced acetylation of PR has also been demonstrated. NLS acetylation domain mutant PR-B, transfected into T47D breast cancer cells, displayed delayed ligand induced phosphorylation, nuclear entry and transcription of PRE containing target genes (Daniel et al., 2010). A second NLS, located within the DBD, can also mediate PR nuclear translocation in response to ligand (Guiochon-Mantel et al., 1992). Nuclear PR dimers bind to a PRE sequence within the promoter region of target genes and regulate transcription via interaction with co-activators which help to control chromatin remodelling and assembly of the transcriptional machinery (Li et al., 2004). The transcriptional activity of PR is partially regulated by phosphorylation; mutation of two key phosphorylation sites, Ser190 and Ser294, leads to a reduction in transcriptional activity of ligand activated PR (Takimoto et al., 1996) and phosphorylation of Ser400 results in increased transcriptional activity of PR in response to progesterone (Pierson-Mullany and Lange, 2004).

1.1.5 Transcriptional activity of PR isoforms

PR-A and PR-B are functionally distinct transcription factors and the gene regulation profiles of the two isoforms have been characterised in breast cancer cells which express either PR-A or PR-B; PR-B was demonstrated to be a more potent activator of transcription, uniquely up-regulating transcription of almost fifteen times as many genes as PR-A in these cells (Richer et al., 2002). A subset of genes was up-regulated by both isoforms, most of which were preferentially up-regulated by PR-B, and some genes were down-regulated by either PR-B only or both PR-A and PR-B (Richer et al., 2002). Ten of the genes shown to be regulated by PR-A or PR-B in these experiments have known roles in mammary gland morphogenesis or tumourigenesis (Richer et al., 2002). Similar patterns of progesterone responsiveness are seen in endometrial stromal cells and endometrial carcinoma cells, with PR-A and PR-B demonstrating overlapping but largely distinct transcriptional activities, and PR-B being the more potent transcription factor (Yudt et al., 2006; Davies et al., 2004). Transfection of T74D breast cancer cells which express only PR-B with a PR-A expression vector led to a reduction in PR-B mediated transcription, demonstrating the stronger activity of PR-B as a transcriptional activator and also the repression of PR-B mediated transcription by PR-A, either through competition for co-factor binding or direct repression in heterodimers (Jacobsen et al., 2002).

PR-B was demonstrated to be a more potent transcriptional activator of a PRE containing reporter gene by transfection of PR negative T74D breast cancer cells with either PR-A or PR-B (Vegeto et al., 1993). PR-A also functions as a dominant negative regulator of PR-B mediated transcription, an interaction termed transrepression; co-transfection of PR-A and PR-B with a PRE reporter gene reduced levels of transcription compared to expression of PR-B alone. Mutations of the PR-A DBD did not abrogate transrepression demonstrating that it is not mediated via competition for PRE binding (Vegeto et al., 1993). Transrepression by PR-A of other structurally related members of the steroid receptor superfamily, including glucocorticoid receptor (GR), mineralocorticoid receptor (MR), androgen receptor (AR) and ER, has also been observed, similar to the repression of PR-B (Abdel-Hafiz et al., 2002; Giangrande et al., 1997; McDonnell et al., 1994; Vegeto et al., 1993). PR dimers were produced in COS-1 cells by forced dimerisation using PR constructs containing the leucine zipper regions of c-fos and c-jun (Mohamed et al., 1994). The transcriptional activity of pure PR-A:B

heterodimers was similar to PR-A:A homodimers and significantly lower than PR-B:B homodimers demonstrating a role for direct interaction between the dimerised PR isoforms in transrepression (Mohamed et al., 1994).

A region responsible for the transrepression of PR-B activity by PR-A was mapped by deletion analysis to the extreme N-terminal region in PR-A, and just after the BUS in PR-B, termed the inhibitory domain (ID). Deletion of this region removed the transrepression of PR-B by PR-A. Furthermore, transrepression could be reversed by creating fusion constructs containing the BUS in addition to AF-1 and ID, suggesting a role for BUS in the modulation of ID (Giangrande et al., 1997). The ID was later demonstrated to also be responsible for autoinhibition, defined as the increase in transcription observed by removal of ID, which occurs in PR-A and not PR-B and may account, in part, for weaker transcriptional activation by PR-A (Hovland et al., 1998). The mechanism controlling transrepression has since been linked to the presence of a binding site for small ubiquitin-like modifier (SUMO)-1 within the N-terminal ID, encoded by exon 1. Mutation of the SUMO-1 binding sequence abolishes both transrepression and autoinhibition, and this sumoylation mediated transrepression and autoinhibition requires progesterone binding to the LBD (Abdel-Hafiz et al., 2002), demonstrating that coordination between distant regions of the full length receptor are necessary for normal function. A putative role in the inhibition of sumoylation was proposed for Ser294 phosphorylation since a serine to alanine mutation at this position was reported to decrease the transcriptional activity and increase ligand dependent sumoylation (Daniel et al., 2007) although the effect of this mutation on sumoylation has recently been disputed (Abdel-Hafiz et al., 2009).

1.1.6 Transactivation domains and PR co-regulators

As well as the autoinhibitory and transrepressive functions assigned above to ID, the differing activity of PR-A and PR-B may also in part be due to the different range of transcriptional co-regulators which interact with AF1, AF2, AF3 and ID in the two isoforms. PR-A and PR-B share two common transactivation domains, AF1 and AF2. Mutagenesis studies have mapped the constitutively active AF1 to an N-terminal region upstream of the DBD. AF2 was similarly mapped to a C-terminal region of the LBD, only becoming active following agonist binding (Meyer et al., 1990). A third

transactivation function (AF3) was initially assigned to the BUS region, unique to PR-B, since co-transfection of BUS containing constructs with a reporter gene could initiate reporter activity independent of the presence AF1 and AF2, but only when constructs contained the PR DBD. This transactivation could be enhanced by the presence of AF1 or AF2, demonstrating the requirement for coordination between distant functional domains for proper PR function (Sartorius et al., 1994).

PR co-regulators can be either co-activators, such as the steroid receptor co-activator (SRC) family, or co-repressors, such as nuclear receptor co-repressor (N-CoR) and silencing mediator for retinoid and thyroid hormone receptor (SMRT). Co-activators bind to ligand activated PR and co-repressors bind to non-ligand bound PR or PR which is bound by an antagonist. The general functions of steroid receptor co-factors include chromatin remodelling, histone acetylation (co-activators)/de-acetylation (co-repressors) and recruitment of transcription factors which form part of the transcriptional initiation complex (McKenna et al., 1999). SRC-1 was characterised as a general co-factor of all steroid hormone receptors, to interact with the C-terminus of PR and to activate increased hormone responsive transcription of a PRE containing reporter gene (Onate et al., 1995). Transcriptional co-factors of steroid receptors work in a coordinate manner as part of multi-protein complexes; CREB-binding protein (CBP) was identified as an activator of ER and PR transcription that when co-expressed with SRC-1 could activate transcription above the levels predicted by additive effect the two receptors working independently (Smith et al., 1996). p300, a structural homologue of CBP, was only recruited after SRC-1 binding to PR and exhibited a similar synergistic activation of transcription, facilitating the re-initiation of transcription of an activated gene. Mutation of the p300 binding site in SRC-1 prevented synergistic activity, demonstrating interaction between PR co-factors (Liu et al., 2001). SRC-2 and SRC-3 have also been demonstrated to interact with, and enhance transcription by PR (Liu et al., 2001). Jun-dimerising protein (JDP)-2, interacts with the PR N-terminus and activates transcription independently of SRC-1. JDP-2 interacts with other co-activators, demonstrating multiple co-activator pathways for regulating PR activated transcription (Wardell et al., 2002). Additionally to recruitment of co-factors which communicate with other transcription factors, PR can interact directly with the transcriptional machinery, including transcription factor (TF)IIB which is involved in recruiting RNA polymerase to the transcriptional start site (Ing et al., 1992).

It has been demonstrated by mammalian two hybrid screening that PR-B interacts more strongly with co-activators SRC-1 and GRIP-1, and that PR-A interacts via ID with the transcriptional co-repressor SMRT at a much higher affinity than PR-B (Giangrande et al., 2000). Because of the competing functions of different co-factors, the ratio of co-activator/co-repressor expression within a cell modulates the response to ligand, providing a potential mechanism for regulating PR function (Liu et al., 2002). Agonist bound PR binds to co-activators such as SRC-1, whereas PR bound to antagonist RU486 binds both SRC-1 and the co-repressor SMRT (Liu et al., 2001).

For all nuclear steroid receptors distinct classes of co-factors exist, those which bind to the constitutive non-conserved N-terminal AF1, and those which interact with the highly conserved, ligand induced, AF2. Additionally, co-factors exist which can interact with both AF1 and AF2. AF-2 binding co-factors differ from AF1 co-factors in that they always possess a leucine rich interaction domain, LXXLL (Warnmark et al., 2003). In PR, AF3 has been shown by sequencing and mutation studies to possess two LXXLL motifs which are essential for function and may mediate intermolecular interactions by binding AF2 (Tung et al., 2001). Direct interaction between PR N-terminal and C-terminal constructs was demonstrated by *in vitro* pull down immobilisation assays, and higher affinity interactions were demonstrated for PR-B N-terminus (BN) than PR-A N-terminus (AN) (Tetel et al., 1999) which may in part be due to the LXXLL motifs present in AF3. The PR N-C terminal interactions were demonstrated by a mammalian two-hybrid assay, using AN/BN and C-terminal constructs, to be enhanced by the PR co-activators SRC-1 and CBP which may function as bridging factors between N-terminal AFs and AF2 (Tetel et al., 1999). Indirect interaction between distant AFs may occur via synergistic binding to different domains of the same co-factor, for example SRC-1 and TIF2 can act as bridging factors between AF1 and AF2 to enhance the activity of both domains (Onate et al., 1998).

1.1.7 B-upstream segment (BUS)

The higher affinity of co-factor binding to PR-B compared to PR-A, and the reverse pattern for co-repressor binding, implicated the PR-B specific AF3 in stabilising a conformation which masked ID function and facilitated the recruitment of co-activators (Giangrande et al., 2000). A role for the conformation of the PR N-terminal regions in mediating co-factor interactions was again suggested by assays to investigate the structure/function relationship of constitutively active N-terminal (NT) constructs; NT-A containing AF-1, ID, DBD and hinge, and NT-B containing BUS, AF-1, ID, DBD and hinge (Bain et al., 2001; Bain et al., 2000). It was demonstrated by protease digestion and peptide sequencing of the fragments that the presence of DBD conferred an organised structure upon AF-1 in both NT-A and NT-B, but that NT-A existed in a wider range of conformations than PR-B. It was suggested therefore that the presence of BUS may help to stabilize NT-B, and by extension natural PR-B, in transcriptionally active conformations, facilitating DNA binding, co-factor interactions with AF-1 and perhaps also intramolecular interactions (Bain et al., 2001; Bain et al., 2000). This idea gained support from a study which demonstrated that functional AF3 is required for interactions with co-factors, intramolecular interactions and transcriptional activation by PR-B or BUS containing constructs, but in PR-A these functions map to regions common to both receptors. Furthermore, mutation of AF3 does not result in a transcriptional activity similar to PR-A, leading to the conclusion that AF3 mediates the formation of intramolecular structures that promote more efficient co-factor binding leading to stronger transactivation by PR-B (Tung et al., 2006).

PR-B specific phosphorylation within BUS was demonstrated to have no effect on the ligand activated transcriptional activity of AF3, or its ability to synergise with AF1 or AF2 (Takimoto et al., 1996). However, more recent studies have demonstrated that increases the expression of endogenous PR target genes in response to progestin stimulation of T47-D breast cancer cells expressing only PR-B could be prevented by mutation of Ser81 within BUS. Cells expressing only PR-A, or PR negative cells did not exhibit progestin responsive transcription of these genes, indicating the role of BUS phosphorylation at Ser81 in regulating gene expression. Mutating Ser81 did not prevent recruitment to PRE, and therefore it was hypothesised that Ser81 phosphorylation may be required for correct co-activator recruitment and interactions (Hagan et al., 2011). Ser294, within the N-terminal region common to both PR-A and PR-B, was

demonstrated to be preferentially phosphorylated in response to ligand in PR-B over PR-A. Unique PR-B conformations due to BUS could alter the accessibility of Ser294 and thus the ability of kinases to phosphorylate at this site, contributing to the observed functional differences between the isoforms (Clemm et al., 2000). Phosphorylation of Ser294 has been linked to transcriptional hyperactivity in response to progesterone, which mediates ubiquitinylation and proteosomal degradation of PR, required for efficient transcriptional activity (Shen et al., 2001; Lange et al., 2000). Therefore, the increased phosphorylation of PR-B Ser294 could mediate hyperactivity of this isoform and increased receptor turnover leading to more efficient transcription compared to PR-A.

1.1.8 Variant PR isoforms

An N-terminally truncated third PR isoform, termed PR-C, was predicted based on the heterogenous nature of the 5′ region of the PR gene, and the abundance of mRNAs which arise, some of which are too short to encode PR-B or PR-A. It was predicted that translation would initiate at Met595 and encode an N-terminally truncated protein of 45-50kDa (Wei et al., 1990). Detection in PR positive breast cancer cells of a 60kDa protein by a C-terminal PR antibody, but not with N-terminal antibodies, targeted to epitopes which would be truncated from the Met595 encoded protein, and could bind to progesterone, supported the existence of PR-C (Wei and Miner, 1994). Cells transfected with PR-C expression vectors, in addition to PR-A/PR-B, were seen to exhibit increased transcriptional responses to progestins compared to cells with just PR-A/PR-B. This led to the hypothesis that PR-C may be able to dimerise with PR-A/PR-B and preferentially recruit co-factors to activate transcription of a PRE reporter gene (Wei et al., 1996). However, it was later demonstrated that whilst the protein encoded by PR-C cDNA could be co-immunoprecipitated using a PR-B specific antibody, suggestive of dimerisation, the presence of PR-C markedly diminished PR-B DNA binding. It was therefore suggested that PR-C may promote PR-A/PR-B activity in heterodimers in a mechanism independent of PRE binding, perhaps by preferential recruitment of co-factors which may themselves activate transcription (Wei et al., 1997). A potential function for PR-C was more recently suggested in the human myometrium at labour. An increase in PR-B is seen in the human myometrial fundus at labour, which is paradoxical since most species exhibit a withdrawal of progesterone signal at the onset of labour. However, a coinciding rise in cytoplasmic detection of PR-C could provide a

mechanism for repressing the function of PR-B in labour, leading to a functional progesterone withdrawal (Condon et al., 2006). A 60kDa cytoplasmic protein, suggested to be PR-C, was also detected in the human amnion, chorion and placenta, suggesting a role in regulating extra-embryonic tissue during pregnancy and parturition (Taylor et al., 2009). However, recent studies have demonstrated that the 60kDa protein identified as PR-C by Western immunoblotting with a C-terminal PR antibody may not be a PR protein (Madsen et al., 2007). The predicted molecular weight and natural occurrence of a protein encoded by PR-C mRNA has also been disputed, predicting a 38kDa protein rather than the previously described 60kDa isoform (Samalecos and Gellersen, 2008).

Mechanisms of alternative splicing, its impact on PR and the potential functionality of PR spliced variants will be discussed in detail later. Briefly, alternative splicing can result in the deletion of one or more exons from PR mRNA, which would either cause a change in reading frame or deletion of specific functional domains, altering PR function. Alternatively, retention of intronic sequences can result in the use of alternative translational initiation sites and generate truncated isoforms which again may display functions which differ from PR-A or PR-B, reviewed by (Cork et al., 2008).

1.1.9 Non-genomic PR signalling

Progesterone signals can be transmitted via PR in a rapid manner that does not involve nuclear translocation and PRE binding, typically occurring within minutes of a progesterone signal, rather than the much longer time scale required for induction of transcription and subsequent protein translation (Gellersen et al., 2009).

The Src/Ras/Raf/MAPK pathway is an important signal transduction pathway involved in transmitting signals from membrane receptors to mediate gene activation, and is involved in growth factor signalling. Ligand binding to a transmembrane receptor leads to Src tyrosine kinase activation, which in turn activates Ras, Raf, MEK and mitogen activated protein kinases (MAPKs) including extracellular signal-regulated kinase (Erk)-1/2 (Cargnello and Roux, 2011). Progesterone was initially demonstrated to activate Erk-1 and Erk-2, and this activation relied on the expression of PR. A similar

PR dependent activation of c-Src was observed, and inhibition of c-Src function blocked the progestin stimulated Erk-2 function (Migliaccio et al., 1998). This activation was demonstrated to be mediated via direct interaction of an N-terminal proline rich region of PR with the Src homology (SH)3 domain of the c-Src tyrosine kinase Hck. Mutation of this proline rich region inhibited interaction and activation (Boonyaratanakornkit et al., 2001). In COS-7 cells transiently transfected with PR-A or PR-B, only PR-B interacted with and activated c-Src, and deletion of NLS did not affect this activation suggesting an extranuclear interaction. The predominant nuclear localisation of PR-A compared to PR-B presumably prevents this cytoplasmic interaction of c-Src with PR-A (Boonyaratanakornkit et al., 2007).

Cross talk of PR with c-Src, and subsequent activation of downstream signalling cascades, provides a mechanism by which a progesterone signal can activate transcription of genes that do not contain a PRE (Boonyaratanakornkit et al., 2007; Proietti et al., 2005). Progesterone stimulation of T47D breast cancer cells enhanced the expression, phosphorylation and activation of the transcription factor signal transducer and activator of transcription (Stat)3. PR interacts directly with Stat3 and the activation of Stat3 in response to progesterone is reliant on c-Src function (Boonyaratanakornkit et al., 2007; Proietti et al., 2005). Potentially, the direct interaction of ligand bound PR with Stat3 mediates the recruitment of c-Src, stimulating Stat3 phosphorylation and activating transcription of non-PRE genes.

Progesterone has also been demonstrated to stimulate activation of the MAPK dependent transcription factor Elk-1 via the PR/c-Src interaction in a reporter gene assay. Progesterone was further able induce PR dependent transcription of the non-PRE gene encoding the cell cycle regulator cyclin-D1, leading to cell cycle induction in quiescent T47D cells. This activation was reliant on PR interaction with the SH3 domain of c-Src. Inhibitors of c-Src, MAPK or phosphatidylinositol 3-kinase (PI3K) attenuated progesterone responsive cyclin-D1 expression, implicating PR in activating multiple signal transduction cascades (Boonyaratanakornkit et al., 2007). Cytoplasmic PR can also activate PI3K, a signal transducer involved in activating transcription via protein kinase-C or regulating apoptosis via Akt activation. Progestin treatment of murine breast cancer cells transfected with PR-B resulted in Akt phosphorylation which

was not affected by transfection with a DBD mutant PR, demonstrating non-transcriptional regulation (Carnevale et al., 2007).

A biphasic pattern of Erk1/2 activation in response to progesterone has been observed in T47D cells expressing only PR-B by Western immunoblotting with a phospho-Erk1/2 antibody. The first phase is due to the rapid interaction of PR and c-Src described above, and the second more sustained period of activation involves PR induced transcription of Wnt-1, a non-PRE gene, and blocking transcription blocks Erk1/2 activation. Wnt-1 is thought to mediate epidermal growth factor receptor (EGFR) activation, which as described activates the c-Src/Erk signalling cascade, which stimulates cyclin-D1 expression. This provides a distinct mechanism for activation of cyclin-D1 transcription which does not require direct PR/c-Src interaction, and provides a link between rapid cytoplasmic and non-classical genomic PR actions (Faivre and Lange, 2007). Progesterone also stimulates a rapid PR-B dependent activation of EGFR in T47-D cells, detected by a phospho-specific antibody. This activation stimulates c-Src which phosphorylates PR at Ser345 (Faivre et al., 2008). Erk1/2 is also phosphorylated in response to c-Src which may account for the first phase of Erk-1 activation described above. Phosphorylation of PR Ser345 was demonstrated to be necessary for PR interaction with the transcription factor Sp1 which could mediate regulation of non-PRE genes via interaction with Sp1 sites within promoter regions (Faivre et al., 2008). This transcriptional PR activity, mediated via Sp1 could account for the processes mediating the second phase of Erk1/2 activation described previously. The progesterone induced cell cycle entry mediated via PR-B in quiescent T47-D breast cancer cells (Boonyaratanakornkit et al., 2007), was also seen in MCF-7 breast cancer cells; however, this activation of the cell cycle required a functional DBD, and mutation of the Src binding domain did not affect transcription of cyclin-D1 in this model suggesting a role for genomic PR signalling independent of Src/MAPK activation (McGowan et al., 2007).

In summary, ligand-bound PR-B is able to bind directly to Stat3 and c-Src. c-Src activates Stat3, activating transcription of non-PRE genes in response to progesterone. Alternatively PR-B/c-Src can activate Erk1/2 MAPKs, which phosphorylate transcription factors to activate non-PRE genes, including cyclin-D1 which promotes

cell cycle entry. Activated MAPKs can also phosphorylate PR Ser345, mediating binding to Sp1 to activate other non-PRE genes with Sp1 sites within their promoters, including EGFR. Wnts, produced in response to progesterone activated transcription of non-PRE genes, can activate EGFR to activate MAPK signalling and induce further transcription. Cyclin-D1 expression is activated both via rapid PR activation of MAPK signalling, and via progesterone induced Wnt activation of EGFR signalling. Cytoplasmic PR can also activate PI3K signalling to activate other intracellular signalling pathways. Some of the pathways leading to progesterone induced transcription of non-PRE genes are summarised in Figure 1.2.

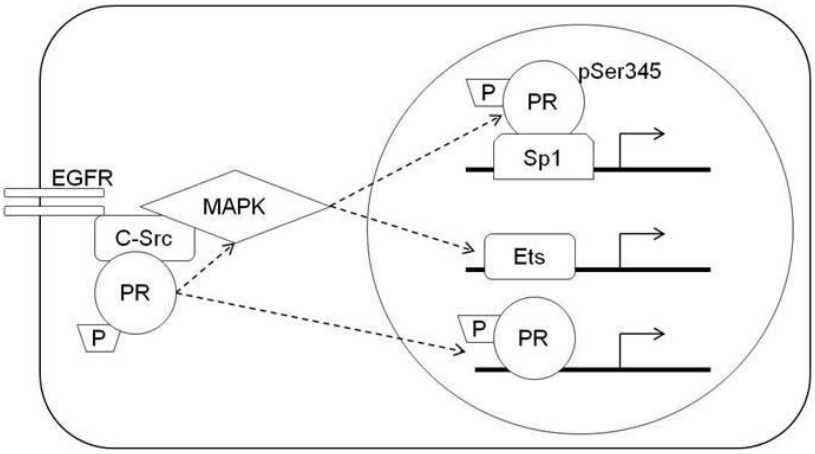


Figure 1.2: Mechanisms of PR signalling.

Ligand activated PR can interact with PRE promoters to regulate gene expression. Alternatively, PR can directly activate Src/MAPK signalling to promote transcription of non-PRE genes. Activated MAPKs can also phosphorylate PR Ser345 to promote interaction with Sp1 and transcription from Sp1 promoters (modified from (Faivre et al., 2008).

1.1.10 Ligand independent activation of PR

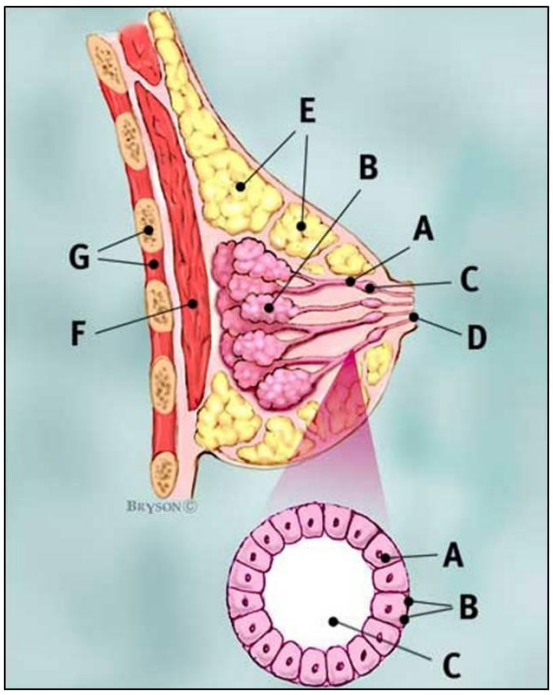
PR can also regulate gene expression in the absence of ligand; induced expression of transfected PR-A or PR-B in PR negative breast cancer cells led to altered gene expression without progesterone stimulation. In contrast to progesterone stimulated transcription, ligand independent transcription was mainly activated by PR-A (Jacobsen et al., 2005). The mechanisms of ligand independent PR activation remain unclear, however Ser400 can be phosphorylated by cyclin dependent kinase (CDK)2 in response to mitogens such as endothelial growth factor (EGF) and insulin-like growth factor (IGF). This phosphorylation stimulates nuclear PR localisation in HeLa cells, and increases transcription of a PRE reporter gene independent of progesterone stimulation (Pierson-Mullany and Lange, 2004). Therefore it is possible that phosphorylation by

CDK2, activated by other mitogenic signals, may be sufficient to induce the PR conformational changes necessary for transcriptional activation in the absence of ligand. This non-classical PR signalling may play a significant role in breast cancer prognosis and therapy in post menopausal women with PR positive tumours and low circulating hormone levels (Pierson-Mullany and Lange, 2004).

1.2 The mammary gland

1.2.1 Anatomy

The adult human breast is made up of 15-25 lobes, consisting of epithelial cell lined ducts which branch into secretory lobules to serve the function of lactation. Mammary ducts drain into lactiferous sinuses behind the nipple which store the milk until release is stimulated. The glandular structures of the breast are surrounded by fatty stromal tissue, providing a network of blood vessels, connective tissue and lymphatic vessels to support mammary gland function. The anatomy of the adult mammary gland is illustrated by the 'Breast profile' in Figure 1.3 (Hinck and Silberstein, 2005; Russo and Russo, 2004; Elston, 1998).



Breast profile:
A. Ducts
B. Lobules
C. Lactiferous sinus
D. Nipple
E. Adipose/stromal tissue
F. Pectoralis major muscle
G. Chest wall/rib cage

Enlargement
A. Normal ductal epithelial cells
B. Basement membrane
C. Lumen

Figure 1.3: Anatomy of the breast and organisation of ductal epithelial cells.
Figure adapted from (breastcancer.org).

1.2.2 Cellular organisation of mammary ducts

At the nipple, the duct consists of squamous epithelial cells which end before the lactiferous sinus. Beyond the lactiferous sinus, mammary ducts consist of a layer of columnar epithelial cells surrounding the central lumen, enclosed within a layer of myoepithelial cells and a collagenous basement membrane. The organisation of luminal epithelial cells within a mammary duct can be seen in the 'Enlargement' in Figure 1.3. These epithelial cells are contained within a layer of delimiting fibroblasts and capillaries within the stroma. In the lobuloalveolar structures, the epithelial cells take on a cuboidal morphology and the cytoplasm contains a high density of organelles associated with the secretory function (Rosen, 2001; Elston, 1998).

1.2.3 Normal mammary gland development

The newborn breast consists of primitive ductal structures which begin to differentiate under the control of steroid hormones at the onset of puberty, forming rounded structures termed terminal end buds (TEBs) (Hinck and Silberstein, 2005; Russo and Russo, 2004). TEBs branch into smaller rounded structures, alveolar buds, forming lobules type 1. Mammary lobular differentiation is illustrated in Figure 1.4. Briefly, the adult breast consists of three types of lobule, types 1, 2 and 3, each more differentiated than the previous type (i.e. containing more branched structures, termed ductules in lobule type 2 and 3). The transition from lobule type 1 to 2 and type 2 to 3 sees an increase in lobule size with a decrease in size of individual structures. If no pregnancy occurs, lobules de-differentiate to the preceding forms after the age of approximately 23 years (Russo and Russo, 2004). During pregnancy, a second major period of mammary development occurs, with further lobuloalveolar differentiation from lobules type 3 to the mature secretory lobules type 4 and the interlobular space is filled by mature secretory acini which serve the function of lactation (Russo and Russo, 2004; Elston, 1998). At the cessation of breast feeding and subsequent termination of lactation, lobules type 4 de-differentiate, however parous women retain a higher proportion of lobule type 3 and type 2 compared to nulliparous women (Russo and Russo, 2004).

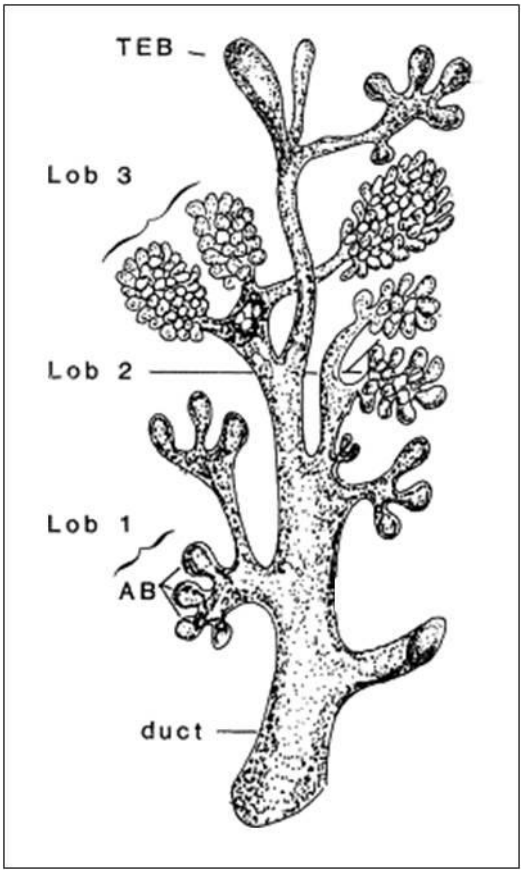


Figure 1.4: **Ductolobular differentiation of the breast**.

Mammary ducts differentiate at puberty to form terminal end buds (TEB) which branch to form alveolar buds (ABs) making up lobules type 1 (lob 1). Further differentiation occurs to lobules type 2 (lob 2) and lobules type 3 (lob 3), each larger and containing an increased number of more mature ductule structures compared to the preceeding lobule type. Figure taken (Russo and Russo, 2004).

1.2.4 Steroid hormones in mammary gland development

The link between mammary gland development and the menstrual cycle, and the role of the ovarian steroid hormones oestrogen and progesterone in mammary development are well established concepts (Topper and Freeman, 1980; Loeb and Hesselberg, 1917); ductal elongation predominates during the follicular phase and alveolar branching occurs during the luteal phase, when serum progesterone levels peak (Topper and Freeman, 1980).

More recent studies in rodents have provided insight into the differing roles of oestrogen and progesterone in mammary development; mice lacking the ERα gene (ER

knockout; ERKO) demonstrate impaired ductal elongation, while mice lacking the PR gene (PR knockout; PRKO) fail to undergo proper lobuloalveolar differentiation (Shyamala et al., 1999; Brisken et al., 1998; Lydon et al., 1995). This suggests that both ERα and PR are capable of stimulating epithelial proliferation, but only PR is capable of stimulating pregnancy associated lobuloalveolar differentiation. The involvement of PR in epithelial differentiation may implicate progesterone signalling in the differentiation associated with tumourigenesis (Shyamala et al., 1999).

In the murine mammary gland nuclear PR mRNA and protein were detectable only in a subpopulation of undifferentiated luminal epithelial cells. These PR positive cells were observed to be intermixed with rapidly dividing PR negative cells in the growing tip of an elongating duct (Silberstein et al., 1996). Cytoplasmic PR protein was also detected in a further subpopulation of luminal epithelial cells. During development, a switch was observed from predominant cytoplasmic expression in pre-pubertal glands, to predominant nuclear expression during early pregnancy, with a mixed population of cytoplasmic and nuclear PR positive cells in the non-pregnant adult gland. This led to the proposition that the cytoplasmic PR positive cells may represent a progenitor population for the mature secretory epithelia, or that PR positive cells may release paracrine factors to stimulate mitosis in nearby PR negative cells (Silberstein et al., 1996). This segregation of PR expression and proliferation in the murine mammary gland was also demonstrated using a model in which the *LacZ* reporter gene was inserted into PR exon 1. This model was able to identify where the PR promoter was active, demonstrating PR to be expressed only in luminal epithelial cells, and only in a subpopulation of cells which are non-proliferative. In the pubertal gland, PR expression was seen in ductal side branches and terminal end buds, whereas in the adult gland PR was concentrated at distal alveolar structures (Ismail et al., 2002).

The same segregation of PR expressing epithelial cells and proliferating epithelial cells exists in the human mammary gland (Russo et al., 1999; Clarke et al., 1997). Immunostaining for ERα and PR protein demonstrated that these receptors are expressed in a small sub-population of luminal epithelial cells in human mammary glands (Clarke et al., 1997) and are expressed at a higher level in the proliferating lobules type 1 with progressively decreased expression in the more differentiated

lobules type 2 and 3 (Russo et al., 1999). ERα and PR were seen to be co-expressed in 96% of steroid receptor positive cells and staining for the proliferation marker Ki67 demonstrated that these steroid receptor expressing cells were non-proliferative, again suggesting that hormone receptors signal in a paracrine manner so stimulate surrounding cells (Russo et al., 1999; Clarke et al., 1997). The paracrine function of PR in murine mammary development was demonstrated by the transplantation of a mixed population of PR positive and PR negative epithelial cells into cleared mammary fat pads leading to the development of lobuloalveolar structures containing PR negative cells. Since PR expression is required for this alveolar differentiation, this demonstrated that a paracrine signal from the PR positive cells stimulates the PR negative cells (Brisken et al., 1998).

It was later demonstrated that ectopic expression of Wnt-1 in PRKO mice was sufficient to rescue lobuloalveolar differentiation and that Wnt-1 released from mammary epithelial cells containing a *Wnt-1* transgene could activate lobuloalveolar differentiation involving surrounding wild-type cells (Brisken et al., 2000). These observations implicated Wnt signalling in the paracrine role attributed to PR in mammary development. Wnt-1 is not normally expressed in mammary tissue, but other Wnt family members are, including Wnt-4 (Brisken et al., 2000). Mammary buds from *Wnt-4* knockout mice were engrafted into cleared mammary fat pads of wild-type mice and failed to undergo normal pregnancy associated lobular differentiation. In wild-type animals Wnt-4 expression was observed specifically in PR expressing luminal epithelial cells, with increased Wnt-4 expression detected after progesterone treatment of the animals. This suggested a role for PR in the release of Wnt-4 and a physiological role for Wnt signalling as a paracrine mediator of progesterone stimulated lobuloalveolar differentiation (Brisken et al., 2000). As previously described in Section 1.1.9, *Wnt-1* is a non-PRE gene which can be activated in response to progesterone in a non-classical PR signalling dependent manner and leads to sustained Erk signalling to activate cyclin-D1 expression, promoting cell cycle entry in response to progesterone (Faivre and Lange, 2007). This non-classical PR signalling to activate Wnt expression provides a potential mechanism by which progesterone releases a PR dependent signal which can act on neighbouring cells to induce cell cycle entry leading to proliferation.

1.2.5 PR isoforms in mammary gland development

Transgenic mice overexpressing PR-A exhibit abnormal mammary gland development, including increased ductal side branching, hyperplasia and disruption of the basement membrane, demonstrating the importance of correct PR isoform balance in mammary gland development (Shyamala et al., 1998). Mice selectively lacking PR-A expression, PR-A knockout (PRAKO), exhibit normal lobuloalveolar differentiation (Mulac-Jericevic et al., 2000), whereas PR-B knockout (PRBKO) mice exhibited impaired lobuloalveolar differentiation (Mulac-Jericevic et al., 2003). Together these observations suggest PR-B is essential for mediating progesterone stimulated lobuloalveolar differentiation. Mice deficient for receptor activator of $NF_{\kappa}B$ ligand (RANKL) exhibit similar developmental abnormalities to those observed in PRBKO mice and mRNA for RANKL was detected to co-localise with PR mRNA in wild-type mice and PRAKO mice, and to be segregated from cyclin-D1 mRNA which was present in neighbouring cells. Furthermore, RANKL and cyclin-D1 mRNA levels were greatly reduced in PRBKO mice suggesting a role for PR-B in paracrine signalling via RANKL in the murine mammary gland (Mulac-Jericevic et al., 2003). Induced expression of RANKL in ER positive cells from PRKO murine mammary glands was sufficient to trigger development of mammary structures analogous to those seen in wild-type mice, confirming RANKL as an important mediator of PR induced mammary differentiation (Mukherjee et al., 2010). The role of PR-B in lobular development was supported by detection of PR-B protein expression specifically in the mammary gland of pregnant mice and not virgin adult glands. PR-A and PR-B were rarely detected in the same cell, and PR-B co-localised with cyclin-D1 in proliferative cells (Aupperlee et al., 2005). Further studies reported that PR-A also co-localised with cyclin-D1 in a subset of proliferative cells and mediated ductal side-branching in early pregnancy, whereas PR-B was again seen to more frequently co-localise with cyclin-D1 and to stimulate lobuloalveolar differentiation. These results suggested that progesterone can directly stimulate PR expressing cells and can exhibit paracrine control over surrounding PR negative cells (Aupperlee and Haslam, 2007).

In contrast to the segregated pattern of PR isoform expression described in the murine mammary gland, PR-A and PR-B are expressed at approximately equimolar levels in all PR positive human mammary epithelial cells (Mote et al., 2002). In the rat mammary gland, a similar pattern to the expression in human was observed, with co-expression of

PR-B in the majority of PR-A positive cells and no significant change in expression at different developmental stages (Kariagina et al., 2007). However, this study also detected a significant subpopulation of PR-B positive cells which did not express PR-A and these cells were detected to be the most proliferative of the PR expressing cells, particularly during pregnancy. All PR-B positive cells also expressed cyclin-D1, but the expression of cyclin dependent kinase inhibitors in cells which also express PR-A is likely to restrict cell cycle progression in these non-proliferative cells (Kariagina et al., 2007). Progesterone can stimulate cyclin-D1 expression via activating transcription of Wnt and RANKL which may mediate the paracrine stimulation of PR negative cells. However, as described in Section 1.1.9, a second mechanism by which progesterone can stimulate cyclin-D1 also exists, via rapid interaction of the PR-B proline rich domain with c-Src to activate MAPK signalling (Boonyaratanakornkit et al., 2007). This PR-B specific autocrine mechanism for activating cell cycle progression could be responsible for cyclin-D1 expression and proliferation of PR-B positive mammary epithelial cells.

The two distinct mechanisms described above for stimulating mammary epithelial cell proliferation, rapid activation of cyclin-D1 mediated mitosis in PR positive cells and a slower activation of PR negative cells, have recently been demonstrated in the murine mammary gland (Beleut et al., 2010). Firstly, it was shown by examining incorporation of 5′bromo-2′deoxyuridine (BrdU) into cells that progesterone was required in addition to oestrogen to stimulate proliferation in the adult gland. Initial proliferation of mammary epithelial cells, within 24 hours of hormone stimulation, showed co-localisation of BrdU with PR and cyclin-D1, resulting from autocrine activation of mitosis in response to progesterone. At later time points, most proliferative cells were PR negative and did not express cyclin-D1. This later phase of proliferation was dependent on RANKL expression, confirming RANKL as a paracrine mediator of PR signalling which functions independently of cyclin-D1 activation of the cell cycle (Beleut et al., 2010).

1.3 Breast cancer

1.3.1 Epidemiology

Breast cancer is now the most commonly diagnosed cancer in the UK with more than 47,000 new cases diagnosed in 2008 (CRUK). Whilst the incidence of breast cancer has risen in the last decade, this is in part attributable to increased awareness and improved detection. The overall survival rate for breast cancer patients has improved significantly in the last 40 years due to the screening programme and improved access to effective treatments. However, breast cancer was still responsible for over 12,000 deaths in 2008 demonstrating the need for continued research to improve the treatment of breast cancer patients (CRUK). The majority of breast cancer deaths occur in women over the age of 50 years, making age a major risk factor for disease development. Other risk factors include a family history of the disease, obesity, alcohol consumption and probably the use of exogenous hormones in hormone replacement therapy (HRT) and some forms of oral contraceptive (CRUK).

1.3.2 Pathology

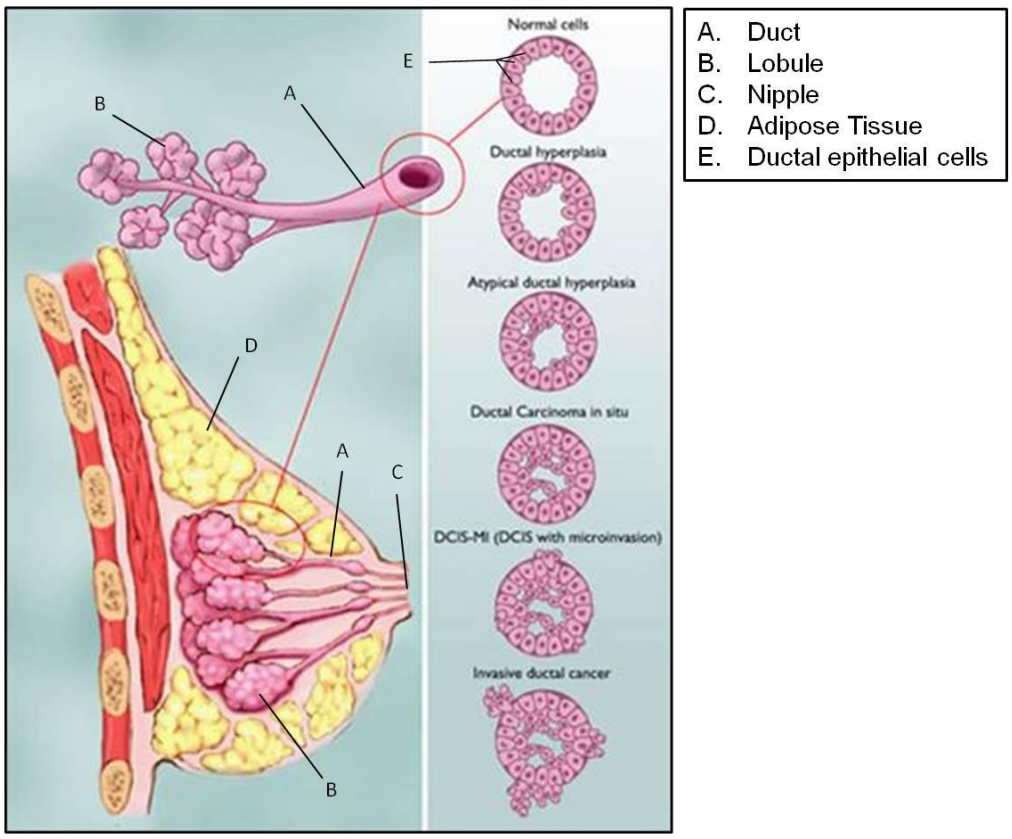
The majority of breast tumours arise from the rapidly dividing epithelial cells lining the ductal lumen. Pre-cancerous atypical ductal hyperplasia can occur and if untreated can progress to a tumour which is limited to the duct in which it originates; ductal carcinoma in situ (DCIS). These tumours can then progress to penetrate the basement membrane surrounding the duct and invade the stroma. Such invasive ductal carcinomas therefore have the potential to spread via the lymphatic system to axillary lymph nodes or other distant sites of metastasis. Ductal carcinomas are classified according to their invasive capacity and distinct patterns of growth and cellular morphology; tubular carcinomas, papillary carcinomas, medullary carcinomas, metaplastic carcinomas and mucinous carcinomas each possess distinct morphological features and are associated with differential prognosis. Tumours that possess some but not all of the features of an individual type of tumour are termed invasive ductal carcinoma: not otherwise specified (NOS). Invasive ductal carcinomas account for around 80% of the total breast cancer diagnoses (Bateman, 2006; Rosen, 2001). Malignant transformations of the lobular epithelium can also occur and again these tumours can be either restricted to the lobule (lobular carcinoma in situ) or invasive (Bateman, 2006; Rosen, 2001). Breast tumours

originating from squamous epithelia, myoepithelia or surrounding stroma can occur but are relatively rare (Bateman, 2006; Rosen, 2001).

Primary breast tumours usually metastasise first to the axillary lymph nodes following invasion from the duct/lobule into the lymphatic system in the breast stroma. Breast cancer cells can also invade blood vessels and metastasise via lymph and blood vessels to distant sites including the brain, liver, lungs and bone where they can form new macroscopic tumours (Bateman, 2006; Rosen, 2001). Breast tumour grading or subtyping can be used to assess prognosis independently and as part of a combined scoring system. Grading is based in the morphology of the cells; grade 1 are well differentiated, grade 2 moderately differentiated and grade 3 are poorly differentiated, with increased grade indicating poorer prognosis. The Nottingham Prognostic Index (NPI) considers the size of the tumour, number of lymph nodes affected and the grade of the tumour and is used to predict the disease prognosis (Bateman, 2006; Galea et al., 1992). Breast tumours can also be assessed using the TNM staging system. A score is given for each of the following disease characteristics: T; the size of the primary tumour, N; the presence of lymph node metastasis and M; the presence of breast cancer cells at distant metastatic sites. The scores for T, N and M are used to classify the stage of a tumour with stage I having the best prognosis and stage IV (the presence of distant metastases) having the poorest prognosis (Bateman, 2006; Rosen, 2001).

1.3.3 Cancer aetiology

Most breast tumours arise from transformation of the rapidly dividing luminal epithelial cells (Russo et al., 2005). As described in Section 1.2.3, progression to more differentiated structures during mammary gland development is associated with decreased cell proliferation. This is thought to contribute to the decreased incidence of breast cancer in parous women who possess fewer undifferentiated, proliferative structures following pregnancy and lactation (Russo et al., 2005). The development of breast cancer from a mammary duct is illustrated in Figure 1.5.



<u>Figure 1.5:</u> Development of ductal carcinoma *in-situ* and invasive ductal carcinoma.

The anatomical structures of the breast are shown in theleft hand panel. Right hand panel shows the organisation of epithelial cells in a mammary duct, developing into invasive ductal carcinoma. Figure adapted from (breastcancer.org(b)).

There are several important steps in the transformation of a normal cell into a cancer cell which are controlled by activating mutations or amplifications of oncogenes, or loss of function mutations in tumour suppressor genes (Lee and Muller, 2010; Croce, 2008). These processes were described by Hanahan and Weinberg as the hallmarks of cancer; a cell must acquire independence from external growth stimulatory signals, become non-responsive to normal anti-growth signals, avoid apoptosis and acquire the potential for unlimited division. These cells must then be supplied by an independent network of blood vessels (angiogenesis) and develop invasive and metastatic capabilities (Hanahan and Weinberg, 2000). To permit the acquisition of these hallmarks, changes within the cell must occur which prevent the normal monitoring of the genome, and increase genomic instability in order to allow the growth and proliferation of cells containing mutations and/or amplifications (Hanahan and Weinberg, 2011).

A tumour consists not only of the transformed cancer cells, but also a diversity of 'normal' cells which are co-opted by the tumour to support its growth (Hanahan and Weinberg, 2011). These include immune inflammatory cells, which may provide growth stimulating signals to cancer cells. Endothelial cells within the tumour microenvironment form the tumour neo-vasculature, supported by a population of pericytes which provide structural support and paracrine stimulation to the endothelial cells. Cancer-associated fibroblasts provide a structural foundation for tumour growth and promote angiogenesis, invasion and metastasis. As well as the rapidly dividing cancer cells, a tumour contains cancer stem cells which exhibit asymmetric division, to form either further cancer stem cells or rapidly dividing cancer cells. These cells are able to initiate development of new tumours when transplanted in mice and since they have mechanisms to evade cancer chemotherapeutic agents, may account for recurrence of disease after treatment and apparent clinical cure (Hanahan and Weinberg, 2011).

1.3.4 Breast cancer oncogenes and tumour suppressors

Somatic activating mutations in the *PIK3CA* oncogene are observed in approximately 25% of primary breast tumours (Lai et al., 2008; Bachman et al., 2004). *PIK3CA* encodes the catalytic subunit of PI3K which is a transducer of growth factor signalling. Constitutively active PI3K due to activation of *PIK3CA* would allow cancer cell growth independent of external stimulus (Lee and Muller, 2010; Dillon et al., 2007). PI3K signalling via Akt kinases has also been demonstrated to be involved in increased cell survival (avoidance of apoptosis), stimulating angiogenesis and increasing the invasive and metastatic potential of breast cancer cells (Dillon et al., 2007). Thus, the activation of the *PIK3CA* oncogene is likely to be vital in the acquisition of several of the hallmarks necessary for initiation and progression of breast cancer. The tumour suppressor phosphatase and tensin homologue (PTEN) functions to antagonise PI3K signalling, leading to repression of growth factor induced cell division, induction of apoptosis and repression of angiogenesis and invasion (Dillon et al., 2007). A high risk of breast cancer development is associated with loss of function germ-line mutations in the *PTEN* gene (Lee and Muller, 2010; Dillon et al., 2007). Another tumour suppressor gene associated with breast cancer is *TP53*, encoding p53. p53 controls the expression of numerous genes, including those involved in stimulating cell cycle checkpoints and response to pro-apoptotic signals. Therefore loss of p53 function would allow

uncontrolled cell cycle entry of cells with potentially harmful mutations (Lee and Muller, 2010).

Most breast tumours arise from somatic mutations and only approximately 1 in 10 has an inherited basis (CRUK). However, in addition to the rare germline mutations described above, inherited mutations in the tumour suppressor genes *BRCA1* and *BRCA2* are associated with an increased risk of developing breast cancer in affected families (Ford et al., 1998). *BRCA1* and *BRCA2* mutations account for a majority of the familial breast cancer risk (Venkitaraman, 2002; Ford et al., 1998). BRCA1 and BRCA2 are involved in maintaining chromosomal stability, and BRCA1 helps control cell cycle checkpoints following DNA damage (Venkitaraman, 2002). A loss of BRCA1 and BRCA2 function would clearly be involved in permitting the cancer associated changes described by Hanahan and Weinberg. BRCA1 is also involved in chromatin remodelling, transcription-coupled DNA repair and targeting aberrant RNA products for degradation, demonstrating the wide range of cellular processes which may be misregulated in *BRCA1/2* mutation carriers, contributing to cancer development (Venkitaraman, 2002).

The breast cancer associated oncogene, *ErbB2*, encodes the human epidermal growth factor receptor (HER)2 which has been seen to be overexpressed in 15-30% of breast cancers. HER2 overexpression is linked to an unfavourable prognosis, including increased lymph node metastases and increased recurrence (Burstein et al., 2007; Slamon et al., 1987). This receptor overexpression potentially makes breast cancer cells responsive to normally non-stimulating levels of hormone, thus contributing to their apparent autonomy (Slamon et al., 1987). HER2 functions via activation of the PI3K and MAPK signalling pathways that promote cell survival and proliferation. However, overexpression of HER2 normally stimulates apoptosis. Therefore, earlier genetic events that protect cells from apoptosis must be required to permit the effects of HER2 overexpression (Menard et al., 2003), in accordance with the necessity for enabling characteristics to allow the development of cancer hallmarks (Hanahan and Weinberg, 2011).

1.3.5 Ovarian hormones in the aetiology of breast cancer

Ovarian control of breast cancer development was first hypothesised by George Beatson in 1896 based on the known contribution of ovarian stimulus to lactation in animals, and observations of the similarities between normal mammary development towards lactation and epithelial proliferation in cancer. Surgical removal of the ovaries was demonstrated to halt progression and even lead to regression of inoperable breast cancers, confirming an element of ovarian control of disease (Beatson, 1983).

HRT is the administration of exogenous steroid hormones to alleviate the symptoms of menopause. Compared to women not receiving HRT mammary epithelial cell proliferation and density was increased in women receiving oestrogen only HRT. A further significant increase in cell proliferation was observed in women receiving combined oestrogen plus progestin HRT. These changes in cell growth patterns are associated with development of pre-cancerous hyperplasia and are concentrated in areas of the breast where most tumours arise (Hofseth et al., 1999). Parallel to these changes in cell growth, an increased incidence of breast cancer is seen in women who have undergone long term HRT (Beral, 2003); oestrogen only HRT increases breast cancer risk and a further increase in risk is associated with combined oestrogen plus progestin HRT, implicating both hormones in breast cancer initiation (Beral, 2003; Persson et al., 1999). Following an apparent initial protection from breast cancer, evidenced by a lower rate of disease within the first two years of HRT, combined oestrogen-progestin HRT was associated with an increased incidence after longer periods of treatment in a randomised, placebo-controlled trial (Rossouw et al., 2002). The risk of breast cancer associated with HRT was immediately decreased following treatment withdrawal (Chlebowski et al., 2009). Combined oestrogen and progestin HRT was also associated with an increased rate of breast cancer mortality in a follow up study of patients from this trial (Chlebowski et al., 2010). A recent reduction in use of combined HRT has seen a parallel decrease in breast cancer incidence in post-menopausal women (Silverman et al., 2011). Together these studies provide strong epidemiological evidence for the involvement of both oestrogen and progesterone signalling in breast cancer initiation.

1.3.6 ER and PR in breast cancer development

Evidence for an involvement of ER signalling in breast cancer initiation comes from the observation of a disruption in the normal pattern of ER expression in the early stages of breast cancer (Clarke et al., 1997). In the normal adult human and rodent mammary gland, ER and PR are largely segregated from the proliferative epithelial cells and hormone induced proliferation is stimulated in a paracrine manner. However in breast cancer a loss of this segregated expression is observed, suggesting a switch from paracrine to autocrine stimulation of mammary epithelial cell proliferation (Clarke et al., 1997). In a model using age matched virgin rats, stimulation with oestrogen and progesterone to mimic parity prevents the development of a hormone receptor positive proliferating epithelial population in response to a chemical carcinogen, which is detected in age control unstimulated animals (Sivaraman et al., 2001b). The protection against breast cancer development associated with early parity is linked to a specific gene expression profile and the decreased proliferative nature of epithelial cells within the structures of the involuted parous gland compared to the proliferative lobules type 1 and lobules type 2 of the virgin gland (Russo et al., 2005). It is suggested therefore, that the proliferative hormone receptor positive population is a target for malignant transformation in the virgin gland. The absence of this population in the parous gland may contribute to this protection (Sivaraman et al., 2001b). Furthermore, rodent models of parity induced breast tumour protection have revealed that several mediators of epithelial proliferation are down regulated, and the tumour suppressor p53 is activated prior to carcinogenic insult in the parous gland (Blakely et al., 2006; Sivaraman et al., 2001a).

The role of progesterone/PR signalling in breast cancer development is less well established than that assigned to oestrogen/ER. To support the epidemiological data suggesting a role for progesterone in breast cancer initiation, described in Section 1.3.4, recent experimental evidence implicates PR in breast cancer development. The use of murine models to dissect the individual contributions of ER and PR signalling to mammary gland development is discussed in Section 1.2. Using the same PRKO mouse model, the importance of progesterone signalling in mammary tumourigenesis was identified; PRKO mice showed a lower incidence of breast tumour development following treatment with a chemical carcinogen, 15% compared to 60% in wild-type mice (Lydon et al., 1999). Culturing of isolated mammary glands from wild-type and

PRKO mice again demonstrated a lack of response to chemical carcinogenesis in the absence of PR, suggesting that the proliferating structures stimulated by progesterone signalling may be the targets for carcinogenesis (Chatterton et al., 2002). Treating mice with ER+/PR+ tumours with an antisense oligonucleotide targeting PR, which decreased both PR-A and PR-B expression, led to decreased proliferation, increased apoptosis and regression of the tumour compared to untreated or control treated animals. The effect was similar to that seen with antiprogestin treatment, supporting a role for PR in the development and progression of breast cancer (Lamb et al., 2005).

1.3.7 Breast cancer therapies

Localised breast cancer is treated by removal of the tumour mass by surgery. This treatment is frequently performed in combination with radiation, adjuvant chemotherapy or endocrine therapy to manage the risk of relapse. Systemic therapies such as chemotherapy and/or endocrine therapy are also used to treat metastatic disease (Makhoul and Kiwan, 2011; Alvarez, 2010; Conzen, 2008).

Chemotherapeutic drugs affect many pathways of cell growth including cell division and DNA synthesis, therefore targeting rapidly dividing cancer cells to a greater extent than normal mammary epithelial cells (MEC). Breast cancer chemotherapeutics include the alkylating agent cyclophosphamide which is often administered in combination with the DNA intercalating antineoplastic drug doxorubicin. These drugs are sometimes administered along with a taxane such as paclitaxel which blocks mitosis, or the antimetabolite fluorouracil which prevents pyrimidine incorporation during DNA synthesis (Berry et al., 2006).

Endocrine therapies used in breast cancer include the targeting of oestrogen/ER signalling to block its mitogenic effect on breast cancer cells. Oestrogen is secreted from the ovaries in pre-menopausal women, in response to luteinising hormone releasing hormone (LHRH). Thus LHRH antagonists can effectively block oestrogen secretion in pre-menopausal patients. Aromatase inhibitors (AIs) are used in post-menopausal patients to block the synthesis of oestrogen from adrenal androgens by aromatase enzymes in breast tissue and peripheral fat. Selective ER modulators

(SERMs), such as the ER antagonist Tamoxifen, are used to block the action of ER in both pre- and post-menopausal patients (Conzen, 2008).

1.3.8 Hormone receptors as breast cancer biomarkers

The assignment of these endocrine therapies is based on the use of ER as a biomarker, i.e. ER+ patients are prescribed the treatments. The ER status of a tumour can also be of significance in predicting the outcome of chemotherapy. Meta-analysis of data from large clinical trials demonstrated that the outcome of a range of adjuvant chemotherapeutic regimes was significantly less for ER+ patients; only a 7% increase in 5 year disease free survival for ER+ patients over patients not receiving chemotherapy was seen, compared to a 22.8% increase for ER- patients (Berry et al., 2006).

Further predictive information for the effectiveness of endocrine therapies can be gained from measurement of the PR status of a tumour. Analysis of survival data from two large patient cohorts, one containing 5427 patients and the other containing 10444 patients, compared patients not receiving any adjuvant endocrine therapy with patients receiving the SERM tamoxifen. A greater relative risk of recurrence was observed in patients with ER+/PR- tumours than ER+/PR+ tumours. The predictive value of PR was shown to be independent of the quantitated level of ER present in a tumour, suggesting that the difference between response in PR+ and PR- patients is not merely due to PR being a surrogate indicator of ER function (Bardou et al., 2003). A clinical trial comparing the AI anastrazole with tamoxifen or a combination of the two drugs, the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial, demonstrated that 7.6% of ER+/PR+ patients receiving the combination therapy suffered relapse compared to 14.8% of ER+/PR- patients. A similar pattern was also evident in the tamoxifen arm of the trial, but not the anastrazole arm, suggesting that the reduced effectiveness of endocrine therapies in PR- patients was restricted to SERMs (Cui et al., 2005). However, analysis of the time to failure of AI treatment for advanced, metastatic breast cancer in 146 ER+ patients revealed an increased time in patients who were also PR+ indicating a potential role for PR status in the assignment of AI therapies and in considering combination therapies containing AIs (Anderson et al., 2011).

As well as serving as predictive markers for the outcome of endocrine therapies, hormone receptor status can also be used in the assessment of disease prognosis. Patients with ER+ tumours have better overall survival than ER- patients. Different metastatic sites are common in patients with different ER status of the primary tumour; bone and soft tissue in ER+, liver and brain in ER-, with the latter associated with worse prognosis (Osborne, 1998). ER+/PR- status is associated with increased metastasis in post-menopausal patients (Balleine et al., 1999a). PR status has also been shown to be inversely associated with tumour stage, i.e. PR expression is a marker for early stage breast cancer but expression is not commonly seen in later stage tumours. This could indicate a loss of PR expression during tumour development but could also indicate the more aggressive nature of PR- primary tumours, making them more likely to metastasise (Coyle et al., 2007). Analysis of the gene expression profile of breast tumours demonstrates an up-regulation of genes involved in cancer associated cell signalling pathways such as the PI3K/Akt pathway in PR- patients compared to PR+, providing a potential mechanism for the increased growth and survival of these cancer cells and thus the poorer prognosis of PR- tumours (Creighton et al., 2009). In a retrospective study, patients with early stage breast tumours which were only positive for one hormone receptor, ER+/PR- or ER-/PR+, were seen to be more likely to suffer local recurrence following surgery than ER+/PR+ patients, demonstrating the prognostic value of both receptors (Albert et al., 2011).

The role of progesterone signalling via PR as a mammary mitogen, discussed in Section 1.2, and therefore its potential contribution to the growth of breast cancer cells makes PR a potential therapeutic target. Mifepristone is a selective PR modulator (SPRM) which has been demonstrated to inhibit the growth of PR+ breast cancer cell lines and to inhibit tumour growth in animal models. Mifepristone has also been reported to be beneficial in preventing tumour progression in patients, especially when administered in combination with a SERM (Im and Appleman, 2010). Three small scale clinical trials reported a positive response to mifepristone treatment in approximately 10-12% of post-menopausal breast cancer patients, either before or after other forms of treatment (Im and Appleman, 2010). A rat model of breast cancer development following chemical carcinogenesis demonstrated a decrease in tumour size and inhibited growth of new tumours following treatment with mifepristone. The same study demonstrated an even greater decrease in tumour growth following treatment with another antiprogestin CDB-

4124, with decreased cell proliferation and increased apoptosis compared to mifepristone (Wiehle et al., 2007). The experimental and clinical data described above point to the potential benefits of SPRMs as breast cancer therapeutics in hormone receptor positive disease and SPRMs could improve response when co-administered with current therapies such as SERMs (Im and Appleman, 2010). Due to the predictive and prognostic value of PR and its potential as a therapeutic target, accurate assessment of PR status is vital in the clinical management of breast cancer patients.

The hormone receptor status of breast tumours is usually measured by immunohistochemistry using specific monoclonal antibodies. ERα protein expression and total PR (PR-A and PR-B) protein expression are both measured using antibodies targeted to N-terminal epitopes. The level of protein expression is quantified either visually or more usually by an automated process by which the intensity of staining and the number of receptor positive cells is accurately quantified to give a score of 0-8 (Hanley et al., 2009; Payne et al., 2008). Approximately 70% of primary breast tumours are reported to be ER+, more than half of which are also PR+. Of the remaining 30%, the majority are ER-/PR-, however a subset (<5%) of tumours appear ER-/PR+ (Osborne, 1998). Development of new methods to identify biomarker expression in breast cancer is ongoing. One such approach is the use of quantitative real-time PCR based assays, with the aim of more accurately and more quickly classifying biomarker expression (Iverson et al., 2009). Although endocrine therapies typically target ER signalling, patients with tumours characterised as ER-/PR+ respond more favourably to these therapies than ER-/PR- patients, indicating the importance of PR expression to the treatment of breast cancer (Osborne, 1998). A recent retrospective study of 5053 patients, analysing the occurrence of ER-/PR+ tumours, which represented approximately 3% of all tumours examined, reported an increased occurrence in patients over 51 years of age. This non-random distribution suggests that this phenotype is not an artificial result of false negative ER status measurements and may be clinically relevant (Rhodes and Jasani, 2009).

As mentioned above, HER2 is a breast cancer associated oncogene which is overexpressed in approximately 20-30% of breast tumours and these tumours tend to have a poor prognosis with a high rate of recurrence and mortality (Slamon et al., 1987).

HER2+ status has recently been demonstrated to be a associated with microinvasion in patients with DCIS, and is therefore of potential importance in the transition from DCIS to invasive ductal carcinoma (Liao et al., 2011). The detection of HER2 in breast cancer and its role as a biomarker are reviewed by (Payne et al., 2008). Briefly, the pathological detection of HER2 in breast cancer is carried out by immunohistochemistry, fluorescence in situ hybridisation or chromogenic in situ hybridisation. HER2 is a predictive biomarker for the potential use of trastuzumab (Herceptin), a humanized monoclonal antibody which targets HER2 protein, preventing its function and stimulating a decrease in HER2 expression, improving the outcome of concurrently administered chemotherapy compared to chemotherapy alone. A combined HER2/EGFR tyrosine kinase inhibitor has also recently been shown to be effective in HER2+ breast cancer. HER2+ tumours may also respond more favourably to some chemotherapeutics and predict resistance to endocrine therapies (Payne et al., 2008).

1.3.9 Molecular subtypes for breast cancer management

Traditional molecular biomarkers can distinguish between tumours with different characteristics, however the heterogeneity of breast tumours means that biomarker expression is not always accurately predictive of prognosis or therapeutic outcome. Advances in molecular biological techniques have enabled the analysis of the gene expression profiles of breast tumours. This can then be used to classify the tumour according to expression of key groups of genes associated with prognosis and response to therapy. Expression levels of genes associated with luminal epithelial cells, such as ER, PR, cytokeratin 8/18, and ER regulated genes define the Luminal subtypes; Luminal A tumours highly express ER/ER regulated genes and have low expression of proliferation related genes such as Ki67, whereas Luminal B tumours have much lower ER/ER regulated gene levels and higher proliferation associated gene expression. Tumours with high HER2 gene expression and low ER/ER regulated gene expression are classified into the HER2 subtype, while tumours negative for HER2 and ER but enriched in genes associated with basal epithelial/myoepithelial cells and proliferation are classified as Basal-like. The Normal-like breast tumour sub-type was defined by a gene expression profile resembling non-epithelial normal breast tissue (Sorlie et al., 2003; Sorlie et al., 2001; Perou et al., 2000). A further subtype, Claudin-low tumours, was more recently suggested, defined as distinct subset of basal-like tumours with reduced expression of genes associated with cell-cell adhesion, such as claudins,

occluding and E-cadherin an increase in genes associated with epithelial-mesenchymal transition and stem cell associated genes (Hennessy et al., 2009). Mutations in the *TP53* tumour suppressor gene are detected at a much higher frequency in luminal subtype B, *HER2* and basal tumours than in luminal subtype A. Additionally, mutations in *BRCA1* were detected specifically in basal tumours, linking *BRCA1* mutations to ER- tumours and poor disease prognosis (Sorlie, 2004). The molecular subtypes described above are summarised in Table 1.1.

Sub-type	Characteristics
Luminal A	Gene expression similar to luminal epithelial cells; enriched for ER and ER regulated genes; Ki67low
Luminal B	Gene expression similar to luminal epithelial cells; reduced expression of ER regulated genes; Ki67high
HER2	HER2 overexpression; gene expression similar to luminal epithelial cells
Basal-like	ER negative; gene expression similar to basal epithalial/myoepithelial cells
Claudin-low	Reduced expression of cell adhesion related genes; enriched for stem cell associated genes
Normal-like	Gene expression similar to non-epithelial normal breast tissue

Table 1.1: Molecular subtypes of breast cancer
Gene expression characteristics defining the molecular sub-types of breast cancer.

These molecular subtypes have been linked to prognostic features of breast tumours (Dunbier et al., 2011; Calza et al., 2006). Luminal A were typically small, low grade tumours which responded favourably to endocrine therapy as expected of ER+ tumours with low levels of proliferation. Luminal B had a poorer disease free survival and poorer response to endocrine therapies than Luminal A, consistent with the lower ER expression predicted from the gene expression profiling. Tumours in the HER2 subtype were usually large, high grade tumours which were ER- and consequently showed limited response to endocrine therapy and had a poor overall prognosis. Basal-like tumours were also high grade and were associated with lymph node metastases, however they were associated with better disease free survival than the HER2 subtype. Tumours classified in the Normal-like subtype were small, low grade and responded well to endocrine therapies (Calza et al., 2006). These observations were supported by a recent study which identified a similar initial response to AI therapy between Luminal

A and Luminal B, however high residual Ki67 staining in Luminal B tumours after 2 weeks of treatment were indicative of a poorer response to therapy. HER2 and Basal-like tumours showed little response to initial AI therapy while Normal-like tumours responded well (Dunbier et al., 2011).

As described, the most favourable prognosis is found in tumours which are ER+, PR+ and HER2-, since these tumours are the least proliferative and likely to respond well to endocrine therapies (Calza et al., 2006). ER+/PR- disease is associated with a worse prognosis and poorer response to endocrine therapies (Dowsett et al., 2005; Balleine et al., 1999a). HER2+ tumours are associated with a poor prognosis but can be effectively treated with Herceptin (Payne et al., 2008). Tumours which are ER-/PR- and HER2-, termed 'triple negative', mostly characterised in the Basal-like molecular sub-type, have been shown by retrospective analysis of breast cancer patients to be associated with poor disease free survival, increased lymphatic invasion and more lymph node metastasis (Mersin et al., 2008; Calza et al., 2006). Analysis of the molecular phenotype of tumours is in its early stages but may play a significant role in the future to accurately predict the characteristics of a tumour and assign appropriate therapies, as well as revealing potential novel therapeutic targets. However, the prognostic and predictive value of ER, PR and HER2 is clearly of great importance. The role of hormone receptors in breast cancer development and progression, particularly the role of PR, will therefore be addressed and is the focus of this thesis.

1.3.10 PR isoforms in breast cancer

The ratio of PR isoform expression is altered early in breast cancer development; in the normal mammary gland PR-A and PR-B are expressed in the same cells at approximately equimolar levels (Graham et al., 1995), whereas in pre-cancerous atypical hyperplasia and DCIS, as well as in invasive breast cancers, a loss of this homogenous PR isoform expression is observed, with a predominance of PR-A over PR-B (Mote et al., 2002), which is due to a reduction in the level of PR-B expression (Mote et al., 2002; Graham et al., 1995). Analysis of gene expression in PR positive T47D breast cancer cells demonstrated that a number of genes became progestin responsive following induced over-expression of PR-A. Many of these genes were involved in cell morphology and adhesion, and cells after PR-A induction showed

decreased adhesion in mono-layer culture and altered focal adhesion protein expression, suggesting an influence on cell mobility (Graham et al., 2005). The altered growth in response to a change in PR isoform expression has been demonstrated using transgenic mice overexpressing PR-A which exhibit a hyperplastic phenotype similar to the early stages of breast cancer development (Shyamala et al., 1998). PR-A transgenic mice contain a population of mammary epithelial cells undergoing excessive proliferation which express elevated levels of cyclin-D1 and decreased expression of p21, a cyclin dependent kinase inhibitor which prevents cyclin-D1 mediated activation of the cell cycle (Chou et al., 2003). Transforming growth factor beta-1 (TGF)β1 maintains MEC quiescence in the adult gland, and decreased TGFβ1 expression is seen in PR-A transgenic mice facilitating the increased proliferation. Increased matrix metalloprotease activity is also observed in PR-A transgenic mice and is associated with decreased laminin and collagen expression and breakdown of the basement membrane, another feature of cancer development (Simian et al., 2009). Inhibiting ER signalling in this model using a synthetic antagonist decreased proliferation and hyperplasia, whereas a PR antagonist had no effect. Oviarectomy of PR-A transgenic mice, to remove both oestrogen and progesterone stimulation, led to decreased proliferation. Co-administration of oestrogen and progesterone to intact PR-A transgenic mice stimulated a much greater incidence of hyperplasia compared to either hormone individually, implicating both hormones, and the ratio of PR isoform expression in regulating mammary epithelial proliferation (Simian et al., 2009). Post menopausal women treated with tamoxifen showed an increased rate of relapse if they exhibited a predominance of PR-A over PR-B relative to other PR+ patients. Thus identification of the PR-A:PR-B isoform ratio may add value to PR status as a predictive marker for use of endocrine therapies in ER+/PR+ breast cancer (Hopp et al., 2004).

1.3.11 PR signalling in breast cancer

A study using DNA microarray screening of changes in gene expression levels in T47D breast cancer cells following progesterone treatment identified several progesterone responsive genes encoding proteins associated with the processes of breast cancer development or progression (Richer et al., 2002). PR signalling via PR-B has been demonstrated *in vitro* to activate transcription of Wnt-1 which mediates EGFR activation, in turn activating the Src/Erk signalling pathway to increase cyclin-D1 expression. This progesterone induced, PR-B dependent, activation of non-PRE

signalling was capable of inducing colony formation of T47D breast cancer cells in 3D culture, a feature analogous to tumour formation, and is a potential mechanism for PR+ breast tumour growth (Faivre and Lange, 2007). Decreased Erk1 and Erk2 phosphorylation was observed in mammary tumours of mice following treatment with mifepristone or antisense oligonucleotides targeted to knockdown PR expression. Upon treatment, regression of the tumours was observed implicating this non-genomic PR signalling pathway in the maintenance of mammary tumours (Lamb et al., 2005). Direct interaction of PR with c-Src has been demonstrated *in vitro* (Boonyaratanakornkit et al., 2001) and this interaction was demonstrated to activate MAPK and PI3K (Boonyaratanakornkit et al., 2007; Carnevale et al., 2007). Progestin treatment of mice injected with murine breast cancer cells transfected with PR-B resulted in formation of lung metastases. However mutation of the N-terminal Src binding domain of the transfected PR resulted in a reduction of progestin induced lung metastases, demonstrating the importance of PR as a non-genomic signal transducer, rather than as a transcription factor, in the development of a metastatic phenotype in PR+ breast cancer (Carnevale et al., 2007).

During normal mammary gland development in a rat model, a population of proliferating PR-B positive cells was detected during pregnancy, expressing cyclin-D1 (Kariagina et al., 2007). It is proposed that hormone receptor positive MECs may function as a stem cell population, since they are usually quiescent and show the ability to differentiate upon the correct stimulus (Clarke, 2005). Therefore, these long living cells may be the targets of accumulated mutations necessary for tumourigenesis (Faivre and Lange, 2007). Consistent with this hypothesis is the suggestion that parity induces a shift from a transformation susceptible stem cell 1, to a transformation refractory stem cell 2, meaning that earlier parity would give less time for accumulation of mutations which can lead to transformation (Russo et al., 2005). It was therefore suggested that an accumulation of transformation inducing mutations in this PR+ MEC population, such as a loss of p53, p21 or p27 expression can re-activate proliferation, contributing to development of PR positive breast tumours (Faivre and Lange, 2007).

1.3.12 PR and growth factor signalling cross-talk in breast cancer

PR phosphorylation at Ser348 was demonstrated to be dependent on functional EGFR and leads to increased PR transcriptional activity at non-PRE promoters (Faivre et al., 2008). EGFR activation also induced the phosphorylation of PR Ser294 which induced hypersensitivity to ligand, increasing ligand dependent down-regulation of PR protein levels. PR Ser294 phosphorylation was necessary for efficient cell proliferation following EGFR activation by EGF (Daniel et al., 2007). Overactivation of EGFR would therefore potentially cause hyperproliferation of PR+ cells. TGF-α is an EGFR ligand and TGF-α transgenic mice are susceptible to the formation of ER+/PR+ mammary hyperplasia, which develop into ER+/PR- tumours, perhaps due to growth factor induced PR hypersensitivity and ligand-dependent PR downregulation (Rose-Hellekant et al., 2007).

HER2 is closely related to EGFR, and HER2 signalling is mediated in part via the Stat family of transcription factors (Yarden, 2001). As previously described, progesterone activated PR can interact directly with Stat3, leading to activation of Stat3 regulated gene expression (Proietti et al., 2005). Heregulin is a HER2 ligand which transactivates PR independent of progesterone by phosphorylation of PR Ser294 through activation of downstream MAPKs. Blocking PR activity prevents Stat3 activation by heregulin which is required for heregulin induced cell proliferation (Proietti et al., 2009).

Fibroblast growth factor receptors (FGFR) are another family of cell surface receptors with tyrosine kinase activity which activate intracellular signal transduction pathways, activating transcription by Stat5. Co-immunoprecipitation demonstrated that both progestin and fibroblast growth factor (FGF)2 stimulated formation of complexes containing PR along with both FGFR2 and Stat5, and immunofluorescent analysis showed nuclear co-localisation of these proteins. These complexes were able to interact with PRE of endogenous genes and increase gene expression in response to FGF2 and PR. Blocking PR activity blocked FGF2 stimulated proliferation of T47D breast cancer cells transfected with FGFR2. T47D cells normally require exogenous hormone stimulation to form tumours when transplanted into mice, however transfection with constitutively active FGFR2 led to hormone independent tumour formation. The nuclear co-localisation of PR with FGFR2 and Stat5 was also reported in some human breast

tumour samples (Cerliani et al., 2011). This cross talk between PR and growth factor signalling could provide a mechanism by which progesterone signalling via PR contributes to the development of breast tumours, and also may implicate PR in enabling the development of hormone independent tumours which grow in response to growth factors secreted by other cells within the tumour microenvironment such as cancer associated fibroblasts.

1.3.13 Ligand independent PR signalling in breast cancer

CDKs contribute to the regulation of cell cycle progression and are frequently activated in cancer due to alterations in tumour suppressor and oncogene expression (Caldon et al., 2010). CDK2 is capable of phosphorylating PR Ser400 and activating transcription of PRE genes independent of progesterone (Pierson-Mullany and Lange, 2004). Gene expression levels were assessed in PR negative T47D-Y cells and compared to T47D cells specifically expressing either PR-A or PR-B in either the presence or absence of external progesterone stimulation. In this system PR-A was demonstrated to be the predominant regulator of transcription in the absence of progesterone treatment compared to PR-B (Jacobsen et al., 2005). The genes regulated in the absence of progesterone treatment were involved in many aspects of tumour cell biology including adhesion, motility, migration, signal transduction, cell cycle progression and apoptosis. ER+/PR-A+ T47D cells exhibit increased adhesion to extra cellular matrix when cultured on matrigel or collagen, and increased migration towards a chemoattractant through transwells coated with collagen compared to ER+/PR-B+ or ER+/PR- T47D cells, characteristics indicative of an aggressive tumour phenotype (Jacobsen et al., 2005). These ligand independent, PR-A mediated effects on cell behaviour may contribute to the poorer outcome of tamoxifen therapy in post-menopausal patients with an overexpression of PR-A relative to PR-B (Hopp et al., 2004).

Phosphorylation of the PR-B specific Ser81 has been linked to the activation of PR-B responsive genes including the breast cancer proliferation associated gene *HSD11β2*. Basal gene activation required Ser81 phosphorylation, as did response to ligand. Ser81 phosphorylation is mediated by casein-kinase (ck)2, and ck2 is widely upregulated in breast cancer. Therefore, this ck2 overexpression could lead to ligand independent

Ser81 phosphorylation and activation of transcription, contributing to hormone independent growth of receptor positive breast tumours (Hagan et al., 2011).

1.3.14 Interaction of PR with BRCA1/2

BRCA1/2 mutations are the most common inherited mutations associated with breast cancer development, and *BRCA1/2* mutations have been shown to disrupt the normal PR expression ratio leading to a predominance of PR-A over PR-B (Mote et al., 2004). Transfection of BRCA1 into PR+ breast cancer cells reduced progestin induced luciferase reporter gene activity and a direct interaction of BRCA1 with both PR-A and PR-B was shown by co-immunoprecipitation (Ma et al., 2006). Further studies used gel shift assays to demonstrate that BRCA1 inhibits PR action by blocking PR binding to PRE, and chromatin immunoprecipitation demonstrated that BRCA1 blocked recruitment of PR co-activators and promoted recruitment of co-repressors to PRE (Katiyar et al., 2009). Mice containing a *BRCA1* mutation encoding a functionally deficient protein undergo increased MEC proliferation in response to progesterone compared to wild-type mice, demonstrating that mutating *BRCA1* removes control over PR signalling (Ma et al., 2006). Inactivation of both *BRCA1* and *p53* in mice, to mimic the genetic deficiencies in human *BRCA1* mutation associated tumours, resulted in increased progesterone induced lobular proliferation and increased tumour formation compared to wild-type mice. Treatment with mifepristone completely blocked tumour formation, demonstrating the contribution of PR signalling to breast cancer development in BRCA1 deficient mice (Poole et al., 2006). These results support the use of anti-progestins in *BRCA1* mutation carriers as prophylactics in an attempt to prevent breast cancer development, or in treating PR+ tumours in pre-menopausal *BRCA1* mutation carriers (Katiyar et al., 2006).

1.3.15 Potential mechanisms of PR loss in breast cancer

Since PR is an oestrogen responsive gene (Kastner et al., 1990), it has been suggested that loss of PR expression could be due to the low circulating oestrogen levels in post-menopausal women being insufficient to stimulate PR expression, or that a non-functional ER may be expressed in breast cancer which may be unable to bind DNA and activate transcription. However, ER from fresh tumour biopsies has been demonstrated

to bind DNA, making this an inadequate explanation of PR loss (Cui et al., 2005). Blocking ER activity with endocrine therapeutics may prevent PR transcription and tumours which initially screen PR+ often appear PR- on subsequent analysis following tamoxifen treatment (Gross et al., 1984).

PR- primary breast tumours can develop, for example by the mechanisms described in the TGF-α transgenic mouse model, via cross talk of PR with overactive growth factor signalling pathways hypersensitising PR to progesterone and leading to ligand-dependent down-regulation (Rose-Hellekant et al., 2007). Indeed HER2 overexpression in breast cancer is more common in ER+/PR- tumours than ER+/PR+, so perhaps the increased growth factor signalling leads to decreased PR expression, leading to PR- tumours (Arpino et al., 2005). Thus the prevalence of growth factor stimulated cell growth may provide an explanation for the tamoxifen resistance of PR- tumours since they may no longer rely on oestrogen/ER for mitogenic stimulus (Cui et al., 2005). Growth factor signalling may also interact with ER and alter its function as a transcription factor, potentially affecting PR expression, and promoting non-genomic ER signalling. This switch may affect the ability of tamoxifen to function as an ER antagonist, again implicating growth factor signalling in both PR loss and tamoxifen resistance in PR- tumours (Cui et al., 2005). This interaction of ER and growth factor signalling in mediating the loss of PR expression in breast cancer highlights the importance of PR as a predictive marker for tamoxifen use; PR- status may reflect ER+ tumours in which excessive growth factor signalling alters ER activity or circumvents the requirement for ER activation, rendering cells resistant to tamoxifen.

Loss of heterozygosity (LOH) mutations on chromosome 11q, which are observed at an increased frequency in breast cancer, may act to prevent high levels of PR expression, since the PR gene is located on chromosome 11q22 (Tomlinson et al., 1996). In ovarian cancer an aggressive phenotype is also associated with LOH at chromosome 11q22 (Launonen et al., 1998). Promoter hypermethylation is an epigenetic mechanism of gene silencing by which a methyl group is added to cytosine/guanine rich sequences within the promoter to either block the assembly of the transcriptional machinery or remodel chromatin to stable conformations which cannot be transcribed (Lapidus et al., 1996). Hypermethylation of the PR promoter has been frequently observed in breast cancer;

40% of PR- breast tumours showed PR promoter methylation which was not seen in PR+ tumours or normal breast tissue (Lapidus et al., 1996). PR-B promoter hypermethylation was later detected in 76% of breast tumours, and has been demonstrated to be linked to PR- tumour status and a number of associated poor prognostic features (Mc Cormack et al., 2008). Another study identified 13 breast tumours showing full promoter hypermethylation, of which 7 showed a loss of PR-B protein, and 19 tumours showing partial methylation, of which 9 showed reduced or lost PR-B expression (Mirza et al., 2007).

The use of alternative promoters either within the normal PR mRNA sequence, such as PR-C described in Section 1.1.8, or promoters resulting from alternative splicing of PR pre-mRNA could lead to the generation of truncated PR isoforms which are not detectable by current screening methods which target N-terminal epitopes (Cork et al., 2008). The alternative splicing of PR pre-mRNA and its potential impact in breast cancer will be discussed later in this thesis (Section 1.4.11).

As described in Sections 1.3.5, 1.3.6 and 1.3.10-1.3.13, progesterone signalling via PR plays a role in the development of breast cancer. Furthermore, the role of PR as a breast cancer biomarker is also described above with PR- tumours associated with poor prognosis and poor response to endocrine therapy. Understanding the mechanisms involved in the loss of PR expression is therefore vital since, as described in this section, PR loss may be a result of cancer associated changes in cell signalling, representing a shift in the mitogenic signals stimulating proliferation of cancer cells. Alternatively a loss of PR expression may represent a lost mechanism of control over tumour growth in PR negative tumours.

1.4 Alternative pre-mRNA splicing

1.4.1 Pre-mRNA splicing

Splicing is the RNA processing event by which non-coding intragenic sequences, introns, are removed from the pre-mRNA transcript and the exons ligated to form the mature protein coding mRNA. Pre-mRNA splicing requires recognition of a 5′ splice site between the upstream exon and the intron and a 3′ splice site between the intron and the downstream exon, followed by cleavage of the splice sites to remove the intron and ligation of the exons. Early reports demonstrated that multi-protein complexes containing small nuclear RNAs, termed small nuclear ribonucleoprotein particles (snRNPs), termed U1, U2, U4, U5 and U6, could be detected in association with RNA at each stage of the splicing reaction and that preventing snRNP function prevented splicing, suggesting involvement of a multiprotein complex in the splicing reaction (reviewed by (Padgett et al., 1986).

Recognition of the 5′ splice site downstream of an exon requires sequence complementarity of an AG|**G**UA/GAGU nucleotide sequence at the upstream exon/intron boundary with the U1 snRNP. Recognition of the 3′ splice site requires cooperative binding of the large subunit of the U2 auxiliary factor, $U2AF_{65}$, to a pyrimidine rich sequence known as the polypyrimidine tract (PPT) adjacent to the AG dinucleotide at the 3′ end of the intron, $U2AF_{35}$ to the 3′ splice site AG, and splicing factor 1 (SF1) with the branch point sequence which contains a conserved adenosine and is located approximately 100 nucleotides upstream of the 3′ splice site. The binding of U2AF at the 3′ splice site and U1 at the 5′ splice site forms the early (E) complex, committing the splice sites to the splicing reaction. U2AF and SF1 recruit U2 snRNP to the branch point sequence (BPS), forming the A complex. The U4/U6-U5 tri-snRNP then binds to form the B complex and following binding rearrangements U1 is replaced at the 5′ splice site by U6 which interacts with U2 and dissociates from U4. Release of U1 and U4 forms the catalytically active pre-spliceosomal B* complex which catalyses the transesterification of the 5′ splice site G (highlighted in bold above) by the branch point adenosine, thus cleaving the upstream exon from the looped intron/downstream exon, termed the lariat. The mature spliceosomal C complex is then formed and catalyses the second transesterification reaction, of the 3′ splice site G by the 3′ hydroxyl group of the upstream exon, excising the intron lariat and ligating the exons

(McManus and Graveley, 2011; Jurica et al., 2002; Makarov et al., 2002; Berglund et al., 1998; Staley and Guthrie, 1998).

Each stage of spliceosome formation and rearrangement subsequent to the E complex is ATP dependent and relies on the ATPase activity of spliceosome associated proteins (Staley and Guthrie, 1998). Mass spectrometric analysis of spliceosomal complexes purified from HeLa cell extract has revealed in excess of three hundred proteins associated with the spliceosome. Some of these are core snRNP proteins, some have essential functions such as mediating/remodelling snRNA/mRNA interactions or ATPase activity and some have as yet unclear roles but associate with spliceosomes *in vitro* (Bessonov et al., 2008; Jurica and Moore, 2003; Makarov et al., 2002; Zhou et al., 2002).

1.4.2 Splice site selection

As described above, initial splice site selection involves the binding of U1 to the 5′ splice site and U2AF to the PPT and 3′ splice site. This process defines the exon by committing the upstream 3′ splice site and downstream 5′ splice sites to spliceosome formation (Berget, 1995). Recruitment of U1 and U2AF involves interaction with *trans*-acting regulatory RNA binding proteins, termed splicing factors, which bind to *cis*-acting regulatory sequences within the pre-mRNA. These regulatory elements either enhance or silence splicing, and can be located within exons or introns, thus they are referred to as either exon/intron splicing enhancers (ESE/ISE) or exon/intron splicing silencers (ESS/ISS), with their function being determined by the range of splicing factors which recognise and bind to the sequences and the location within the pre-mRNA. The presence of competing *cis*-elements and the balance of expression of splicing factors therefore combine to determine splice site selection and the pattern of alternative splicing (reviewed by (Hertel, 2008; Matlin et al., 2005).

1.4.3 Serine/arginine rich (SR) proteins

SR proteins are a family of splicing factors which bind to ESEs via N-terminal RNA recognition motifs (RRM) and contain a C-terminal arginine/serine rich (RS) domain, which mediates protein/protein interactions (Long and Caceres, 2009; Shepard and Hertel, 2009). SR protein family members were initially named based on their function or approximate molecular weight, and a group of classical SR proteins which were detectable by a phospho-specific antibody was identified. Additional SR-proteins were also identified and proteins with similar structures were termed SR-related proteins (SRrp) (Long and Caceres, 2009; Shepard and Hertel, 2009). The structure of the seven classical SR proteins and three SR-related proteins is illustrated in Figure 1.6. A new system of nomenclature has recently been proposed, defining SR proteins as containing at least one N-terminal RRM or RRM homology (RRMH) domain and a C-terminal RS domain of at least 50 nucleotides consisting of at least 40% R/S or S/R dinucleotides. Under this new system the classical SR proteins and SR-related proteins depicted in Figure 1.6 are renamed, along with two additional SR-related proteins which are re-classified as SR proteins. The new system terms the proteins, and the genes which encode them, as SR splicing factors (SRSF), and the proteins/genes are numbered based on the order of discovery (Manley and Krainer, 2010). This new nomenclature will be used for the remainder of this thesis and the new protein/gene names are listed, along with the traditional terminology for both genes and proteins, in Table 1.2.

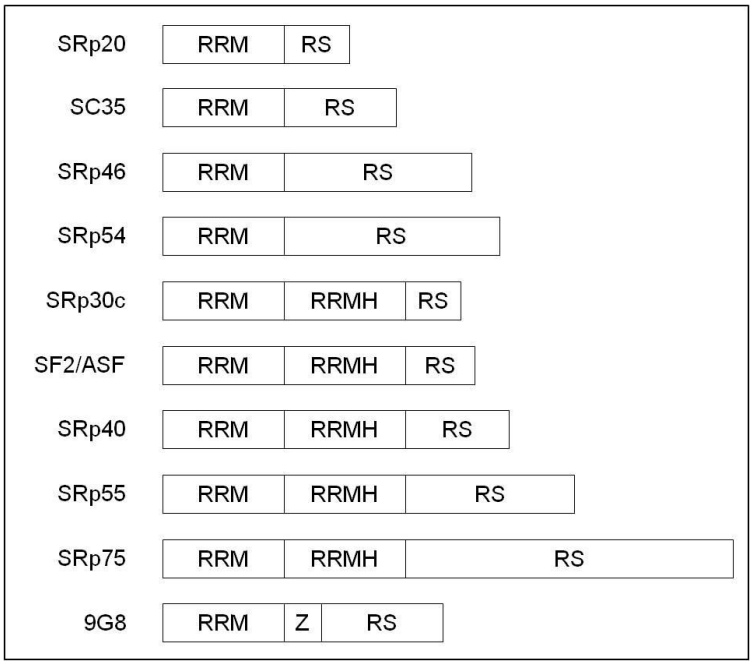


Figure 1.6: Schematic representation of the domain structure of human SR proteins.

RRM: RNA recognition motif; RRMH: RRM homology; Z: zinc knuckle; RS: argenine/serine-rich domain. Figure adapted from (Graveley, 2000).

Old gene symbol	Old protein names	New gene/protein symbol
SFRS1	SF2/ASF, SRp30a	SRSF1
SFRS2	SC35, PR264,SRp30b	SRSF2
SFRS3	SRp20	SRSF3
SFRS4	SRp75	SRSF4
SFRS5	SRp40, HRS	SRSF5
SFRS6	SRp55, B52	SRSF6
SFRS7	9G8	SRSF7
SFRS2B	SRp46	SRSF8
SFRS9	SRp30c	SRSF9
SFRS13A	TASR1, SRp38, SRrp40	SRSF10
SFRS11	p54, SRp54	SRSF11
SFRS13B	SRrp35	SRSF12

Table 1.2: Comparison of old and new nomenclature for SR proteins.

Table adapted from (Manley and Krainer, 2010).

Initial splice site recognition is thought to occur across an exon, in a process termed 'exon definition' in which U2AF binds to the upstream 3′ splice site and U1 to the downstream 5′ splice site either side of an exon (Berget, 1995). Addition of individual SR proteins (SRSF1, SRSF2 or SRSR6) to a splicing reaction mixture of nuclear extract and pre-mRNA was sufficient to commit the substrate to splicing, implicating SR proteins in initial splice site selection (Fu, 1993). Yeast two-hybrid screening revealed interaction of SRSF1 and SRSF2 with both $U2AF_{35}$ and the U1 protein U1-70k (Wu and Maniatis, 1993) and interaction of the RS domains of SRSF6 or SRSF7 were later demonstrated to be required for recruitment of $U2AF_{65}$ to a weak PPT which was dependent on $U2AF_{35}$ (Graveley et al., 2001). Therefore, a model of SR protein involvement in normal splice site recognition is that SR proteins bound to ESEs recruit U2AF and U1 to splice sites during E complex formation. Furthermore SR proteins interacting with $U2AF_{35}$ can strengthen a 3′ splice sites with a weak PPT by stabilising $U2AF_{65}$ binding (Graveley et al., 2001). The reported interactions of ESE bound SR proteins with U1 and U2AF may also occur across introns after initial splice site recognition to bring distant 5′ and 3′ splice sites into proximity for pairing during spliceosome assembly in order to ligate upstream and downstream exons (Long and Caceres, 2009).

SR proteins have also been reported to be involved in spliceosome formation after initial splice site selection. Spliceosome assembly from purified A complex and U4/U6-U5 tri-snRNP required the addition of SRSF1, SRSF2, SRSF4, SRSF5 or SRSF6. This function of SR proteins was reliant on phosphorylation of the RS domain and suggests an interaction of the SR proteins with the tri-snRNP components to stabilise tri-snRNP binding to the 5′ and 3′ splice site (Roscigno and Garcia-Blanco, 1995). The requirement for SRSF1 or SRSF7 to mediate efficient splicing of an mRNA construct containing only 1-3 nucleotides of exon indicated that SR proteins could also function independent of ESE binding, and the requirement for a functional RRM to mediate this splicing suggested interaction with intronic sequences (Hertel and Maniatis, 1999). The RS domain of SRSF1 has since been reported to interact with the branch point in the A complex and to interact with the 5′ splice site and facilitate U6 binding during spliceosome maturation, suggesting a potential mechanism of SR protein dependent tri-snRNP recruitment and demonstrating the importance of SR proteins in all stages of the splicing reaction (Shen and Green, 2004; Shen et al., 2004).

1.4.4 Alternative pre-mRNA splicing

Alternative splicing is the differential removal of exons or retention of introns during pre-mRNA splicing, generating different mRNAs which encode structurally and functionally distinct protein isoforms. This process allows the generation of protein diversity from a relatively compact genome (Nilsen and Graveley, 2010) and it is estimated that approximately 95% of multi-exon genes may undergo alternative splicing (Pan et al., 2008). This enables the tissue specific or developmentally regulated expression of functionally different proteins by alterations in splicing patterns, and facilitates response to changes in external stimuli (Moroy and Heyd, 2007; Stamm, 2002). Changes in the pattern of alternative splicing can also contribute to disease development which will be discussed in Section 1.4.7 (reviewed by (Caceres and Kornblihtt, 2002).

Spliceosome formation, as described above, requires the interaction of spliceosome component snRNPs with splicing factors to direct recruitment to the 5′ and 3′ splice sites. In addition to splice sites at constitutive exon/intron boundaries, alternative splice sites can exist within exons and introns, or exons may exist with weak splice sites which are not normally used (Solis et al., 2008; Kanopka et al., 1996). Silencing of splice sites can lead to exon skipping (Figure 1.7b) (Solis et al., 2008). The deletion of an exon often causes a change in reading frame and premature termination of translation. However a cassette exon is an exon which can be deleted without causing a change in reading frame (Richer et al., 1998). Multiple exons can be deleted from a pre-mRNA by alternative splicing, allowing further diversity of protein generation (Springwald et al., 2010). Mutually exclusive exons also exist, whereby one exon is retained and one deleted, but never both retained or deleted, generating two different isoforms (Jumaa and Nielsen, 1997) (Figure 1.7c). Cassette exonic sequences with weak splice sites can also exist within introns and can be retained upon activation of the weak splice sites (Pollard et al., 2002) (Figure 1.7d). Silencing of constitutive splice sites either side of an intron can lead to retention of the intronic sequence (Figure 1.7e). Weak cryptic splice sites can exist within an exon or intron (Balleine et al., 1999b); activation of splicing at a cryptic 5′ splice site within an exon would lead to deletion of the 3′ end of the upstream exon, and an alternative 3′ splice within an exon would lead to deletion of the 5′ end of the downstream exon (Figure 1.7f). Cryptic splice sites can also exist within an intron leading to retention of parts of the intronic sequence.

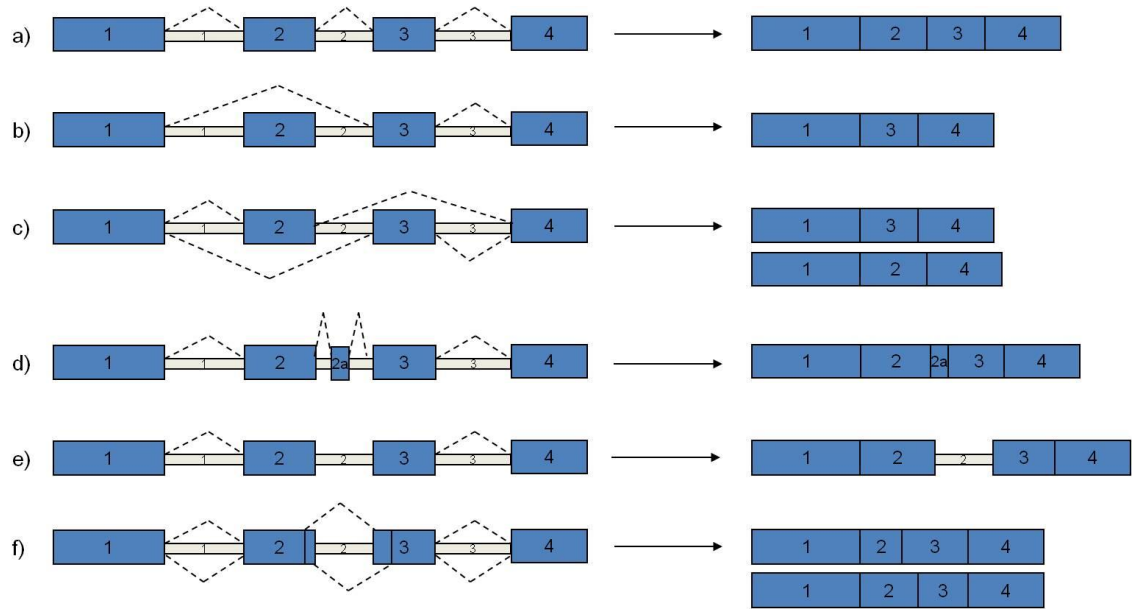


Figure 1.7: Patterns of constitutive and alternative splicing.

a) Constitutive splicing; the removal of introns and ligation of exons. **b)** Exon skipping; silencing of constitutive splice sites upestream and downstream of exon 2 leading to removal of the exon and ligation of exon 1 to exon 3. **c)** Exons 2 and 3 are mutually exclusive; either exon 2 or exon 3 is always removed generating two distinct mRNAs. **d)** Cryptic exons within intronic sequences; exon 2a is normally reomved along with intron 2, activation of the weak splice sites flanking exon 2a leads to its inclusion in the mRNA. **e)** Intron retention; silencing of the 5′ splice site downstream of exon 2 and the 3′ splice site upstream of exon 3 leads to the inclusion of intron 2 in the mRNA. **f)** Cryptic splice sites; activation of a weak 5′ splice site within exon 2 leads to deletion of the 3′ end of the exon. Activation of a weak 3′ splice site within exon 3 leads to the deletion of the 5′ end of exon 3.

1.4.5 Non-sense mediated decay

Deletion of exonic sequences or retention of intronic sequences often disrupts the normal reading frame and the mRNA therefore contains a premature translational termination codon (PTC) which makes the mRNA a target for degradation by non-sense mediated decay (NMD) (Nicholson et al., 2010). An estimated one third of all alternatively spliced mRNA are degraded in this way which serves to protect cells from the deleterious effects that may be associated with alternatively translated proteins (Lewis et al., 2003). NMD is a pathway by which non-sense mRNA resulting from non-sense mutations which generate a PTC and nucleotide insertions or alternative pre-mRNA splicing which cause a change in reading frame is degraded. PTCs are recognised by a complex of proteins which catalyse mRNA degradation (Nicholson and Muhlemann, 2010). During splicing, a multi-protein complex is added approximately 20-24 nucleotides upstream of the exon/exon junction, termed the exon junction complex (EJC). This complex functions in nuclear export of the mRNA and translation.

The 5′ end of a nascent mRNA becomes associated with the cap-binding complex (CBC) and the 3′ end is polyadenylated. During the initial round of translation ribosomes move along the mRNA, displacing the EJCs until they encounter a termination codon, upon which they stall and eukaryotic poly-peptide release factor (eRF)3 interacts with the poly-A binding protein (PABP) which mediates polypeptide release from the ribosome. The CBC is then replaced by eukaryotic initiation factor which appears to protect mRNA from NMD (Gardner, 2010). The exact mechanisms by which NMD targeted mRNAs are recognised and degraded have not been fully elucidated but current understanding is reviewed by (Nicholson et al., 2010); when the ribosome encounters a PTC, the interaction of eRF3 with PABP is less efficient due to the increased distance from the poly-A sequence allowing distinction to be made between normal and premature termination codons. The presence of EJCs downstream of the PTC and the long 3′ UTR which is associated with a PTC also appear to promote NMD. The interaction of the ribosome with PABP is further inhibited by binding of the essential *trans*-acting NMD factor RENT1 with eRF3. RENT1 is then phosphorylated by RENT2 and this phosphorylation mediates binding of RENT1 to SMG proteins which stimulate mRNA degradation (Nicholson et al., 2010). An example of alternative splicing resulting in a PTC and targeting of the mRNA for NMD is the gene encoding murine exportin 4 which contains an exon which is alternatively included in adult and embryonic tissue, inclusion of which introduces a PTC. Exportin 4 is involved in the nuclear import/export of factors essential for embryonic development and the exon is preferentially included in adult tissue but excluded in embryonic tissue. The PTC introduced by exon inclusion was demonstrated to induce NMD, thereby regulating protein expression in different developmental states through coupling of alternative splicing and NMD (Barash et al., 2010). The expression and function of NMD factors is regulated in response to cellular stresses such as hypoxia and generation of reactive oxygen species which are common in cancer and function to inhibit NMD. In the absence of NMD, mRNA containing a PTC may be able to encode proteins in cancer cells with functions which differ from the normal protein (Gardner, 2010).

1.4.6 Regulation of alternative pre-mRNA splicing

The binding of SR proteins to ESEs in proximity to weak splice sites can lead to selection of the weak splice site; SR proteins can enhance recruitment of U1 to a poor consensus 5′ splice site, or stabilise the interaction of $U2AF_{65}$ to a short PPT at a weak

3′ splice site (Graveley et al., 2001). Whilst SR proteins usually function as splicing enhancers, this function is dependent on the context of the sequence to which they are bound. For example SRSF1, SRSF2 and SRSF6 bind to an ISS sequence close to a weak 3′ splice site in an adenovirus mRNA to repress splicing by preventing the binding of U2 snRNP. The function of this SR protein binding sequence was switched to a splicing activator when moved to an exonic context and similarly an ESE sequence became inhibitory when moved to an intron, demonstrating the importance of the context in which an SR protein binds to pre-mRNA in determining its function as a splicing regulator (Kanopka et al., 1996). Silencing of a proximal 3′ splice site in an mRNA construct resulted in suppression of a distal 3′ splice site by downstream ESE sequences, and when these sequences were moved to an intronic context, again they repressed splicing. This repression of splicing was reliant on SR protein (SRSF2/SRSF7) binding, supporting the earlier observation of context dependent SR protein function (Ibrahim et al., 2005). As well as demonstrating that SR proteins can function as both enhancers or silencers of splicing depending on context, this latter study also reported that the binding of SR proteins to ESE sequences prevents exon skipping by both activating proximal and silencing distal 3′ splice sites (Ibrahim et al., 2005).

The effect of SR proteins as splice site enhancers is opposed by heterogenous ribonucleoprotein (hnRNP) A1, as demonstrated by co-transfection of the adenovirus E1A gene into HeLa cells along with either hnRNP A1 or SRSF1 cDNA. hnRNP A1 overexpression favoured distal 5′ splice site selection whereas SRSF1 overexpression favoured proximal 5′ splice site selection (Caceres et al., 1994). hnRNP A1 has been demonstrated to bind non-specifically to pre-mRNA and prevent U1 binding to 5′ splice sites by competitive inhibition, a role opposing that described above for SR proteins. Therefore another mechanism by which SR proteins are required for splice site selection and spliceosome formation is by removing the hnRNP block of U1 binding. Prevention by hnRNP A1 of U1 recruitment by SR proteins may result in the use of upstream, intron distal, 5′ splice sites (Eperon et al., 2000). High affinity binding of hnRNP A1 to a specific ESS sequence in exon 3 of HIV-1 *tat* pre-mRNA enhanced hnRNP binding to other exonic sequences and prevented splicing in HeLa cell extracts. Depleting extracts of hnRNP A/B proteins activated splicing and addition of SRSF1 but not SRSF2 to hnRNP A1 supplemented S100 could restore splicing, demonstrating both that SRSF2 binding is inhibited by hnRNP A1, and that SRSF1 can out compete hnRNP A1 for

binding to the pre-mRNA preventing splicing inhibition, again demonstrating a mechanism for the antagonistic properties of hnRNP A1 and SR proteins (Zhu et al., 2001). This antagonistic function is again demonstrated by the altered inclusion of exon 3 of the gene encoding the G-protein $G\alpha_S$ upon overexpression of SRSF1 or hnRNP A1 in HeLa or myometrial smooth muscle cells, which promoted proximal or distal 5′ splice site selection, respectively (Pollard et al., 2002).

The mechanisms by which individual SR proteins mediate splice site recognition and spliceosome formation appear to be similar in the *in vitro* studies described above and there appears to be a degree of functional redundancy among SR family members. SRSF2, SRSF4, SRSF5 or SRSF6 could all restore the splicing of human β-globin pre-mRNA in splicing deficient HeLa cell extract, S100 (Zahler et al., 1993). siRNA knockdown of the genes encoding five SR proteins in *C. elegans* resulted in embryonic lethality, whereas knocking down individual SR genes resulted in no developmental defects, suggesting the proteins serve similar roles in development and can compensate for depletion of individual family members (Kawano et al., 2000). However, when alternative splice sites were present in pre-mRNA, addition of different SR proteins to S100 extract resulted in different patterns of splicing, demonstrating different roles for the individual SR proteins. Analysis of SR protein expression in different calf tissues demonstrated tissue specific patterns of expression of SRSF2, SRSF4, SRSF5 and SRSF6 (Zahler et al., 1993). Furthermore, tissue specific gene knockdown of SRSF1 or SRSF2 in mice had differing effects on cardiac development; SRSF1 ablation resulted in lethality at around 6-8 weeks old which was linked to a change in splicing of a protein involved in calcium metabolism, CaMKIIδ, whereas SRSF2 deleted mice showed no obvious defect until a much later stage and no alteration in splicing of CaMKIIδ was observed. SRSF2 deleted mice also demonstrated reduced T-cell maturation which was linked to a change in splicing of CD45. These mouse models demonstrate critical tissue and developmental state specific roles for individual SR proteins in splicing regulation (Moroy and Heyd, 2007). The importance of alternative splicing in developmental regulation is also seen in the myometrium during pregnancy. The myometrium is the uterine smooth muscle responsible for contractions during labour and the upper and lower segments of the myometrium are functionally distinct, showing a temporal change in function during pregnancy and labour. Spatial and temporal variation in SRSF1 and hnRNP A1 expression in the myometrium are

observed during pregnancy (Pollard et al., 2000) and SRSF5 regulates the temporal switch in alternative splicing of cyclic AMP-response element modulator (CREM) in the myometrium during pregnancy (Tyson-Capper et al., 2005).

Since different SR proteins regulate different splice sites, the ratio of expression of SR proteins is important in regulating splicing. For example the ratio of SRSF2 and SRSF4 added to S100 extract determined the pattern of alternative 3′ splice site usage in a simian virus 40 pre-mRNA construct; SRSF2 alone favoured the proximal 3′ splice site, whereas SRSF4 alone favoured the distal 3′ splice site, and when added in combination an alteration in the ratio of SRSF2 and SRSF4 resulted in a parallel shift in the ratio of proximal and distal 3′ splice site mRNA products (Zahler et al., 1993).

SR proteins have also been demonstrated to exhibit antagonistic functions, again demonstrating the importance of the expression ratio of different splicing factors in the regulation of alternative splicing. The mutually exclusive chicken β-tropomyosin exons 6A and 6B are retained in mRNA in different tissue types. Recognition of the 5′ splice site downstream of exon 6A specifically requires SRSF1 binding to an ISE sequence and this activation of splicing by SRSF1 is inhibited by addition of SRSF2 but not SRSF7 demonstrating the antagonistic functions of specific SR proteins binding to regulatory sequences (Gallego et al., 1997). Antagonism between SRSF1 and SRSF2 has also been reported in the skipping of exon 3 from human growth hormone pre-mRNA; SRSF1 binds to an ESE within exon 3 to promote exon inclusion, whereas SRSF2 binds to a different sequence to repress splicing and promote exon skipping (Solis et al., 2008). Another example of SR proteins exhibiting antagonistic functions is the autoregulation of SRSF3. Inclusion of SRSF3 exon 4 alters the reading frame encoding a truncated protein lacking the RS domain. SRSF3 binds to an ESE within exon 4 and enhances a weak 3′ splice site leading to inclusion of the exon, whereas SRSF1 silences the weak 5′ splice site downstream of exon 4 promoting exon skipping (Jumaa and Nielsen, 1997). SRSF10 and SRSF12 are described as SR-repressor proteins. Neither SRSF10 nor SRSF12 are capable of enhancing splicing in S100 extract and both demonstrate a dominant negative regulation over SRSF1 and SRSF2 mediated splicing in S100. This antagonism may be via direct competition for binding to ESEs or

interaction with specific silencer sequences since the RRM of SRSF10 and SRSF12 were critical for antagonism (Cowper et al., 2001).

Another inhibitory splicing factor is hnRNP I, also known as PPT binding protein (PTB), which binds to the PPT and inhibits the binding of $U2AF_{65}$ and subsequent recruitment of U2 during early pre-spliceosome formation. In the neuronal specific alternative splicing of *src* exon N1, hnRNP I binds to ISS sequences and represses U2 binding, however a neuronal specific hnRNP I variant promotes binding of hnRNP F, hnRNP H and SRSF1 to an ESE, promoting splicing (reviewed by (Matlin et al., 2005). As well as alterations in the expression of ubiquitous splicing factors such as hnRNPs and SR proteins, tissue specific splicing factors have been described. One such example is the Nova family of proteins which function by binding to specific *cis*-acting sequences which can be either exonic or intronic; when bound to exonic sequences Nova proteins block U1 binding whereas when bound to intronic sequences they enhance U1 binding (Ule et al., 2006). Nova family members are expressed specifically in brain and control the neuronal specific splicing of several genes (Buckanovich and Darnell, 1997). Other splicing factors which bind to ISE sequences include Fox family members and CELF proteins, expression of which is developmentally and spatially regulated to differentially regulate alternative splicing (Venables, 2007). Fox proteins bind to conserved sequences within introns, and the position of the binding site in relation to the exon determines the effect on exon inclusion; Fox binding downstream of an exon promotes exon inclusion by activating the 5′ splice site, whereas Fox binding upstream of the exon promotes exon skipping by silencing the 3′ splice site (Venables et al., 2009).

The differing and antagonistic functions of different splicing factors in enhancing/silencing splice site selection which are described above demonstrate that the expression ratio of these different splicing factors in a particular tissue, developmental state or disease process, as well as the presence and context of the *cis*-acting regulatory elements to which they bind combine to regulate alternative patterns of pre-mRNA splicing. It has been reported that ESE and ESS sequences cannot always be considered as independent elements with opposing functions, instead positive and negative splicing regulators can compete for binding to overlapping sequences, making prediction of

splicing patterns difficult. These sequences, termed composite exonic regulatory elements of splicing (CERES) were first identified in exon 12 of the *CFTR* gene, encoding the cystic fibrosis transmembrane conductance regulator. It was demonstrated by site directed mutagenesis that missense mutations of bases in close proximity within two regions of the exon could have differing effects on exon inclusion, and that the change in splicing was not predicted by the effect on splicing factor binding sequences. It was therefore proposed that these regions represented CERES sequences (Pagani et al., 2003b). RNA binding precipitation assays were performed using the RNA sequence of a *CFTR* exon 12 CERES or site directed CERES mutants; both SR proteins and hnRNPs were capable of interactions with this region and mutations that altered splicing in favour of exon inclusion interacted most strongly with SR proteins. SRSF1 and SRSF6 interacted directly with CERES and these SR proteins induced the greatest level of exon inclusion compared to SR proteins that did not bind CERES when expression vectors were co-transfected into HeLa cells along with a mini-exon 12 CERES construct. Similarly, siRNA knockdown of hnRNP A1 which also bound CERES was able to rescue exon inclusion in this mini-exon model. Neither SR proteins nor hnRNP A1 could affect the splicing of constructs which were deleted for the CERES sequence demonstrating that both silencing and enhancing functions are mediated via direct interaction with this element (Haque et al., 2010). A similar juxtaposition/overlap of ESE and ESS functions has also been observed in *CFTR* exon 9 (Pagani et al., 2003a), survival of motor neuron 1 (*SMN1*) exon 7 (Vezain et al., 2010) and exon 10 of the gene encoding luteinising hormone receptor type 2 (Gromoll et al., 2007).

In order to consider the combined effect of competing splicing regulatory elements within pre-mRNA, a splicing code has been formulated using mouse tissue to predict tissue specific alternative splicing events. A computer program was developed by inputting sequences known to bind *trans*-acting splicing regulatory proteins, sequences overrepresented close to alternatively spliced exons and information on the inclusion/exclusion of an exon in a particular tissue type to determine the combinations of *cis*-regulatory sequences which regulate splicing of different pre-mRNA in different tissues. This code was then applied to various genes to predict exon inclusion/exclusion which was validated by a PCR assay (Barash et al., 2010). The development of this code may be the first step in identifying the sequences/features responsible for tissue specific alternative splicing regulation. The effect of transcription and RNA secondary

structures on alternative splicing must also be considered, and the splicing code is predicted to be species specific since there is only an estimated 20% conservation in splicing regulation from mouse to human (Tejedor and Valcarcel, 2010).

1.4.7 Alternative pre-mRNA splicing in cancer

It is estimated that approximately 50% of point mutations that cause genetic disease affect alternative splicing, either by silencing a splice site, creating a new splice site, strengthening an existing cryptic splice site or altering a *cis*-regulatory element (Ward and Cooper, 2010). Diseases such as cystic fibrosis and spinal muscular atrophy can be linked to mutations which affect splicing regulatory elements within *CFTR* exon 12 and *SMN* exon 7 respectively, which alter the patterns of exon inclusion (Vezain et al., 2010; Pagani and Baralle, 2004). Numerous alternatively spliced variants are detected specifically in cancer tissue leading to the suggestion of an increase in alternative splicing in cancer cells. However, a recent report suggested that there was actually a decrease in overall alternative splicing in cancer compared to normal tissue, specifically associated with a decrease in exon skipping, whereas an increase in the use of alternative splice sites and intron retention was observed in cancer (Kim et al., 2008). From these observations it is clear that there is a misregulation of splicing associated with cancer development. Whether the observed cancer associated changes in alternative splicing are a cause or a consequence of transformation remains to be fully elucidated. As will be discussed below, cancer associated changes in alternative splicing may provide novel biomarkers and perhaps therapeutic targets in the management of cancer patients (reviewed by (Ward and Cooper, 2010; Pajares et al., 2007; Srebrow and Kornblihtt, 2006).

The importance of changes in oncogene and tumour suppressor expression in cancer development has been described in Section 1.3.4. p53 is a tumour suppressor which normally functions to prevent cell cycle entry of cells harbouring mutations, to stimulate apoptosis and maintain genetic stability (Lee and Muller, 2010). Nine alternative p53 isoforms were identified resulting from alternative promoters or alternative pre-mRNA splicing including two alternatively spliced variants (ASVs), p53β and p53γ, resulting from retention of a cryptic exon within intron 9, termed exon 9b (Bourdon et al., 2005). The resulting C-terminally truncated variant proteins lack the tetramerisation domain,

and conflicting reports exist over the function of the variant proteins; suggesting that the proteins are unable to interact with p53 and are non-functional (Graupner et al., 2009) and conversely that p53β was able to enhance the pro-apoptotic effect of p53 when co-expressed in p53 null cells (Bourdon et al., 2005). p53β but not p53γ has been detected in both normal and breast tumour tissue (Bourdon et al., 2005), and expression of p53β has recently been linked to poorer recurrence free survival in ovarian cancer (Hofstetter et al., 2010). Three further novel p53 ASVs were recently reported resulting from point mutations in normal splice sites; p53ζ, p53δ, p53ε. All of the p53 ASVs were detected in ovarian cancer tissue except p53γ. p53δ expression correlated with poorer disease free and overall survival, increased rate of resistance to chemotherapy and an increased rate of early recurrence following chemotherapy (Hofstetter et al., 2010). It is therefore possible that the expression of these p53 ASVs may modulate normal p53 function or represent a loss of p53 function and contribute to disease progression.

The *BRCA1* gene encodes a protein with tumour suppressor function, and mutations in *BRCA1* are associated with a high risk of breast and ovarian cancer as described in Section 1.3.4. An analysis of 607 nucleotide variations within *BRCA1* which had previously been reported in either breast or ovarian cancer patients identified a frequent colocalisation of these mutations within predicted ESE sequences identified based on the match to consensus SR protein binding sequences. These mutations resulted in the weakening or loss of ESE sequences (210/607 mutations studied), as well as strengthening or gain of ESEs (131/607 mutations studied) (Pettigrew et al., 2005). The G5199T substitution within exon 18 has been previously reported to encode a stop codon within exon 18, but to result in skipping of exon 18 which would encode a non-functional protein (Mazoyer et al., 1998), and was identified to disrupt an ESE binding sequence for SRSF1 (Pettigrew et al., 2005). The G231T substitution in *BRCA1* causes exon 6 skipping by creating an ESS, sequence and varying the levels of SRSF1, SRSF2, SRSF7 and hnRNP A1 alters the inclusion or skipping of this exon (Raponi et al., 2011). Exon 23 is another *BRCA1* exon which shows altered patterns of inclusion as a result of a mutation affecting splicing regulation; the C5434G substitution, identified in an ovarian cancer patient, resulted in increased exon 23 skipping, encoding a non-functional protein. The C5434T polymorphism at the same location has limited effect on exon inclusion/skipping. Whilst both substitutions disrupt SRSF1 and SRSF6 binding ESEs, C5434G, which increases exon 23 skipping, creates an hnRNP A1

binding ESS (Gaildrat et al., 2010). Mutations which were reported to affect ESEs in *BRCA1* included previously uncharacterised variants which had no obvious affect on the encoded protein sequence (Pettigrew et al., 2005). The loss of function of proteins encoded by the exon deleted *BRCA1* variants demonstrates the importance of predicting the effect of mutations on splicing regulatory elements and the potential role of altered *BRCA1* splicing in the development of breast and ovarian cancer.

Alternative splicing of oncogenes, generating constitutively active protein isoforms, has also been reported in cancer. For example, small deletions have been observed spanning the intron 10/exon 11 boundary of the *KIT* gene which encodes a receptor tyrosine kinase which is constitutively active in gastrointestinal stromal tumours (GIST) (Corless et al., 2004). This deletion destroys the 3′ splice site upstream of exon 11, resulting in the use of an alternative AG within the exon. The resulting deletion of the beginning of exon 11 encodes a protein lacking the region targeted by autophosphorylation and autoinhibition leading to constitutive KIT activity in mutation carrying GIST patients (Chen et al., 2005).

Alternative splicing of the three exon *Bcl-x* gene encodes an anti-apoptotic long isoform (Bcl-x_L) and use of a 5′ splice site within exon 2 deletes a coding region present in Bcl-x_L encoding a shorter pro-apoptotic isoform (Bcl-x_S) (Boise et al., 1993). Overexpression of the antiapoptotic Bcl-x_L is common in many cancers including breast (Olopade et al., 1997), prostate and lung (Ward and Cooper, 2010). Three *cis*-regulatory elements, B2, B3 and SB1, have been identified within the alternatively spliced exon 2 of *Bcl-x*; examining the effect of systematic deletions within the exon identified region B2 immediately downstream of the Bcl-x_S 5′ splice site which enhanced Bcl-x_S splicing and region B3 immediately upstream of the Bcl-x_L 5′ splice site which enhanced Bcl-x_L splicing (Garneau et al., 2005). hnRNP F and hnRNP H bind to B2 and siRNA knockdown of these hnRNPs reduced Bcl-x_S splicing (Garneau et al., 2005). SRSF9 binds to two regions within B3 to promote Bcl-x_L splicing and a third region within B3 contains two cryptic splice sites which bind U1 and repress splicing at the Bcl-x_L 5′ splice site (Cloutier et al., 2008). A third exonic sequence upstream of the Bcl-x_S 5′ splice site, termed SB1, was identified to be responsible for the increased Bcl-x_S splicing following inhibition of protein kinase C (PKC) (Revil et al., 2007). Therefore

PKC signalling has been implicated in activating a splicing repressor which binds to SB1 to promote expression of the anti-apoptotic Bcl-x_L and promote cell survival. A role antagonistic to that of PKC is assigned to DNA damage response pathways which promote Bcl-x_S expression and apoptosis via either the inactivation of the PKC induced repressor or activation of pro-splicing factors (Shkreta et al., 2011). The presence of these three *cis*-elements demonstrates the combinatorial regulation of splice site selection and the interplay of cell signalling and splicing regulation which could be altered in cancer to provide a survival advantage to cancer cells.

Vascular endothelial growth factor-A (VEGF-A) is involved in the process of angiogenesis, vital for tumour growth and survival (Pajares et al., 2007). Alternative spicing of the 8 exon *VEGFA* gene can result in nine angiogenic isoforms, the most widely expressed of which are $VEGF_{165}$ resulting from the skipping of exon 6, $VEGF_{121}$ resulting from the skipping of exons 6 and 7 and $VEGF_{189}$ which contains both exons 6 and 7. These three VEGF-A isoforms are pro-angiogenic and are widely expressed in tumour tissue in response to hypoxia, inducing formation of the tumour neovasculature (reviewed by (Pajares et al., 2007). Use of an alternative 3′ splice site within the 3′ UTR of exon 8 results in the encoding of an alternative 6 amino acid sequence at the C-terminus in the $VEGF_{165}$b isoform. This isoform was demonstrated to be widely expressed in normal kidney tissue, to inhibit VEGF-A function *in vitro* and to be down-regulated in the majority of kidney tumour samples analysed, suggesting an anti-angiogenic function which is removed by alternative splicing in cancer (Bates et al., 2002). Decreased VEGFb isoform expression has since been demonstrated in a range of other tumour tissues including lung and pancreatic (Pajares et al., 2007). The ratio of VEGF/VEGFb isoforms is regulated by growth factors in primary epithelial cells; insulin-like growth factor 1 and tumour necrosis factor-α increased proximal 3′ splice site selection, increasing VEGF expression, whereas TGFβ1 induced distal 3′ splice site selection via activation of p38 MAPK and the SR protein kinase Clk. SRSF1 and SRSF5 overexpression favoured VEGF expression whereas SRSF6 overexpression favoured VEGFb isoform. This study implicated growth factor signalling in the regulation of splicing factor function in cancer associated alternative splicing events (Nowak et al., 2008).

CD44 is a cell surface glycoprotein with roles in cell-cell contact, adhesion and migration. A complex pattern of alternative splicing involving the differential inclusion of ten variable exons has been described for CD44, showing tissue specific variation in exon inclusion (reviewed (David and Manley, 2010). CD44 splicing shows cancer specific variations, potentially providing therapeutic targets or molecular biomarkers (Pajares et al., 2007; Srebrow and Kornblihtt, 2006; Venables, 2004). The Ras/Erk signalling pathway has been implicated in inclusion of v6 (Cheng et al., 2006) and v5 (Matter et al., 2002). v5 inclusion is mediated by the RNA-binding protein Sam68 binding to an ESE within the exon and Sam68 is phosphorylated by Erk (Matter et al., 2002). The regulation of CD44 splicing by the Ras/Erk pathway provides a positive feedback mechanism for Ras signalling since v6 containing CD44 is a receptor which can activate Ras (Cheng et al., 2006).

Altered expression of several splicing factors has also been reported in cancer compared to normal tissues (Jia et al., 2010; Piekielko-Witkowska et al., 2010; Karni et al., 2007). Genetic amplification was demonstrated to account for increased expression of SRSF1 (Karni et al., 2007) and SRSF3 (Jia et al., 2010) and the genes encoding these splicing factors are within chromosomal regions commonly amplified in cancer. SRSF1 (Karni et al., 2007) and SRSF3 (Jia et al., 2010) have recently been described as proto-oncogenes whilst increased SRSF1 has been linked to poorer disease free survival in small cell lung cancer (Ezponda et al., 2010), and increased SRSF3 expression is observed from normal ovarian tissue to non-invasive ovarian tumours and to invasive ovarian tumours, being associated with increased tumour grade (He et al., 2011). Overexpression of SRSF1 in fibroblasts protected cells from apoptosis, and induced transformation; increasing anchorage independent growth in soft agar, and promoting tumour formation when cells were transplanted into nude mice (Karni et al., 2007). SRSF1 overexpression led to decreased expression of the tumour suppressor BIN1 and promoted the splicing of oncogenic isoforms of the ribosomal translational regulator S6 Kinase 1 and of MAPK interacting kinase 2 which constitutively activates eIF4E independent of its normal regulation by PI3K signalling (Karni et al., 2007). In lung cancer cells, knockdown of SRSF1 led to apoptosis associated with decreased expression of survivin and SRSF1 was reported to stabilise survivin mRNA and increase translation by inactivating negative regulators of the mTOR complex to (Ezponda et al., 2010). Similarly, SRSF3 knockdown with siRNA increases apoptosis

(He et al., 2011) and decreases proliferation in a range of cancer cell lines (Jia et al., 2010). SRSF3 overexpression stimulated anchorage independent growth in culture and induced tumour formation when transfected MEF 323 mouse fibroblasts were transplanted into nude mice (Jia et al., 2010).

Described above are mutations in tumour suppressors and oncogenes which alter splicing patterns and protein functions in cancer and the designation of splicing factors as oncogenes due to their capacity to induce transformation, protect against apoptosis and alter tumour-suppressor/oncogene expression. These mutations and the altered expression of splicing factors early in cancer development suggest that changes in alternative splicing may contribute to the initiation of cancer. The examples of Bcl-x, VEGF-A, CD44 and fibronectin demonstrate how growth factor and cell signalling pathways which, as described in Section 1.3, are misregulated in cancer can contribute to splicing. Therefore, changes in cell signalling, which may be a cause or consequence of cancer, may result in an altered pattern of alternative splicing. The products of these cancer associated changes in splicing can in turn contribute to cell survival, angiogenesis or metastasis which are processes required for tumour progression. Aberrant and alternative splicing may therefore be considered both a cause and a consequence of cancer initiation and a contributor to cancer progression.

1.4.8 Global identification of cancer specific alternative splicing events

Indentifying the effect of altered splicing of individual genes or of altered expression of splicing factors is useful but may not fully reveal the contribution of altered splicing patterns to the development of cancer. To more fully determine the changes in splicing, global methods to identify cancer associated alternative splicing are being utilised. High throughput sequencing has revealed novel potential alternative splicing events (ASEs) as well as tissue specific variations in regulated ASEs. This technology could be used to compare cancer and normal tissue to identify cancer specific splicing signatures (Blencowe et al., 2009; Pan et al., 2008; Wang et al., 2008). Exon junction microarrays use probes spanning each putative exon junction to identify alternative patterns of splicing, as well as probes located within an exon to measure changes in exon inclusion between different tissues. Initial use of these arrays identified tissue specific variations in alternative splicing patterns. These methods enable the identification of known as

well as novel ASEs which can be validated by quantitative real-time PCR (Johnson et al., 2003). The application of exon-junction microarrays to cancer has identified distinct patterns of alternative splicing between metastatic and non-metastatic murine breast cancer cells following transplantation into immunodeficient mice (Bemmo et al., 2010). Distinct patterns of ASEs were also observed which could differentiate between normal mammary epithelial cell lines, and breast cancer cell lines of different molecular phenotypes (Lapuk et al., 2010). In non-small cell lung cancer 330 genes were identified with a different pattern of alternative splicing compared to normal lung tissue (Langer et al., 2010) and splicing patterns were observed to differ between early and late stage neuroblastoma (Guo et al., 2011).

A compromise between global and functional studies is provided by the layered and integrated system for splicing annotation (LISA) platform, in which expression of 600 cancer associated genes is examined using primers spanning each potential exon junction to identify ASEs and calculate a value for the frequency of exon inclusion (percent spliced in, PSI) (Klinck et al., 2008). PSI values from LISA identified 48 ASEs which were specific for ovarian cancer (Klinck et al., 2008) and 41 ASEs which were specific to breast cancer (Venables et al., 2008) compared to normal ovarian or normal breast. Cancer specific ASEs could be used to accurately distinguish normal from cancer samples in a blinded study illustrating the potential for LISA to be used for cancer diagnosis (Klinck et al., 2008; Venables et al., 2008). For breast cancer, ASEs were also observed which were specific for ER+ tumours and others which could distinguish tumour grade (Venables et al., 2008). Analysis of the PSI profiles for breast and ovarian cancer revealed tissue specific differences, however a common signature of cancer associated ASEs was observed between the two types of cancer (Venables et al., 2009).

1.4.9 Therapeutic strategies targeting alternative splicing

Identification of cancer specific ASEs can provide novel therapeutic targets. Antisense oligonucleotides can be used to bind either to splice sites in order to block spliceosome formation, or to target *cis*-regulatory elements and prevent binding of splicing factors. Antisense oligonucleotides have been successfully used in the treatment of non-malignant diseases such as Duchenne's muscular dystrophy and spinal muscular

atrophy (van Alphen et al., 2009) and have been trialled for the blocking of mature mRNA translation of protein kinase C in non-small cell lung cancer (Pajares et al., 2007). Cancer specific exon skipping could potentially be corrected by the administration of peptide tagged oligonucleotides which bind ESEs and promote splice site selection, thus restoring normal exon inclusion (Cartegni and Krainer, 2003).

Cancer specific protein isoforms which result from ASEs may also provide therapeutic targets; if the ASE results in cancer specific exon inclusion, a cancer specific amino acid-sequence will be encoded which could be targeted with antibodies. Antibodies against cancer specific CD44v6 have been trialled but resulted in toxicity, and a tenascin-C antibody, targeting a brain tumour specific ASV has been tested in animals (Pajares et al., 2007). These antibodies may directly prevent function of the cancer specific protein or could be co-administered with cytotoxic drugs which could be targeted to cells expressing the cancer specific isoform (Pajares et al., 2007). An antibody targeted to fibronectin extra domain A (EDA) specifically targets tumour neovasculature, providing a potential vehicle for cancer targeted therapies (Rybak et al., 2007).

Administration of alternatively spliced isoforms may also provide a therapeutic strategy. As described, $VEGF_{165}b$ is predominantly expressed in normal tissue and inhibits $VEGF_{165}$, but a cancer specific shift to predominant expression of other spliced isoforms leads to angiogenesis. Therefore, administration of $VEGF_{165}b$ could inhibit angiogenesis (Pajares et al., 2007). SR protein function requires phosphorylation of the serine rich RS domains, thus drugs which target the SR proteins kinases (SRPKs) may remove the oncogenic effects of these proteins (van Alphen et al., 2009). Inhibition of SRPKs has been reported to switch VEGF splicing in favour of the anti-angiogenic $VEGF_{165}b$ isoform and reduce angiogenesis in mouse models (Nowak et al., 2010).

Finally, targeting spliceosomal components to prevent functional spliceosome formation could again prevent oncogenic ASEs. Inhibiting SF3b, a component of U2 snRNP, has been demonstrated to have anti-cancer properties *in vitro* (van Alphen et al., 2009). However, the widespread and essential role of SR proteins and the spliceosome in

normal cell processes would make targeting of these latter two strategies to specifically affect cancer associated ASEs difficult.

1.4.10 Alternative splicing of breast cancer biomarkers

As has been described, the identification of cancer specific splicing signatures can be used to accurately distinguish normal from cancer tissue, and could therefore provide a diagnostic tool in cancer management (Klinck et al., 2008; Venables et al., 2008). The identification of cancer specific spliced variants has also been described to be associated with prognosis, such as p53δ expression which correlated with poorer survival and increased resistance to chemotherapy (Hofstetter et al., 2010). Detection of cancer specific protein regions, such as fibronectin EDA can also provide specific biomarkers for cancer identification; EDA was detected at high levels in liver metastases neovasculature in mice but was undetectable in normal tissue, allowing identification of metastatic tumours (Rybak et al., 2007).

The ERα status of a breast tumour is currently used to predict the outcome of endocrine therapies. mRNA for several exon deleted ERα spliced variants have been reported in breast cancer tissue and examples are described below. Exon 2 deleted ERα (ERΔ2) has been observed in both normal and breast tumour but no specific association with breast cancer is reported (Herynk and Fuqua, 2004). ERΔ3 which lacks the DNA binding domain and is unable to bind DNA is able to activate the VEGF promoter via interaction with Sp1 transcription factors (Koduri et al., 2006). Increased ERΔ5 mRNA and protein expression was detected in breast cancer tissue compared to normal breast (Desai et al., 1997; Leygue et al., 1996b). A similar association of ERΔ5 mRNA with cancer is reported by an observed overexpression of ERΔ5 in tumour adjacent prostate tissue compared to benign prostate tissue (Taylor et al., 2010). Conflicting reports exist regarding ERΔ7 expression, with reports of high expression in breast cancer (Herynk and Fuqua, 2004) and reports of decreased expression in breast cancer compared to normal breast tissue (van Dijk et al., 2000) and of decreased expression in breast tumours with lymph node metastasis compared to non-metastatic tumours (Hsiao et al., 2006). Increased detection of multi-exon deleted ERα variants and a variant resulting from use of a cryptic splice site within exon 7 have been reported in breast cancer compared to normal breast tissue (van Dijk et al., 2000). Detection of exon deleted ER

variants in breast cancer could provide insight into the likely response to endocrine therapy as well as provide novel biomarkers of disease.

ERα antibodies targeting the LBD detect three predominant bands reported to correspond to the 66kDa ERα and two spliced variants, a 46kDa protein termed ERα46 and 36kDa protein termed ERα36. ERα46 results from the splicing of an exon within the 5′ UTR onto exon 2 encoding a protein lacking the N-terminal AF1 which is a dominant negative regulator of wild-type ER (Flouriot et al., 2000). ERα36 results from similar splicing of an exon within the 5′ UTR onto exon 2 resulting in skipping of exon 1, and the splicing of exon 6 onto a downstream exon within the 3′ UTR. This mRNA therefore encodes a unique C-terminal sequence but results in a protein lacking AF2 and part of the LBD, as well as AF1 (Wang et al., 2005). Despite the lack of so many functional domains, many studies have reported the activation of cytoplasmic ERα signalling pathways in cells expressing the membrane associated ERα36 isoform (Zhang et al., 2011; Kang et al., 2010; Lin et al., 2010; Wang et al., 2006). Furthermore, ERα36 overexpression has been reported in apocrine and adenoid cystic breast carcinomas compared to normal tissue (Vranic et al., 2011). The presence of a unique C-terminal sequence in ERα36 provides a potential target for specific antibodies for detection of this protein and use as a cancer biomarker.

HER2 is commonly overexpressed in breast cancer and provides a therapeutic target for the Herceptin antibody (Payne et al., 2008). A natural spliced variant retaining intron 8, termed Herstatin is described as a dominant negative regulator of HER2 (Doherty et al., 1999) and is expressed in both normal breast and breast tumour tissue (Koletsa et al., 2008). Expression of HER2 was higher in normal tissue and expression in breast tumour tissue may prevent therapeutic benefit of HER2 targeted therapies (Koletsa et al., 2008). Oligonucleotides to switch HER2 splicing in favour of exon 15 skipping generate the Δ15HER2 variant. Treatment of the HER2 overexpressing breast cancer cell line SK-BR-3 with the splice switching oligonucleotides or with exogenous Δ15HER2 resulted in down-regulation of full length HER2 expression, inhibited proliferation and induced apoptosis. Therefore it was proposed that these splice switching oligonucleotides or administration of HER2 spliced variants may provide novel therapeutics in HER2 positive breast cancer (Wan et al., 2009).

1.4.11 Alternative splicing of progesterone receptor pre-mRNA

The structure and function of PR-A and PR-B isoforms has been described in detail in Section 1.1. The PR status of a breast tumour is reported to have value, independent of ERα status, in predicting the outcome of endocrine therapy (Bardou et al., 2003). Furthermore, the ratio of PR-A:PR-B detected in a tumour is reported to be of prognostic significance (Hopp et al., 2004). Therefore, the accurate identification of PR isoforms, including those resulting from alternative pre-mRNA splicing is important in the management of breast cancer patients.

A 78kDa protein was detected by PR antibodies in approximately 26% of breast tumours examined (Graham et al., 1996) and this protein was later demonstrated to be N-terminally truncated and likely to be a variant of PR-A since it was detected by antibodies targeting PR-A and PR-B but not the BUS (Yeates et al., 1998). Whole exon deletions of exon 2, exon 3, exon 4, exon 6 and exon 7, and multiple exon deletions involving various combinations of internal exons have been previously reported in a range of human tissues and cell lines (Springwald et al., 2010; Misao et al., 2000; Balleine et al., 1999b; Misao et al., 1998; Richer et al., 1998; Leygue et al., 1996a). Exon 4 deleted mRNA (PRΔ4) and exon 6 deleted mRNA (PRΔ6) were detected more frequently in breast tumour than normal breast tissue, PRΔ6 was detected more frequently in breast tumours characterised as expressing a low level of PR than a high level of PR (Nagao et al., 2003) and deletions of exons 3+6 and exons 5+6 were detected specifically in breast tumour tissue (Leygue et al., 1996a). PRΔ4 and PRΔ6 were also detected in vascular smooth muscle, specifically in pre-menopausal women. The post menopausal patients examined had all undergone HRT, leading to the suggestion that the exogenous oestrogens administered in HRT may affect the alternative splicing of PR (Hodges et al., 1999).

PR exons 3, 4 and 6 are cassette exons, therefore deletion of the exon would result in a protein lacking just the functional domain encoded by that exon, whereas deletion of other exons or multiple exons often changes the reading frame and would prematurely truncate the C-terminal end of any encoded protein. A study to examine the function of exon deleted PR variants reported that PRΔ2, PRΔ4 and PRΔ5,6 were unable to bind DNA, whereas PRΔ6 exhibited constitutive DNA binding in the absence of ligand.

Furthermore PRΔ5,6 and PRΔ6 were dominant negative regulators of PR-A and PR-B function (Richer et al., 1998). However, the low level of exon deleted PR mRNA detected, PRΔ6 was reported at just 8% of wild-type PR mRNA levels, has led to the suggestion that any encoded proteins are likely to be expressed at very low levels and may not be functionally relevant (Balleine et al., 1999b).

As described in Section 1.4.7, the use of alternative cryptic splice sites is observed more frequently in cancer than normal tissue (Kim et al., 2008). Use of two cryptic 3′ splice sites have been reported for PR, within exon 6 an exon 4; PRΔ6/2 mRNA with an out of frame deletion of 52bp from the start of exon 6 was detected more frequently in breast cancer than normal breast tissue and would lead to protein truncation within the DBD. PRΔ4/2 mRNA with an in frame deletion of 126bp at the start of exon 4 was detected in breast cancer tissue (Balleine et al., 1999b) and normal endometrium (Marshburn et al., 2005). This mRNA would encode a protein lacking the hinge region and NLS but possessing the remaining functional domains.

Another mode of alternative splicing common in cancer is intron retention (Kim et al., 2008). Several PR isoforms have been reported to result from the retention of intronic exons. Sequencing of genomic DNA and screening of the testicular cDNA library identified a sequence within intron 3 which was spliced onto the start of exon 4. mRNA containing this exon, termed exon S, was detected at a high level in spermatozoa and also in the endometrium (Hirata et al., 2000). Further screening of the testicular cDNA library identified another retained sequence from intron 3, termed exon T (Hirata et al., 2002). Neither exon S nor exon T encode protein and these mRNA are predicted to generate identical proteins from the first initiation codon within exon 4 (Met692) (Hirata et al., 2003). Retention of a 232bp intronic sequence between exons 4 and 5 was identified in endometrial tissue by RT-PCR. Sequencing of genomic DNA identified two exons termed i45a (123bp) and i45b (109bp). i45a encodes an in frame termination codon so would generate a protein truncated within the LBD (Yamanaka et al., 2002). Another insertion between exons 3 and 4, termed exon M, was identified by screening cDNA libraries and cDNA containing this sequence was later identified in PR+ T47D and PR- T47D-Y breast cancer cells. Exon M contains an initiation codon so the PR-M protein is predicted to possess 16 unique N-terminal amino acids encoded by exon M

along with the normal PR hinge region, NLS, LBD, DD and AF2, with a theoretical molecular weight of 38kDa. A protein of appropriate molecular weight was detected in the membrane fraction of T47D-Y cells by Western immunoblotting with the C-terminal PR antibody C19 (Saner et al., 2003) and within the cytoplasm of breast cancer cells by immunofluorescence using the same antibody (Price et al., 2005).

A recent study using expression vectors containing each initiation codon predicted to generate an N-terminally truncated PR protein, i.e. Met692 (PR-S/T), MetPR-M, or Met595 (PR-C), reported that none of these lead to translation unless a consensus Kozak sequence is added, and suggested that a protein translated from Met595 would be 38kDa not the previously reported 60kDa PR-C protein (Samalecos and Gellersen, 2008). This study also suggested that commercial C-terminal PR antibodies are non-specific, supporting the earlier identification by mass-spectrometry of cytoskeletal proteins detected by the C19 antibody, including the protein commonly reported as PR-C (Madsen et al., 2007). These studies challenge the existence of N-terminally truncated PR variant isoforms and the use of current antibodies to identify them. This controversy necessitates the study of the expression profile and activity of PR spliced variants in breast cancer cell lines and tissue with the aim of improving the understanding of PR function. The improved detection of PR proteins with novel antibodies could contribute to the identification of variant proteins both in the lab and in clinicopathological breast cancer screening, contributing to the value of PR as a clinical biomarker. Furthermore, the identification of apparently breast cancer specific PR ASEs such as deletion of exons 3+6 or exons 5+6 (Leygue et al., 1996a), and the increased occurrence of other ASEs in breast cancer such as deletion of exon 4, exon 6 (Nagao et al., 2003) or usage of the exon 6 cryptic splice site (Hisatomi et al., 2003), suggest that a full identification of the PR splicing signature may provide molecular markers for breast cancer diagnosis. These cancer specific variant PR mRNAs also demonstrate a requirement for investigation into the regulation of PR alternative splicing to identify the factors which may regulate cancer specific ASEs.

1.5 Hypotheses and aims

1.5.1 Hypotheses

Breast cancer PR screening uses N-terminal antibodies (Hanley et al., 2009) which may detect truncated proteins resulting from exon deletion that are functionally distinct from wild type PR, or may fail to detect N-terminally truncated proteins. Evidence exists for the detection of truncated PR proteins by C-terminally targeted PR antibodies (Condon et al., 2006), however the identity of these truncated proteins and the specificity of commercial PR antibodies remains controversial (Samalecos and Gellersen, 2008; Madsen et al., 2007). These truncated PR proteins may result from alternative splicing of PR pre-mRNA and ASEs involving PR mRNA have been extensively characterised in the PR positive MCF-7 and T74D breast cancer cell lines. However, until recently (Springwald et al., 2010), no evidence of PR expression had been reported in the MDA-MB-231 breast cancer cell line which is characterised as PR negative. These ASEs have the potential to encode proteins with functions that differ from PR-A and PR-B. Therefore the hypotheses of this project are:

1. The PR negative breast cancer cell line MDA-MB-231 may express alternatively spliced PR mRNA which could encode potentially functional PR variant proteins.

2. Variant PR proteins resulting from alternative splicing of PR pre-mRNA may be expressed in breast cancer and contribute to misclassification of a breast tumour PR status in screening using N-terminally targeted antibodies.

3. The lack of specificity associated with commercial non N-terminal PR antibodies may make them unsuitable for clinical PR screening and necessitates the development of novel antibodies.

4. The use of novel PR screening antibodies which detect truncated PR variant proteins may add to the predictive value of PR status for administration of endocrine therapies.

1.5.2 Aims

To investigate the hypotheses described above the specific aims of this study are to:

1. Develop an RT-PCR based gene walking assay using primers directed to each exon of the PR gene to identify PR ASEs. PR positive MCF-7 cells will be used to validate this assay.

2. Use the PR gene walking assay and primers directed to previously identified PR spliced variants to characterise PR ASEs in MDA-MB-231 cells.

3. Identify PR proteins detected by commercial PR antibodies by Western immunoblotting in breast cancer cell lines.

4. Validate the specificity of commercial N- and C-terminal PR antibodies using siRNA knockdown of total PR expression and Western immunoblotting with each PR antibody.

5. Investigate the potential functionality of PR proteins present in MDA-MB-231 cells using a range of assays including ligand blotting, co-immunoprecipitation with a PR co-factor and DNA affinity precipitation.

6. Characterise PR ASEs in a small cohort of breast tumour samples using the PR gene walking assay.

7. Develop novel non N-terminal PR antibodies and undertake preliminary validation of their specificity for use in Western immunoblotting, co-immunoprecipitation and immunofluorescent analysis of FFPE breast tumour tissue.

8. Undertake a preliminary investigation into the regulation of PR alternative splicing by examining the effect of SR protein knockdown by siRNA on exon inclusion.

2 Materials and Methods

2.1 Source of tissue

Frozen breast tumour tissue was obtained from the Newcastle University/Gateshead Queen Elizabeth Hospital Breast Cancer Biomarkers tissue bank, and used in accordance with Gateshead Local Research Ethics Committee Ref 52/02. Samples were taken from consenting patients and snap frozen in liquid nitrogen before storing at -80°C. Samples from the same tumour were formalin fixed and embedded in paraffin wax. Formalin fixed paraffin embedded (FFPE) breast tumour tissue was used on site at Gateshead Queen Elizabeth Hospital Histocytopathology Department in accordance with local ethics permission.

2.2 Cell culture materials

2.2.1 Cell lines

MCF-7 (MD Anderson) and T47-D (American Type Culture Collection, (ATCC)) are ER+/PR+ human breast cancer cell lines. MDA-MB-231 (MD Anderson and ATCC) is an ER-/PR- human breast cancer cell line. MCF-10A (ATCC) is a non-tumourigenic mammary epithelial cell line. Characteristics of these different cell lines can be seen in Table 2.1.

2.2.2 Cell culture media

Dulbecco's Modified Eagles medium (DMEM) without phenol red (Sigma-Aldrich), was supplemented with 10% Foetal Bovine Serum (FBS; Sigma-Aldrich), 20mM L-Glutamine (Sigma-Aldrich) and 1% penicillin/streptomycin (pen/strep; Sigma-Aldrich,). RPMI-1640 without phenol red (Sigma-Aldrich) was supplemented with 10% FBS, 20mM L-glutamine and 1% pen/strep. DMEM/F12 (Invitrogen, Life Technologies) was supplemented with 5% horse serum (Invitrogen, Life Technologies), 20ng/ml EGF (Peprotech), 0.5mg/ml hydrocortisone (Sigma-Aldrich), 100ng/ml cholera toxin (Sigma-Aldrich), 10μg/ml insulin (Sigma-Aldrich) and 1% pen/strep. Cryoprotective media consisted of FBS+ 5% dimethyl sulphoxide (DMSO, Sigma-Aldrich; MCF-7, T47-D,

MDA-MB-231) or complete DMEM/F12 + 7.5% DMSO (MCF-10A). Details of cell culture media can be seen in Table 2.1.

Cell line	ER	PR	Tumorigenic	Origin	Growth media	Cryo-media
MCF-7	+	+	+	Plural effusion from adenocarcinoma of the breast	DMEM	FBS + 5% DMSO
T47-D	+	+	+	Plural effussion from infiltrating ductal carcinoma	RPMI-1640	FBS + 5% DMSO
MDA-MB-231	-	-	+	Pleural effusion from metastatic site in patient with adenocarcinoma of the breast	DMEM	FBS + 5% DMSO
MCF-10A	-	-	-	Fibrocystic disease of the breast	DMEM/F12	DMEM/F12 + 7.5% DMSO

Table 2.1: Cell lines and cell culture media used in this study.

2.3 Cell culture methods

2.3.1 General cell maintenance

MCF-7 and MDA-MB-231 cells were routinely cultured in complete DMEM without phenol red. T47-D cells were cultured in complete RPMI-1640 without phenol red. MCF-10A cells were cultured in complete DMEM/F12. All cell lines were routinely cultured in 75cm^2 or 25cm^2 tissue culture flasks (Greiner) at 37°C in a humidified atmosphere containing 5% CO_2. Cells were passaged by removing growth media, washing with PBS and stripping from the surface of the flask by incubation with 2mM trypsin-EDTA (Sigma-Aldrich). Trypsin-EDTA was neutralised by addition of complete growth media and cells were pelleted by 5 minutes centrifugation at 1500rpm, before resuspending and re-seeding in complete growth media. Cells were passaged at a ratio of 1:2-1:4 when they reached approximately 70-80% confluency.

2.3.2 Cryo-storage

For long term storage cells were pelleted as above, and resuspended in an appropriate cryoprotective media (Table 2.1). Cells were stored in 2ml cryovials (Simport) in liquid nitrogen or at -80°C.

2.3.3 Hormone stimulation of cultured cells

MCF-7, MDA-MB-231 and MCF-10A cells were seeded at a density of $5x10^4$ cells per well in 24 well culture plates and cultured until approximately 60% confluency before serum starving by replacing complete growth media with media without FBS. After 24 hours serum starving, media was replaced with complete media supplemented with a range of concentrations of β-oestradiol (Sigma-Aldrich) and cells were cultured for a further 24 hours. β-oestradiol was dissolved in absolute ethanol and added to the media at a range of concentrations from 0.001-10μM. An 'ethanol only' control (labelled 0μM was also included for each experiment. Experiments were repeated twice.

2.3.4 Transfecting cultured cells with siRNA to knockdown gene expression

siRNA was purchased from Ambion, Applied Biosystems. Conditions for efficient siRNA knockdown were optimised by varying the cell density per well and the concentrations of siRNA and transfection reagent within the ranges recommended in the Ambion *Silencer* siRNA Starter Kit manufacturer's guidelines. For each siRNA, 2.5μl siPORT *NeoFX* Transfection Agent (Ambion, Applied Biosystems) was added to 100μl DMEM without FBS and antibiotics, and incubated for 10 minutes to allow equilibration of the transfection agent to a working temperature. 2.5μl of 2μM siRNA in 100μl DMEM without FBS or pen/strep was added to the siRNA and incubated for a further 20 minutes to allow formation of siRNA/transfection agent complexes. siRNA/siPORT mixes were added to the wells of 12 well culture plates along with MCF-7 cells at a density of $1x10^5$ cells per well in DMEM without antibiotics. Media was replaced after 24 hours and cells cultured for a further 48hours prior to harvesting.

PR siRNA was used to knockdown total PR expression and validate antibody specificity by Western immunoblotting and immunofluorescence. siRNA targeted to knockdown expression of the RNA-binding proteins SRSF1, SRSF2, SRSF5, and SRSF6 was used to examine the effect on PR alternative splicing events (ASEs) by RT-PCR. Cells were also transfected with a positive transfection control GAPDH siRNA and a negative control scrambled sequence siRNA in each experiment. A transfection agent only control was also performed and untreated cells were cultured in parallel for comparison. Each siRNA knockdown was performed in duplicate and repeated twice. Details of all siRNAs used can be seen in Table 2.2.

siRNA target	Sense (5'-3')	Antisense (5'-3')
PR	GGUUUUCGAAAGUUACAUAtt	UAUGUAAGUUUCGAAAACCtg
SRSF1	GGAUAACACUAAGUUUAGAtt	UCUAAACUUAGUGUUAUCCag
SRSF1	GCAUCUACGUGGGUAACUUtt	AAGUUACCCACGUAGAUGCgg
SRSF2	GCACUAUCCUCUUAGAGAAtt	UUCUCUAAGAGGAUAGUGCat
SRSF2	GCACGAAGGUCCAAGUCCAtt	UGGACUUGGACCUUCGUGCgg
SRSF5	CCACCUGUAAGAACAGAAAtt	UUUCUGUUCUUACAGGUGGag
SRSF5	GGACGAUACUCUGACCGUUtt	AACGGUCAGAGUAUCGUCCtc
SRSF6	CAAAUGAGGGUGUAAUUGAtt	UCAAUUACACCCUCAUUUGtt
SRSF6	CGUUCUCGAUCAAAAGGCAtt	UGCCUUUUGAUCGAGAACGtg
GAPDH	Sequence not provided	
Negative control	Sequence not provided	

Table 2.2: Targets and RNA sequences of siRNA used in this study.

2.4 Preparation of protein samples

Whole cell lysates from MCF-7 and MDA-MB-231 cells were produced by two methods depending on the analysis method to be used. SDS Lysis Buffer was used to harvest cells for Western immunoblotting. Radioimmunoprecipitation assay (RIPA) buffer was used to extract lysates for shrimp alkaline phosphatase (SAP) treatment and ELISAs. Both lysis methods and details of how the lysates were used are detailed below.

2.4.1 SDS lysis buffer

0.125M Tris-HCl pH6.8

2% SDS

10% glycerol

10% β-mercaptoethanol

0.1% bromophenol blue

2.4.2 Preparation of whole cell lysates using SDS lysis buffer

Growth media was removed and cells washed twice in cold 1 x phosphate buffered saline (PBS). Cells in 12 well culture plates were then incubated for 5 minutes at room temperature in 100µl SDS lysis buffer before scraping using a P200 pipette tip and transferring to a 1.5ml centrifuge tube. Following aspiration of growth media and washing in PBS, cells in 75cm^2 culture flasks were scraped into 1ml 1xPBS using a cell scraper and pelleted by 5 minutes centrifugation at 600rpm in a refrigerated centrifuge at 4°C. Cell pellets were resuspended in 750µl SDS lysis buffer.

2.4.3 RIPA buffer

25mM Tris-HCl pH7.6

150mM NaCl

1% NP-40

0.1% sodium deoxycholate

0.1% SDS

2.4.4 Preparation of whole cell lysates using RIPA buffer

MCF-7 and MDA-MB-231 cells were cultured in 25cm^2 tissue culture flasks and stimulated with 1µg/µl β-oestradiol, with or without subsequent stimulation with 0.1µg/µl progesterone or an ethanol only control as previously described. Before use, cold RIPA buffer was supplemented with 20µl protease inhibitor cocktail (Sigma-Aldrich) per 1ml of buffer. Control cells were lysed with RIPA buffer without protease inhibitors. Cells were washed 3 times in cold 1xPBS before incubation on ice with 330µl cold RIPA buffer for 5 minutes under gentle agitation. Lysate was collected using a cell scraper, transferred to a 1.5 ml microcentrifuge tube and centrifuged for 15 minutes at 14,000xg to pellet cell debris. Supernatant was removed and stored for quantification and further use.

2.4.5 Preparing subcellular fractions: Protocol 1

Nuclear and cytoplasmic fractions were prepared from cultured MCF-7 and MDA-MB-231 cells using the Universal Magnetic Co-IP kit (Active Motif) following the manufacturer's protocol. Briefly, cells were washed in PBS containing protease inhibitors. Cells were then scraped into 1x complete hypotonic buffer (containing protease inhibitors, phosphatise inhibitors and PMSF) and pelleted by centrifugation at 4°C. The supernatant (cytoplasmic fraction) was removed and stored at -20°C. The pellet was resuspended in complete digestion buffer (containing protease inhibitors, phosphatase inhibitors, PMSF and detergent) and incubated on ice for 20 minutes. Enzymatic shearing cocktail was added and incubated for 10 minutes at 37°C with occasional vortexing. EDTA was added to inactivate the enzymatic shearing complex and cells pelleted by centrifugation at 4°C. The supernatant (nuclear fraction) was removed and stored at -80°C.

2.4.6 Preparation of subcellular fractions: Protocol 2

Cytosolic, membrane, nuclear and cytoskeletal fractions were prepared from MCF-7 and MDA-MB-231 cells using the ProteoExtract Subcellular Proteome Extraction Kit (Calbiochem) following a modified version of the protocol provided. Briefly, cells in 25cm^2 culture flasks treated with 1µM β-oestradiol, 1µM β-oestradiol followed by 0.1µM progesterone or an ethanol only control were washed twice in cold Wash Buffer.

Cells were then gently scraped from the surface of the flask using a cell scraper into cold Extraction Buffer I + protease inhibitors, transferred to a pre-cooled microcentrifuge tube and incubated at 4°C on a rotary shaker for 10 minutes, before pelleting by 10 minutes centrifugation at 4°C, 1000xg. The supernatant (cytosolic fraction) was removed and stored at -20°C. The pellet was resuspended in cold Extraction Buffer II + protease inhibitors and incubated at 4°C for 30 mins on a rotary shaker before pelleting by 10 minute centrifugation at 4°C, 6000xg. The supernatant (membrane fraction) was removed and stored at -20°C. The pellet was resuspended in cold Extraction Buffer III + protease inhibitors and Benzonase nuclease, incubated for 10 minutes at 4°C on a rotary shaker, then pelleted by 10 minutes centrifugation at 4°C, 6800xg. The supernatant (nuclear fraction) was removed and stored at -20°C. The pellet, resuspended in room temperature Extraction Buffer IV formed the cytoskeletal fraction and was stored at -20°C.

2.4.7 Protein quantification

Protein concentrations of cell fractions and whole cell lysates in RIPA buffer were quantified using the BioRad DC Protein Assay following the manufacturer's protocol. Absorbances were read using a SpectraMax 190 plate reader (Molecular Devices), and protein concentrations calculated using SoftMax Pro software (Molecular Devices).

2.4.8 Protein storage

Whole cell lysates in SDS lysis buffer were stored at -20°C. Following protein quantification, RIPA whole cell lysates and nuclear and cytoplasmic fractions from protocol 1 (2.4.5) were divided into 100µg aliquots and stored at -20 or -80°C. Proteome fractions from protocol 2 (2.4.6) were divided into 10µg aliquots and stored at -20°C.

2.5 Shrimp Alkaline Phosphatase treatment

Whole cell protein lysates in RIPA buffer were prepared as previously described from MCF-7 and MDA-MB-231 cells stimulated with 1µg/µl β-oestradiol or an ethanol only control. 30µg of protein was used for each 30µl reaction with 10µl shrimp alkaline

phosphatase (SAP) and an appropriate volume of 10x SAP reaction buffer. Control reactions contained 10µl RIPA buffer instead of SAP. Reactions were incubated for 1 hour at 37°C before heat inactivation of the enzyme at 65°C for 15 minutes. Samples were analysed by Western immunoblotting as described below following addition of 30µl 2xSDS lysis buffer.

2.6 Co-Immunoprecipitation

Co-immunoprecipitation (Co-IP) was carried out using the Universal Magnetic Co-IP kit (Active Motif), following the manufacturer's protocol. Briefly, 50µg nuclear or cytoplasmic extract was incubated for 4 hours at 4°C on a rotary shaker with a capture antibody, or an appropriate IgG control, in a total volume of 250µl of complete Co-IP buffer. 12.5µl of Protein G magnetic beads (Active Motif) were added to each tube and incubated for 1 hour at 4°C on a rotary shaker. Antibody/protein complexes bound to protein G beads were precipitated using a magnetic tube stand, and the beads washed by resuspending in complete Co-IP buffer and re-precipitating 4 times. The beads were then resuspended in 20µl SDS lysis buffer. Protein/antibody/bead complexes were denatured by heating to 95°C for 5 minutes. Proteins in SDS lysis buffer were then removed from the beads and stored at -20°C. A list of antibodies used can be seen in Table 2.8.

2.7 DNA affinity precipitation assay (DAPA)

2.7.1 Oligonucleotides

Biotinylated single strand DNA oligonucleotides were purchased from Invitrogen, Life Technologies. Sequences for the sense strands of the consensus PRE, and mutated PRE sequence oligonucleotides are shown below. Bases altered in the mutated sequence are underlined in bold font:

Consensus PRE: GATCCTGTCACGGATGTTCTAGCTACA

Mutated PRE: GATCCT**CA**ACAGGAT**CA**TCTAGCTACA

Unbiotinylated (cold) PRE oligonucleotides of the same sequences were purchased from Santa Cruz Biotechnology.

2.7.2 Annealing buffer

10mM Tris-HCl pH 7.5,

50mM NaCl

1mM EDTA

2.7.3 Annealing oligonucleotides

Complementary oligonucleotides were resuspended to the same molar concentration in annealing buffer. Equal volumes of each oligonucleotide were mixed in a 1.5ml microcentrifuge tube and placed in a heat block at 95°C. The heat block containing the annealing reactions was removed from the apparatus and allowed to cool to room temperature on the bench to allow complete annealing of complementary oligonucleotides.

2.7.4 DAPA binding buffer

12mM HEPES pH7.9

4mM TRIS-HCl

60mM KCl

5% glycerol

0.5mM EDTA

2.7.5 DAPA protocol

Nuclear and cytoplasmic fractions were isolated from MCF-7 and MDA-MB-231 cells following hormone stimulation using the Universal Magnetic Co-IP kit as described previously. 50µg of nuclear or cytoplasmic extract was incubated in 400µl DAPA binding buffer for 1 hour on ice with gentle agitation, with 0.25µg synthetic biotinylated double stranded oligonucleotide for a consensus PRE sequence or a mutated PRE sequence. 40µl streptavidin magnetic beads were equilibrated in DAPA binding buffer and added to each reaction before incubation at 4°C for 2 hours on a rotator to prevent

settling of the beads and allow formation of protein/PRE/bead complexes. Unbound proteins were removed by washing six times in 1ml of DAPA binding buffer. Beads were then resuspended in 30µl SDS lysis buffer to dissociate protein/PRE complexes. Samples were heated to 95°C for 5 minutes to denature proteins before removing from the beads by magnetic separation. Samples were stored at -20°C for future analysis by Western immunoblotting.

2.8 SDS-PAGE and Western immunoblotting

2.8.1 Materials

2.8.1.1 Polyacrylamide gel

Stacking gel: 4% acrylamide (BioRad), 125mM Tris-HCl pH6.8, 1% sodium dodecyl sulphate (SDS, Sigma-Aldrich), 0.1% TEMED (Sigma-Aldrich) and 0.5% ammonium persulphate (APS; Sigma-Aldrich).

Resolving gel: 10% acrylamide, 375mM Tris-HCl pH8.8, 0.1% SDS, 0.1% TEMED and 0.1% ammonium persulphate.

2.8.1.2 Electrode buffer

1x electrode buffer contained 25nm Tris-HCl, 190mM glycine and 0.1% SDS.

2.8.1.3 Transfer buffer

1x transfer buffer contained 25nM Tris-HCl, 0.15M glycine and 10% methanol.

2.8.1.4 Phosphate buffered saline (PBS)

137mM NaCl

2.7mM KCl

8.1mM $Na_2HPO_4.H_20$

1.76mM KH_2PO_4

2.8.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate proteins for Western immunoblot analysis. Samples were loaded into wells in a 4% stacking gel and separated by electrophoresis using a 10% acrylamide resolving gel. Protein samples were prepared as described previously. Proteins in RIPA buffer and cell fractions were denatured by addition of an appropriate volume of 2xSDS lysis buffer, and all samples were heated to 95°C for 5 minutes prior to loading. PageRuler Prestained Protein Ladder (Fermentas) was used to determine protein size and track migration. Electrophoresis was performed in 1x electrode buffer at 200V using a BioRad mini-PROTEAN Tetra System.

2.8.3 Protein transfer

Proteins were transferred onto nitrocellulose membrane (BioRad) by electrophoresis in 1x transfer buffer at 90V for 2 hours.

2.8.4 Ponceau S staining

Equal protein loading was assessed by Ponceau S staining (data not shown). Membranes were incubated for 1 minute in Ponceau S solution (Sigma-Aldrich) and washed in distilled water. Membranes were de-stained by washing in 1xPBS.

2.8.5 Blocking of non-specific proteins

Unbound sites on membranes were blocked by incubation in a blocking solution of 10% non-fat milk powder (Marvel) in 1xPBS for 1 hour at room temperature.

2.8.6 Primary antibody incubation

Primary antibody incubations were performed at an appropriate dilution in a solution of non-fat milk powder in 1xPBS or 1xPBS + 0.01% Tween20 (PBST), either for 1 hour at room temperature or overnight at 4°C on a rocking platform. Details of primary antibody concentrations can be seen in Table 2.8. Membranes were washed twice for 5 minutes and once for 20 minutes in 1xPBS following primary antibody incubation. The membranes were then incubated with a horseradish peroxidise (HRP) conjugated secondary antibody (Dako), diluted 1:3000 in 1xPBS, for 1 hour at room temperature. Washes were repeated after secondary antibody incubation. The approximate regions of the PR protein targeted by each commercial PR antibody can be seen in Figure 2.1.

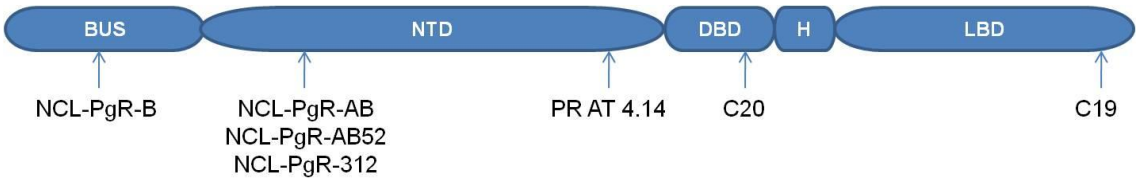


Figure 2.1: Approximate regions targeted by each commercial PR antibody used in this study.

BUS= PR-B specific B-upstream segment, A/B= PR-A and B N-terminal region, C= DBD, D= hinge/NLS, E= LBD/DD.

2.8.7 Detection and analysis

Proteins were detected using ECL Western Blotting Substrate (Pierce), or SuperSignal West Femto (Pierce) and exposure to x-ray film (KODAK). Autoradiographs were developed using an X-ograph Compact X4 X-ray film processor. Densitometric analysis of protein bands detected following Western immunoblotting was performed using Intelligent Quantifier (BioImage Systems, Inc).

2.9 Ligand blot assay

2.9.1 Cell fractionation

Cytosolic, membrane, nuclear and cytoskeletal fractions were prepared using the ProteoExtract Subcellular Proteome Extraction Kit as previously described.

2.9.2 Protein separation

7µg of each fraction was diluted with an appropriate volume of 2xSDS lysis buffer and denatured by heating to 95°C. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membrane as described previously.

2.9.3 Ligand blotting

Ligand blotting was carried out following a published protocol (Jang and Yi, 2005). Briefly, membranes were incubated for 30 minutes in ligand blot buffer (1x PBS + 3% Nonidet-P40), then blocked by incubating for 2 hours in a blocking buffer (1x PBS + 0.3% bovine serum albumin (BSA) + 0.1% Tween20) and washed for 10 minutes in 1x PBS + 0.1% Tween 20. Membranes were then incubated overnight at 4°C in blocking buffer + 1µM peroxidase conjugated progesterone (P-POD; Fitzgerald). Membranes were washed 3 times in 1xPBS + 0.1% Tween20 before developing using SuperSignal West Femto and X-ray film as described above.

2.10 Reverse-transcriptase polymerase chain reaction (RT-PCR)

2.10.1 RNA extraction

RNA was extracted from cultured cells or frozen breast tumour tissue using the SV Total RNA Isolation Kit (Promega) following the manufacturer's Spin Protocol. Briefly, cells cultured in a 75cm^2 culture flask were harvested directly into 700µl RNA Lysis Buffer. 175µl cell lysate was used for RNA extraction. Following a DNase I treatment step, RNA was eluted in nuclease free water provided with the kit. RNA was extracted

from cells in 12 well culture plates after siRNA treatment following the same protocol, with cells harvested into 175µl RNA Lysis Buffer per well. Approximately 30mg frozen breast tumour tissue was homogenised in liquid nitrogen using a pestle and mortar and lysed using 175µl RNA Lysis Buffer. RNA was extracted following the manufacturer's Spin Protocol. RNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (ThermoScientific). RNA samples were stored at -80°C.

2.10.2 Reverse transcription

First strand cDNA was synthesised from total RNA by reverse transcription using SuperScript III Reverse Transcriptase (ssRT; Invitrogen, Life Technologies). 1µg total RNA was mixed with 10µl 5x First Strand Synthesis Buffer (Invitrogen, Life Technologies), 2µl 40mM PCR nucleotides (Promega), 1µl 0.5µg/µl oligo dT_{20} primers (Invitrogen, Life Technologies), 1µl 0.1M DTT (Invitrogen, Life Technologies), 1µl RNAse OUT Recombinant Ribonuclease Inhibitor (Invitrogen, Life Technologies) and 1µl ssRT. Reaction volumes were adjusted to 50µl using nuclease free water (Invitrogen, Life Technologies). Reactions were performed at 50°C for 1 hour and cDNA was stored at -20°C. A no RT reaction was set up in parallel to each cDNA synthesis reaction as a control for genomic DNA contamination; 1µl nuclease free water was added in place of ssRT.

2.10.3 Design of the PR gene walking assay

A PR gene walking assay was developed using primers directed to each of the eight exons of the PR gene. Primers were designed based on the PR mRNA sequence (NCBI Accession: NM_000926.4) and synthesised by Invitrogen, Life Technologies. Sense primers directed to the PR-B and PR-A promoter regions, regions within exon 1 common to PR-A and B, and to each of the other 7 exons. Antisense primers were designed to each of the 8 exons, and 3 within exon 1. Primers were used in combination to examine different regions of the PR gene and assess any alternative splicing which may be occurring by comparing the size of PCR product to the predicted size base on the PR gene sequence (NCBI Accession: NM_000926.4). The predicted sizes of PCR products produced using each primer combination in the PR gene walking assay can be

seen in table 2.3. PCR reactions were performed using all combinations of primers giving products ≤1kb using MCF-7 and MDA-MB-231 cDNA.

	A10	A9	A8	A7	A6	A5	A4	A3	A2	A1
S1	517	907	1386	1650	1836	1936	2231	2371	2500	2778
S2	270	660	1139	1403	1598	1689	1984	2124	2253	2531
S3		407	886	1150	1336	1436	1731	1871	2000	2278
S4			497	761	949	1047	1342	1482	1611	1889
A				40	226	326	621	761	890	1168
S6					203	303	598	738	867	1145
S7					51	151	446	586	715	1003
S8							317	457	586	864
S9								160	289	567
S10									147	425
S11										295

Table 2.3: Predicted PCR product size (bp) of each primer combination in the PR gene walking assay.

Combinations with product sizes in the unshaded area (products <1kbp) were used for PCR with MCF-7 and MDA-MB-231 cDNA. S=sense primer, A=antisense primer.

2.10.4 PCR reactions for gene walking assay

25µl PCR reaction mixes consisted of 2µl cDNA template, 0.25µl of each primer (sense and antisense, 0.5µg/µl), 12.5µl PCR Master Mix (Promega) and 10µl DNase free water (Promega). All PR primers were designed to be used at an annealing temperature of 55°C. 25-30 cycles of PCR were initially performed using RNA from MCF-7, T47-D and MDA-MB-231 cells. PCR cycle number was increased to 45 to detect low levels of truncated PCR products in MDA-MB-231 cells for cloning and sequencing. PCR reaction conditions were as follows: an initial denaturation step of 95°C for 5 minutes, followed by 30-45 cycles consisting of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 30-60 seconds depending on the size of PCR product to be generated. This was followed by a final extension at 72°C for 5 minutes and cooling to 4°C for storage. To demonstrate that no contamination was present in PCR master mix or primers, a no template control (NTC) was included for each primer set. Reaction mixtures contained an extra 2µl DNase free water instead of cDNA. As a control for genomic DNA contamination, PCR reactions using GAPDH primers were performed using the no RT control template described in Section 2.10.2. Primer sequences used in the PR gene walking assay can be seen in Table 2.4.

Primer name	Target	Sequence (5'-3')	Annealing temp. (°C)
Sense 1	PR exon 1 (BUS)	AGCTGAAGGCAAAGGGTC	55
Sense 2	PR exon 1 (BUS)	CGCATATTCCAGAGCTGA	55
Sense 3	PR exon 1	TGCAAGGTTGGAGACAGC	55
Sense 4	PR exon 1	ACGGTGATGGATTTCATCC	55
Sense A	PR exon 1	CCCTATCTCAACTACCTGA	55
Sense 6	PR exon 2	GGATTCAGAAGCCAGCCA	55
Sense 7	PR exon 3	AGCACAACTACTTATGTGC	55
Sense 8	PR exon 4	AGTTCAATAAAGTCAGAGTTG	55
Sense 9	PR exon 5	GAAACTTACATATTGATGACC	55
Sense 10	PR exon 6	GCGGATGAAAGAATCATCA	55
Sense 11	PR exon 7	CCTTTGGAAGGGCTACGA	55
Antisense 10	PR exon 1	GCTGTCTCCAACCTTGCA	55
Antisense 9	PR exon 1	GTGGATGAAATCCATCACC	55
Antisense 8	PR exon 1	GCTTTGTACAGGATGCAC	55
Antisense 7	PR exon 2	TGGCTGGCTTCTGAATCC	55
Antisense 6	PR exon 3	CGGATTTTATCAACGATGC	55
Antisense 5	PR exon 4	CACAACTCTGACTTTATTGAA	55
Antisense 4	PR exon 5	GGTCATCAATATGTAAGTTTC	55
Antisense 3	PR exon 6	TGATGATTCTTTCATCCGC	55
Antisense 2	PR exon 7	TCGTAGCCCTTCCAAAGG	55
Antisense 1	PR exon 8	AGAAGGGGTTTCACCATC	55

Table 2.4: Sequences and annealing temperatures of the PCR primers used in the PR gene walking assay.

2.10.5 PR intron retention PCR

To identify the presence of mRNA for previously identified PR isoforms resulting from retention of intronic sequences, primers directed to introns M, S, T and i45b were used. A sense primer was designed based on the published PR-M intronic sequence (Saner et al., 2003) and previously published sense primers were used to detect PR-S and PR-T mRNA (Hirata et al., 2002; Hirata et al., 2000). A previously reported antisense primer directed to PR-i45 intron b was also used (Yamanaka et al., 2002). Primer sequences can be seen in Table 2.5. Primers were synthesised by Invitrogen, Life Technologies. PCR was performed using the same reaction conditions described for the PR gene walking assay.

Primer name	Target	Sequence (5'-3')	Annealing temp. (°C)
Sense S	PR intronic exon S	CAGGAGAGTGGGTGCTC	55
Sense T	PR intronic exon T	TCTGCAGGTCATCCCAC	55
Sense M	PR intronic exon M	GGGCTGGCAAACAGATG	55
Antisense i45	PR intronic exon 45b	CTTCCTACTTTCCCACGGA	55

Table 2.5: Sequences and annealing temperatures of the PCR primers used to examine intron retention.

2.10.6 RT-PCR to characterise PR mRNA in frozen breast tumours

PCR conditions were identical to those used in the PR gene walking assay for breast cancer cell lines. 45 cycles of PCR were performed with each 25µl reaction containing 4µl of tumour cDNA and the volume of water accordingly adjusted to 8µl. Positive control reactions used 2µl MCF-7 cell cDNA and no-template controls were included for each primer set.

2.10.7 SRSF PCR

To confirm successful knockdown of splicing factors by siRNA treatment of MCF-7 cells, PCR was performed using primers directed to each of the targeted splicing factor genes. Primers for *SRSF1, SRSF2, SRSF5 and SRSF6* were synthesised by Invitrogen, Life Technologies and used to examine siRNA knockdown of these splicing factors. GAPDH primers were used to ensure equal loading. Semi-quantitative PCR reactions were performed using the same reaction conditions described above using the annealing temperature and cycle number detailed for each primer set in Table 2.6.

Primer name	Target	Sequence (5'-3')	Annealing temp. (°c)	Cycle number
SRSF1 F	SRSF1	GCCCCGCAGGGAACAACGAT	60	26
SRSF1 R		CGTCTCGCGGGTCCTCGAAC	60	
SRSF2 F	SRSF2	AGTCTCGGTCCCGCACTCGT	60	28
SRSF2 R		CCAAAGGTGAGTAACCTCCGAGCAG	60	
SRSF5 F	SRSF5	TTGCGGATGCACACCGACCT	60	24
SRSF5 R		TCCGGCTCCTGCTCCTGCTT	60	
SRSF6 F	SRSF6	CTTGCGGTCCGCCGTTCGAC	60	26
SRSF6 R		GTATCCACCTCCACCACTGCGG	60	
GAPDH F	GAPDH	CTGCCGTCTAGAAAAACC	55	25
GAPDH R		CCAAATTCGTTGTCATACC	55	

Table 2.6: Sequences and annealing temperatures of PCR primers used to assess SRSF knockdown.

2.10.8 PCR to investigate the regulation of PR exon 4 and 6 splicing

Primers from the PR gene walking assay spanning exons 3 to 5 (S7/A4) and exons 5 to 7 (S9/A2) were used to identify the effect of SRSF knockdown in MCF-7 cells on the alternative splicing of exon 4 (S7/A4) or exon 6 (S9/A2). PCR reactions were performed following the same conditions as described for the PR gene walking assay. To allow semi-quantitative analysis of the levels of PCR products detected, 25 cycles of PCR were performed to study full length PCR products and 35 cycles to study exon deleted PCR products.

2.10.9 Ethidium bromide/agarose gel electrophoresis

Orange G DNA loading dye (Sigma-Aldrich) was added to PCR products which were then separated by electrophoresis through a 2% agarose gel (Sigma-Aldrich) in TBE buffer (Sigma-Aldrich) containing 1.5µl ethidium bromide (10mg/ml; Sigma-Aldrich) per 50ml agarose gel. PCR products were visualised under UV light using a BioSpectrum Multi Spectral Imager System (UVP). Approximate molecular weight was determined using Quick-Load 100bp DNA Ladder (New England Biolabs).

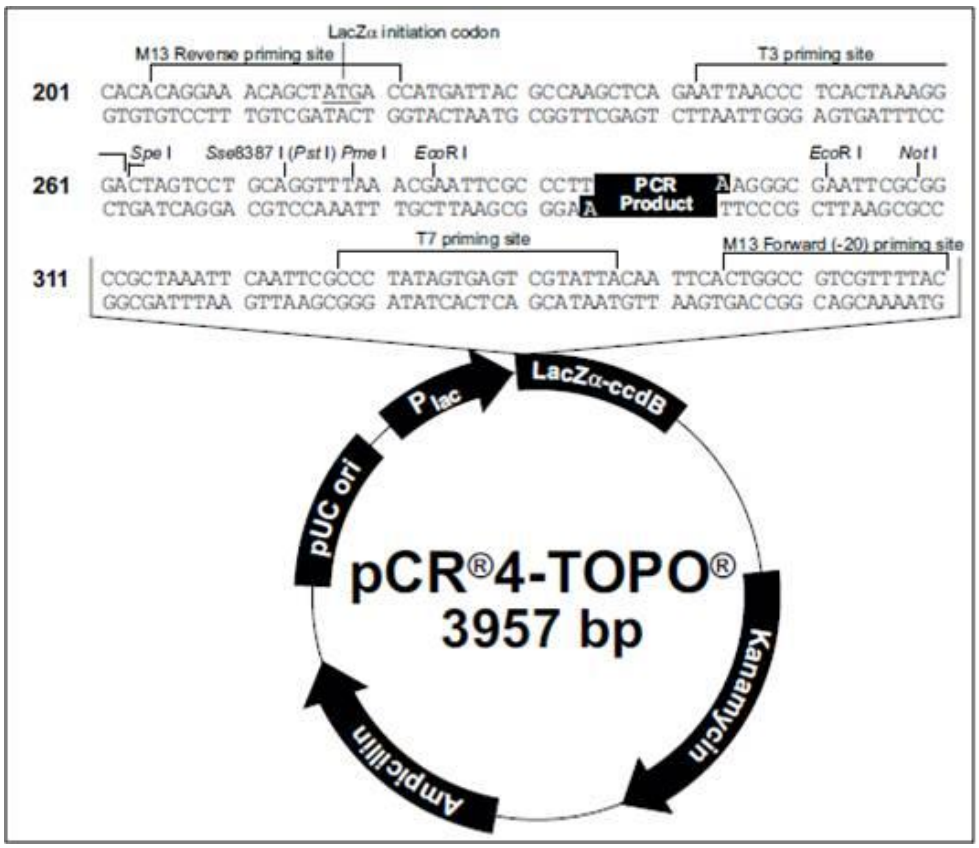
2.11 Cloning and sequencing of PCR products

2.11.1 Excision and purification of PCR products from agarose gels

Following electrophoresis and visualisation under UV light, PCR products were excised from the agarose gel using a sterile scalpel. PCR products were purified from gel slices using the Wizard SV DNA Purification Kit (Promega) following the manufacturers protocol. Briefly, excised gel slices were melted by incubation at 60°C in binding buffer, added to a spin column and incubated for 1 minute at room temperature to allow DNA to bind to the column membrane. The column was then centrifuged and washed twice by centrifugation with membrane wash solution. DNA was eluted into 25µl nuclease free water provided with the kit.

2.11.2 Ligation into TOPO TA vector

4µl purified DNA/PCR product was incubated with 1µl pCR4-TOPO plasmid (TOPO TA Cloning Kit for Sequencing; Invitrogen, Life Technologies) and 1µl salt solution provided with the kit for 30 minutes at room temperature. A map of the pCR4-TOPO plasmid can be seen in Figure 2.2.



<u>Figure 2.2:</u> Map of the pCR4-TOPO plasmid used for cloning PCR products.

PCR products were ligated into the plasmid in the region indicated. Ligation of the PCR product disrupted the *LacZα* gene preventing production of the toxic gene product, and allowed transcription of the ampicillin resistance gene enabling selection for positive clones on LB-agar plates containing 100µg/ml ampicillin. *EcoR1* restriction sites are located both sides of the PCR product insert and were used to digest the plasmid DNA for analysis of the size of DNA insert before sequencing. The location of sequencing primer sites are indicated in the sequence and M13F was routinely used for sequencing reactions by SourceBiosciences. Figure extracted from TOPO TA Cloning Kit for Sequencing User Manual (Invitrogen, Life Technologies).

2.11.3 Transformation of competent *E.coli* cells

2µl ligation mix was incubated for 20 minutes on ice in a vial of chemically competent *E.coli* (TOPO TA Cloning Kit for Sequencing; Invitrogen, Life Technologies) before heat shocking (42°C, 30 seconds), addition of 250µl room temperature SOC media (TOPO TA Cloning Kit for Sequencing; Invitrogen, Life Technologies), and 1 hour incubation at 37°C in an orbital shaker at 225rpm.

2.11.4 Lysogeny broth (LB)

10g tryptone

5g yeast extract

10g NaCl

Dissolved in 950ml distilled water, pH adjusted to 7.5 and volume made up to 1L with distilled water before autoclaving.

2.11.5 Growth of *E.coli* colonies on selective agar plates

Selective 1.5% LB-agar plates containing 100µg/ml ampicillin (Sigma-Aldrich) were produced and 125µl transformation mixture was spread onto each plate. The TOPO TA vector contains the *LacZα* gene which can is disrupted by the ligated DNA insert. When the vector is circularized, plasmid without an insert will transcribe the LacZ gene, the product of which is toxic to the cell. Non-circularized plasmid DNA will be unable to transcribe the ampicillin resistance gene, allowing selection of positive clones. Plates were incubated overnight at 37°C to allow formation of ampicillin resistant colonies containing the plasmid DNA. Isolated colonies were picked from the agar plates, added to 4ml LB media containing 100µg/ml ampicillin, and cultured overnight at 37°C in an orbital shaker at 225rpm.

2.11.6 Small scale preparation of plasmid DNA

Plasmid DNA was isolated from overnight cultures of *E.coli* using the PureYield Plasmid DNA Mini Prep Kit (Promega) following the manufacturer's protocol. Bacteria were lysed using alkaline Cell Lysis Buffer, neutralized using Neutralization Solution containing acetate and centrifuged to remove cell debris, proteins and bacterial genomic DNA. Plasmid DNA in solution was then added to a spin column, incubated for 1 minute at room temperature to allow DNA binding to the column membrane, before centrifugation to remove buffers and washing of the column to remove contaminants such as salts and endotoxin. Plasmid DNA was eluted into 25µl nuclease free water (Invitrogen, Life Technologies).

2.11.7 Restriction analysis

Restriction digest reactions consisted of 5µl plasmid DNA, 0.2µl BSA, 2µl Buffer H, 1µl *EcoR1* (Promega) and 11.8µl nuclease free water. Reactions were incubated overnight at 37°C in a water bath without shaking. Restriction digest fragments were analysed by ethidium bromide gel electrophoresis to assess the size of DNA insert in the plasmid DNA. Plasmid DNA with the correct size insert was selected for DNA sequencing.

2.11.8 DNA sequencing

Plasmid DNA sequencing was performed using M13F sequencing primers by Source BioSciences, Nottingham, UK.

2.11.9 Analysis of DNA sequences

DNA sequences were compared to the PR mRNA sequence (NCBI Accession: NM_000926.4) using the ALIGN program (http://xylian.igh.cnrs.fr/bin/align-guess.cgi) to determine the region removed by alternative splicing. The predicted amino acid sequence resulting from sequenced mRNA was determined by deleting the appropriate region (as identified by ALIGN) from the PR mRNA sequence and inputting the truncated sequence into the ExPASy Translate tool (http://web.expasy.org/translate/). This predicted amino acid sequence was then used with the ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/) to predict the molecular weight of the protein produced from each sequenced mRNA.

2.12 Bioinformatic analyses of splicing factor binding sites within the PR cDNA sequence

PR exon 4 and exon 6 mRNA sequences (NCBI Accession: NM_000926.4) were entered into the ESEfinder (Cartegni et al., 2003; http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) and Splicing Rainbow databases (http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8) to predict the presence of potential binding sites for a range of RNA binding splicing factors. ESEfinder uses a functional

SELEX method to predict the presence of binding sites; briefly, potential exon splice enhancers (ESEs) were determined by the ability of sequences to enhance splicing in a pool of mini-genes. The mini-gene splicing was then carried out in the presence of individual SR proteins (SRSF1, SRSF2, SRSF5 and SRSF6), and sequences bound to SR proteins were aligned to identify a consensus motif (Liu et al., 1998). The frequency of individual nucleotides at particular positions within the sequence was used to create a scoring system to identify how likely a sequence is to bind to an SR protein (Cartegni et al., 2003). Splicing Rainbow analyses query sequences for the presence of previously identified sequences known to be binding sites for a wide range of RNA binding proteins (RBP) including those analysed by ESE Finder, and provides a score indicating how closely each potential binding site matches the consensus for each RBP (Stamm et al., 2006).

2.13 Immunofluorescent staining

2.13.1 Cultured cells

To determine the effect of hormone stimulation on PR localisation, MDA-MB-231 cells were seeded at a density of $0.3x10^6$ cells per well in 6 well culture plates containing glass coverslips (Scientific Laboratory Supplies), cultured to 60% confluency, serum starved as previously described, and stimulated with 1μM β-oestradiol, 0.1μM progesterone or an ethanol only control. Alternatively MCF-7 cells were seeded at a density of $0.3x10^6$ cells per well in the presence of PR targeted siRNA or a negative control siRNA and cultured for 72 hours. Cells on coverslips were fixed by immersion in cold methanol and incubated at -20°C for 20 minutes. Coverslips were then air dried at room temperature before being mounted onto microscope slides (Erie Scientific) using clear nail polish.

2.13.2 Tris-buffered saline (TBS)

0.1M TRIS,

0.05M NaCl

Dissolved in distilled water, pH adjusted to 7.6.

2.13.3 Blocking of non-specific binding

Non-specific binding was blocked by the addition of a solution of 10% normal goat serum (NGS; Jackson ImmunoResearch Laboratories) in TBS and incubation at room temperature for 10 minutes.

2.13.4 Primary antibody Incubation

Primary antibody C19 was added to the coverslips at a dilution of 1:300 in 1xTBS + 1% NGS and incubated at 4°C overnight. Following removal of the primary antibody, cells were washed 3 times in TBS. Staining specificity was confirmed by a no-primary antibody control (data not shown). Control cells were treated in the same way as experimentally stained cells; fixed, blocked and incubated in the absence of primary antibody before addition of the secondary antibody.

2.13.5 Secondary antibody incubation

Rhodamine conjugated goat anti rabbit secondary antibody (Jackson ImmunoResearch Laboratories) was added to each slide at a dilution of 1:40 in 1xTBS and incubated at room temperature for 30 minutes in the dark.

2.13.6 Analysis of staining by confocal microscopy

Glass coverslips were mounted over the cells using DAPI Vectashield mounting media (Sigma-Aldrich). DAPI provided nuclear staining of the cells. Coverslips were secured using clear nail polish. Analysis of immunofluorescent staining was performed using a Leica TCS SP2 UV confocal microscope at the Newcastle University Bioimaging Facility.

2.14 Production of antisera against PR

2.14.1 Synthesis of internal and C-terminal PR MAPs

A series of potential peptide sequences were identified within the regions of the full length PR protein encoded by exons 2 to 8. These peptides were selected based on their specificity for PR and to differ from the regions targeted by commercial PR antibodies. Peptide sequences were sent to Alta Bioscience and analysed for potential immunogenicity. The two peptide sequences predicted to be the most immunogenic were selected and used by Alta Bioscience to synthesise multiple antigenic peptides (MAPs) A3 (internal) and A4 (C-terminal). Peptide sequences are detailed in Table 2.7.

MAP	Sequence	Amino acids
A3	KVFFKRAMEGQHNYL	588-602
A4	PLEGLRSQTQFEEMRSSYIR	831-851

Table 2.7: Sequences of MAPs used to produce novel PR antibodies.
MAP A3 and A4 amino acid sequences and locations within the full length PR-B sequence.

2.14.2 Production of antisera

MAP A3 and MAP A4 were diluted in Freund's adjuvant and used by PTU/BS Scotland to inoculate, by injection, a rabbit and a sheep respectively. Ethical approval for the use of animals was obtained by PTU/BS Scotland; PPL 60/3464 in accordance with the Home Office Animals (Scientific Procedures) Act. Serum was taken from a bleed pre-immunisation, and a boost immunisation was administered four weeks after the initial immunisation, with the first serum taken from a bleed after five weeks. Further boost immunisations were administered after eight and twelve weeks, with serum taken from bleeds after nine and thirteen weeks. Serum from the final thirteen week bleed was used for antibody purification.

2.14.3 Affinity purification of PR antisera

IgG was purified from 20ml rabbit serum diluted in 20ml 1xPBS pH7 (Sigma-Aldrich) using a protein A column (GE Scientific) and sheep serum (20ml diluted in 20 ml 1xPBS) using a protein G column (GE Scientific). IgG was eluted using glycine pH2.5 into a neutralising solution of TRIS-HCl pH8. Rabbit IgG was concentrated and rediluted into 1xPBS using a 6ml Vivaspin centrifuge tube (Sartorius Stedim Biotech). Sheep IgG was dialysed into 1xPBS using a Slide-A-Lyzer Dialysis Cassette (Thermo Scientific).

Affinity columns containing MAPs A3 and A4 were provided by Alta Biosciences and affinity purification was performed according to the protocol provided. Purified IgG in 1xPBS was added to the appropriate column (rabbit to A3 and sheep to A4) and the eluate was again added to the column to ensure maximum binding of specific IgG. Final eluate was stored at -20°C. Purified A3 or A4 IgG was eluted into 5 aliquots using glycine pH2.5 into a neutralising solution of 10xPBS pH7. The aliquots were combined and the concentration of IgG determined using the BioRad DC protein assay. The novel PR antisera will hereon in this thesis be termed Rabbit PRA3 and Sheep PRA4.

2.15 Validation of PR antisera by Western immunoblotting

Initial Western immunoblotting experiments were performed using whole cell lysates in SDS Lysis Buffer from MCF-7 and MDA-MB-231 cells and conditions were optimised using a range of concentrations of each novel primary antibody; PRA3 1:2000-1:500 and PRA4 1:250-1:100. The conditions for incubation with the anti-sheep secondary antibody were optimised by varying the antibody concentration (1:5000 and 1:3000), dilution in the presence or absence of 0.1% Tween20 and diluting the secondary antibody in solutions of either milk or BSA. Western immunoblotting was also performed using whole cell lysates in SDS lysis buffer from MCF-7 cells treated with PR siRNA (and controls as described above), for each of the novel antibodies to validate specificity to PR using the optimised Western immunoblotting conditions.

2.16 Validation of PR antisera by Co-Immunoprecipitation

Co-IPs were performed as described earlier using commercial PR antibodies, A3 and A4 as capture antibodies, and A3 and A4 as detection antibodies in Western immunoblotting. Details of antibodies used can be seen in Table 2.8.

2.17 Validation of PR antisera by enzyme linked immunosorbant assay (ELISA)

2.17.1 ELISA coating buffer

35mM $NaHCO_3$,

15mM Na_2CO_3

Dissolved in distilled water, pH adjusted to 9.6

2.17.2 ELISA developing solution

0.43mg/ml o-phenylenediamine (OPD) in 0.1M citrate buffer (pH5.0) + 0.012% H_2O_2

2.17.3 ELISA protocol

Immulon ELISA plates (ThermoFisher Scientific) were coated with 1µg/µl or 10µg/µl of primary antibody diluted in ELISA coating buffer. 100µl diluted antibody was added per well, with ELISA coating buffer used in blank wells. Plates were incubated at 4°C overnight to allow antibody binding to the plate. Details of the antibodies used can be seen in Table 2.8. Plates were then washed 5 times using 1xPBST and blocked by 2 hours incubation at room temperature with a 5% solution of non-fat milk powder in 1xPBST. Plates were again washed in 1xPBST and then incubated for 2 hours at room temperature with HRP conjugated secondary antibodies (Dako), diluted 1:3000 in 5% non-fat milk in 1xPBST. The wash steps were repeated before addition of ELISA developing solution. After sufficient colour development 2M H_2SO_4 was added to each well as a stop solution. Absorbances were read at 490nm using a plate reader.

2.18 Clinicopathological analysis of breast tumour tissue sections

2.18.1 Optimisation of immunohistochemistry

Epitope retrieval was optimised by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital; high and low pH retrieval solutions were used with varying incubation times. Antibodies PRA3 and PRA4 were optimised for use in staining of FFPE breast tumour sections by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital. 4µm sections were cut from three FFPE blocks of tumour samples previously characterised as strongly PR positive, weakly PR positive or PR negative, deparaffinised in xylene and rehydrated in graded alcohol and water. A range of dilutions of each antibody, from 1:500-1:50, was used for immunohistochemical staining in an automated two step polymer method using the EnVision Flex detection system (Dako) on an Autostainer Plus Link (Dako) to determine optimal conditions.

2.18.2 Analysis of tumour PR expression by immunohistochemistry

Following optimisation of conditions, staining was performed using PRA3 and PRA4 on 4µm sections from fifteen FFPE breast tumour blocks matched to the frozen samples used for RT-PCR analysis of PR expression. One sample used for RT-PCR was a core biopsy and therefore matched FFPE tissue was not available. Sections were initially analysed by haematoxylin and eosin staining to confirm the presence of tumour tissue (data not shown). The optimised antibody dilutions used for immunohistochemical staining are detailed in Table 2.8. Sections were incubated for 40 minutes with primary antibody and analysed using the Autostainer Plus Link. Sections were counterstained with haematoxylin (EnVision Flex, Dako) to highlight cell architecture and identify nuclear staining.

2.18.3 Scoring of tumour PR status

Tumour PR status was scored depending on the clarity, intensity and proportion of nuclear staining observed during microscopic analysis using a modification of the Quickscore method for hormone receptors. The Quickscore method assigns a score of 0-3 for intensity of staining, and 0-6 for the proportion of nuclei stained. The two scores

are multiplied to give a maximum score of 18. The modified scoring system was developed by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital. Tumours were scored negative if ≤10% of nuclei stained positive (Quickscore 0-2), and positive if >10% of nuclei stained positive (Quickscore 3-6). PRA3 staining intensity was not as strong as commercial antibodies and was therefore not considered in the modified Quickscore method.

2.19 Antibodies

Antibody	Company (product code)	Target protein (epitope)	Dilution for Western immunoblotting	Dilution for immunofluorescence (cell lines)	Dilution for immunohistochemistry	ELISA coating concentration	Amount for coIP capture
NCL-PgR-B	Novocastra	PR (BUS)	1:1000 PBS+1%milk	-	-	-	-
NCL-PgR-AB	Novocastra	PR (N-terminal)	1:1000 PBS+1%milk	-	-	-	2µg
NCL-PgR-AB52	Novocastra	PR (N-terminal)	1:300 PBS+1%milk	-	-	-	-
NCL-PgR-312	Novocastra	PR (N-terminal)	1:20000 PBS+1%milk	-	-	-	-
C20	Santa Cruz (sc-539)	PR (internal)	1:200 PBS+1%milk	-	-	1µg/ml and 10µg/ml	1µg
PR-AT 4.14	Abcam (ab2764)	PR (internal)	1:400 PBS+1%milk	-	-	-	-
C19	Santa Cruz (sc-538)	PR (C-terminal)	1:200 PBS+1%milk	1:300 TBS+1% normal goat serum	-	1µg/ml and 10µg/ml	-
PRA3	-	PR (DBD)	1:500 PBS+1%milk	-	1:100	1µg/ml and 10µg/ml	5µg
PRA4	-	PR (C-terminal)	1:100 PBST+5% milk	-	1:50	1µg/ml and 10µg/ml	0.6µg
PSF	Sigma-Aldrich (P2860)	PSF	-	-	-	-	0.1µg
ERK-1	BD Transduction Laboratories (610030)	ERK-1	1:1000 PBST+5% milk	-	-	-	-
c-jun	Santa Cruz (sc-7481)	c-jun	1:200 PBS+1%milk	-	-	-	-
Vimentin	Santa Cruz (sc-6260)	Vimentin	1:100 PBST+5% milk	-	-	-	-
GAPDH	Santa Cruz (sc-25778)	GAPDH	1:1000 PBS	-	-	-	-
α-tubulin	Sigma-Aldrich (T6074)	α-tubulin	1:5000 PBST+5% milk	-	-	-	-

Table 2.8: Details of the antibodies used in this study.

3 Characterisation of PR expression in breast cancer cell lines

3.1 Introduction

ASEs involving deletion of various internal PR exons have been described (Balleine et al., 1999b; Hodges et al., 1999; Misao et al., 1998; Leygue et al., 1996a). In addition, retention of intronic sequences in PR mRNA has also been reported (Saner et al., 2003; Hirata et al., 2002; Yamanaka et al., 2002; Hirata et al., 2000). The previously reported alternative splicing of PR is described in depth in Section 1.4.11 of this thesis, and could give rise to mRNA which would encode truncated proteins which may impact on PR screening in breast cancer. PR screening in breast cancer has traditionally used antibodies targeted to N-terminal epitopes (Hanley et al., 2009). N-terminally truncated PR variants may result from alternative splicing or alternative promoter usage and would not be detected by PR screening. Alternatively, exon deleted PR mRNA may encode functionally different PR proteins with N-terminal regions identical to full length PR, making them indistinguishable from full length PR using current screening antibodies. Therefore, a breast cancer cell line characterised as PR negative, MDA-MB-231, was assessed for the presence of PR spliced variants and to predict the structures of the potential proteins which may be encoded. The potential functionality of the putative variant proteins and the impact of variant protein expression on breast cancer screening were then predicted. Expression of previously reported N-terminally truncated PR spliced variants was also assessed in this cell line. Recent controversy over the specificity of non N-terminal commercial PR antibodies (Pang and Thomas, 2011; Madsen et al., 2007), and over the existence of truncated PR isoforms (Samalecos and Gellersen, 2008) has necessitated firstly the characterisation of proteins detected by commercial PR antibodies, and secondly validation of the specificity of these antibodies.

Small interfering RNAs (siRNAs) are short, non-coding, double stranded, antisense RNA molecules which are produced by the cleavage of longer RNAs (Elbashir et al., 2001; Hamilton and Baulcombe, 1999). siRNA reduces gene expression through the RNA interference pathway; the two siRNA strands separate and one of the strands is integrated into the RNA induced silencing complex (RISC). RISC binds via the siRNA strand to a complementary sequence of mRNA and the active component of RISC,

Argonaute, then cleaves the mRNA preventing translation (Liu et al., 2004). siRNA molecules are designed to be complementary for a specific gene. Therefore transfection of siRNA into cultured cells prevents translation of a specific mRNA. This reduced protein expression was used to validate the specificity of antibodies by observing knockdown of proteins detected by the antibody.

3.2 Aims

The aims of this chapter were to:

1. Develop a PR gene walking assay and use this to identify ASEs and characterise PR mRNA expression in a range or breast cancer cell lines with different phenotypes.

2. Characterise PR protein expression in MCF-7 and MDA-MB-231 breast cancer cell lines and select which commercially available antibodies to use for further experimentation.

3. Determine the effect of β-oestradiol stimulation on the expression of PR in breast cancer and normal mammary epithelial cell lines.

4. Examine the subcellular localisation of PR in breast cancer cell lines and the effect of hormone stimulation on subcellular localisation.

5. Validate the specificity of commercial PR antibodies.

3.3 Results

3.3.1 Characterising PR mRNA expression in breast cancer cell lines

3.3.1.1 Development of an RT-PCR based PR gene walking assay

An RT-PCR based PR gene walking assay was developed using sense and antisense primers targeted to each of the 8 exons of the PR gene. This assay was used to characterise PR mRNA expression in the breast cancer cell lines MCF-7, T47-D and MDA-MB-231. Details of these cell lines can be seen in Table 2.1. The size of PCR product detected was compared to the predicted size based on the PR gene sequence (NCBI Accession: NM_000926.4) to determine potential regions which may undergo alternative splicing (see Table 2.4 for predicted PCR product size). This assay was initially validated using cDNA from MCF-7 cells. Each pair of primers predicted to give a PCR product size less than 1kbp was used for PCR. The gene walking assay is illustrated in Figure 3.1; sense primers directed to exon 3 (S7), exon 4 (S8), exon 5 (S9), exon 6 (S10) and exon 7 (S11) were used for PCR in combination with antisense primers directed to exon 5 (A4), exon 6 (A3), exon 7 (A2) and exon 8 (A1). As the distance between the two exons to which the sense and antisense primers are directed is increased, an increased size of PCR product was detected. For example S7/A2 (exon 3 to 7) gives a larger PCR product than S7/A3 (exon 3 to 6) (Figure 3.1).

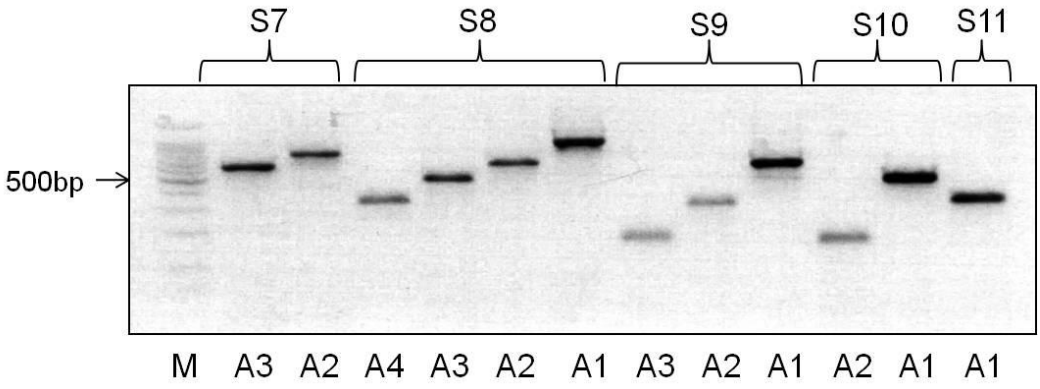


Figure 3.1: Validation of the PR gene walking assay using MCF-7 cDNA.

RT-PCR was carried out for 30 cycles using MCF-7 cDNA and a range of primers from the PR gene walking assay. M=100bp DNA ladder; S=sense: S7=exon 3, S8=exon 4, S9=exon 5, S10=exon 6, S10=exon 7; A=antisense: A1=exon 8, A2=exon 7, A3= exon 6, A4=exon 5. Correct molecular weight of PCR product was confirmed by comparison to the size predicted by analysis of the PR mRNA sequence (NCBI Accession: NM_000926.4).

3.3.1.2 PR mRNA is detected in breast cancer cell lines using a gene walking assay

After initial validation, the PR gene walking assay was then used to characterise PR mRNA expression in the PR+ breast cancer cell lines MCF-7 and T47-D, and the reportedly PR- breast cancer cell line MDA-MB-231. PR mRNA was detected in all three cell lines. Figure 3.2 shows an example of the PCR products detected using this gene walking assay. Primers spanning exons 3 to 4 and 3 to 5 detected PR mRNA in MCF-7 and T47-D cells. Additionally, primers spanning exons 3 to 6 also detected PR mRNA in MDA-MB-231 cells which are characterised as PR- (Figure 3.2a). The gene walking assay further detected truncated PCR products in MCF-7 and T47-D cells which may be indicative of ASEs occurring in breast cancer cell lines (Figure 3.2b). Primers spanning exons 3 to 4 detect only the full length PCR product, whereas primers spanning longer regions also detect the truncated product, suggesting alternative splicing of exon 4 may be occurring.

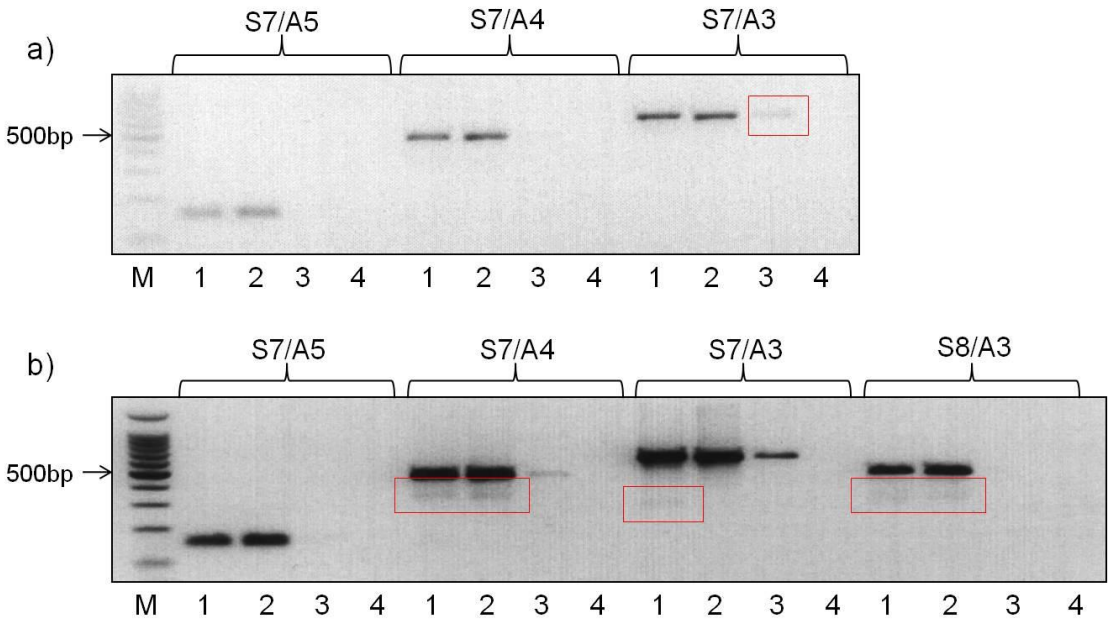
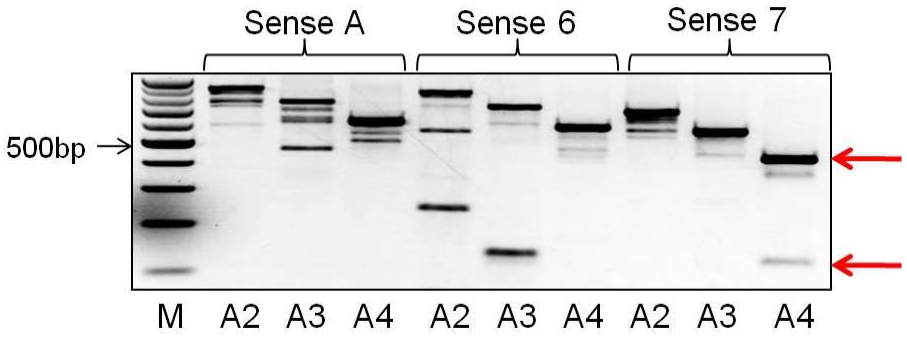


Figure 3.2: PR mRNA expression in breast cancer cell lines.

RT-PCR carried out for 35 cycles using MCF-7 (lane 1), T47-D (lane 2) and MDA-MB-231 (lane 3) cDNA detects PR mRNA in each of these cell lines using primers from the gene walking assay. A no template control PCR reaction (NTC, lane 4) was also performed. Molecular weights were approximated using a 100bp DNA ladder (M). **a)** PR mRNA is detected by the PR gene walking assay in MCF-7 and T47-D cells using a sense primer in exon 3 (S7) with antisense primers in either exon 4 (A5) or exon 5 (A4) after 30 cycles of PCR. Primers spanning exon 3 (S7) to exon 6 (A3) also detect PR mRNA in MDA-MB-231 cells. **b)** A longer exposure of the same gel additionally reveals PR mRNA detected by primers spanning exon 3 (S7) to exon 5 (A4) in MDA-MB-231 cells, and reveals the presence of additional PCR products in MCF-7 and T47-D cells.

3.3.1.3 Detection of alternatively spliced PR mRNA in MDA-MB-231 cells

To enable increased detection of low levels of PR mRNA in MDA-MB-231 cells, the PCR cycle number was increased from 35 to 45 cycles. In addition to the predicted full length band, the gene walking assay detected a range of PCR products in MDA-MB-231 cells ranging from 100bp to almost 700bp smaller than the predicted full length size (Figure 3.3). This suggests that not only do reportedly PR- MDA-MB-231 cells express PR, but that PR may be undergoing ASEs in this breast cancer cell line. Comparison of the size of PCR product with the predicted size based on the PR gene sequence suggests both whole and partial exon deletion may be occurring.



<u>Figure 3.3:</u> Truncated PR mRNA is detected in MDA-MB-231 cells using the gene walking assay.

PCR performed for 45 cycles using MDA-MB-231 cDNA generates the predicted full length product and a range of products from 100bp-700bp below the predicted size, e.g. the upper arrow highlights the full length product of S7/A4 (446bp), and the lower arrow indicates a truncated PCR product approximately 300bp smaller than full length. PCR product size can be compared to the full length size to predict the regions which may be deleted. PCR was performed with sense primers: A (exon 1), 6 (exon 2) and 7 (exon 3); in combination with antisense primers: A2 (exon 7), A3 (exon 6) and A4 (exon 5). Molecular weights were approximated using a 100bp DNA ladder (M).

3.3.1.4 PR intron retention in breast cancer cell lines

As discussed in the introduction to this thesis, and in Section 3.1, previous studies have demonstrated the retention of intronic sequences in PR mRNA. These include intronic exons S, T and M between exons 3 and 4 (Saner et al., 2003; Hirata et al., 2002; Hirata et al., 2000), and intronic exons i45a and b between exons 4 and 5 (Yamanaka et al., 2002). The locations of the intronic primers and the PR gene walking assay primers used to detect intron retained PR mRNA are shown in Figure 3.4.

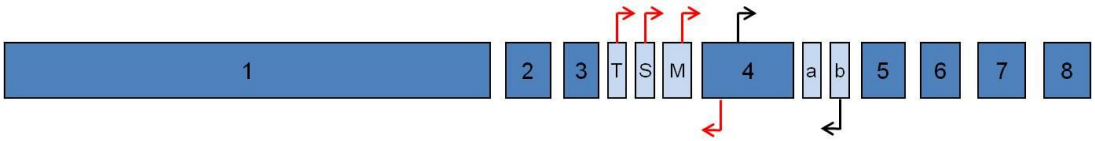


Figure 3.4: Location of PCR primers used to investigate intron retention in breast cancer cell lines.

Sense primers directed to intronic exons S, T and M were used in combination with an antisense primer directed to exon 5 (A5). An antisense primer to the intronic exon i45b was used in combination with a sense primer directed to exon 4 (S8).

Expression of intron retained PR mRNA was compared between the PR+ MCF-7 cell line which was predicted to express these isoforms, and the reportedly PR- MDA-MB-231 cell line, which has been demonstrated by the gene walking assay in Section 3.3.1.3 to potentially express PR mRNA which may undergo ASEs. PR-T, PR-S and PR-i45 mRNA was detected specifically in PR+ MCF-7 cells. However, PCR products representing mRNA for the PR-M isoform was also detected in MDA-MB-231 cells (Figure 3.5).

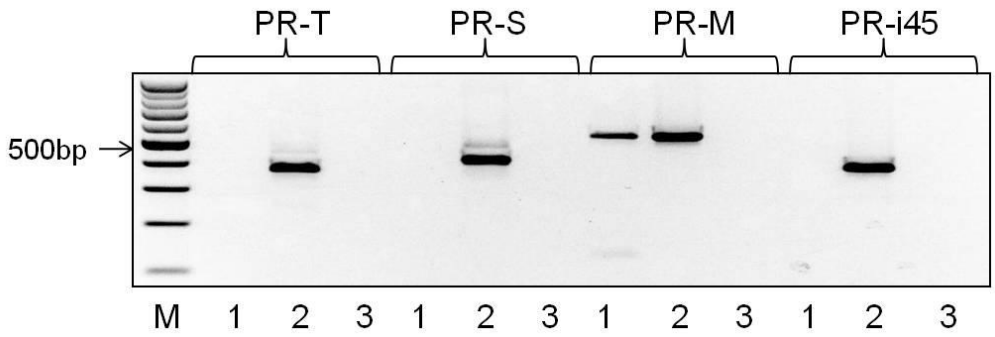


Figure 3.5: Intron retention in breast cancer cell lines.

RT-PCR was undertaken using sense primers directed to the previously described intronic exons T, S and M in combination with an antisense primer directed to exon 4, and an i45a antisense primer with an exon 4 sense primer. PCR products are detected for all isoforms in MCF-7 cells (lane 2). PR-M is also detected in MDA-MB-231 cells (lane 1). M=100bp DNA ladder, 1=MDA-MB-231, 2=MCF-7, 3=NTC.

3.3.1.5 Alternative splicing of PR in MDA-MB-231 breast cancer cells

All combinations of primers in the PR gene walking assay were used for PCR with MDA-MB-231 and MCF-7 cDNA. Cloning and sequencing of PCR products focused on reportedly PR- MDA-MB-231 cells which have been previously demonstrated in this chapter to express PR mRNA which undergoes ASEs (Figures 3.3 and 3.5). Computational analysis of DNA sequences obtained from cloning and sequencing of MDA-MB-231 PCR products was performed using ALIGN and ExPASy Translate as described in Section 2.11.9. This analysis revealed several modes of ASE that may be affecting PR mRNA in MDA-MB-231 cells. Figure 3.6 is a diagrammatic representation of the regions removed/retained by the alternative splicing of PR in MDA-MB-231 cells, and the potential structures of putative protein products of these alternatively spliced mRNAs. Examples of DNA sequencing analysis using ALIGN can be seen in Appendix 1.

Firstly, several partial exon deletions were detected, potentially resulting from the use of cryptic splice sites within the exons (Figure 3.6b). Deletions of large regions within exon 1 and the start of exon 2 usually cause a change in reading frame and truncation within the DBD, so are unlikely to generate functional PR proteins and may be targeted for NMD (Figure 3.6b). However, one of the observed deletions, designated Δp1c would remain in the original reading frame, potentially generating a protein with a

truncated N-terminal domain lacking the constitutive AF1, but retaining the other functional domains including DBD, NLS, hinge region and LBD (Figure 3.6b).

Whole exon deletions of exon 2 (Δ2) and the cassette exons 3 (Δ3) and 4 (Δ4) were also observed in MDA-MB-231 cells (Figure 3.6c). Deletion of exon 2 predicts a change in reading frame, leading to a truncated protein which would be unlikely to function. The predicted protein product of the PRΔ3 mRNA would lack the second zinc finger of the DBD encoded by this cassette exon, but since the mRNA would remain in the original reading frame, would possess the normal C-terminal regions so could potentially be functional either non-genomically or by competition with wild-type PR for ligand binding or dimerisation, thereby modulating PR function. Deletion of the cassette exon 4 would generate a protein lacking, the hinge region and NLS, but possessing the other functional domains. Therefore PRΔ4 protein may be structurally distinct from wild-type PR and segregated from the nucleus, but again may be able to function either non-genomically or through modulation of wild-type PR function (Figure 3.6c). Several partial exon deletions within the region of exon 4 to 6 would generate proteins lacking parts of the LBD, DD ligand dependent AF2 (Figure 3.6d).

Intron retention was demonstrated in this cell line by presence of mRNA specific for the intronic exon M. As previously reported (Saner et al., 2003), PR-M mRNA would encode an N-terminally truncated protein also lacking a DBD. This protein would contain a functional LBD and a short N-terminal sequence of hydrophobic amino acids encoded by exon M (Figure 3.6e).

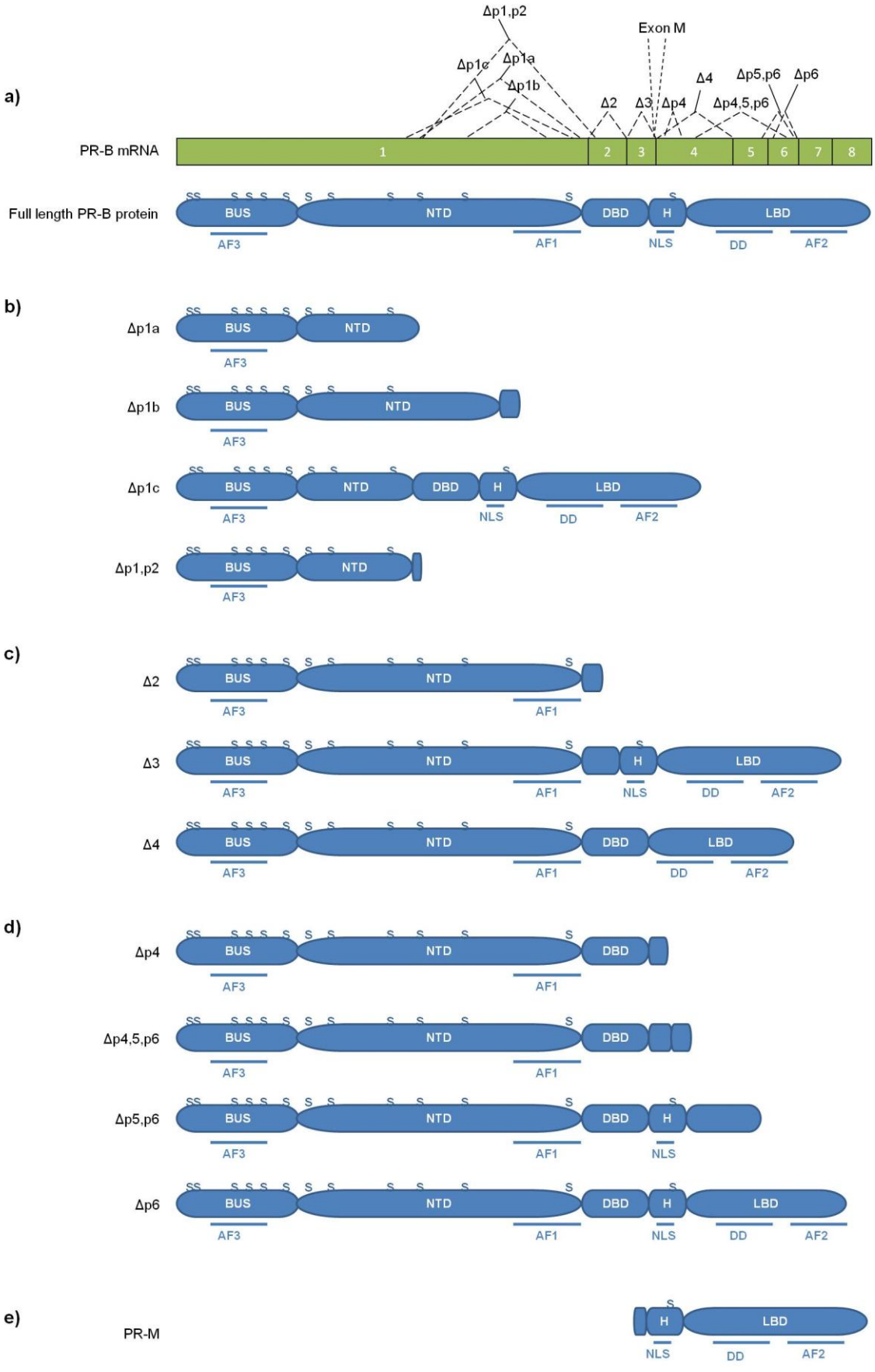

a)
Δp1,p2
Exon M
Δp1c
Δp1a
Δp1b
Δ4
Δp5,p6
Δp6
Δ2
Δ3
Δp4
Δp4,5,p6
PR-B mRNA
1
2
3
4
5
6
7
8
Full length PR-B protein
BUS
NTD
DBD
H
LBD
AF3
AF1
NLS
DD
AF2
b)
Δp1a
Δp1b
Δp1c
Δp1,p2
c)
Δ2
Δ3
Δ4
d)
Δp4
Δp4,5,p6
Δp5,p6
Δp6
e)
PR-M

Figure 3.6: Predicted structures of alternatively spliced PR variants detected in MDA-MB-231 cells.

Computational analysis of sequenced PCR products from MDA-MB-231 cells was used to analyse the regions spliced out of the mRNA and predict the structures of proteins produced. ALIGN was used to identify the region spliced out of the mRNA by comparison with the published PR gene sequence (NCBI Accession: NM_000926.4), and ExPASy Translate, to determine where a STOP codon would be inserted. **a)** Diagram of the 8 exon PR gene sequence showing the ASEs detected in MDA-MB-231 using the gene walking assay. Dashed lines indicate the regions deleted/retained by each ASE. The domain structure of the full length PR-B protein is shown, indicating the domains encoded by each exon. PR variants have been named Δ for deletion and the number of the exon deleted (e.g. Δ4 for exon 4 deleted). Where partial exons have been deleted this is indicated by p before the exon number (e.g. Δp4). When multiple deletions within an exon are observed, a letter is assigned to each deletion e.g. (Δp1a). **b)** Predicted proteins produced from mRNAs with deletions of parts of exon 1 and exon 2. **c)** Predicted proteins produced from mRNAs with whole exon deletions of exons 2, 3 or 4. **d)** Predicted proteins produced from mRNAs with partial or multiple exon deletions in the region exon 4 to 6. **e)** Potential protein produced from mRNA retaining the intronic exon M.

3.3.2 Characterisation of PR protein expression in breast cancer cell lines

PR protein expression was compared in breast cancer cell lines with different reported PR statuses, MCF-7 (PR+) and MDA-MB-231 (PR-). PR expression in these cell lines was determined by Western immunoblotting using whole cell lysates and a range of commercially available PR antibodies, targeted to epitopes in different regions of full length PR protein, as detailed in Figure 2.1 and Table 2.8.

Using N-terminally targeted PR antibodies, proteins corresponding to the full length nuclear PR isoforms PR-B (~120kDa) and PR-A (~80kDa) were detected specifically in the MCF-7 cell line (Figure 3.7a and b). However, N-terminally targeted PR antibodies did not detect protein in the reportedly PR negative MDA-MB-231 cells (Figure 3.7c). The C-terminally targeted PR antibody C19 detected a range of low molecular weight proteins, between 30 and 50kDa, in addition to a protein of approximately 80kDa in MDA-MB-231 and MCF-7 cells. A 60kDa protein, previously described as PR-C, was also detected by this antibody, though the identity of PR-C remains controversial as will be discussed. The presence of these low molecular weight proteins could indicate that alternatively spliced PR proteins are expressed in this cell line, as predicted from the mRNA expression previously described.

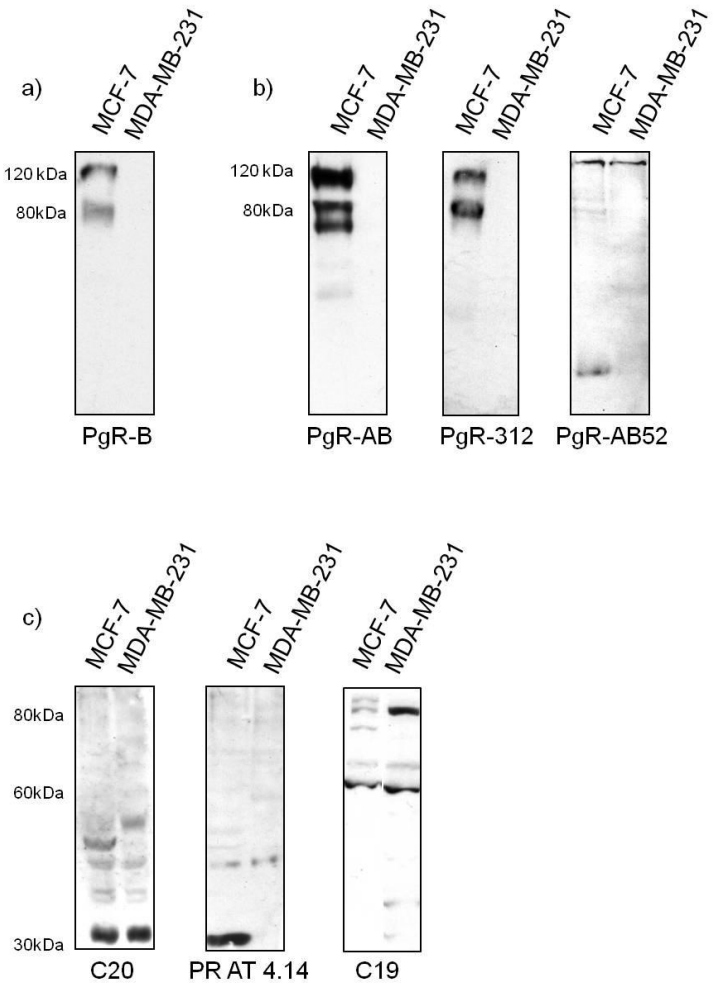


Figure 3.7: Characterising PR protein expression in MCF-7 and MDA-MB-231 cells.

SDS-PAGE and Western immunoblotting was carried out using whole cell lysates from cultured breast cancer cell lines and a range of commercially available PR antibodies. **a)** Proteins are detected at 120kDa (PR-B) and 80kDa in MCF-7 cells using NCL-PgR-B targeted to the B upstream segment. No protein is detected in MDA-MB-231 cell lysates. **b)** N-terminal antibodies NCL-PgR-AB and NCL-PgR-312 detect protein at 120kDa (PR-B) and 80kDa (PR-A) in MCF-7 cells. No protein is detected in MDA-MB-231 cells. NCL-PgR-AB52 failed to adequately detect PR proteins in either cell line. **c)** C20, targeted to an internal epitope, and C19, targeted to a C-terminal epitope detect low molecular weight proteins in MCF-7 and MDA-MB-231 cells. Another internal PR antibody PR-AT 4.14 failed to adequately detect PR proteins.

3.3.3 Selection of antibodies for use in future studies

N-terminal antibodies NCL-PgR-B and NCL-PgR-AB were selected for use in future studies as they appeared to demonstrate the greatest specificity of binding to PR-A and PR-B in MCF-7 cells. Since C19 was able to detect low molecular weight proteins with minimal background in the MDA-MB-231 cells, this C terminal antibody was selected for use in assays to determine the effect of β-oestradiol and progesterone stimulation on PR expression in breast cancer cell lines. The specificity of all of these antibodies will be further analysed in Section 3.3.6.

3.3.4 Hormone responsive PR expression in breast cancer cell lines

Since the PR gene is oestrogen responsive, containing a partial ERE sequence in both the PR-A and PR-B promoters (Kastner et al., 1990), and β-oestradiol has previously been reported to influence alternative splicing (Zhang et al., 2007; Zhu et al., 2005), a hormone response assay was developed to determine if β-oestradiol had an effect on PR expression in breast cancer cell lines. Cells were cultured in the presence of a range of concentrations of β-oestradiol (0-10μM) for 24 hours prior to harvesting. The antibodies selected from PR characterisation experiments were used for Western immunoblotting to assess the effect of β-oestradiol on PR expression.

3.3.4.1 Effect of β-oestradiol on PR-A and B expression

As previously described (Section 3.3.2) the N-terminal antibodies, NCL-PgR-B and NCL-PgR-AB, detect protein specifically in the MCF-7 (PR+) cell line. As expected of products of oestrogen responsive genes, low doses of β-oestradiol stimulate an increased expression of both PR-B (~120kDa) (Figure 3.8a and b) and PR-A (~80kDa) detected by these N-terminal PR antibodies (Figure 3.8b). Figure 3.8 provides further validation that these N-terminal PR antibodies do not detect PR protein in MDA-MB-231 and also in the MCF-10A non-tumorigenic MEC line.

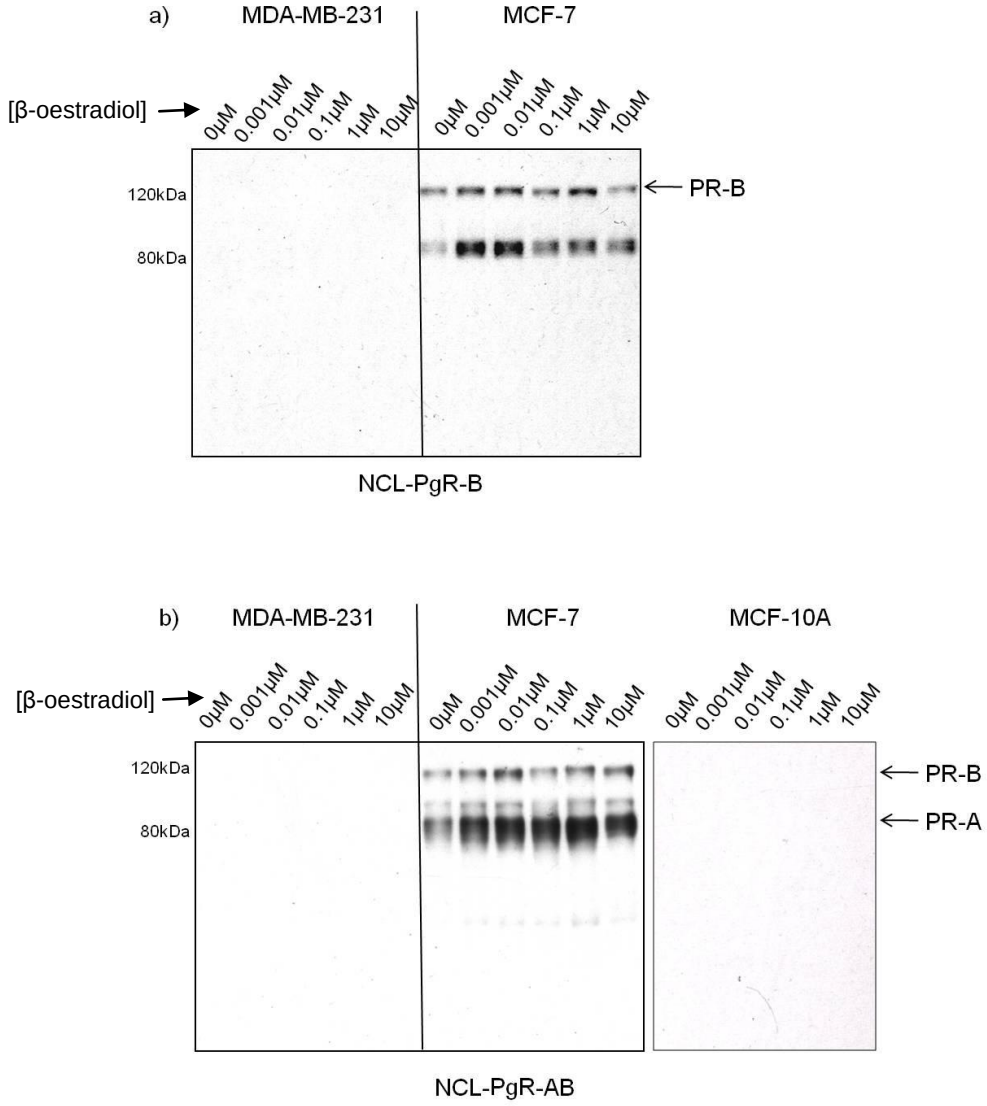


Figure 3.8: β-oestradiol stimulates increased expression of PR- B and PR-A in MCF-7 cells.

MDA-MB-231, MCF-7 and MCF-10A cells were cultured in the presence of a range of concentrations of β-oestradiol (0-10μM, 0μM=ethanol only control) for 24 hours prior to lysing for SDS-PAGE and Western immunoblotting. All experiments were replicable and carried out at least twice. **a)** Western immunoblotting using NCL-PgR-B detects increased PR-B expression at low doses of β-oestradiol. **b)** Western immunoblotting using NCL-PgR-AB detects increased expression of PR-A and PR-B at low doses of β-oestradiol. No protein was detected by either antibody in MDA-MB-231 or MCF-10A cells.

3.3.4.2 β-oestradiol stimulates cell line specific changes in PR expression

The C-terminal PR antibody C19 detected a range of low molecular weight proteins in breast cancer and non-tumourigenic MEC lines. Distinct changes in the patterns of PR expression were observed between the cell lines in response to changing concentrations of β-oestradiol (Figure 3.9).

In MCF-7 cells low doses of β-oestradiol (0.01μM) appear to stimulate an increased expression of proteins in the range 70-90kDa, possibly including PR-A, supporting the observations made with N-terminal PR antibodies (Figure 3.9). However, in MDA-MB-231 cells no effect on the approximately 80kDa protein was observed, and an increased expression of several low molecular weight proteins, between 30-60kDa was seen (Figure 3.9). In contrast to this metastatic breast cancer cell line, the non-tumorigenic mammary epithelial MCF-10A cell line exhibits decreased expression of the low molecular weight proteins around 30-40kDa and 60kDa, and an increased expression of the 80kDa protein in response to increasing β-oestradiol concentration (Figure 3.9).This demonstrates that not only may β-oestradiol be having a cell line specific effect on the level of PR expression, but also on the ASEs affecting PR in these cell lines.

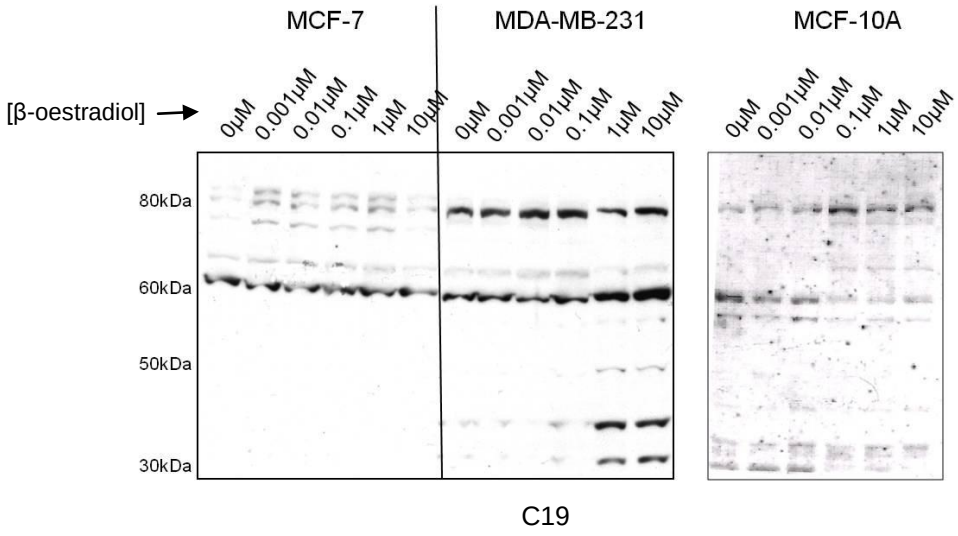


Figure 3.9: β-oestradiol stimulates cell line specific changes in PR expression in breast cancer and non-tumourigenic mammary epithelial cells.

Lysates from MCF-7, MDA-MB-231 and MCF-10A cells stimulated for 24 hours with a range of concentrations of β-oestradiol (0-10μM, 0μM=ethanol only control) were used for SDS-PAGE and Western immunoblotting using the C-terminal PR antibody C19. Cell line specific effects on PR expression were observed. Experiments were reproducable and repeated twice.

3.3.5 Effect of hormone stimulation on subcellular localisation of PR in MDA-MB-231 cells

β-oestradiol was demonstrated to increase PR expression in MDA-MB-231 cells in the hormone response assay (Section 3.3.4). The β-oestradiol concentration which stimulated the greatest effect on PR expression in MDA-MB-231 cells was selected to stimulate cells to examine the subcellular distribution of PR. MDA-MB-231 cells were cultured in the presence of 1μM β-oestradiol for 24 hours. Alternatively, MDA-MB-231 cells were stimulated with 0.1μM progesterone to examine the effect of ligand binding on the translocation of PR in this cell line. Immunofluorescent staining was performed using the C-terminal PR antibody C19 to localise PR within the cells. Treated cells were compared to a vehicle only control (ethanol) and untreated cells. Untreated and ethanol treated cells exhibit a similar distribution of both cytoplasmic and nuclear PR (Figure 3.10).

The classical mechanism of PR action is discussed in Section 1.1; ligand binding causes receptor dimerisation phosphorylation, leading to nuclear translocation. Increased nuclear and perinuclear staining for PR in MDA-MB-231 cells following progesterone treatment suggests PR is translocating in response to progesterone signal (Figure 3.10), indicating that PR is potentially functional in these cells. β-oestradiol stimulation increased expression of both nuclear and cytoplasmic PR (Figure 3.10). This supports the observed β-oestradiol induced increase in PR expression detected by Western immunoblotting using C19.

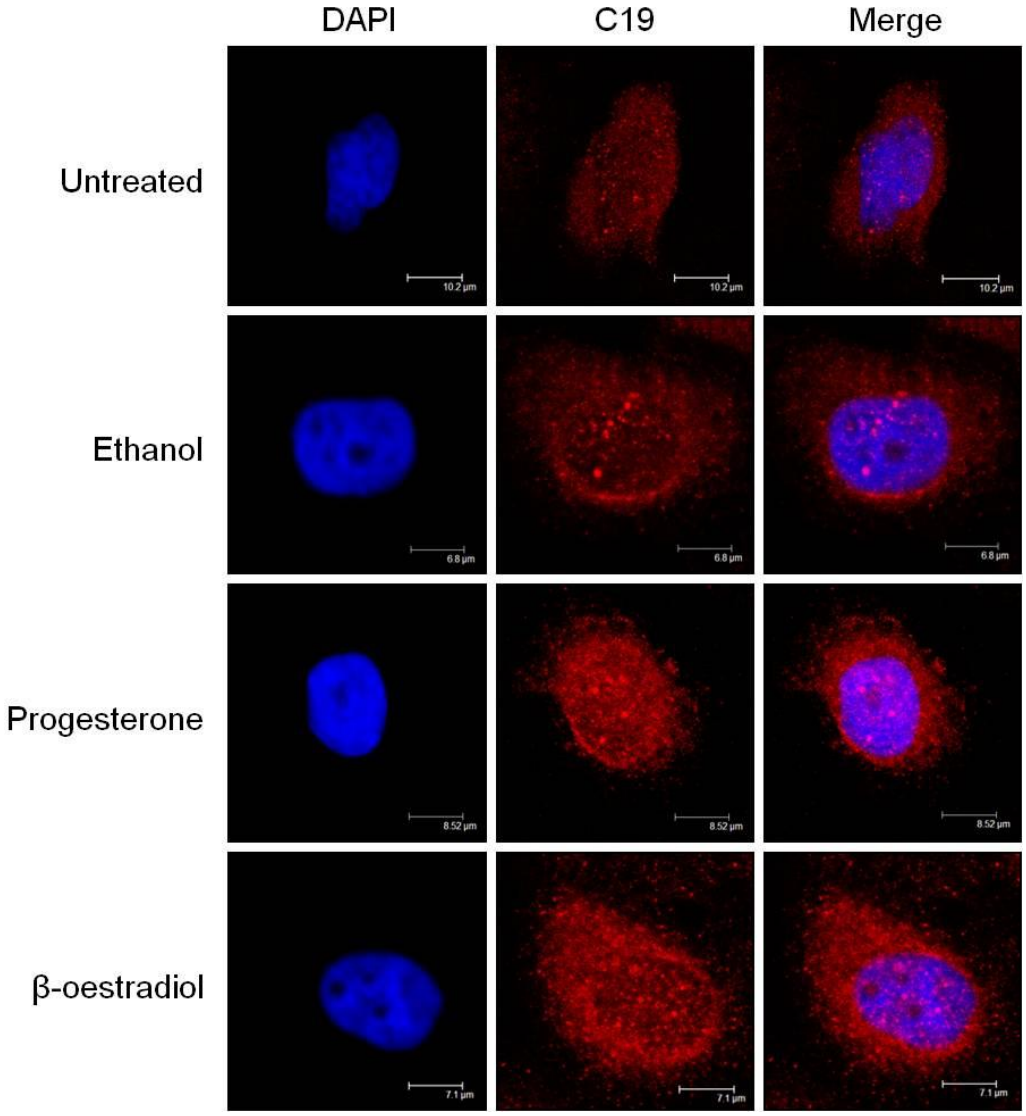


Figure 3.10: Hormone responsive subcellular localisation of PR in MDA-MB-231 cells.

MDA-MB-231 cells were cultured on glass coverslips in the presence of either 1μM β-oestradiol or 0.1μM progesterone for 24 hours prior to fixing. Cells were also treated with an ethanol only control. Nuclear staining was performed using DAPI and PR staining using the C-terminal antibody, C19. Untreated cells were also stained for comparison. All staining was analysed by confocal microscopy using identical settings.

3.3.6 Validation of commercial PR antibodies

As discussed in Section 1.4.11, recent publications since the start of this study have questioned the existence of truncated PR variants, and controversy persists over the specificity of commercial non N-terminal PR antibodies (Samalecos and Gellersen, 2008; Madsen et al., 2007). Since the commercially available PR antibodies have been shown in this thesis to detect a range of proteins (Sections 3.3.2 and 3.3.4), validation of the specificity of these antibodies was necessary to confirm the identity of the PR proteins detected.

3.3.6.1 Knockdown of PR expression demonstrates the specificity of NCL-PgR-B, NCL-PgR-AB and C19

siRNA knockdown of total PR expression was performed in MCF-7 cells, and analysed by Western immunoblotting using the N-terminal PR antibodies NCL-PgR-B and NCL-PgR-AB and the C-terminal PR antibody, C19. Using the PR-B specific antibody NCL-PgR-B a significant knockdown of the approximately 120kDa PR-B protein was observed (Figure 3.11a and b). However, no knockdown was seen of the second band at approximately 80kDa (Figure 3.11a). Since the epitope for this antibody is within the PR-B specific B-upstream segment, detection of PR-A is not expected. Therefore, while NCL-PgR-B does detect PR-B, this data also suggests it detects a smaller non-PR protein. MCF-7 cells were also treated with a GAPDH siRNA as a positive transfection control and reduced GAPDH levels can be observed in these cells compared to untreated, vehicle, negative control siRNA or PR siRNA treated cells (Figure 3.11a).

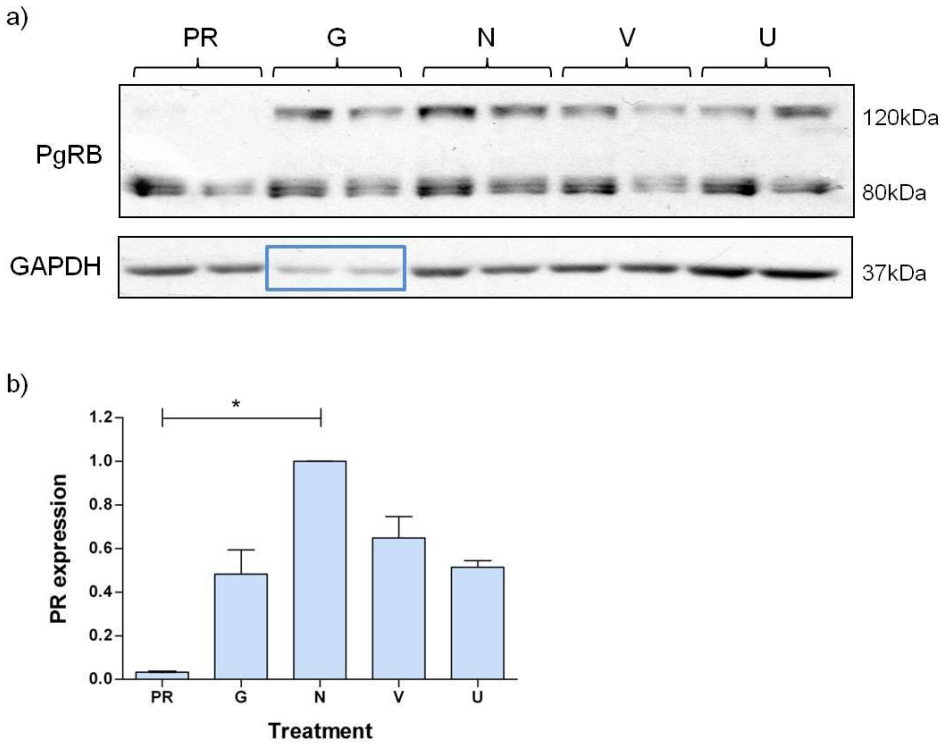


Figure 3.11: Validation of PR binding by NCL-PgR-B.

a) MCF-7 cells in culture were treated with siRNA for 72 hours prior to lysing for SDS-PAGE and Western immunoblotting using NCL-PgR-B. PR= PR targeted siRNA, G= GAPDH targeted siRNA (positive transfection control), N= negative control (untargeted) siRNA, V= transfection reagent (vehicle) only control, U= untreated cells. GAPDH knockdown is highlighted in the blue box. **b)** Histogram showing levels of PR knockdown relative to the negative control siRNA treated cells. Pixel density of the 120kDa band was measured using Intelligent Quantifier, and values normalised to the negative control siRNA bands which were always assigned a value of 1. Data represents average PR expression relative to N for three experiments. Statistical significance of PR knockdown was calculated using a student's t-test. $*$ = $p<0.0001$.

Knockdown was observed of both the approximately 120kDa (PR-B) and the approximately 80kDa (PR-A) proteins using the N-terminal antibody NCL-PgR-AB (Figure 3.12a and b). However, this antibody also detected another protein band close to the molecular weight of PR-A and no knockdown of this protein was observed (Figure 3.12a). NCL-PgR-AB is a cocktail of an antibody with an epitope common to PR-A and PR-B, and an antibody targeted to the same epitope as NCL-PgR-B, therefore this band could result from detection of the same non-specific protein detected by NCL-PgR-B (Figure 3.12a). MCF-7 cells were again treated with the positive transfection control GAPDH siRNA and reduced GAPDH levels can be observed in these cells compared to untreated, vehicle, negative control siRNA or PR siRNA treated cells (Figure 3.12a).

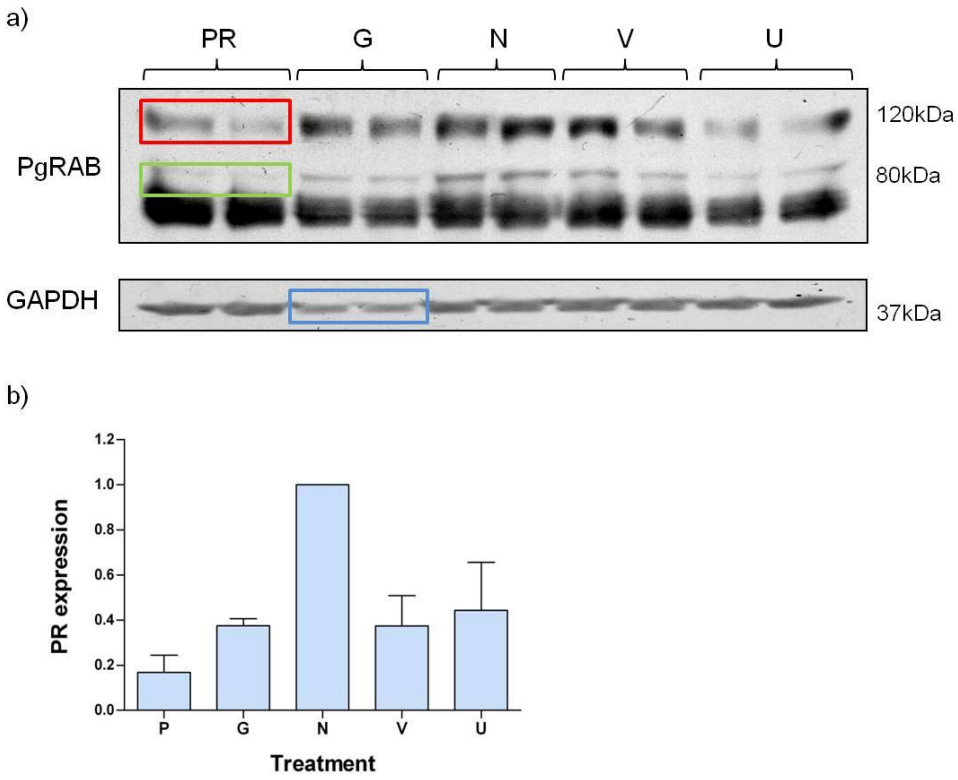


Figure 3.12: Validation of PR binding by NCL-PgR-AB.

a) MCF-7 cells in culture were treated with siRNA for 72 hours prior to lysing for SDS-PAGE and Western immunoblotting using NCL-PgR-AB. PR= PR targeted siRNA, G= GAPDH targeted siRNA positive transfection control, N= negative control untargeted siRNA, V= transfection reagent (vehicle) only control, U= untreated cells. PR-B knockdown is highlighted in the red box, PR-A knockdown in the green box and GAPDH knockdown is highlighted in the blue box. **b)** Histogram showing levels of PR knockdown relative to the negative control siRNA treated cells. Pixel density of the 120kDa bands was measured using Intelligent Quantifier, and values normalised to the negative control siRNA bands which were always assigned a value of 1. Data represents average PR expression relative to N for two experiments.

Using the C-terminal PR antibody C19, knockdown of an approximately 80kDa protein (highlighted red), potentially PR-A, and several low molecular weight proteins between 30-50kDa (highlighted green) is observed, confirming that these are PR proteins. However, no knockdown of the abundant 60kDa protein is seen (Figure 3.13).

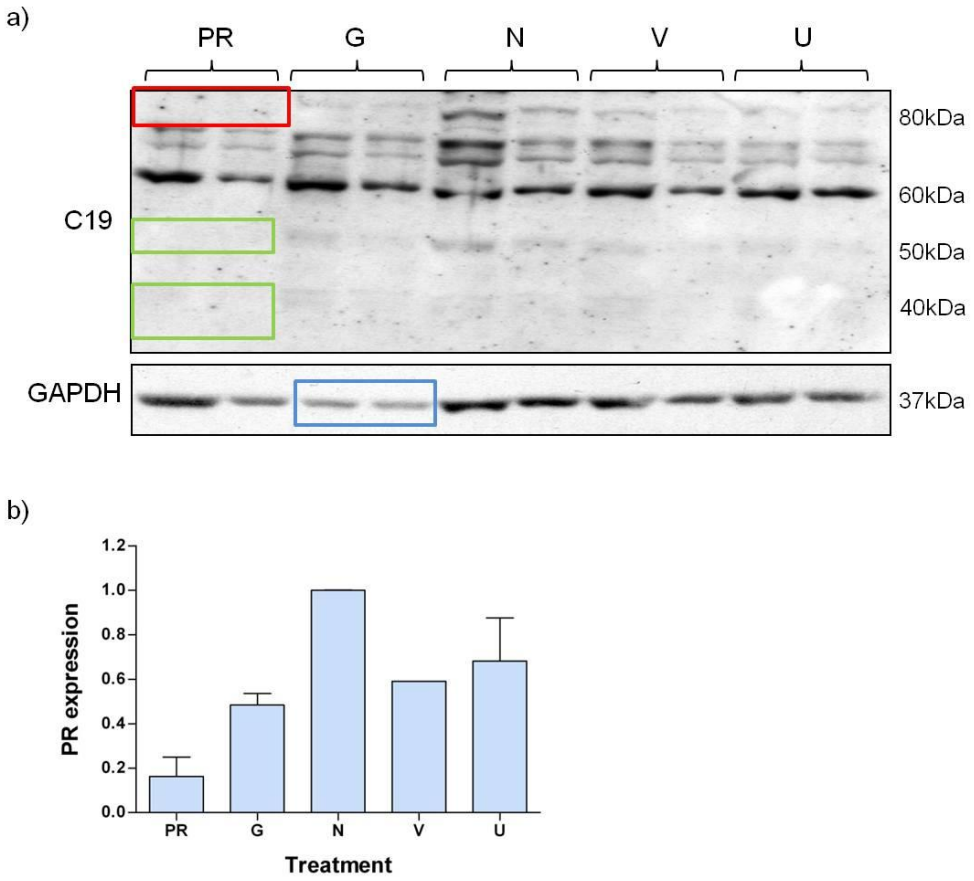


Figure 3.13: Validation of PR binding by C19.

a) MCF-7 cells in culture were treated with siRNA for 72 hours prior to lysing for SDS-PAGE and Western immunoblotting using C19. PR= PR targeted siRNA, G= GAPDH targeted siRNA positive transfection control, N= negative control untargeted siRNA, V= transfection reagent (vehicle) only control, U= untreated cells. PR-A knockdown is highlighted in the red box, low molecular weight PR knockdown in the green boxes and GAPDH knockdown is highlighted in the blue box. **b)** Histogram showing levels of PR knockdown relative to the negative control siRNA treated cells. Pixel density of the 80kDa bands was measured using Interligent Quantifier, and values normalised to the negative control siRNA bands which were always assigned a value of 1. Data represents average PR expression relative to N for two experiments.

3.3.6.2 Validation of the C-terminal PR antibody C19 for use in immunofluoresent staining

An abundant non-specific protein at approximately 60kDa was detected by the C-terminal PR antibody C19 by Western immunoblotting after PR knockdown using siRNA. Therefore siRNA treatment of MCF-7 cells was repeated, culturing the cells on glass coverslips in 6 well culture plates in the presence of PR or negative control siRNA for 72 hours before fixing. These cells were used for validation of the C-terminal antibody C19 for immunofluorescent detection of PR. Immunofluorescent staining of PR siRNA treated cells was performed with C19 and compared to cells which were untreated or treated with negative control siRNA. Little reduction of overall staining was observed with C19 after treatment with PR siRNA compared to untreated or negative control siRNA treated cells (Figure 3.14a). Successful PR knockdown in these cells was confirmed by Western immunoblotting with the N-terminal PR antibody NCL-PgR-B, using lysates of residual cells from the well in which the coverslips were cultured (Figure 3.14b).

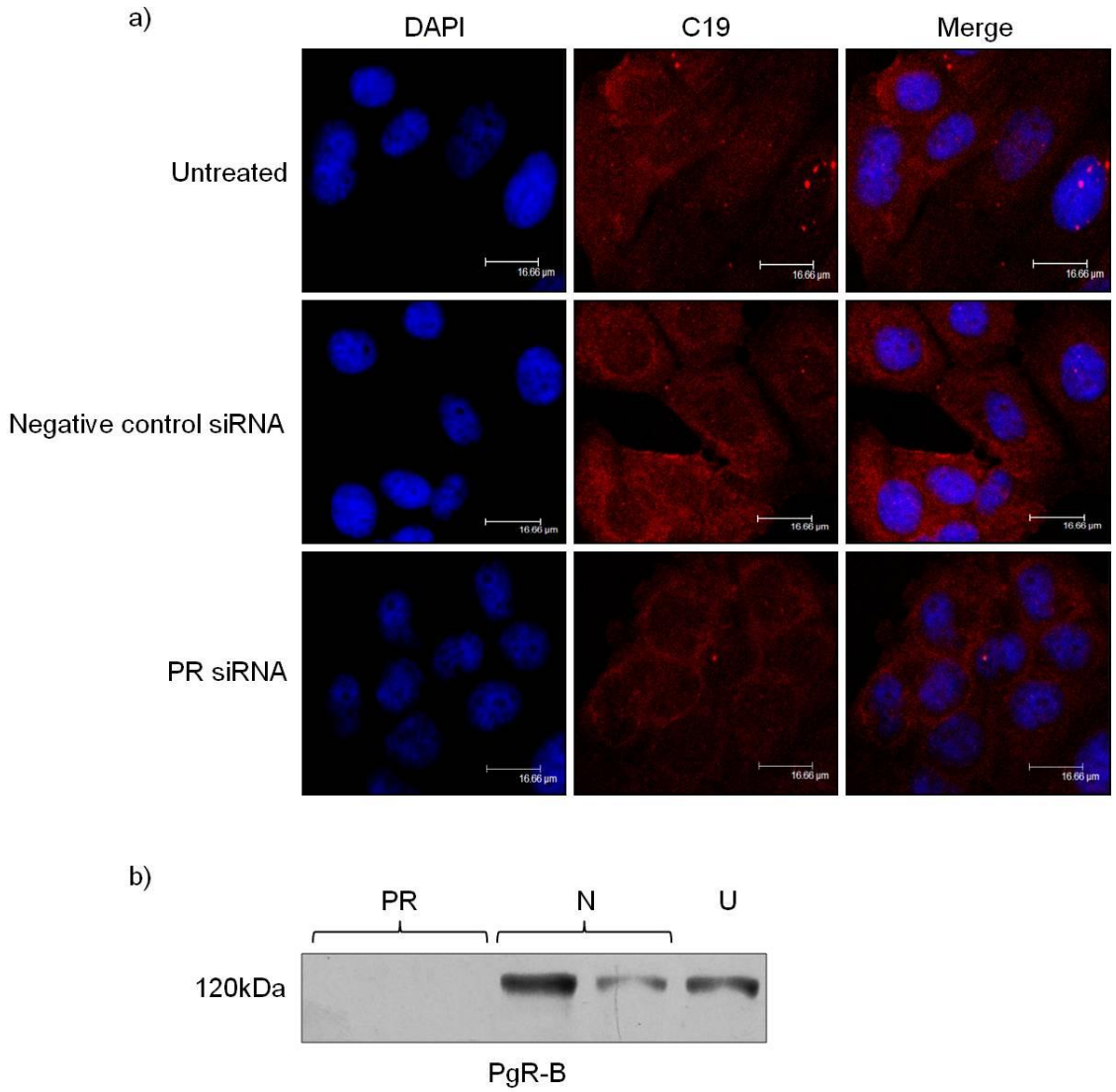


Figure 3.14: Immunofluorescent analysis of PR knockdown in MCF-7 cells.

MCF-7 cells were cultured on glass cover slips in 6 well plats and treated with PR or untargeted (negative control) siRNA for 72 hours prior to fixing. **a)** Cells on coverslips were stained using DAPI (nuclear staining) and C19 (PR staining). Untreated cells were also stained for comparison. All staining was analysed by confocal microscopy using identical settings. **b)** Cells from the culture plate around the periphery of the coverslips were lysed after 72 hours for use in SDS-PAGE and Western immunoblotting using NCL-PgR-B to confirm efficiency of PR knockdown. PR= PR targeted siRNA, N= negative control untargeted siRNA, U= untreated cells.

3.3.6.3 Analysis of PR proteins detected after treatment with protease inhibitors

To validate that the low molecular weight proteins detected by the PR antibodies were not the products of protein degradation, cells were lysed in RIPA buffer with or without protease inhibitors. When cell lysates were treated with protease inhibitors, no change was observed in the overall pattern of proteins detected with NCL-PgR-B (Figure 3.15a), NCL-PgR-AB (Figure 3.15b) or the C-terminal PR antibody C19 (Figure 3.15c). This confirms that the low molecular weight proteins detected are not the products of degradation of full length PR in either MCF-7 or MDA-MB-231 cells. Hormone stimulation experiments were repeated, stimulating cells with 1μM β-oestradiol or an ethanol only control. PR expression was compared to untreated cells, and no effect of hormone stimulation on protein degradation was seen compared to previous experiments (Figure 3.15d).

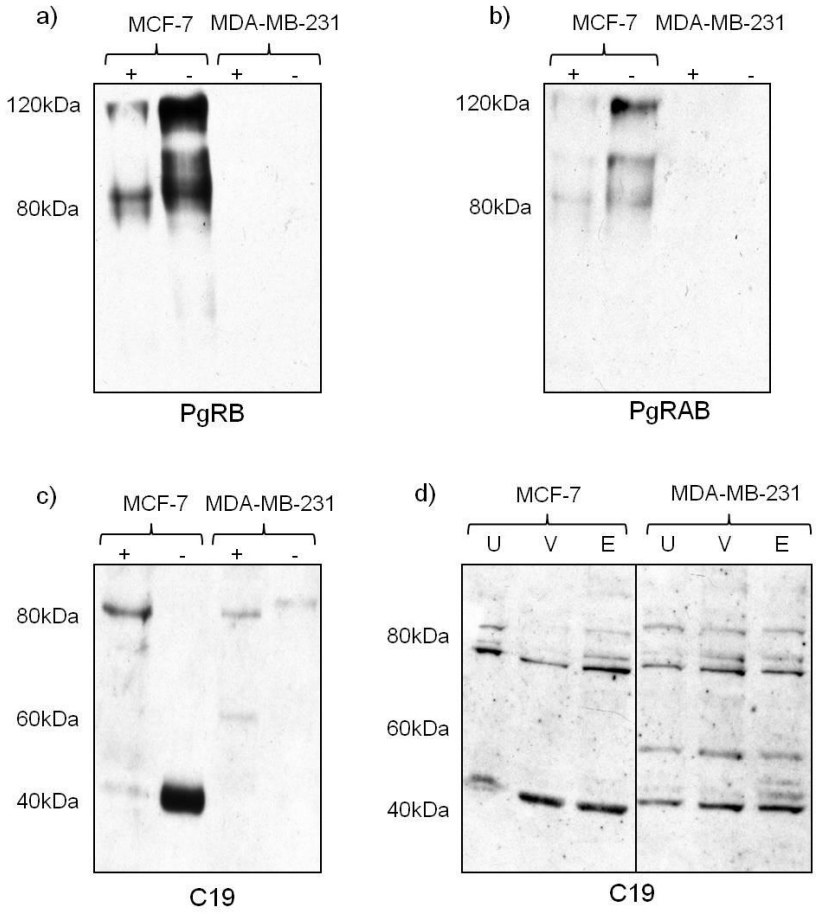


Figure 3.15: Expression of low molecular weight PR proteins detected by commercial PR antibodies is not due to protein degradation.

MCF-7 and MDA-MB-231 cells were lysed in RIPA buffer in the presence or absence of a protease inhibitor cocktail. Paired lysates were then used for SDS-PAGE and Western immunoblotting. **a)** PR proteins detected by NCL-PgR-B in MCF-7 cell lysates with or without protease inhibitors. **b)** PR proteins detected by NCL-PgR-AB in MCF-7 cell lysates with or without protease inhibitors. **c)** PR proteins detected by C19 in MCF-7 and MDA-MB-231 cell lysates with or without protease inhibitors. **d)** PR proteins detected by C19 in lysates, with protease inhibitors, from MCF-7 and MDA-MB-231 cells after stimulation with ethanol (V), 1μM β-oestradiol (E). Untreated cells (U) were used for comparison.

3.3.6.4 Analysis of the phosphorylation of proteins detected by C19

As discussed previously in Section 1.1, PR can be phosphorylated at a number of different sites (Figure 1.1) and this phosphorylation is important for all mechanisms of PR action. Therefore, the effect of phosphorylation on the range of proteins detected by the C-terminal PR antibody C19, and the effect of hormone stimulation on PR phosphorylation was assessed. Western immunoblotting was carried out using cell lysates treated with SAP following stimulation of MCF-7 and MDA-MB-231 cells with 1μM β-oestradiol or an ethanol only control. The pattern of proteins detected was compared to lysates without SAP treatment; a change in molecular weight post SAP treatment would indicate a change in phosphorylation status. Whilst the level of expression of some low molecular weight proteins seemed to differ after SAP treatment, the overall pattern of protein expression remained similar in both cell lines (Figure 3.16). In MDA-MB-231, whilst most protein bands remained the same size +/- SAP treatment, there was an apparent shift in molecular weight of one protein after dephosphorylation with SAP (Figure 3.16). However, this protein, highlighted by red arrows, may be the non-specific 60kDa protein previously described. This experiment demonstrates that although low molecular weight PR proteins do undergo phosphorylation, the detection of a range low molecular weight proteins in MDA-MB-231 and MCF-7 breast cancer cells is not due to differential phosphorylation of a single protein isoform.

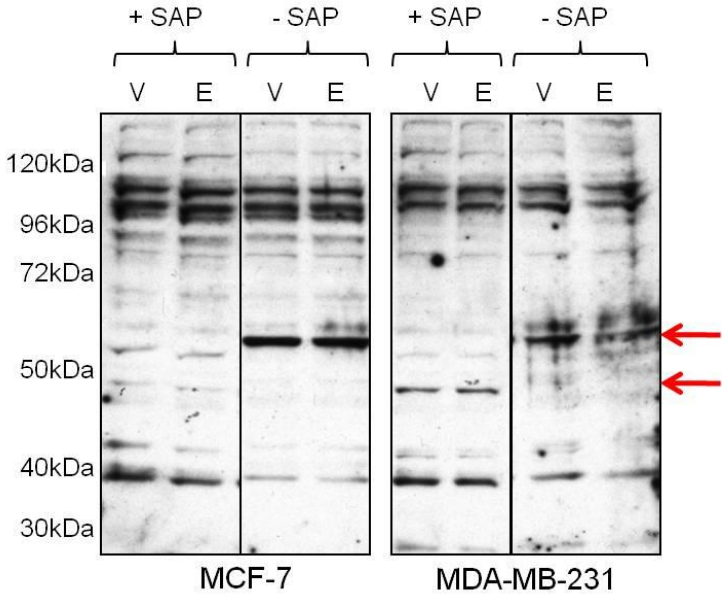


Figure 3.16: Low molecular weight PR proteins are not the products of differential phosphorylation.

MCF-7 and MDA-MB-231 cells were treated with 1μM β-oestradiol (E) or an ethanol control (V) for 24 hours prior to lysing in RIPA buffer containing protease inhibitors. Lysates were incubated in the presence or absence of Shrimp Alkaline Phosphatase (SAP) to dephosphorylate proteins. SAP +/- lysates were used for SDS-PAGE and Western immunoblotting using the C-terminal PR antibody C19.

3.4 Discussion

This chapter uses breast cancer cell lines as a model to characterise PR expression and identify PR ASEs which may have an impact on PR detection in breast cancer. Breast cancer cell lines were selected for the reported hormone receptor status. MCF-7 and T47-D are both reported to be ER+/PR+. Both cell lines were initially used and provided similar results allowing use of just one of the cell lines, MCF-7, in further experiments. MDA-MB-231 cells are a commonly used ER-/PR- breast cancer cell line, so were selected for analysis of potential PR expression in a PR- cell line. Use of this cell line for analysis of PR expression is supported by recent controversy over the detection of PR mRNA in this cell line, described in publications since the start of this study (Pang and Thomas, 2011; Springwald et al., 2010). MDA-MB-231 and MCF-7 were both isolated from metastatic breast cancer. In order to compare these cell lines representing advanced stages of breast cancer with PR expression in normal mammary cells, a non-tumourigenic mammary epithelial cell line MCF-10A was used. A second ER-/PR- breast cancer cell line (SK-BR-3) and another non-tumourigenic mammary epithelial cell line (184A1) were also used for some studies (data not shown), providing similar results to MDA-MB-231 and MCF-10A respectively, supporting use of the selected cell lines.

Normal PR signalling, as discussed in Section 1.1, involves ligand binding, dimerisation, nuclear localisation, interaction with PRE and transcriptional co-regulators to modulate gene expression. Alternatively, the N-terminus of ligand activated PR can rapidly interact with cytoplasmic signal transduction molecules such as c-Src (Boonyaratanakornkit et al., 2007) or PI3K (Carnevale et al., 2007) to activate signalling cascades and regulate non-PRE containing genes. Ligand activated PR can also interact with Stat transcription factors to mediate their activation (Proietti et al., 2005). Activated MAPKs can phosphorylate PR, leading to tethering of PR with Sp1 transcription factors and transcription of Sp1 regulated genes (Faivre et al., 2008).

An RT-PCR based PR gene walking assay detected truncated PR variant mRNA in the MDA-MB-231 breast cancer cell line (Figure 3.3). None of the PR mRNA variants detected by this assay have been previously reported in this cell line. However, exon 7 and exon 6+7 deleted PR mRNA was recently detected in this cell line using primers

spanning distal exon/exon boundaries (Springwald et al., 2010). Furthermore, alternatively spliced PR mRNA has been reported in the PR negative T47D-Y breast cancer cell line (Saner et al., 2003), and PR mRNA was detected using primers to the exons encoding the LBD in PR negative MDA-MB-468 breast cancer cells (Pang and Thomas, 2011). Whilst this observation was not discussed in the report by (Pang and Thomas, 2011), a low level of PR mRNA is clearly visible in MDA-MB-468 cells in the results presented. Together with these previous studies the results presented in this chapter support the existence of PR variants which are not detected by common screening methods leading to characterisation as PR negative.

Sequencing of the PCR products generated by the PR gene walking assay identified several modes of alternative splicing occurring in MDA-MB-231 cells to generate truncated PR mRNA including; whole, partial and multiple exon deletions, and intron retention (Figure 3.6). Examples of sequencing analysis can be seen in Appendix 1. Each of the ASEs identified by sequencing could affect either PR-B or PR-A pre-mRNA so potentially generate two distinct variants which would not be distinguished by the gene walking assay. The structures of the potential proteins resulting from the identified ASEs in PR-B are shown in Figure 3.6. Proteins encoded by spliced variants of PR-A mRNA would be identical except for the absence of the BUS. Table 3.1 details the region deleted from the PR-A or PR-B mRNA, the number of amino-acids each mRNA would encode and the theoretical molecular weight of the protein before post-translational modifications.

Variant	Deletion	In frame	Premature STOP	No. amino acids encoded ΔPR-B/ΔPR-A	Predicted ΔPR-B/ΔPR-A molecular weight (kDa)
full length PR-B/PR-A	n/a	n/a	n/a	933/769	99/82
Δp1a	649bp within exon 1	no	exon 1	328/164	33/16
Δp1b	329bp within exon 1	no	exon 2	462/298	47/30
Δp1c	669bp within exon 1	yes	no	710/546	77/60
Δp1,p2	682bp end of exon 1, 27bp start of exon 2	no	exon 2	331/167	34/17
Δ2	exon 2	no	exon 3	556/392	57/40
Δ3	exon 3	yes	no	895/731	95/78
Δ4	exon 4	yes	no	831/667	87/70
Δp4	76bp within exon 4	no	exon 4	657/493	67/50
Δp4,5,p6	214bp end ofexon 4, exon 5, 14bp start of exon 6	no	exon 6	688/524	71/54
Δp5,p6	26bp end of exon 5, 105bp start of exon 6	no	exon 6	783/619	81/64
Δp6	96bp within exon 6	yes	no	901/737	95/78

Table 3.1: Predicted structures of alternatively spliced PR variants detected in MDA-MB-231 cells.

Computational analysis of sequenced PCR products from MDA-MB-231 cells reveals the potential structure and molecular weight of putative PR proteins resulting from alternative splicing in this cell line. ALIGN was used to determine the region removed by the ASE, ExPASy Translate to determine the presence of a premature termination codon (STOP) and ExPASy Compute pI/Mw was used to predict the molecular weight of the unmodified protein predicted based on the amino acid sequence encoded from each alternatively spliced mRNA. Full length PR-B and PR-A molecular weights were calculated by the same means and variant sizes resulting from alternative splicing of both full length mRNAs were predicted.

Δp1a and Δp1b mRNAs are the result of deletions of different regions within exon 1, and Δp1,p2 mRNA results from deletion of the 3′ end of exon 1 and the 5′ end of exon 2. These variant mRNAs would cause a change in reading frame and if translated would encode a protein lacking a varying portion of exon 1, in each case disrupting AF1, and Δp1,p2 would also lack the start of exon 2. The altered reading frame of each mRNA would encode a PTC within either exon 1 or exon 2. Therefore none of these proteins would be likely to function as a progesterone receptor. The PTC encoded by these deletions makes the mRNA likely targets for NMD. Deletion of exon 2 (Δ2 mRNA) has been previously reported in MCF-7 and T47-D breast cancer cells (Richer et al., 1998). This deletion would again cause a change in reading frame encoding a PTC within exon 3, so the mRNA may be targeted for NMD. A potential protein encoded by this mRNA

would however contain the full sequence of the N-terminal domain and therefore may still be capable of interactions with PR co-regulators and cytoplasmic signalling molecules. C-terminally truncated AR variants have been reported to be constitutively active (Hu et al., 2009; Dehm et al., 2008). These variants contained at least the first zinc finger of the DNA binding domain and were functional as nuclear transcription factors, an unlikely function for PRΔ2 based on the functional domains it contains. However, these reports do demonstrate the potential for expression of C-terminally truncated hormone receptor proteins. However, analysis of the structure of N-terminal PR constructs containing the N-terminal AFs and DBD, and which similar to the truncated AR variants were constitutively active, demonstrated that in the absence of DBD the NTD is not stabilised in a functional conformation (Bain et al., 2001; Bain et al., 2000). Therefore, Δ2 mRNA, similarly to the partial exon 1 deleted mRNAs, is unlikely to encode a functional protein.

The previously unreported Δp1c mRNA results from an in-frame deletion of 469bp within exon 1 which would encode a protein lacking 223 amino acids including AF1. This protein would however retain the DBD, hinge, LBD and DD, as well as the ligand inducible AF2 and the PR-B specific AF3 in PR-BΔp1c. Therefore this protein may be able to function through classical PR signalling mechanisms similarly to wild-type PR-A and PR-B. Additionally, PR-AΔp1c would also posses ID, so may be capable of regulating the function of full length PR. The N-terminal cystine rich domain involved in interaction of PR with c-Src (Boonyaratanakornkit et al., 2001) would also be intact, suggesting a possible role for PR-BΔp1c in activating rapid PR signalling.

PR Δ3 mRNA was previously detected at low levels in breast cancer tissue (Balleine et al., 1999b). PR exon 3 encodes the second zinc finger of the DNA binding domain, so the protein encoded by Δ3 mRNA is unlikely to function through the classical mechanism of PRE gene regulation. However, the presence of the other functional domains suggest PR-BΔ3 may be able to function through rapid PR signalling pathways, and that PR-AΔ3 may be able to regulate PR function in a similar manner to full length PR-B and PR-A respectively. Exon 4 of the PR gene encodes the hinge region, containing a nuclear location signal and phosphorylation and acetylation domains which are vital for normal nuclear localisation and gene activation (Daniel et al., 2010), and

exon 4 also encodes the N-terminal part of the LBD (Misrahi et al., 1993; Misrahi et al., 1987). Therefore, deletion of exon 4 in the Δ4 ASV makes normal genomic PR function unlikely. However, it is possible that the delayed nuclear entry and delayed gene activation observed upon mutation of the hinge region acetylation domain (Daniel et al., 2010) may still occur in the absence of this region. Furthermore, this variant may still be functional through other previously described rapid PR signalling pathways. PRΔ4 mRNA has been reported previously in a range of human tissues and cell lines (Misao et al., 2000; Hodges et al., 1999; Misao et al., 1998; Richer et al., 1998; Leygue et al., 1996a). A previous report of the function of PRΔ4 protein translated from a cDNA expression vector suggested that this variant is non-functional as either a transcriptional activator or a regulator of wild-type PR function (Richer et al., 1998). However this study was conducted before recent extensive characterisation of the non-genomic mechanisms of PR signalling and therefore focussed only on classical PR signalling. The expected structure of the PRΔ4 protein predicts that this variant would most likely function through non-genomic pathways.

Previously unreported out of frame deletions were observed within exon 4 (Δp4), spanning the 3′ end of exon 4, exon 5 and the 5′ end of exon 6 (Δp4,5,p6) and spanning the 3′ end of exon 5 and the 5′ end of exon 6 (Δp5,p6). Δp4 lacks the hinge region and generates a PTC within exon 4 encoding a protein also lacking the LBD. Both Δp4,5,p6 and Δp5,p6 contain PTC within exon 6, encoding a protein with a truncated LBD. The presence of the PTC may target these mRNA for NMD. If a protein were encoded, the lack of a functional LBD would make function as a ligand activated transcription factor unlikely. However, ligand independent PR functions have been reported, regulating genes in the absence of ligand stimulation (Jacobsen et al., 2005) via activation by cross-talk with other signalling pathways (Proietti et al., 2009; Pierson-Mullany and Lange, 2004), and these functions may be particularly important in breast cancer development (Cerliani et al., 2011; Hagan et al., 2011). Furthermore, C-terminally truncated ER proteins encoded from mRNAs containing an insertion between exons 4 and 5 are truncated after exon 4, thus lacking the LBD. These proteins were reported in the mouse to be constitutively transcriptionally active nuclear proteins (Ishii et al., 2011), demonstrating the potential functionality of hormone receptor proteins lacking LBD, suggesting that if encoded PRΔp4, Δp4,5,p6 or Δp5,p6 proteins may be able to function independent of ligand activation and contribute to breast cancer development.

A novel in-frame deletion of 96bp was observed within exon 6, generating a protein lacking just 32 amino acids from within the LBD. Therefore the encoded protein may lack a functional LBD and function through other PR signalling mechanisms as discussed for other variants, or the small deletion may leave the LBD functional and thus the protein may function in a similar manner to wild-type PR.

Retention of intronic sequences within PR mRNA has previously been reported as described in Section 1.4.11. Exons S and T from intron 3 are spliced onto exon 4 and may encode a protein from the first initiation codon within exon 4 (Hirata et al., 2002; Hirata et al., 2000). Exon M which contains a consensus Kozak translation initiation site is also located within intron 3 and spliced onto exon 4 (Saner et al., 2003). These mRNA would therefore encode N-terminally truncated proteins containing the normal hinge and C-terminal domains. Exons i45a and i45b are located within intron 4 and are spliced onto the end of exon 4 (Yamanaka et al., 2002). i45a contains an in-frame termination codon so this mRNA would encode a protein similar in structure to those discussed for Δp4, Δp4,5,p6 and Δp5,p6, containing a short i45a specific sequence of amino acids. PCR primers directed to these intronic exons detected mRNA for each isoform in the PR positive MCF-7 breast cancer cell line, and mRNA for PR-M was also detected and sequenced in the reportedly PR negative MDA-MB-231 cell line (Figure 3.5). The presence of mRNA encoding an N-terminally truncated PR protein supports the initial hypothesis that expression of such proteins may lead to misclassification as PR negative. Exon M would encode a unique N-terminal sequence of 16 hydrophobic amino acids and therefore the function of PR-M was initially suggested to be as a membrane associated receptor (Saner et al., 2003). However, later analysis of the expression of green fluorescent protein (GFP) tagged PR-M identified the protein to be expressed predominantly in the cytoplasm (Price et al., 2005). This suggests a potential mechanism of PR-M protein function through interaction with other cytoplasmic signalling pathways. However, the absence of the N-terminal domain which has been implicated in interactions with Src and other signalling molecules suggests other domains my mediate these interactions for PR-M. The expression and function of N-terminally truncated ER variants was recently examined in the mouse (Ishii et al., 2011), with a cytoplasmic expression and an inability to activate ERE being observed. The observation that both N-terminally truncated PR and ER proteins are expressed in the cytoplasm suggests that the PR-M mRNA detected in the MDA-MB-

231 cell line may encode a protein which could function cytoplasmically via non-genomic signalling pathways.

Analysis of the potential splice sites associated with each deletion reveals that none of the partial exon deletions (Δp1a, Δp1b, Δp1c, Δp1,p2, Δp4, Δp4,5,p6, Δp5,p6 or Δ6) were associated with a classical GT 5′ splice site or AG 3′ splice site. Further analysis of the sequences is necessary to identify potential ESE sequences which could serve to strengthen weak non-consensus splice sites within the exons and silencer elements which may inactivate the normal splice sites. Skipping of exons 2 and 3 are possible through silencing of the 3′ splice sites at the end of the preceding intron, although again further analysis of the sequences is required to identify silencer elements which may mediate this mechanism of splicing. Exon 4 is 306bp long, longer than the ideal length of <300bp necessary for efficient exon definition (Berget, 1995). At the 5′ end of exon 4 is a GT dinucleotide, although the rest of the 5′ sequence differs from the splice site consensus. The 3′ end of exon 4 is an AG dinucleotide, with a short PPT located just upstream, and a consensus BPS approximately 100bp upstream of the 3′ AG. Whilst the weak 5′ splice site and short PPT mean that utilisation of these splice sites may require enhancer sequences, the presence of the 5′ and 3′ splice sites together with the length of the exon suggest that exon 4 is likely to undergo enhancer dependent alternative splicing. Preliminary investigation of the regulation of PR exon 4 splicing and analysis for the presence of enhancer elements within the sequence will be undertaken in Chapter 6.

The PR gene walking assay, using RT-PCR, agarose gel electrophoresis and DNA sequencing, enabled the identification of novel ASEs which would not have been possible by real-time quantitative (q)RT-PCR which would be unable to distinguish between full length and exon deleted mRNA when quantifying expression. Exon deletion ASEs can be assessed by qRT-PCR using splice specific primers (as described by Springwald *et al* (Springwald et al., 2010). This method however requires prediction of ASEs and design of primers for each predicted ASE whereas the conventional RT-PCR and agarose gel electrophoresis method employed in this study removes this requirement. And does not allow comparison of the levels of expression of exon deleted and full length mRNA. The relative costs of the two methods for application to a

discovery screen such as the gene walking assay further support the use of conventional RT-PCR. Capillary electrophoresis is a method of separating differently sized PCR products by which product size is analysed by UV absorbance. This method allows quantification of the relative expression levels of differently sized PCR products within a sample and is an alternative to the agarose gel electrophoresis used in this study. However, PCR products separated by agarose gel electrophoresis can be purified, cloned and sequenced to identify the regions deleted. Thus, despite strengths of individual alternative methods, the RT-PCR and agarose gel electrophoresis method used in this study is the most appropriate for a gene walking discovery assay.

Clinicopathological screening of PR status in breast tumour tissue uses antibodies targeted to N-terminal epitopes common to both PR-A and PR-B (Hanley et al., 2009). These epitopes would also be present in the proteins produced from the exon deleted PR mRNA detected by the gene walking assay, so this screening would be unable to distinguish between differently functional PR isoforms. Additionally, these antibodies would be unable to detect N-terminally truncated proteins such as PR-M. Using N-terminal antibodies similar to those used clinically, PR protein was detected specifically in PR positive MCF-7 cells (Figure 3.7a) and the specificity of these antibodies was confirmed by siRNA knockdown of PR expression (Figures 3.11 and 3.12). However, oestrogen inducible expression of low molecular weight PR proteins was also detected in MDA-MB-231 cells using the C-terminally targeted C19 antibody (Figure 3.9). PR is an oestrogen responsive gene containing a partial ERE sequence within the promoters for both PR-A and PR-B (Kastner et al., 1990) so the increase in PR expression following β-oestradiol stimulation observed in Figure 3.8 was expected. Oestrogen has also been shown to influence alternative splicing; β-oestradiol treatment decreased inclusion of an alternatively spliced exon in the rat Slo channel mRNA (Zhu et al., 2005), and increased exon 11 skipping in the liver scavenger receptor B (SR-B) (Zhang et al., 2007). Liver SR-B exon 11 skipping was regulated by SRSF1, and the SR-related proteins Tra2α and Tra2β, with the expression of Tra2β regulated by β-oestradiol suggesting that β-oestradiol may regulate exon skipping by regulating SR-related protein expression (Zhang et al., 2007). The increase in expression of low molecular weight proteins following β-oestradiol stimulation of MDA-MB-231 breast cancer cells (Figure 3.9), and the apparent switch in isoform expression observed in MCF-10A non-

tumourigenic MECs (Figure 3.9) following stimulation may indicate a similar effect of β-oestradiol on PR alternative splicing to that previously reported for Slo and SR-B.

siRNA knockdown of total PR expression in MCF-7 cells confirmed that low molecular weight proteins detected by C19 are PR proteins, although non-specific proteins were also detected (Figure 3.13). Expression of proteins of approximately 80kDa, 50kDa and a doublet between 30kDa and 40kDa was knocked down by PR siRNA, expression of which was not due to protease degradation of full length PR (Figure 3.15) or differential phosphorylation (Figure 3.16). A truncated PR protein has previously been reported with molecular weights of 60kDa resulting from use of an alternative promoter and initiation of translation at Met595 (Wei and Miner, 1994). Expression of this protein has since been reported using the C19 antibody in human tissue (Condon et al., 2006). However, recent reports suggest that the 60kDa protein detected by C19 is in fact a cytoskeletal protein (Madsen et al., 2007), and that the PR-C cDNA is likely to give rise to a protein smaller than originally predicted, around 38kDa (Samalecos and Gellersen, 2008). These reports are supported by the results presented in this chapter demonstrating that expression of the 60kDa protein detected by C19 is not reduced by siRNA targeting PR (Figure 3.13). This validation of the specificity of the C19 antibody was performed in MCF-7 cells whereas analysis of PR alternative splicing was performed predominantly in MDA-MB-231 cells. Transfection of MDA-MB-231 cells with siRNA was attempted; the transfection conditions used in MCF-7 cells did not produce knockdown of PR or GAPDH and extensive optimisation using three different transfection reagents at a range of concentrations with a range of siRNA concentrations failed to achieve efficient knockdown. Transfection of MDA-MB-231 cells cultured on glass coverslips with GFP labelled GAPDH siRNA and analysis by confocal microscopy did not demonstrate any uptake of the siRNA or any reduction in the level of GAPDH when cell lysates were analysed by Western immunoblotting (data not shown). Since both MCF-7 and MDA-MB-231 cells express a similar range of proteins detected by C19 (Figures 3.7c, 3.15 and 3.16) and PR appears to undergo a similar pattern of alternative splicing in the two cell lines, validation of the antibody was performed in just the MCF-7 cells and the PR proteins identified in this cell line can be compared to the proteins predicted from the ASEs identified in MDA-MB-231 cells.

Table 3.1 shows the predicted molecular weight of proteins resulting from each of the detected ASEs affecting either PR-B or PR-A. From these predicted molecular weights several potential identities for each of the knocked down PR proteins can be determined based on the pattern of splicing detected by the PR gene walking assay. Full length PR-A (82kDa), PR-AΔ3 (78kDa), PR-AΔp6 (78kDa), PR-BΔp1c (77kDa), PR-BΔ4 (87kDa) or PR-BΔp5,p6 (81kDa) could all account for the PR protein detected at approximately 80kDa although the latter is unlikely to be a functional PR protein. A truncated PR isoform of 78kDa has previously been reported to be detected in approximately 26% of breast tumours (Graham et al., 1996). This protein could be detected by an antibody targeting both PR-A and PR-B, but not by an antibody targeting the B-upstream segment, and it was therefore suggested that the protein was N-terminally truncated, perhaps resulting from alternative splicing of PR-A pre-mRNA. This 78kDa PR protein was also demonstrated to bind ligand (Yeates et al., 1998). If the approximately 80kDa protein knocked down by PR siRNA in this study is the previously identified 78kDa protein then the PR-AΔ3 mRNA is most likely to give rise to an appropriately sized protein with ligand binding potential.

An approximately 50kDa PR protein was also identified by siRNA knockdown and Western immunoblotting using C19. PR-AΔp4 (50kDa), PR-AΔp4,5,p6 (54kDa), PR-BΔp1b (47kDa) or PR-BΔ2 (57kDa) mRNAs could give rise to appropriately sized proteins. PR-AΔ2 (40kDa) or the more likely functional PR-M (39kDa) could give rise to proteins represented by the 40kDa band. Alternatively, as suggested previously (Samalecos and Gellersen, 2008), PR-C cDNA may encode a 38kDa protein which could be the protein detected by C19 at approximately 40kDa. The lower band of the doublet could be the protein product of either PR-AΔp1b, PR-BΔp1a or PR-BΔp1,p2 mRNAs, again none of which are predicted to give rise to functional PR proteins. The low molecular weight PR proteins detected by C19 undergo extensive phosphorylation; dephosphorylating the proteins increased the intensity of the 50kDa band (Figure 3.16). This phosphorylation could provide a mechanism by which non-ligand binding proteins which may be produced from the ASV mRNAs could be activated. However, the detection of these proteins by C19 in MDA-MB-231 cells which appear PR negative when analysed using N-terminal antibodies suggests that the proteins are unlikely to be exon deleted variants of either PR-A or PR-B. Furthermore, as previously suggested, (Samalecos and Gellersen, 2008; Yeates et al., 1998), the levels of exon deleted mRNAs

detected by RT-PCR relative to wild-type mRNA are unlikely to give rise to highly expressed proteins. Therefore, it is more likely that the truncated proteins detected in MDA-MB-231 and MCF-7 cells, and knocked down by siRNA targeting PR in MCF-7 cells, are N-terminally truncated, resulting from alternative promoters or intron retention.

Expression of non-functional exon deleted PR variant proteins which would be detected by PR screening and not distinguished from PR-A or PR-B could lead to patients being characterised as PR positive but lacking PR function. Alternatively, variant PR proteins such as PRΔ3 or PRΔ4 may function differently to wild-type PR and could alter the progesterone response of a tissue without altering the overall appearance as PR positive since they would be detected by N-terminal antibodies in PR screening. N-terminally truncated proteins such as PR-M or PR-C would not be detected by N-terminal antibodies, but would be able to bind progesterone and therefore may function as PR in tissues described as PR negative.

The C19 antibody was demonstrated by Western immunoblotting following siRNA knockdown of PR expression in MCF-7 cells to detect PR proteins in addition to some non-specific proteins. When used for immunofluorescent analysis of PR expression in MDA-MB-231 cells, C19 detects abundant nuclear and cytoplasmic proteins and detects increased nuclear staining following progesterone stimulation of the cells (Figure 3.10). However, upon siRNA knockdown of total PR expression in MCF-7 cells, which was confirmed by Western immunoblotting, C19 still detects protein by immunofluorescence (Figure 3.14). Therefore, it can be concluded that the non-specific binding observed using C19 for Western immunoblotting makes this antibody unsuitable for detecting PR by immunofluorescence, by which different proteins cannot be distinguished by size. C19 has previously been used to detect PR-C in human tissue (Condon et al., 2006) and to detect the PR-M protein in PR negative T47D-Y breast cancer cells (Price et al., 2005). The observation that C19 detects protein in cells following PR knockdown suggests that these immunofluorescent studies may not be identifying PR protein. This is supported by the predominant membrane staining observed using C19 in T47D-Y cells, suggested to be PR-M, but a cytoplasmic localisation of GFP tagged PR-M (Price et al., 2005), suggesting that the antibody is not detecting the protein that was expected. Therefore the development of novel, more

specific, non N-terminal PR antibodies could improve both laboratory investigation of PR expression and function, as well as clinicopathological detection of PR in breast cancer.

4 Validating the functionality of truncated PR isoforms in MDA-MB-231 cells

4.1 Introduction

MDA-MB-231 cells are described as a PR- breast cancer cell line (Neve et al., 2006) and it has been demonstrated in this thesis that Western immunoblot analysis using PR specific N-terminal antibodies, similar to those used in breast cancer screening, does not detect PR protein in lysates from these cells (Section 3.3.2). It was also demonstrated by the PR gene walking assay in Section 3.3.1 of this thesis that alternatively spliced PR mRNA is present in MDA-MB-231 cells which may encode truncated, potentially functional proteins. Further to this, Western immunoblotting and hormone response assays have demonstrated that truncated PR proteins can be detected using the C-terminal PR antibody C19 in MDA-MB-231 cells (Sections 3.3.2 and 3.3.4). A similar range of protein bands was observed in Western immunoblotting of MCF-7 and MDA-MB-231 whole cell lysates by C19 and siRNA knockdown of total PR expression in MCF-7 cells identified 80kDa, 55kDa and 30-40kDa PR proteins detected by this antibody (Section 3.6). An analysis of the potential for the proteins detected by C19 to function as classical ligand activated nuclear transcription factor PR isoforms in MDA-MB-231 cells was therefore undertaken to examine ligand binding, dimerisation, co-factor binding and DNA binding.

Nuclear progesterone receptor isoforms PR-A and PR-B function as ligand activated nuclear transcription factors. The first stage of the classical mechanism of PR action is the binding of progesterone to the LBD of PR in the cytoplasm. Progesterone binding has previously been assessed using a ligand blot assay (Jang and Yi, 2005). Proteins are separated by SDS-PAGE and transferred to a nitrocellulose membrane. Membranes are then incubated with P-POD which binds to any progesterone binding proteins present and can be detected by developing using ECL and x-ray film. Rapid non-genomic progesterone responses can be mediated through cytoplasmic PR-B (Boonyaratanakornkit et al., 2007) and potentially through other truncated variant PR isoforms which result from alternative splicing (Price et al., 2005) or the use of alternative promoters (Taylor et al., 2009; Condon et al., 2006). Membrane bound progesterone receptors (mPRs) are the products of different genes to the 8 exon gene

encoding PR-A and PR-B. Interaction of the intracellular domain of ligand activated mPRs with other signalling pathways can activate cellular responses to progesterone that are independent of DNA binding (Thomas, 2008). Response to cell-impermeable progesterone has been reported in the PR negative T47D-Y breast cancer cell line suggesting that membrane associated PR proteins, potentially mPR, are expressed in this cell line which is characterised as nuclear PR negative (Lee and Muller, 2010). In this chapter, the presence of non-genomic progesterone binding proteins in MDA-MB-231 cells was assessed by ligand blotting using fractionated lysates from breast cancer cells.

Following ligand binding, nuclear receptors undergo dimerisation and nuclear translocation. PR homo- or heterodimers then interact with transcriptional co-regulators. The interaction of PR with other proteins can be examined by Co-IP using highly specific antibodies. If PR proteins are in complex with the targeted proteins they can then be detected by Western immunoblotting (Tyson-Capper et al., 2009). PPT-binding protein associated splicing factor (PSF) is a multi-functional nuclear protein possessing both RNA and DNA binding domains. PSF was first characterised bound to the PPT in intronic RNA and associates with spliceosome components involved in the second catalytic step of splicing, with a role in 3′ splice site selection. PSF has also been demonstrated to bind to the DBD of thyroid hormone and retinoid X receptors and prevent their DNA binding in the absence of ligand. The DNA binding capacity of PSF was demonstrated by binding to insulin-like growth factor response elements (IGFRE) and preventing transcription of the porcine P450scc gene (Shav-Tal and Zipori, 2002). In human cells, murine (m)VL30 induces the transcription of the human homologue of P450scc and other IGFRE containing genes, including the oncogene 24p3. PSF was demonstrated to complex with the PSF binding tracts (pbt) of mVL30 and ptb was demonstrated to prevent the binding of PSF to IGFRE, thus preventing the repression of gene expression by PSF (Song et al., 2004). This indicates a potential role for PSF in the repression of cancer associated gene expression. Furthermore, repression of PSF by interaction with pbt induced progesterone synthesis in mice, demonstrating a role for PSF in regulating steroidogenesis (Song et al., 2004). Due to these established roles of PSF in cancer, hormone synthesis and splicing, in addition to a previously demonstrated interaction of PSF with PR-A and PR-B in myometrial cells (Tyson-Capper et al., 2009),

PSF was chosen as a candidate PR co-factor to investigate the potential functionality of truncated PR proteins by Co-IP.

Activated PR dimers then bind to PREs in promoter regions of progesterone controlled genes to regulate gene expression. A biotinylated synthetic PRE oligonucleotide sequence was used to bind to any PR proteins which possess DNA binding capacity present in cell lysates. These protein/PRE complexes can then be precipitated using streptavidin magnetic beads and the PR proteins present can be analysed by Western immunoblotting (Tyson-Capper et al., 2009).

4.2 Aims

The aims of this chapter are to:

1. Identify the proteins detected by the C-terminal PR antibody C19 in different cell fractions from MDA-MB-231 breast cancer cells by Western immunoblotting.

2. Determine the progesterone binding capacity of proteins in MDA-MB-231 cells using a ligand blot assay.

3. Assess the detection of PR proteins in MDA-MB-231 cells by two commercial non N-terminal PR antibodies by Co-IP using C20 and detection of proteins by Western immunoblotting with C19, and to use this assay to assess the potential presence of PR dimers in MDA-MB-231 cells.

4. Identify interactions of proteins detected by C19 in MDA-MB-231 cells with the nuclear PR co-factor PSF by Co-IP.

5. Determine if truncated PR proteins detected by C19 in MDA-MB-231 cells can interact with PREs by DAPA using biotinylated PRE.

4.3 Results

4.3.1 Low molecular weight progesterone binding proteins exist in MDA-MB-231 cells

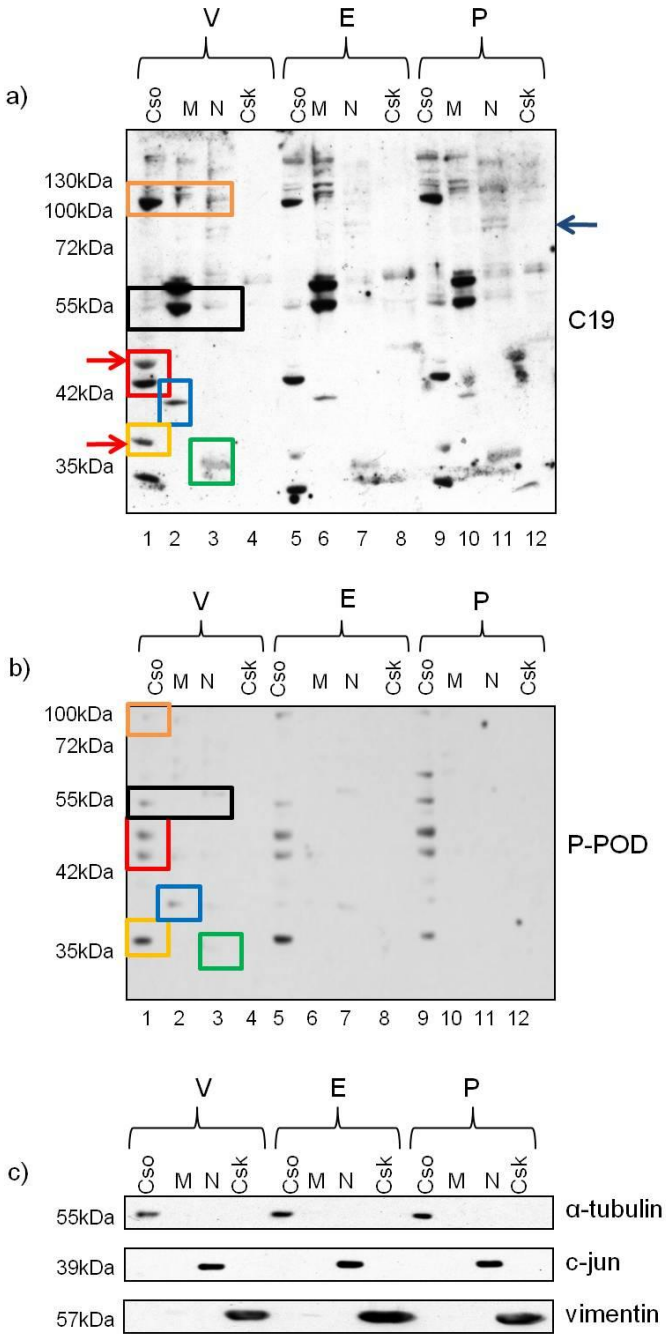
The C-terminal PR antibody C19 detects a range of proteins in breast cancer cells, some of which have been demonstrated in this thesis to be PR proteins by siRNA knockdown of total PR expression in MCF-7 breast cancer cells (Section 3.3.6). To establish the potential for the proteins detected by C19 to function as progesterone receptors, their ability to bind progesterone was examined. MDA-MB-231 cells were stimulated with 1μM β-oestradiol, the optimal concentration for stimulating low molecular weight PR protein expression (Section 3.3.4) or an ethanol control and cultured for 24 hours to allow the effects of stimulation on protein expression to develop. Following β-oestradiol stimulation some cells were stimulated with 0.1μM progesterone (a concentration optimised to stimulate PR expression) and cultured for a further 24 hours. Following hormone stimulation cells were fractionated to isolate cytosolic, membrane, nuclear and cytoskeletal proteins. Proteins in the four cell fractions were then separated by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were used for either Western immunoblotting with C19, or ligand blotting using P-POD.

Western immunoblotting using C19 detected a range of proteins in each of the cell fractions of both hormone stimulated and unstimulated cells (Figure 4.1a). A similar pattern of proteins was observed with β-oestradiol or progesterone treatments, and in the control ethanol treated cells. However, an apparent decrease in expression of some low molecular weight proteins was observed in the cytosolic fraction following β-oestradiol stimulation (Figure 4.1a, lanes 5 and 9, indicated with red arrows). An increased detection of nuclear proteins by C19 can be observed following progesterone stimulation suggesting nuclear translocation (Figure 4.1a, lane 11).

The proteins detected by C19 were compared to proteins present in the MDA-MB-231 cell fractions which bound to peroxidase conjugated progesterone in the ligand blot assay (Figure 4.1b). Proteins of approximately 100kDa in the cytosolic fraction (orange box), 55kDa in the cytosolic and nuclear fractions (black box), two proteins between 42 and 50kDa in the cytosolic fraction (red box), a protein in the membrane fraction just

under 42kDa (blue box), and two different proteins of approximately 35kDa in the cytosolic and nuclear fractions were all detected by both assays. Progesterone binding proteins were not detected in the cytoskeletal fraction of MDA-MB-231 cells.

The purity of the individual fractions was confirmed by detection of α-tubulin specifically in the cytosolic fraction, c-jun in the nuclear fraction and vimentin in the cytoskeletal fraction (Figure 4.1c). Purity of the membrane fraction was concluded based on the purity of the other three fractions and the unique pattern of proteins detected in this fraction by Western immunoblotting and ligand blotting.



<u>Figure 4.1:</u> Detection of progesterone binding proteins in MDA-MB-231 cells.

MDA-MB-231 cells were stimulated with 1µM β-oestradiol (E), 1µM β-oestradiol followed by 0.1µM progesterone (P) or an ethanol only control (V). Cells were then fractionated to separate cytosolic (Cso), membrane (M), nuclear (N) and cytoskeletal (Csk) fractions. Proteins within the fractions were separated by SDS-PAGE and transferred to nitrocellulose membranes. **a)** Western immunoblotting using the C-terminal PR antibody C19. The blue arrow indicates an 80kDa nuclear protein detected by C19 in nuclear fractions (lanes 3, 7 and 11). **b)** Ligand blotting using peroxidase conjugated progesterone (P-POD). Coloured boxes in **a** and **b** highlight where proteins of similar molecular weight are detected by C19 and P-POD, representing progesterone binding PR proteins as discussed above. **c)** Purity of the fractions was confirmed by Western immunoblotting using antibodies against α-tubulin (Cso), c-jun (N) and vimentin (Csk).

4.3.2 Identifying proteins detected by the C-terminal PR antibody C19 in MDA-MB-231 cells

MDA-MB-231 cells were fractionated to separate nuclear and cytoplasmic proteins following stimulation with 1μM β-oestradiol, 1μM β-oestradiol followed by 0.1μM progesterone or an ethanol only control. These fractions were initially used for Western immunoblotting using the C-terminal PR antibody C19 to identify the proteins detected by this antibody in each fraction prior to using C19 as the detection antibody in future Co-IP and DAPA experiments. C19 detected a range of different molecular weight proteins in both nuclear and cytoplasmic fractions of MDA-MB-231 cells (Figure 4.2a). Proteins of approximately 100kDa and 80kDa were detected specifically in the nuclear fraction along with a doublet of proteins detected at 35kDa (Figure 4.2a, lanes 1-3). A 55kDa protein was detected in the nuclear and cytoplasmic fractions (Figure 4.2a). In the cytoplasmic fraction, a protein of 42kDa was also detected and expression of the 42kDa and 55kDa cytoplasmic proteins was induced by β-oestradiol stimulation (Figure 4.2, lane 5). The purity of the fractions was confirmed by the detection of c-jun specifically in the nuclear fractions and ERK-1 in the cytoplasmic fractions (Figure 4.2b).

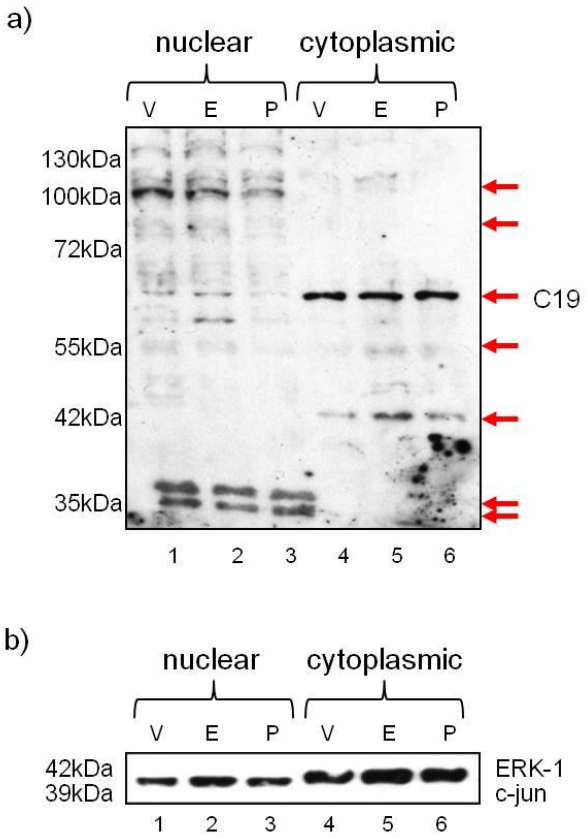


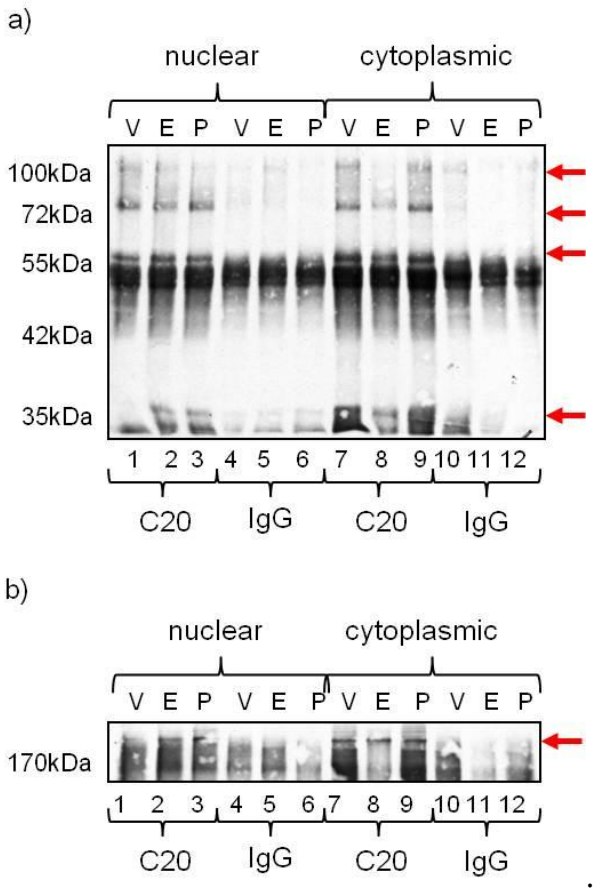
Figure 4.2: Detection of PR proteins in MDA-MB-231 nuclear and cytoplasmic fractions using C19.

MDA-MB-231 cells were stimulated with 1µM β-oestradiol (E), 1µM β-oestradiol followed by 0.1µM progesterone (P) or an ethanol only control (V) and cultured for 24 hours before harvesting and fractionating to separate nuclear and cytoplasmic proteins. **a)** Western immunoblotting of MDA-MB-231 nuclear and cytoplamic fractions using C19 to. **b)** Western immunoblotting of MDA-MB-231 nuclear and cytoplasmic fractions to confirm the purity of the fractions using c-jun (nuclear) and ERK-1 (cytoplasmic).

4.3.3 PR proteins can form heterodimers in MDA-MB-231 cells

As described, C19 detects a range of potential PR proteins in MDA-MB-231 cells. To identify whether these proteins, which are not detected by N-terminal antibodies, can also be detected by another non N-terminal antibody and to identify the ability of the proteins detected by C19 to dimerise, co-immunoprecipitation experiments were performed using nuclear and cytoplasmic fractions from hormone stimulated and unstimulated MDA-MB-231 cells. The PR antibody C20, targeted to an epitope within the DBD, was used as the capture antibody and proteins bound to this antibody were precipitated using Protein G magnetic beads. Immunoprecipitated proteins were separated by SDS-PAGE and used for Western immunoblotting with C19 as the detection antibody.

Immunoprecipitated proteins were detected by C19 in both nuclear and cytoplasmic fractions at approximately 100kDa, 80kDa, 55kDa and 35kDa (Figure 4.3a, lanes 1-3 and 7-9, indicated by red arrows), demonstrating that the proteins detected by C19 are also bound by the other PR antibody. These proteins were not detected in IgG control experiments. In addition to this, Western immunoblotting using C19 following Co-IP with C20 detects protein bands in excess of 170kDa, suggesting that proteins detected by C19 may be in complex with other PR proteins detected by C20, and that these heteromeric complexes may be resistant to denaturation during sample preparation (Figure 4.3b lanes 1-3 and 7-9, indicated by red arrow).

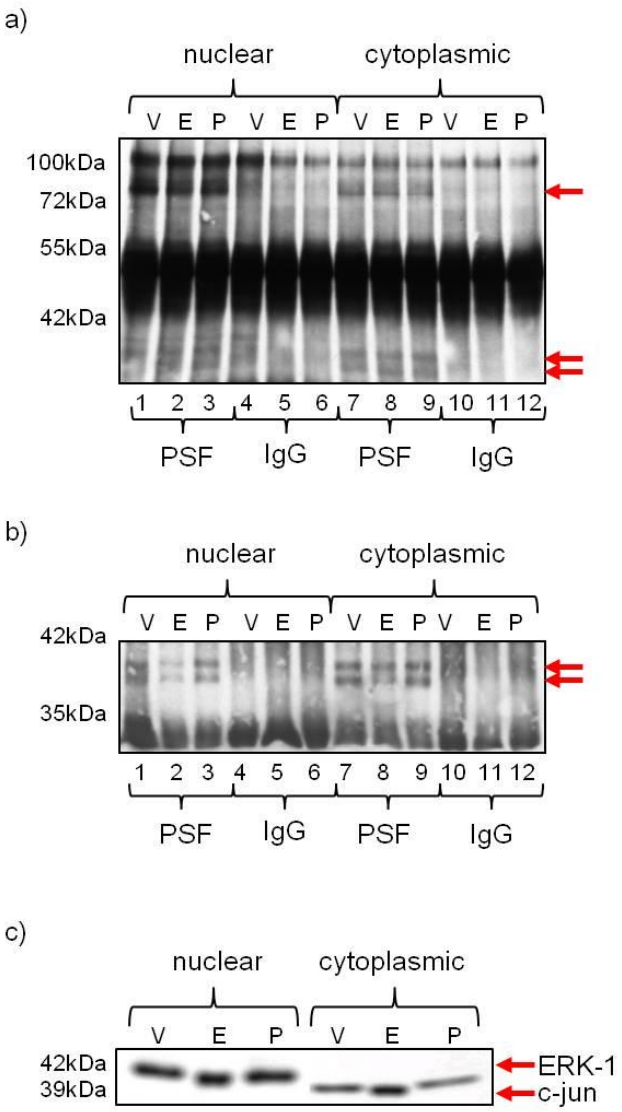


<u>Figure 4.3:</u> C19 detects proteins that are immunoprecipitated using C20, including high molecular weight heterodimers.

Nuclear and cytoplasmic fractions of MDA-MB-231 cells stimulated with 1μM β-oestradiol (E), 1μM β-oestradiol followed by 0.1μM progesterone (P) or an ethanol only control (V) were used for Co-IP using the PR antibody C20 as the capture (lanes 1-3 and 7-9) and a rabbit IgG as control (lanes 4-6 and 10-12). **a)** Western immunoblotting of immunoprecipitated proteins using C19 detects PR proteins in MDA-MB-231 nuclear and cytoplasmic fractions. **b)** Western immunoblotting with C19 detects potential high molecular weight heterodimers in MDA-MB-231 cells.

4.3.4 Nuclear and cytoplasmic PR proteins interact with the PR co-regulator PSF

As described in Section 4.1, PSF is a nuclear receptor co-factor with previously described roles in alternative splicing, hormone synthesis and cancer which has previously been demonstrated to interact with PR-A and PR-B. These roles relating to the subject of this thesis therefore make PSF an ideal co-factor for preliminary analysis of the functionality of low molecular weight PR proteins. To identify interaction of the low molecular weight PR proteins detected by C19 in MDA-MB-231 cells with the nuclear PR co-factor PSF, Co-IP experiments were performed using MDA-MB-231 nuclear and cytoplasmic fractions and a highly specific PSF monoclonal antibody, which has previously been validated for use in Co-IP experiments (Tyson-Capper et al., 2009), as the capture antibody. C19 was used for Western immunoblotting to detect PR proteins which were precipitated with PSF and identify whether any truncated PR isoforms can interact with this PR co-factor in a similar manner to wild-type PR. Western immunoblotting with C19 detected a protein at approximately 80kDa predominantly in the nuclear fraction which was not detected in IgG control experiments (Figure 4.4a, upper red arrow). C19 also detected a predominantly cytoplasmic doublet of proteins at around 35-40kDa which was co-immunoprecipitated with PSF (Figure 4.4a, lower red arrows). In MCF-7 breast cancer cells, a similar protein doublet was also detected at relatively high levels in the nuclear fraction by the N-terminal PR antibody NCL-PgR-AB (Figure 4.4b). The purity of MCF-7 fractions was confirmed by the detection of ERK-1 specifically in cytoplasmic fractions and c-jun predominantly in nuclear fractions (Figure 4.4c).



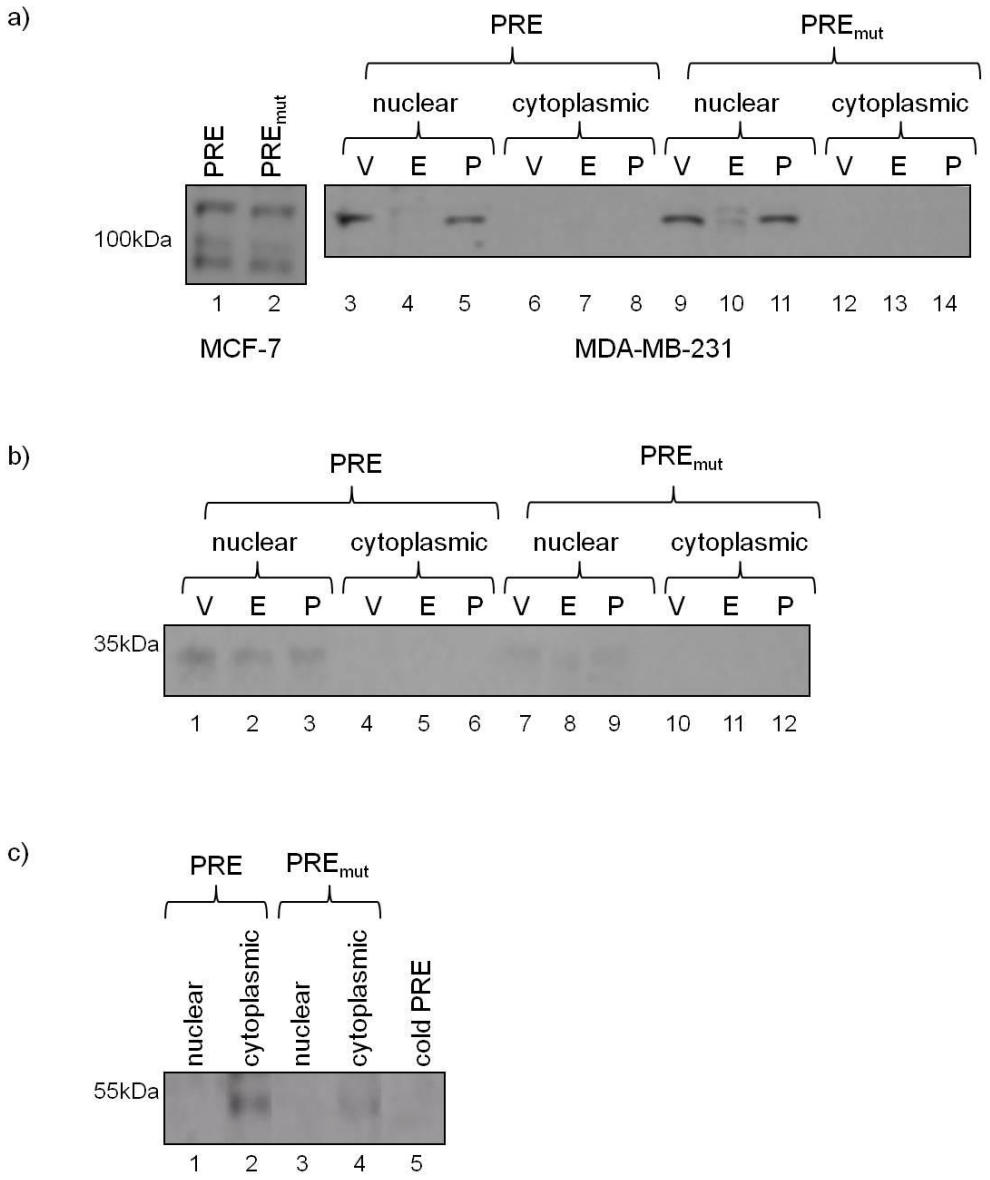
<u>Figure 4.4:</u> Low molecular weight PR proteins can form complexes with PSF in MDA-MB-231 and MCF-7 breast cancer cells.

MDA-MB-231 and MCF-7 breast cancer cells were stimulated in culture with 1µM β-oestradiol (E), 1µM β-oestradiol followed by 0.1µM progesterone (P) or an ethanol only control (V). Nuclear and cytoplasmic fractions from these cells were used for Co-IP using an anti-PSF antibody. **a)** Western immunoblotting of co-immunoprecipitated proteins from MDA-MB-231 cells using C19. Red arrows indicate proteins co-immunoprecipitated specifically with PSF (lanes 1-3 and 7-9) and not the IgG controls (lanes 4-6 and 10-12).**b)** Western immunoblotting of co-immunoprecipitated proteins in MCF-7 cells using NCL-PgR-AB. **c)** Western immunoblotting of MCF-7 nuclear and cytoplasmic fractions using fraction controls c-jun (nuclear) and ERK-1 (cytoplasmic).

4.3.5 PRE binding proteins are present in MDA-MB-231 cells

To assess the interaction of proteins detected by C19 in MDA-MB-231 cells with PRE, nuclear and cytoplasmic extracts were incubated with double stranded biotinylated oligonucleotides. These oligonucleotides were either a consensus PRE or a mutated sequence (detailed in Section 2.7.2). PRE binding proteins were precipitated using streptavidin magnetic beads and analysed by Western immunoblotting with C19 (Figure 4.5). In MCF-7 cells, a PRE binding nuclear protein which migrated slightly faster than the 100kDa molecular weight marker was assumed to be PR-A, and a larger protein nearer to the 130kDa marker was assumed to be PR-B (Figure 4.5a, lane 1). In MDA-MB-231 cells, a PRE binding protein was detected specifically in the nuclear fraction (Figure 4.5a, lanes 3-8) and did not migrate to the same size as either of the PR-A or PR-B proteins detected in MCF-7 cells. However, perhaps due to insufficient mutation of the consensus PRE sequence, PR-A, PR-B (Figure 4.5a, lane 2), and the MDA-MB-231 specific 100kDa protein (Figure 4.5a, lanes 9-11) could also bind PRE_{mut}. In MDA-MB-231 cells, a nuclear protein which migrated slightly faster than the 35kDa molecular weight marker bound to PRE with a higher affinity than PRE_{mut} (Figure 4.5b, lanes 1-3 and 7-9).

In addition to these nuclear PRE binding proteins present in MDA-MB-231 cells, under these artificial conditions, where PRE is accessible to cytoplasmic proteins, a protein that migrated slightly faster than the 55kDa marker was detected specifically in the cytoplasmic fraction (Figure 4.5c, lane 2). This protein did not bind to PRE_{mut} (Figure 4.5c, lane 4), and the binding to biotinylated PRE could be successfully prevented by incubation of the cell extract with an unlabelled PRE prior to incubation with the biotinylated PRE (Figure 4.5c, lane 5).



<u>Figure 4.5:</u> MDA-MB-231 cells express proteins capable of binging PRE.

MDA-MB-231 and MCF-7 breast cancer cells were stimulated in culture by 1µM β-oestradiol (E), 1µM β-oestradiol followed by 0.1µM progesterone (P) or an ethanol only control (V). Nuclear and cytoplasmic fractions from these cells were used for DAPA reactions using a biotinylated synthetic PRE oligonucleotide or a biotinylated mutated sequence PRE (PRE_{mut}). Biotinylated oligonuleotides, and any proteins which bind them, were precipitated using streptavidin magnetic beads. Precipitated proteins were analysed by Western immunoblotting using C19. **a)** Oligonucleotide binding proteins in vehicle treated MCF-7 nuclear extract and MDA-MB-231 nuclear and cytoplasmic fractions. **b)** Oligonucleotide binding proteins in nuclear and cytoplasmic fractions of oestrogen stimulated MDA-MB-231 cells. Cytoplasmic extract was also pre-incubated with unbiotinylated PRE (cold PRE). **c)** Oligonucleotide binding proteins in MDA-MB-231 nuclear and cytoplasmic fractions.

4.4 Discussion

The C-terminal PR antibody C19 detects a range of proteins in cytosolic, membrane and nuclear fractions of MDA-MB-231 cells. Expression of these proteins appears to be responsive to β-oestradiol; low molecular weight proteins of around 37kDa and 40-45kDa in the cytosolic fraction appear reduced in expression following stimulation whereas there is increased detection of a range of proteins in the nuclear fraction following β-oestradiol stimulation (Figure 4.1a). Several of the proteins detected by C19 in MDA-MB-231 cells were demonstrated by the ligand blot assay to bind progesterone (Figure 4.1b). Three proteins between 35kDa and 42kDa were detected by C19 and were able to bind progesterone; one in the nuclear fraction, one in the membrane fraction and one in the cytosolic fraction. These proteins correlate to the size of proteins knocked down by PR siRNA in MCF-7 cells in Section 3.3.6 of this thesis. The detection of three proteins rather than the two previously detected by C19 (Figure 3.13) could result from the fractionation of the cells enabling distinction between similarly sized proteins which may appear as a single band in Western immunoblots of whole cell lysates. The presence of progesterone binding proteins which are also detected by C19 and may be knocked down by PR targeted siRNA supports the existence of PR proteins in the MDA-MB-231 cell line which is reported to be PR negative.

MDA-MB-231 cells were separated into nuclear and cytoplasmic fractions and again a range of proteins were detected in each fraction by Western immunoblotting using C19. Bands corresponding to the size of proteins previously analysed by siRNA and Western immunoblotting in MCF-7 cells were again detected. In cells fractionated by this method a doublet of proteins at around 35kDa was evident in the nuclear fraction (Figure 4.2a) which was previously observed in whole cell lysates (Section 3.3.6) and in the nuclear fraction in Figure 4.1a as a single band. As described in Section 1.1 and demonstrated in Section 3.3.6, PR undergoes extensive phosphorylation so it is possible that the extra band is a phosphorylated version of the 35kDa protein and this differential phosphorylation may due to the different sample preparation methods used. The protein detected just above the 42kDa marker in the cytoplasmic fraction (Figure 4.2a) and as the lower band in the cytosolic doublet in Figure 4.1a (red box) may correlate to the upper protein of the doublet knocked down by PR siRNA in MCF-7 cells in Section 3.3.6. Similarly to the pattern observed in Figure 3.9, expression of this 42kDa cytoplasmic protein is increased by β-oestradiol stimulation. The third protein detected

by C19 that was knocked down by PR siRNA in MCF-7 cells was approximately 50kDa and a protein of appropriate size (50-55kDa) is detected in both nuclear and cytoplasmic fractions of MDA-MB-231 cells, with cytosolic expression increasing in response to β-oestradiol (Figure 4.2a). In the ligand blot assay a progesterone binding protein of this molecular weight was detected in the nuclear and cytosolic fractions which was also detected in these fractions by C19. The final protein described as PR following siRNA knockdown in MCF-7 cells was approximately 80kD and protein of this molecular weight was detected at a very low level in the nuclear fraction of MDA-MB-231 cells following fractionation by both methods (Figure 4.1a and Figure 4.2a). The 60kDa protein which was previously described as PR-C (Condon et al., 2006) but demonstrated to be non-specific binding by PR siRNA knockdown in this thesis was present in the cytoplasmic fraction of MDA-MB-231 cells (Figure 4.2a).

Co-IP using the nuclear and cytoplasmic fractions of MDA-MB-231 cells with the C20 antibody, targeted to an internal PR epitope, followed by Western immunoblotting using the C-terminal PR antibody C19 allowed comparison of proteins detected by these two non N-terminal PR antibodies and also assessment of the potential presence of multimeric PR complexes. A protein at approximately 100kDa was detected in both nuclear and cytoplasmic fractions (Figure 4.3a) which had not been identified as PR by siRNA knockdown in MCF-7 cells. It is possible that this protein is a cell type specific isoform and it may therefore still be a PR protein since it is detected by both antibodies and a progesterone binding protein was detected at approximately this size in MDA-MB-231 cell cytosolic fractions (Figure 4.1b). Proteins at approximately 80kDa, 55kDa and 35kDa were also detected by C19 after immunoprecipitation with C20 ((Figure 4.3a). These proteins were detected in both nuclear and cytoplasmic fractions whereas they were detected specifically in the nuclear fraction by Western immunoblotting of the nuclear and cytoplasmic fractions before Co-IP (Figure 4.2a). The protein concentrations of the individual fractions allowed loading of 30µg of nuclear or cytoplasmic extract for SDS-PAGE, whereas 50µg was used for Co-IPs. Immunoprecipitation using C20 would lead to an increased concentration of specific proteins being loaded for SDS-PAGE allowing detection of proteins expressed at a much lower level than were detectable by initial Western immunoblotting of cell fractions. All of these proteins detected by both C19 and C20 correspond to proteins previously described in this thesis as PR following siRNA knockdown in MCF-7 cells

and the 55kDa and 35kDa proteins correspond to the molecular weight of progesterone binding proteins identified in MDA-MB-231 cells by ligand blot experiments.

High molecular weight protein bands (greater than 170kDa) were detected using C19 following Co-IP by C20 (Figure 4.3b) and potentially represent PR dimers which may contain truncated PR proteins detected by C19 in complex with other PR proteins detected by C20. This supports the possibility that PR proteins detected by these antibodies may be able to function similarly to wild-type PR which dimerise following ligand binding.

The nuclear protein PSF has previously been reported to interact with PR-A and PR-B (Tyson-Capper et al., 2009). Co-IP using a PSF antibody and MDA-MB-231 nuclear and cytoplasmic extracts demonstrated that the 80kDa and 35kDa PR proteins detected by C19 were capable of interacting with PSF in a similar manner to full length PR (Figure 4.4a). PSF is a nuclear protein (Shav-Tal and Zipori, 2002) and the interaction of these PR proteins with PSF was predominantly nuclear. PSF has been demonstrated to interact with the DBD of thyroid hormone receptors and retinoid X receptors (Shav-Tal and Zipori, 2002). These receptors are members of the same receptor superfamily as PR so it is possible that PR also interacts with PSF via the DBD, meaning that the 80kDa and 35kDa proteins reported herein to interact with PSF may require at least part of this domain. A 35kDa protein was previously demonstrated to be PR in MCF-7 cells in Section 3.3.6 of this thesis and a similarly sized protein is shown in Figure 3.1b to bind progesterone in MDA-MB-231 cells. Interestingly, a protein doublet at 35kDa was also detected by Western immunoblotting using the N-terminally targeted PR antibody NCL-PgR-AB following Co-IP of MCF-7 nuclear and cytoplasmic fractions using a PSF antibody (Figure 4.4b). The proteins detected in MCF-7 cells must therefore possess the normal PR-A/B N-terminal domain which would be predicted to be lacking from the proteins in MDA-MB-231 cells since NCL-PgR-AB fails to detect protein in this cell line. Therefore these results suggest that different 35kDa PR proteins may be expressed in the two cell lines.

This chapter has assessed the potential for interaction of low molecular weight PR proteins with PSF to investigate the potential for co-factor regulation. The choice of PSF as the co-regulator to be studied has been described in Section 4.1 and again in Section 4.3.4. Other co-factors such as the SRC family can also interact with PR, as described in Section 1.1.6, and the interaction of these proteins with low molecular weight PR is a potential area of study. However, the aim of this chapter was to make a preliminary assessment of several aspects of PR function; ligand binding, dimerisation, co-factor binding and DNA binding. Therefore only one potential co-factor of low molecular weight PR was studied.

DNA affinity precipitation assays have demonstrated that a nuclear protein of approximately 100kDa which is specifically detected in MDA-MB-231 cells, and is neither PR-A nor PR-B, is capable of binding to PRE (Figure 4.5a). A protein of this size was also demonstrated to be able to bind to progesterone in the ligand blot assay (Figure 4.1b) and was detected by both C19 and C20 antibodies as demonstrated by Co-IP experiments (Figure 4.3a). If the proteins identified in the three assays described are the same protein then a potentially functional protein may be expressed which must possess both the LBD and DBD. Based on the PR variant mRNAs detected by the PR gene walking assay in Section 3.3.1, PRΔ4 could generate a protein with these functional domains mostly intact. The previously reported PRΔp4$_{126}$ (Marshburn et al., 2005; Balleine et al., 1999b) with a deletion of 126bp from the start of exon 4 would generate a protein lacking the hinge region but which may still be able to function as a nuclear receptor. The predicted molecular weights of PRΔ4 and PRΔp4$_{126}$ resulting from alternative splicing of PR-B mRNA are 87kDa and 94kDa respectively. Since post-translational modifications mean that full length PR-B with a theoretical molecular weight of 99kDa migrates close to 120kDa in SDS-PAGE, it is possible that these two exon 4 alternatively spliced variants may generate proteins which are detected at 100kDa. However, if these proteins were expressed resulting from alternative splicing of PR-B mRNA then detection with N-terminal PR antibodies would be observed, suggesting that any proteins observed in MDA-MB-231 cells resulting from exon deleted mRNA must also have an altered N-terminal domain. Therefore whilst these results support the existence of functional truncated PR proteins in a reportedly PR negative breast cancer cell line, the identity of this protein remains undefined.

A 35kDa nuclear protein in MDA-MB-231 cells was demonstrated to bind PRE (Figure 4.5b) in addition to binding progesterone. This protein which is likely to be different to the 35kDa protein knocked down by PR siRNA in MCF-7 cells, is likely to be a PR since it is capable of binding progesterone and PRE and interacting with the PR co-factor PSF. Interaction of PSF with other nuclear receptors is mediated via the DBD (Shav-Tal and Zipori, 2002) however it is not known whether PR interacts with PSF in the same way, so the interaction of this protein with PSF may not require the DBD. Binding of this protein to PRE again suggests that the DBD must be present. However, if the protein was capable of dimerisation then it may be precipitated by PRE due to interaction with another DBD containing protein, such as the 100kDa protein described above, and the potential presence of PR dimers in MDA-MB-231 cells is demonstrated by Co-IP using C20. The 35kDa PR protein detected in MDA-MB-231 cells may therefore contain the C-terminal LBD and DD encoded by exons 4 to 6.

A cytoplasmic protein of 50-55kDa was demonstrated by DAPA in MDA-MB-231 cells to be able to interact with PRE (Figure 4.5c). Whilst the cytoplasmic location of this protein suggests that it is unlikely to usually function as a genomic PR, DNA binding ability does indicate that the DBD may be present, so this protein could potentially function through DNA binding if nuclear localisation could be induced following ligand binding. Expression of a protein of this size was induced by β-oestradiol stimulation of MDA-MB-231 cells (Figure 4.2a) and expression of a similar sized protein was knocked down by PR targeted siRNA in MCF-7 cells previously in this thesis (Section 3.3.6). A protein of this molecular weight was also demonstrated in MDA-MB-231 nuclear and cytosolic fractions to bind progesterone in the ligand blot assay. The progesterone binding and DNA binding ability of this protein suggest that both LBD and DBD are present. However, as described above, DNA binding may be mediated through dimerisation with other functional proteins. The functions described for the 55kDa PR protein, together with its detection specifically by non N-terminal PR antibodies in MDA-MB-231 cells, suggests that this protein may be an N-terminally truncated PR variant encoded by exons 2 to 8 or exons 4 to 8. Two truncated ER-α proteins, ERα46 and ERα36, have been identified resulting from the alternative splicing of non-coding leader exons from the 5′ UTR onto exon 2 resulting in use of promoters within exon 2 (Wang et al., 2005; Flouriot et al., 2000). ERα46 contains sequences encoded by the remainder of exons 2 to 8 (Flouriot et al., 2000), whereas ERα36 results

from alternative splicing of the 3′ end of exon 6 onto a novel exon 9 within the 3′ UTR (Wang et al., 2005). Alternative splicing of non-coding exons onto PR exon 4 has previously been reported (Hirata et al., 2002; Hirata et al., 2000) and if leader exons exist within the 5′ UTR of PR which are spliced onto exon 2 in a similar manner to that described for ERα then this could provide a mechanism of alternative splicing which could encode N-terminally truncated proteins which contain the exon 2 to 8 sequence. The exon 1 sequence encodes 56kDa of the full length PR-B protein predicted using ExPASy Compute pI/MW. Therefore the predicted molecular weight of a putative protein encoded by just exons 2 to 8 would be approximately 43kDa depending on the leader exon sequence and the translational initiation site used, and could potentially be the protein detected at 50-55kDa in MDA-MB-231 cells .

The 42kDa cytoplasmic protein detected by C19 in MDA-MB-231 cells, which was demonstrated to bind progesterone in the cytosolic fraction was unable to interact with PSF or PRE and may therefore function through non-genomic pathways. PR-M mRNA is predicted to encode a 39kDa protein that localises in the cytoplasm (Price et al., 2005; Saner et al., 2003) and this mRNA was detected in MDA-MB-231 cells in Section 3.3.1 of this thesis. As described above, PR-B has a theoretical molecular weight of 99kDa but migrates close to 120kDa in SDS-PAGE, so it is therefore possible that PR-M mRNA could encode a protein with a molecular weight greater than the predicted 39kDa, possibly this 42kDa cytoplasmic protein or the 55kDa protein described above.

A 60kDa protein detected by C19 has previously been described as the N-terminally truncated PR-C isoform, resulting from initiation of translation at Met595 (Condon et al., 2006; Wei et al., 1996). Detection of this protein in MCF-7 cells was demonstrated in Section 3.3.6 of this thesis to be the result of non-specific binding since expression was not reduced by siRNA knockdown of total PR expression. In MDA-MB-231 cells this protein was detected in the cytoplasmic fraction but was unable to bind progesterone, PSF or PRE in the assays described in this chapter. Therefore the results presented in this thesis support previous reports (Madsen et al., 2007) that the 60kDa protein detected by C19 is not a PR isoform.

In summary, siRNA knockdown of PR expression in MCF-7 cells reported earlier in this thesis has demonstrated that C19 detects PR proteins of approximately 80kDa, 50kDa and two proteins between 30-40kDa (Section 3.3.6). Progesterone binding proteins of these sizes were detected by a ligand blot assay in MDA-MB-231 cell fractions. Whilst results described in this chapter suggest that differences may exist between proteins detected by C19 in MCF-7 and MDA-MB-231 cells, the functional data presented above suggests that the MDA-MB-231 cells do contain PR proteins. The 35kDa nuclear progesterone binding protein detected in MDA-MB-231 cells was also able to interact with the nuclear co-factor PSF and with PRE. An 80kDa nuclear protein expressed at a low level in this cell line was able to bind progesterone and interact with PSF and so potentially function as a classical nuclear PR, although specific interaction of this protein with PRE was not observed. A third nuclear protein was detected specifically in MDA-MB-231 cells and was able to bind progesterone and interact with PRE. This protein of approximately 100kDa was a different size to either PR-A or PR-B. Exon deletion events characterised in Section 3.3.1 of this thesis could generate both the 80kDa and 100kDa potential PR proteins in MDA-MB-231 cells. However, since neither protein is detected by N-terminal antibodies in MDA-MB-231 cells, the normal N-terminal region must be truncated or altered in these proteins in addition to any internal exon deletions. Deletions within PR exon 1 were identified in MDA-MB-231 cells by the gene walking assay in Section 3.3.1 of this thesis. PR exon 1 is an extremely long exon (1637bp) and the PR gene walking assay used primers spanning exons and therefore only identified deletions towards the 3′ end of exon 1, in regions in which the primers could be paired with primers to downstream exons and generate PCR products of less than 1kbp. However exon 1 contains many putative 5′ and 3′ splice sites, suggesting that this extremely long exon may be alternatively spliced which could generate proteins lacking the N-terminal region such as those described in this chapter.

A cytoplasmic protein of around 50-55kDa was able to bind progesterone and interact with PRE, suggesting that it may be able to function as a nuclear PR if it is able to dimerise and translocate. Alternatively the presence of this protein in breast cancer cells, along with other low molecular weight progesterone binding proteins such as the cytoplasmic 42kDa protein, could indicate the presence of non-genomic PR isoforms. This chapter has focussed on the potential of PR proteins in MDA-MB-231 cells to function through the classical genomic PR signalling pathway in which PR binds

progesterone, dimerises and translocates to the nucleus before interacting with co-regulators and PREs within target genes. The presence of cytoplasmic proteins which are capable of binding progesterone suggests that PR proteins which function through non-genomic pathways may also exist in the reportedly PR negative MDA-MB-231 breast cancer cell line. Future studies may focus on the potential for non-genomic signalling by low molecular weight PR proteins by assessing interaction with known PR non-genomic binding partners and activation of rapid progesterone response pathways in cells expressing these low molecular weight proteins.

5 Characterisation of PR alternative splicing events in breast tumours

5.1 Introduction

Current breast cancer PR screening uses N-terminally targeted antibodies which detect, but do not distinguish between, both PR-A and PR-B (Hanley et al., 2009). Alternative splicing of PR pre-mRNA may encode proteins which are undetectable by current breast cancer screening using these N-terminally targeted antibodies, or which are functionally distinct but indistinguishable from full length PR. An RT-PCR based gene walking assay in Section 3.3.1 of this thesis has demonstrated that alternatively spliced PR mRNA is expressed in the reportedly PR negative MDA-MB-231 breast cancer cell line. Furthermore PR proteins are detected in this cell line using a C-terminally targeted antibody, which are not detected using N-terminally targeted antibodies (Section 3.3.2). The proteins present in MDA-MB-231 cells which are detected by this C-terminal antibody, and may result from alternative splicing of PR pre-mRNA, have been demonstrated to have the potential to function as PR proteins (Chapter 4).

To validate the results of the cell line studies presented in Chapters 3 and 4, primers from the PR gene walking assay were used to identify PR ASEs in a small cohort of frozen breast tumour tissue and predict the potential functionality of the proteins that may be encoded. RNA was extracted from fifteen frozen breast tumour tissue samples obtained from a Newcastle University/Gateshead Queen Elizabeth Hospital Breast Cancer Biomarkers tissue bank. The tissues used had been previously characterised for ER and PR status at the time of sample collection using N-terminal antibodies for both hormone receptors. RT-PCR was performed using the PR gene walking assay and DNA sequencing analysis was used to identify PR ASEs in the breast tumour tissue. Computational analysis was then performed to predict the functionality of the expected protein product of each alternatively spliced PR mRNA.

The breast cancer cell line studies detailed in Section 3.3.6 demonstrated that the C-terminal PR antibody C19, which detects PR proteins by Western immunoblotting (Section 3.3.6.) is unsuitable for immunofluorescent analysis and

immunohistochemistry. The potential expression of N-terminally truncated PR variants in breast cancer, resulting from alternative splicing (Saner et al., 2003; Hirata et al., 2002; Hirata et al., 2000) or alternative promoter usage (Wei and Miner, 1994) suggests that non N-terminal antibodies should also be considered for use in breast cancer PR screening. The lack of specificity of commercially available non N-terminal PR antibodies has been previously reported (Samalecos and Gellersen, 2008; Madsen et al., 2007) and makes them unsuitable for use in breast cancer screening, necessitating the development of novel non N-terminal PR antibodies. Therefore, the PR antibodies A3 and A4 were developed, targeting an internal and a C-terminal epitope respectively. The specificity of the antibodies was assessed by Western immunoblotting, Co-IP and siRNA knockdown of PR expression. The PR status of the breast tumours used for the PR gene walking assay was then reassessed using the novel antibodies. Further information on the HER2 status, tumour grade, lymph node involvement and Nottingham prognostic index was also available from original histology reports for the tumours analysed.

5.2 Aims

The aims of this chapter were to:

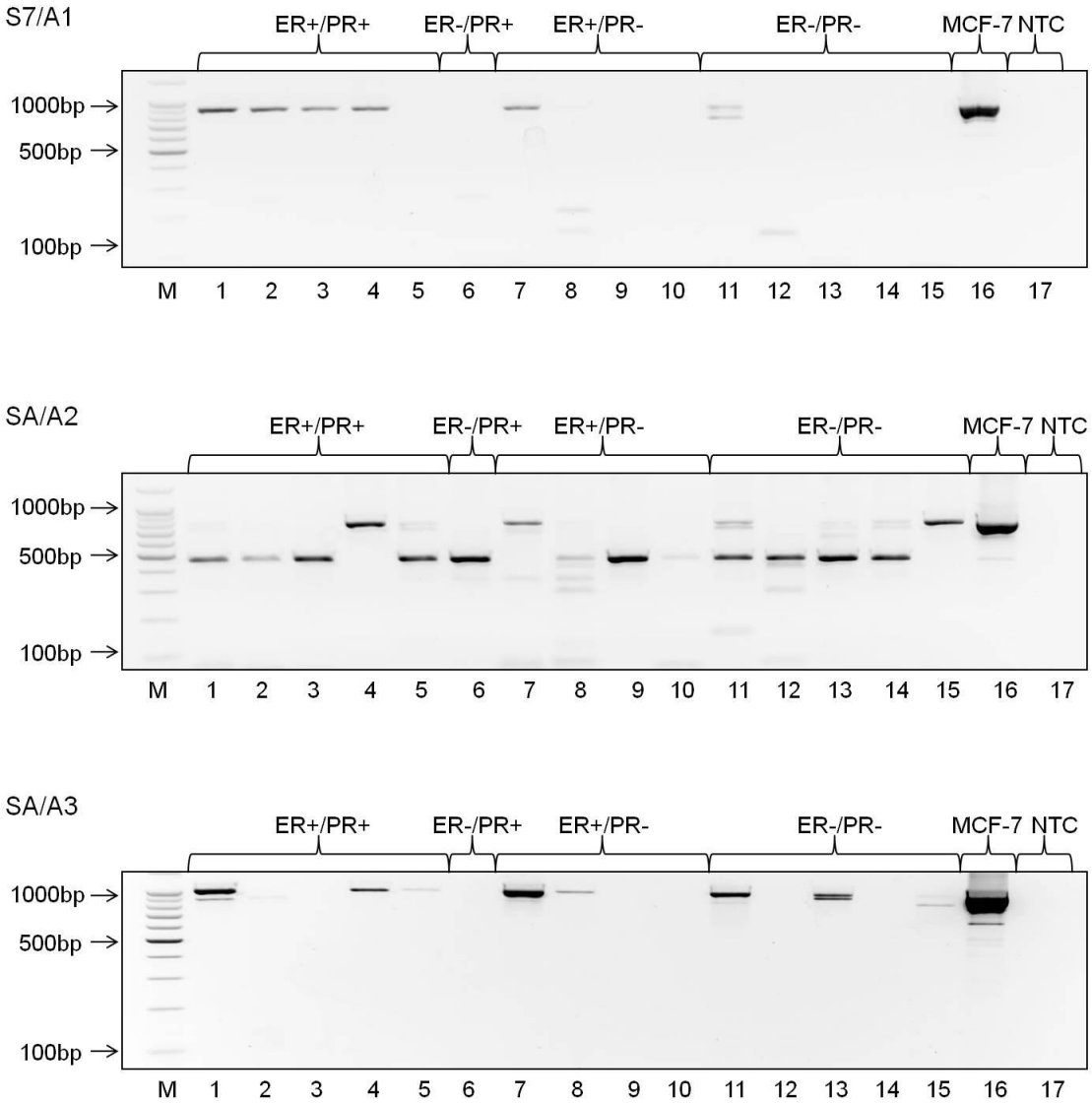
1. Analyse PR mRNA expression in a small cohort frozen breast tumour samples of previously characterised PR status and identify PR ASEs occurring in breast tumours.
2. Purify novel PR antibodies PRA3 and PRA4
3. Optimise PRA3 and PRA4 for use in:
    - Western immunoblotting
    - Immunohistochemical staining of FFPE breast tumour tissue
4. Validate the specificity of PRA3 and PRA4 by Western immunoblotting following Co-IP or siRNA knockdown of PR expression in MCF-7 cells.
5. Use PRA3 and PRA4 to characterise the PR status of FFPE breast tumours, and compare this to the reported PR status determined using N-terminal PR antibodies.

5.3 Results

5.3.1 Characterisation of PR mRNA expression in breast tumour tissue

Alternative splicing of PR has been shown in this thesis to generate mRNA which may encode truncated protein isoforms which are functionally distinct from wild-type PR-A and B in breast cancer cell lines. The presence of these isoforms in breast tumour tissue may have implications in understanding the role of PR in breast cancer development and in the use of PR as a predictive marker for assigning breast cancer therapies.

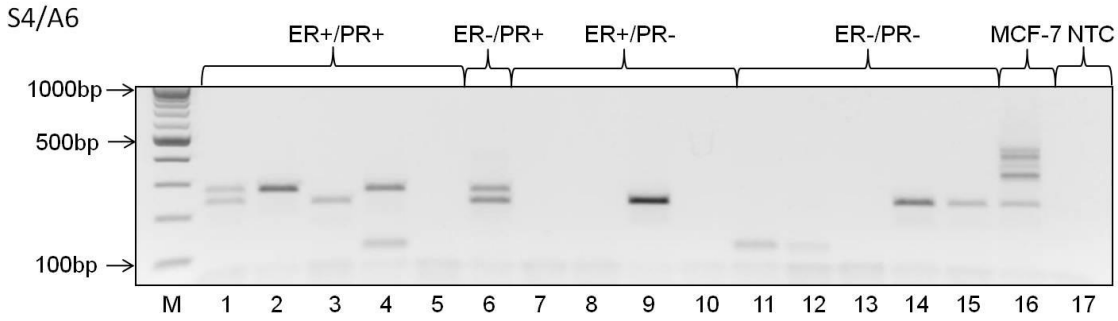
The presence of alternatively spliced PR mRNA was examined in fifteen frozen breast tumour tissue samples. RNA was extracted from five ER+/PR+, one ER-/PR+, four ER+/PR- and five ER-/PR- breast tumours and used for RT-PCR using primer combinations from the PR gene walking assay which had previously been demonstrated in cell line studies to detect ASEs (Section 3.3). Initially, RT-PCR was performed using primers spanning long regions of the PR gene; sense primers to exon 2 (S7) or exon 1 (SA) and antisense primers to exon 8 (A1), exon 7 (A2) or exon 6 (A3). The full length PCR product was seen in the positive control MCF-7 lane (Figure 5.1, lane 16). However, whilst full length PCR products were detected in some samples, primers spanning these long regions (PCR products >750bp) failed to consistently detect full length product, potentially demonstrating degradation of the RNA in the frozen breast tumour samples (Figure 5.1, lanes 1-15).



<u>Figure 5.1:</u> RNA degradation in frozen breast tumour samples may prevent detection of long PCR products.

RT-PCR was performed for 45 cycles on RNA from frozen breast tumour samples using sense primers to exon 2 (S7) or exon 1 (SA) and antisense primers to exon 8 (A1), exon 7 (A2) or exon 6 (A3). 15 breast tumour samples were analysed with the following previously characterised hormone receptor status; lanes 1-5= ER+/PR+, lane 6= ER-/PR+, lanes 7-10= ER+/PR-, lanes 11-15= ER-/PR-. MCF-7 cDNA was used in a positive control reaction (lane 16) and a no-template control reaction was performed for each primer pair (NTC, lane 17). M=100bp DNA ladder.

Since primer pairs spanning long regions of the PR gene had failed to consistently detect PR mRNA, primer sets were used to focus specifically on the regions which were previously demonstrated by the PR gene walking assay to undergo ASEs in breast cancer cell lines. Firstly, primers spanning exon 1 (S4) to exon 3 (A6) were used to examine ASEs occurring within exons 1 and 2 (Figure 5.2). No full length PCR product was detected in any of the tumour samples. Two differently sized PCR products were detected between approximately 250 and 300bp suggesting that ASEs may be occurring in this region of PR in breast tumour samples. The larger product (~300bp) was detected specifically in PR+ tumours (Figure 5.2, lanes 1, 2, 4, 6). In two of these tumours the smaller product (~250bp) was also detected (Figure 5.2, lane 1 and 6). The larger product was not detected in PR- tumours but the smaller band was detected in some of these samples (Figure 5.2 lanes 9, 14 and 15). A smaller PCR product of approximately 150bp was also detected using these primers in one PR+ and one PR- tumour, potentially representing mRNA with large deletions within exon 1 and exon 2 (Figure 5.2, lanes 4 and 11).



<u>Figure 5.2:</u> Different patterns of exon 1 and 2 alternative splicing exist between PR+ and PR- breast tumours.

RT-PCR performed for 45 cycles using RNA from frozen breast tumour samples and primers spanning exon 1 (S4) to exon 3 (A6). PCR products were compared between 5 ER+/PR+ tumours (lanes 1-5), 1 ER-/PR+ tumour (lane 6), 4 ER+/PR- tumour (lanes 7-10), 5 ER-/PR- tumours (lanes 11-15) and a positive control (MCF-7 cells, lane 16). A no template control reaction (NTC, lane 17) was also included. M=100bp DNA ladder.

ASEs involving exon 3 were then examined using primers spanning exon 2 (S6) to exon 4 (A5). The full length PCR product was detected in the positive control MCF-7 cells (Figure 5.3, lane 16) and in all PR+ tumours and the majority of PR- tumours (7 out of 9 tumours examined, Figure 5.3). No truncated PCR product was detected, indicating that exon 3 does not appear to be alternatively spliced in breast tumours (Figure 5.3).

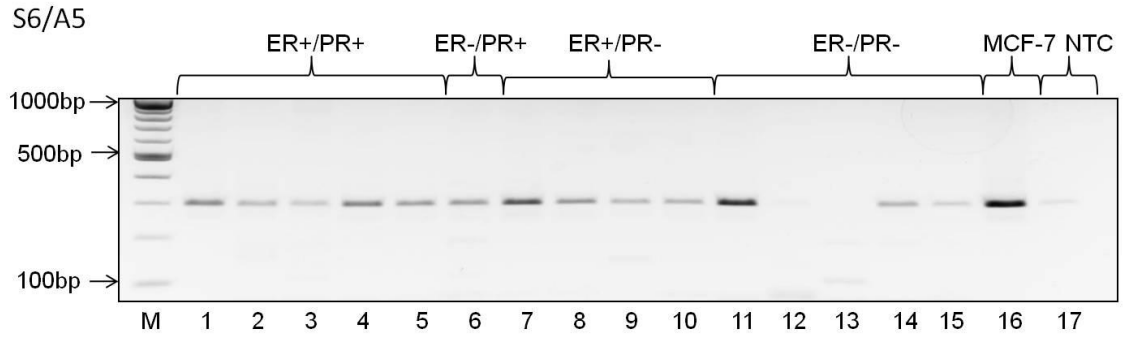


Figure 5.3: PR exon 3 does not undergo alternative splicing in breast tumour tissue.

RT-PCR performed for 45 cycles using RNA from frozen breast tumour samples and primers spanning exon 2 (S6) to exon 4 (A5). PCR products were compared between 5 ER+/PR+ tumours (lanes 1-5), 1 ER-/PR+ tumour (lane 6), 4 ER+/PR- tumour (lanes 7-10), 5 ER-/PR- tumours (lanes 11-15) and a positive control (MCF-7 cells, lane 16). A no template control reaction (NTC, lane 17) was also included. M=100bp DNA ladder.

In breast cancer cell lines the PR gene walking assay detected extensive ASEs involving both whole exon (Δ4) and partial exon deletion (Δp4) of exon 4 (Section 3.3.1.4). Primers spanning exon 3 (S7) to exon 5 (A4) were used to detect Δ4 or Δp4 mRNA. Full length PCR product was detected in the positive control MCF-7 cells (Figure 5.4a, lane 16), and also in many PR+ and PR- tumours (Figure 5.4a). In most tumours in which PR mRNA was detected, a truncated PCR product, potentially representing Δ4 mRNA was also detected (Figure 5.4a). In some tumours, both PR+ and PR-, only the truncated Δ4 PCR product was detected (Figure 5.4a, lanes 3, 10, 12, 14). The truncated PCR product was detected at higher levels in breast tumour samples compared to the positive control, suggesting that more Δ4 mRNA may be present in these tumours than in the MCF-7 breast cancer cell line (Figure 5.4a). The presence of a PCR product, potentially representing Δ4 mRNA was confirmed by two further PCR experiments; the first using primers spanning exon 1 (SA) to exon 4 (A5) in which only a full length PCR product was detected in breast tumours and the positive control (Figure 5.4b, lanes 1-6)), and the second using primers spanning exon 1 (SA) to exon 5 (A4) in which both a full length and truncated PCR product were detected in both breast tumours and MCF-7

cells (Figure 5.4b, lanes 7-12). Again, the Δ4 PCR product was more prevalent in breast tumours than MCF-7 cells, and was detected as the main product in one of the tumours examined (Figure 5.4b, lane 8). Primer sets S7/A4 and SA/A4 both detected, at a lower level, another PCR product with a smaller deletion, potentially indicating the presence of mRNA with a deletion of part of exon 4, such as Δp4 (Figure 5.4a and b).

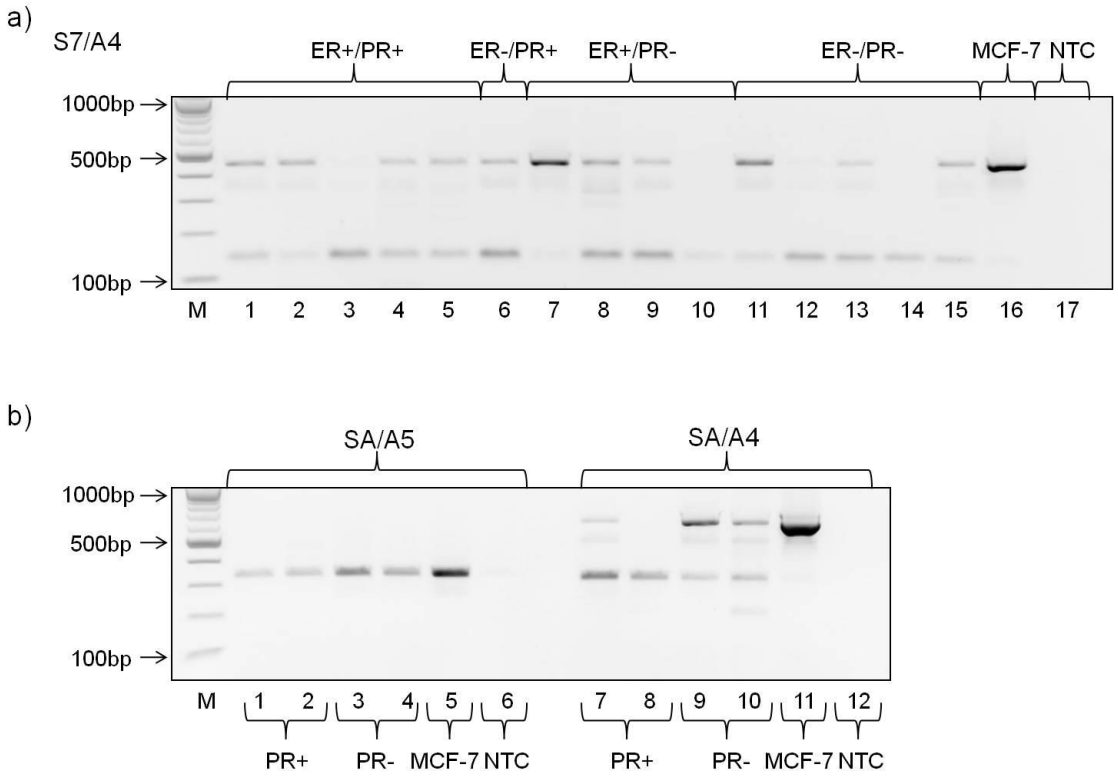


Figure 5.4: Exon 4 deleted mRNA is detected in breast tumours at a higher level than MCF-7 breast cancer cells.

RT-PCR performed for 45 cycles using RNA from frozen breast tumour samples. **a)** PCR using primers spanning exon 3 (S7) to exon 5 (A4). PCR products were compared between 5 ER+/PR+ tumours (lanes 1-5), 1 ER-/PR+ tumour (lane 6), 4 ER+/PR- tumour (lanes 7-10), 5 ER-/PR- tumours (lanes 11-15) and a positive control (MCF-7 cells, lane 16). A no template control reaction (NTC, lane 17) was also included. M=100bp DNA ladder. **b)** PCR using primers spanning exon 1 (SA) to exon 4 (A5) (lanes 1-6); and primers spanning exon 1 (SA) to exon 5 (A4) (lanes 7-12. PCR products were compared between 2 PR+ tumours (lanes 1,2 and 7, 8), 2 PR- tumours (lanes 3, 4 and 9, 10) and a positive control (MCF-7 cells, lanes 5 and 11). No template control reactions were included for each primer set (NTC, lanes 6 and 12). M=100bp DNA ladder.

Exon 6 deleted mRNA (Δ6) was not detected by the gene walking assay in MDA-MB-231 cells, but has previously been reported (Balleine et al., 1999b; Richer et al., 1998). RT-PCR was performed using the primer set S9/A2, spanning exon 5 to exon 7, to examine the presence of Δ6 mRNA in the breast tumour samples. The full length PCR product was detected in the positive control (Figure 5.5, lane 16) and in all PR+ tumours (Figure 5.5, lanes 1-6). A truncated (Δ6) PCR product was also detected in half of the PR+ tumours (Figure 5.5, lanes 4-6). Full length PCR product was detected in some PR- tumours (Figure 5.5, lanes 7, 8, 11, 12, 13, 12 and 15), along with Δ6 PCR product in lanes 7, 8, 12 and 15. Δ6 was the only PCR product detected in one PR- tumour (Figure 5.5, lane 9) and another PCR product, potentially representing an mRNA with a deletion of part of exon 6, was detected in addition to full length and Δ6 in one PR- tumour (Figure 5.5, lane 15).

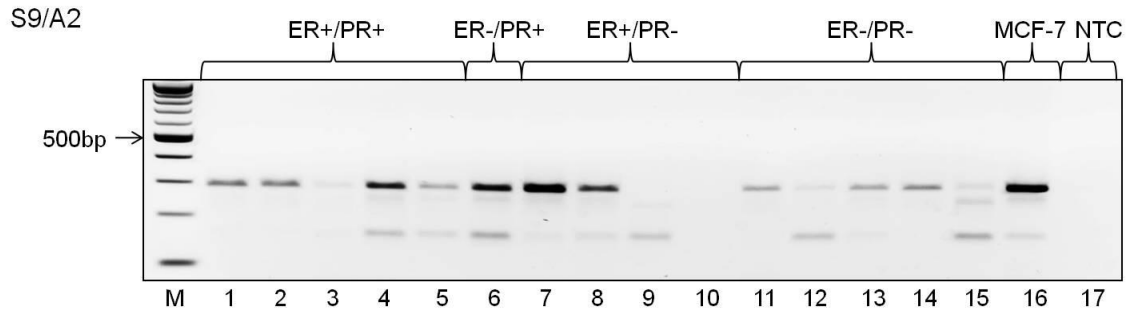


Figure 5.5: Exon 6 deleted mRNA was detected in breast tumours.

RT-PCR performed for 45 cycles using RNA from frozen breast tumour samples and primers spanning exon 5 (S9) to exon 7 (A2). PCR products were compared between 5 ER+/PR+ tumours (lanes 1-5), 1 ER-/PR+ tumour (lane 6), 4 ER+/PR- tumour (lanes 7-10), 5 ER-/PR- tumours (lanes 11-15) and a positive control (MCF-7 cells, lane 16). A no template control reaction (NTC, lane 17) was also included. M=100bp DNA ladder.

PR-M is a previously reported potential PR isoform which is the result of retention of an intronic sequence between exons 3 and 4. PR-M mRNA would encode a protein lacking the normal N-terminal domain and DBD, with an N-terminal signal sequence of hydrophobic amino acids (Saner et al., 2003). PR-M mRNA was detected in the reportedly PR- MDA-MB-231 breast cancer cell line using a sense primer to the intronic exon M (M) and an antisense primer to exon 4 (A) (Section 3.3.1.3). Sense primer M was used in combination with an antisense primer directed to exon 5 (A4) to examine the presence of PR-M mRNA in breast tumour samples. A PCR product representing PR-M mRNA was detected infrequently in both PR+ and PR- tumours (Figure 5.6a). PCR was repeated to confirm the presence of PR-M mRNA using the tumour samples in which PR-M mRNA had previously been detected (Figure 5.6b).

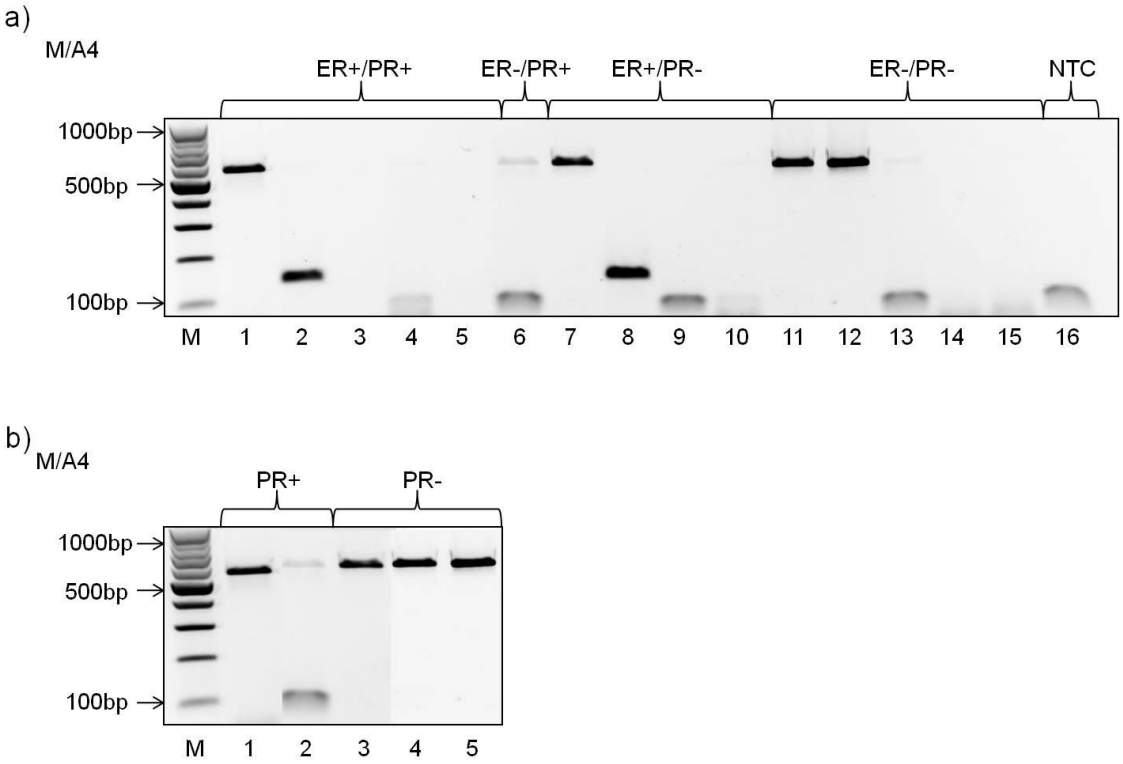


Figure 5.6: PR-M mRNA was detected in a small number of breast tumours.

RT-PCR performed for 45 cycles using RNA extracted from frozen breast tumour samples and primers spanning intronic exon M (sense, M) to exon 5 (antisense, A4). **a)** PCR products were compared between 5 ER+/PR+ tumours (lanes 1-5), 1 ER-/PR+ tumour (lane 6), 4 ER+/PR- tumour (lanes 7-10) and 5 ER-/PR- tumours (lanes 11-15). A no template control reaction (NTC, lane 16) was also included. M=100bp DNA ladder. **b)** PCR was repeated using previously PR-M positive samples. PR+ tumours (lanes 1 and 2), PR- tumours lanes 3, 4, and 5). M=100bp DNA ladder.

5.3.2 Identifying potential PR proteins resulting from ASEs in breast tumours

Truncated PCR products detected in breast tumour samples were cloned and sequenced to validate the regions deleted in each mRNA. Figure 5.7 is a diagrammatic representation of the potential protein that would be generated by each of the truncated mRNAs which were identified by sequencing. Figure 5.7a shows the eight exon structure of the PR gene and the full length protein which is encoded, identifying the main functional domains associated with the role of PR-B as a ligand activated nuclear transcription factor. Deletions/retentions within the PR mRNA are indicated by the dashed lines in Figure 5.7a. A deletion of a large region at the end of exon 1, along with the start of exon 2 was identified by sequencing of a PCR product generated in a PR- tumour using primers spanning exon 1 to exon 3 (Figure 5.2, lane 9). A similar deletion was also seen in several other tumours, both PR+ and PR- (Figure 5.2). This deletion had previously been identified in Section 3.3.1 of this in MDA-MB-231 breast cancer cells and the protein product would lack all of the internal and C-terminal functional domains so is unlikely to have any function (Figure 5.7b, Δp1,p2). A deletion of part of exon 2, the whole of exon 3 and part of exon 4 was identified by sequencing of a PCR product generated using primers spanning exon 1 to exon 7 in a PR+ tumour (Figure 5.1, lane 3). This deletion (Δp2,3,p4) would lead to truncation of the protein in the region encoded by exon 4 so would lack most PR functional domains (Figure 5.7).

Two deletions were sequenced from breast tumour samples which would not change the translational reading frame; a deletion of the whole of the cassette exon 4 (Δ4) and of the first 126 base pairs of exon 4 ($\Delta p4_{126}$). Δ4 was generated using primers spanning exon 1 to exon 5 in a PR+ tumour (Figure 5.4b, lane 7). This deletion was also detected in MDA-MB-231 cells and in a number of other PR+ and PR- tumours and using primers spanning exon 1 to exon 6 and exon 3 to exon 5. The Δ4 protein would lack the hinge region and part of the LBD encoded by this exon but possess normal N-terminal and C-terminal functional domains (Figure 5.7c). $\Delta p4_{126}$ was sequenced from a PCR product detected by primers spanning exon 3 to exon 4 in a PR- tumour (Figure 5.4a, lane 8) and by primers spanning exon 3 to exon 8 and exon 1 to exon 7 in both PR+ and PR- tumours. $\Delta p4_{126}$ differed from the partial exon 4 deletion (Δ4) mRNA identified in MDA-MB-231 cells and would encode a protein lacking most of the hinge and NLS but

would possess completely undisrupted C-terminal domains along with normal N-terminal functional domains (Figure 5.7c).

Deletion of the entire sequences of exons 5 and 6 (Δ5,6) was sequenced from a PCR product generated by primers spanning exon 1 to exon 7 in a PR- tumour and would encode a protein lacking a functional LBD, but possessing the DBD, hinge region, NLS and N-terminal functional domains (Figure 5.7d). A deletion of a 52bp section at the start of exon 6 was detected using primers spanning exon 3 to exon 8 in a PR- tumour. This deletion (Δp6.2) differs from the deletion in exon 6 which was observed in MDA-MB-231 cells (termed Δp6, Chapter 3). Δp6.2 would cause a shift in reading frame and truncation within the LBD encoded by exon 6 (Figure 5.7d).

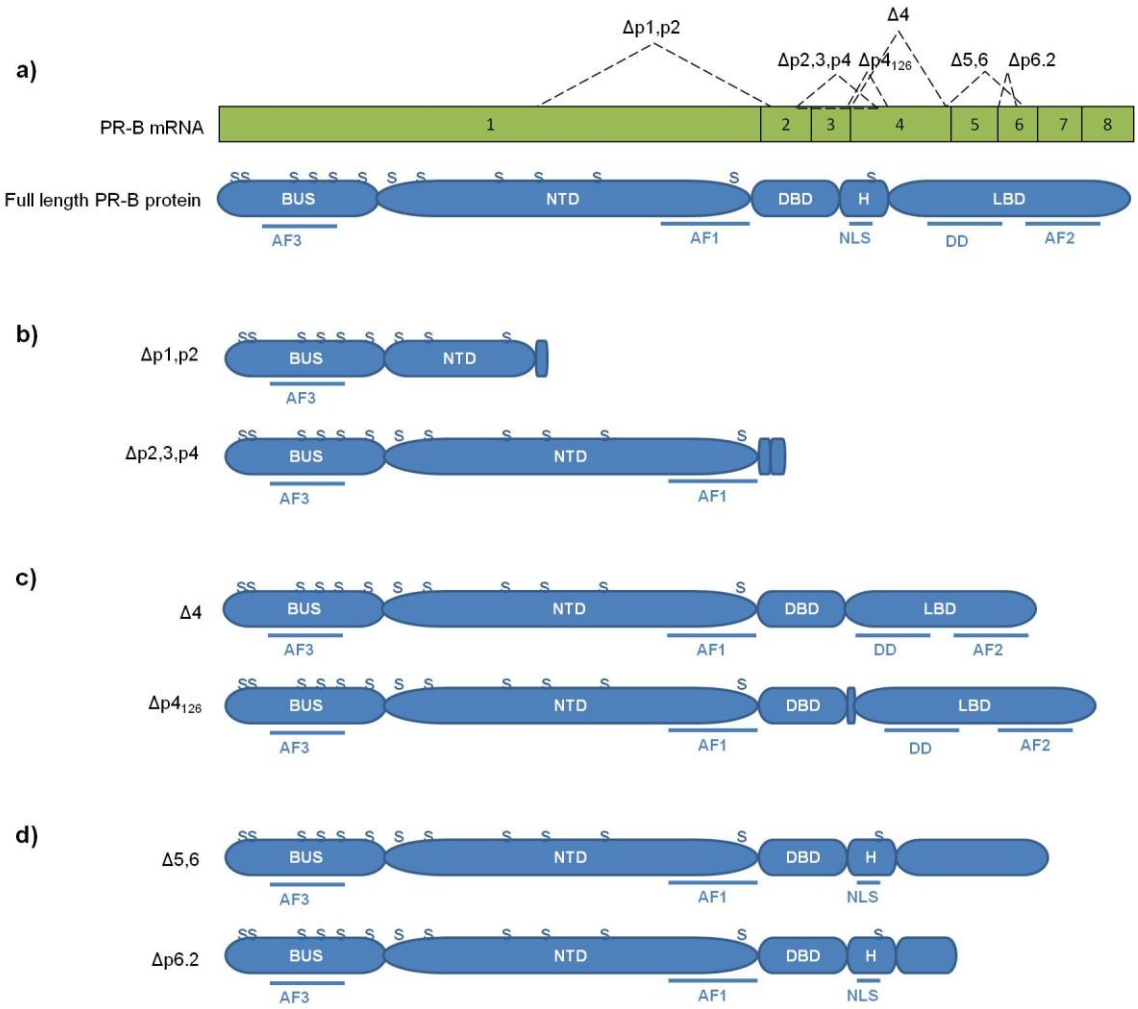


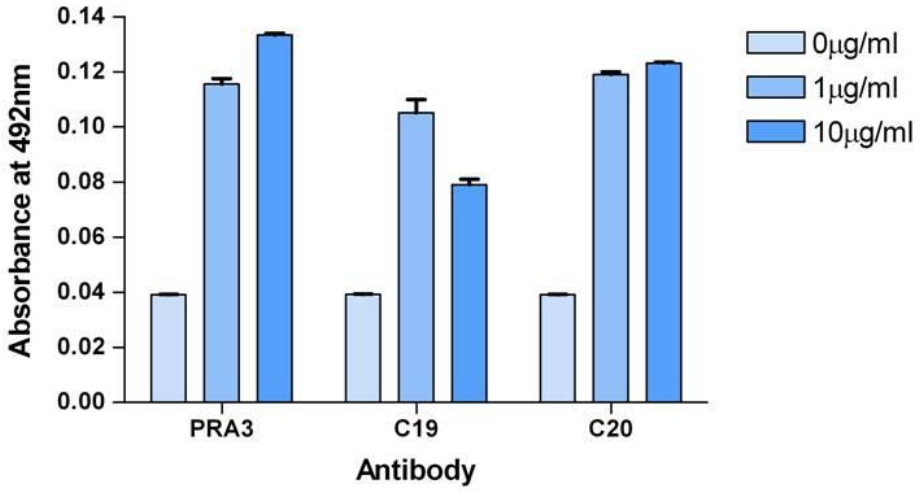
Figure 5.7: Structures of potential truncated PR proteins in breast tumour tissue.

RT-PCR products generated using primers from the PR gene walking assay and cDNA from breast tumour samples were cloned and sequenced. DNA sequences were compared to the published PR gene sequence (NCBI Accession: NM_000926.4) using the ALIGN database to identify the regions deleted from the mRNA represented by each PCR product. The protein sequence generated by each truncated mRNA was determined using the ExPASy Translate Tool. **a)** Eight exon PR gene sequence with the regions deleted from each sequenced mRNA indicated by dashed lines. The functional domains of the full length protein encoded by each exon are shown. **b)** Deletions of parts of multiple exons in the region exons 1 to 4 cause truncation of the protein. **c)** Deletions that do not cause a change in reading frame encode proteins lacking specific functional domains. **d)** A deletion at the start of exon 6 causes a change in reading frame and truncation of the LBD.

5.3.3 Affinity purification of novel PR antibodies

MAPs consisting of individual peptides targeted to different non N-terminal PR epitopes were designed for PR specificity, to differ from the epitopes targeted by commercial antibodies and for maximum immunogenicity. MAP A3 targeted to an internal epitope within the DBD, encoded by exons 2 and 3, and MAP A4 targeted to a C-terminal epitope encoded by exon 7, were used to inoculate a rabbit and a sheep respectively and produce novel PR antibodies. Serum extracted from the animals thirteen weeks post-immunisation was used to affinity purify the novel PR antibodies which will be referred to as PRA3 and PRA4 in reference to the MAP used to immunise the animal. PRA3 was purified at a concentration of 1mg/ml in 20ml 1xPBS, whereas a much lower quantity of PRA4 was successfully purified, only 117µg/ml in 6ml 1xPBS.

The presence of purified PRA3 antibody was assessed by adsorbing antibody onto an ELISA plate and comparing the levels present to two commercial non N-terminal PR antibodies, C19 and C20. The amount of antibody adsorbed was detected using an anti-rabbit IgG secondary antibody and OPD, and determined by reading the absorbance at 492nm. All three antibodies adsorbed at similar efficiencies, confirming successful PRA3 antibody purification, and no major differences were seen in the amount adsorbed when 1µg/ml and 10µg/ml of antibody were added to the plate (Figure 5.8).



<u>Figure 5.8:</u> PRA3 can be detected, bound to ELISA plates, at a similar level to commercial PR antibodies by an anti-rabbit IgG secondary antibody.

PRA3, C19 and C20 were used to coat ELISA plates at concentrations of 1µg/ml and 10µg/ml. No antibody controls (0µg/ml) were included along with each antibody. Antibody bound to the plate was detected using an HRP-conjugated anti-rabbit IgG secondary antibody and developed using OPD. The average absorbance at 492nm from two experiments is plotted for each antibody concentration.

5.3.4 Optimisation of novel PR antibodies for Western immunoblotting

Following affinity purification and quantification of novel PR antibodies, initial Western blotting experiments were performed using MCF-7 and MDA-MB-231 whole cell lysates to determine the optimal conditions for use. A range of dilutions of PRA3 (1:500-1:2000) were used for Western immunoblotting to optimise conditions (Figure 5.9). Best detection of low molecular weight proteins was seen at the highest concentration, 1:500, with minimal background. This dilution was selected for future Western immunoblotting experiments. These initial experiments demonstrated that PRA3 could detect high molecular weight proteins, of around 100kDa (PR-A/PR-B) and lower molecular weight proteins (Figure 5.9). The identity of these proteins requires further investigation.

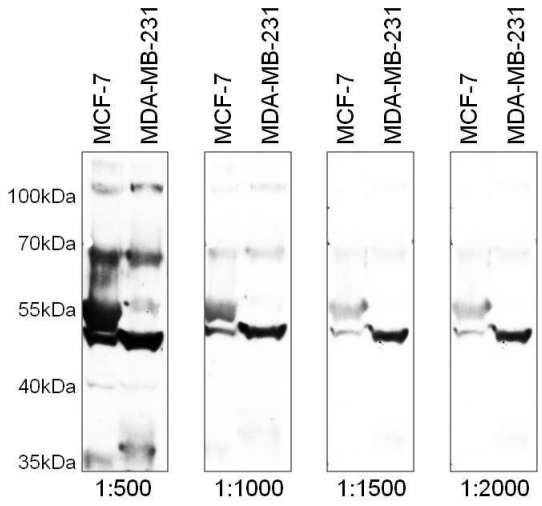


Figure 5.9: Optimisation of PRA3 for use in Western immunoblotting.

Western immunoblotting of MCF-7 and MDA-MB-231 whole cell lysates using PRA3. A range of concentrations of PRA3 were used; 1:2000 - 1:500 in 1xPBS + 1% non-fat milk powder.

Similar Western immunoblotting experiments were performed to optimise the conditions for use of PRA4. A range of concentrations of PRA4 were used, with two concentrations of anti-sheep IgG secondary antibody. None of the concentrations of PRA4 detected any protein in either MCF-7 or MDA-MB-231 cells (Figure 5.10). Both concentrations of secondary antibody produced high levels of background; therefore further optimisation of the secondary antibody conditions was necessary.

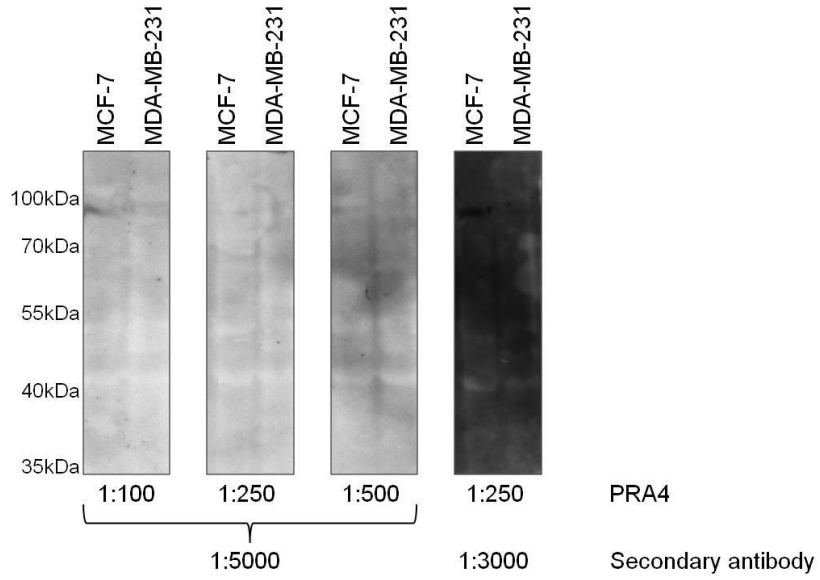


Figure 5.10: Optimisation of PRA4 for Western immunoblotting.

Western immunoblotting of MCF-7 and MDA-MB-231 whole cell lysates using PRA4. A range of concentrations of PRA4 1:500-1:100 in 1xPBS + 1% non-fat milk powder. Two concentrations of anti-sheep IgG secondary antibody were used, 1:5000 and 1:3000 in 1xPBS.

Different concentrations of anti-sheep IgG secondary antibody were used for Western immunoblotting of MCF-7 whole cell lysates using PRA4 at a dilution of 1:100. The best conditions for use of PRA4 were with the secondary antibody diluted 1:5000 adding Tween20 to the secondary antibody to reduce background. PRA4 detects a range of low molecular weight proteins under these conditions (Figure 5.11).

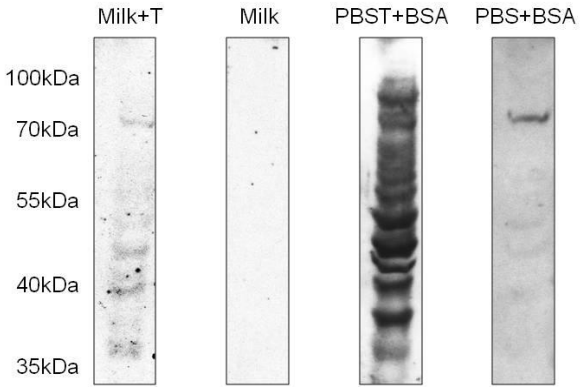


Figure 5.11: Optimisation of rabbit anti-sheep IgG secondary antibody for use with PRA4 for Western immunoblotting.

Western immunoblotting of MCF-7 whole cell lysates using PRA4 (1:100 in 1x PBS+1% non-fat milk powder) and anti-sheep IgG secondary antibody under different conditions. Milk= 1x PBS + 1% non-fat milk powder. Milk+T= 1x PBS + 1% non-fat milk powder + 0.1% Tween20. PBST= 1xPBS + 0.1% Tween20. BSA= 1% bovine serum albumin.

5.3.5 PRA3 detects similar proteins to commercial PR antibodies C20 and C19

Whilst also detecting non-specific proteins, the commercial C-terminal PR antibody C19 has been demonstrated to detect some PR proteins. Therefore, to validate the specificity of the novel PR antibodies PRA3 and PRA4, they were used as detection antibodies for Western immunoblotting following Co-IP of MCF-7 whole cell lysates using C19 and C20. PRA3 and PRA4 were also used as the capture antibodies for Co-IP of MCF-7 whole cell lysates with detection using the other novel antibody.

Western immunoblotting using PRA3 of co-immunoprecipitated proteins demonstrates that PRA3 is capable of detecting similar proteins to both of the commercially available antibodies studied, suggesting that these may be PR proteins. Proteins greater than 100kDa and around 35-40kDa that were immunoprecipitated specifically by C20 and

C19, and not the IgG control, were detected by Western immunoblotting with PRA3 (Figure 5.12a). However, using PRA4 as the capture antibody, PRA3 did not detect any immunoprecipitated proteins (Figure 5.12a). Western immunoblotting of proteins immunoprecipitated using C20, C19 or PRA3 using PRA4 did not detect any proteins (Figure 5.12b), demonstrating that PRA4 may not detect any PR proteins.

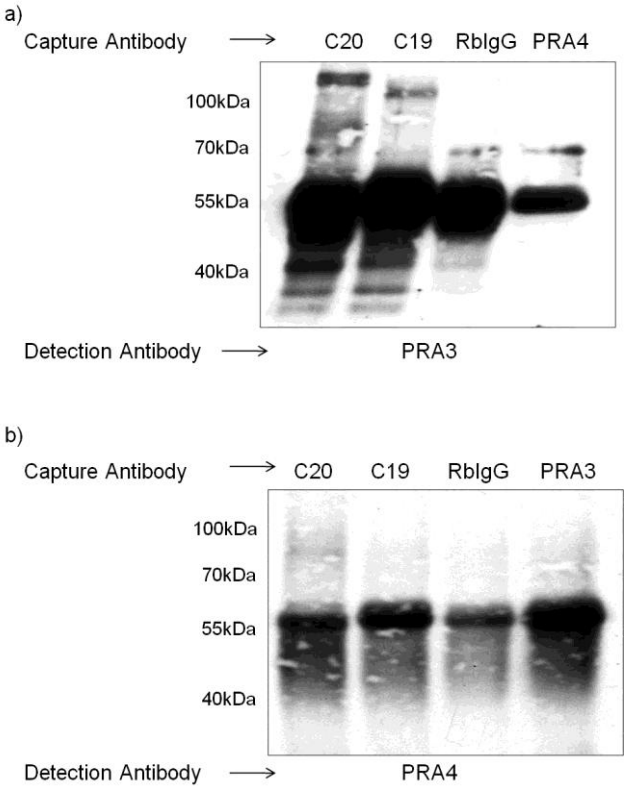


Figure 5.12: PRA3 detects similar proteins to commercial PR antibodies but PRA4 fails to detect any proteins following co-immunoprecipitation.

MCF-7 whole cell lysates were used for co-immunoprecipitation using C20, C19, PRA3 and PRA4. A Rabbit IgG control was also used. **a)** Western immunoblotting using PRA3 of proteins immunoprecipitated by C20, C19 a rabbit IgG control (RbIgG) or PRA4. **b)** Western immunoblotting using PRA4 of proteins immunoprecipitated by C20, C19, RbIgG or PRA3.

5.3.6 Validation of PRA3 and PRA4 by siRNA knockdown of PR expression

The novel PR antibody PRA3 appears to behave as a specific PR antibody in Co-IP and Western immunoblotting. The second novel PR antibody, PRA4, fails to detect any protein following immunoprecipitation using PR antibodies. To validate the specificity of these antibodies for Western immunoblotting, siRNA knockdown of total PR expression was carried out to determine whether a reduction in the level of proteins detected by the novel antibodies could be observed. Using PRA3 for Western immunoblotting following siRNA knockdown of PR expression, a slight reduction in the levels of protein detected at 100kDa and 40kDa can be seen suggesting that this novel non N-terminal PR antibody is detecting truncated PR proteins (Figure 5.15a, lanes 1 and 2). No knockdown of any of the proteins detected by PRA4 was observed (Figure 5.15b), suggesting that this antibody may be detecting non-specific proteins by Western immunoblotting. To confirm that the siRNA knockdown of PR had been successful, lysates were used for Western immunoblotting using NCL-PgR-B, the specificity of which was demonstrated in Section 3.3.6 of this thesis. Knockdown of PR-B was observed in cells treated with PR siRNA (Figure 5.15c, lanes 1 and 2). Knockdown of GAPDH with the positive transfection control GAPDH siRNA was observed in each experiment (Figure 5.15a, b and c, lane 3 lower panels).

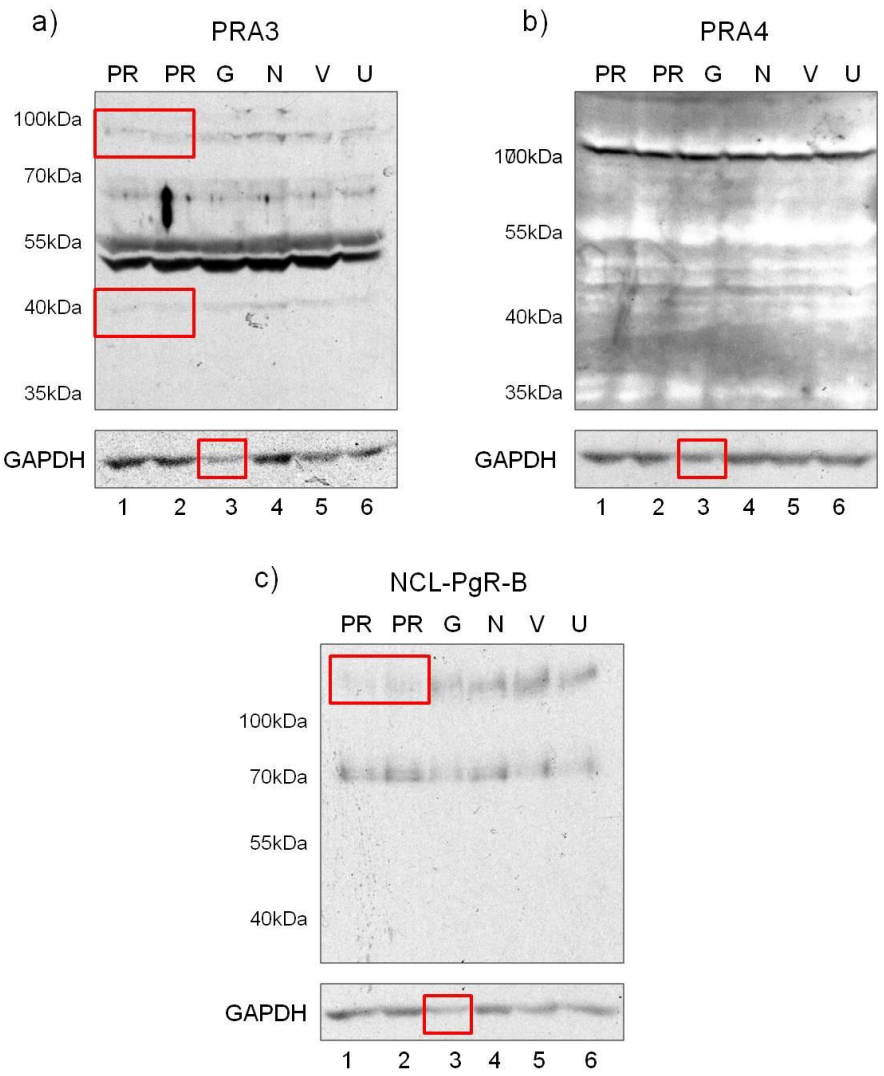


Figure 5.13: PR knockdown is not observed using PRA3 or PRA4.

MCF-7 cells were treated in culture with siRNA for 72 hours prior to lysing for SDS-PAGE and Western immunoblotting. PR= PR targeted siRNA, G= GAPDH targeted siRNA (positive transfection control), N= negative control (untargeted siRNA), V= transfection reagent (vehicle) only control, U= untreated cells. Specific PR and GAPDH knockdown is highlighted in red boxes. **a)** Western immunoblotting using: Upper panel= PRA3, Lower panel= GAPDH. **b)** Western immunoblotting using: Upper panel= PRA4, Lower panel= GAPDH. **c)** Western immunoblotting using: Upper panel= NCL-PgR-B, Lower panel= GAPDH.

5.3.7 Clinicopathological screening of PR status in FFPE breast tumour sections using novel PR antibodies

4µm sections from FFPE breast tumour blocks were stained using PRA3 and PRA4. Optimisation of immunohistochemical staining conditions demonstrated that PRA4 did not produce staining of sufficient quality for assessment of PR status as determined by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital. Therefore, only PRA3 was used to stain the fifteen breast tumour samples analysed by RT-PCR (Section 5.3.1). Sections from each tumour block were initially analysed by H&E staining to assess the presence of tumour tissue (data not shown). This analysis revealed that insufficient tissue was present to reassess sample 15 (Table 5.1).

The remaining fourteen samples were demonstrated by H&E staining to be tumour tissue and were therefore used for re-assessment of PR status by immunohistochemical staining with PRA3. The original PR status was determined by assessment of nuclear staining using an N-terminal PR antibody and the positive/negative tumour status, as well as the Quickscore (when available), is detailed in Table 5.1. The tumour PR status was re-assessed by immunohistochemistry using PRA3 and a modification of the Quickscore method at the Histocytopathology Department, Gateshead Queen Elizabeth Hospital where the original assessment of PR status had been performed. The modified Quickscore method is described in Section 2.18.3. A tumour was deemed PR positive if >10% of nuclei were stained. The positive or negative status determined using PRA3 for each tumour is detailed in Table 5.1.

Sample	Tumour grade	Original PR	PRA3
1	Lobular 2	18	+
2	Ductal 2	6	+
3	Ductal 2	15	+
4	Ductal 2	+	+
5	Ductal 3	15	+
6	Ductal 3	12	-
7	Ductal 1+2	0	-
8	Ductal 2	0	-
9	Ductal1	0	-
10	DCIS	0	+
11	Ductal 3	0	-
12	DCIS	-	-
13	Ductal 3	0	-
14	DCIS	-	-
15	Ductal 2	0	N/A

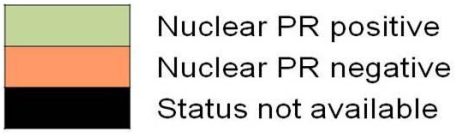
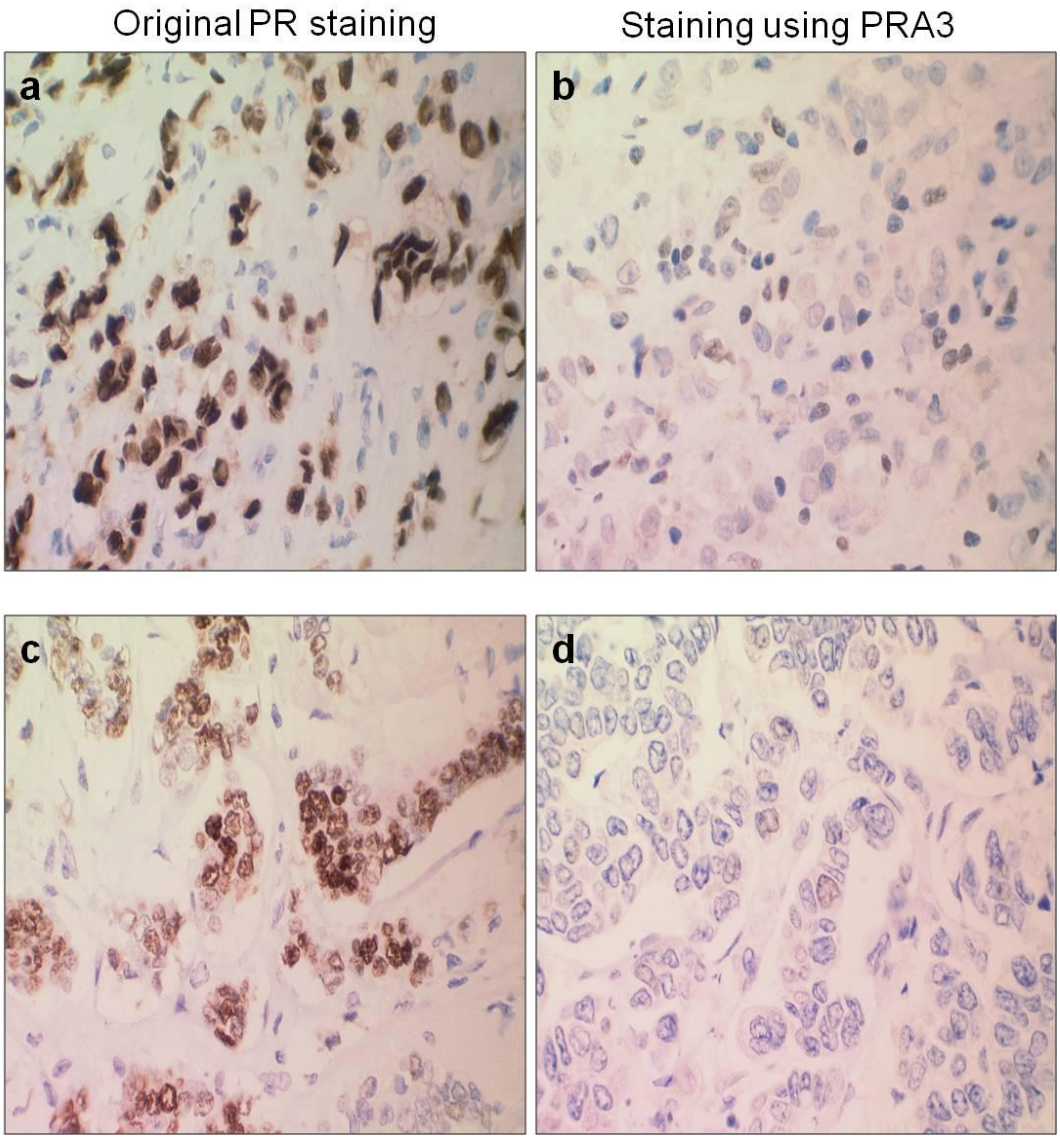


Table 5.1: Comparison of the original PR status of fifteen breast tumour samples with the status determined using PRA3.

Original PR status was assessed by immunohistochemistry using an N-terminal antibody and the Quickscore method, as indicated by the number under 'Original PR'. When Quickscore was not detailed in the histology report, a positive/negative status is indicated (+/-). FFPE breast tumour sections were analysed by H & E staining for the presence of cancer cells. Sample 15 was discounted from re-assessment of PR status using PRA3 due to insufficient tissue being present. Re-screening by immunohistochemistry using PRA3 analysed the presence or absence of nuclear PR staining to determine positive (+) or negative (-) status. Sample 7 is from a patient with multiple tumours classed as grade 1 and grade 2; information detailing from which tumour the research tissue was taken is not available.

Comparison of the nuclear PR status determined using PRA3 with the original PR status, determined using the N-terminal antibody, shows that PRA3 confirms the status as positive or negative for twelve of the fourteen tumours examined (Table 5.1). Original PR staining with the N-terminal antibody and PR staining using PRA3 is shown in Figure 5.14 for two samples for which PR positive status was confirmed; sample 3 (Figure 5.14a and b) and sample 1 (Figure 5.14c and d). Nuclear PR staining is visible in both samples using both the N-terminal antibody (Figure 5.14a and c) and PRA3 (Figure 5.15b and d).



<u>Figure 5.14:</u> Immunohistochemical staining of breast tumour sections using PRA3 confirms the previously reported PR status.

Breast tumour sections were analysed for PR expression by immunohistochemistry using PRA3 and compared to the PR expression detected using N-terminal antibodies during initial screening. Images at x40 magnification. **a)** and **b)** PR expression in a grade 2 invasive ductal carcinoma specimen. **a)** PR expression detected by the N-terminal antibody during original tumour screening; intense nuclear staining (brown) indicates PR expression. **b)** PR expression detected by PRA3; intense nuclear staining (brown) is visible in some cells confirming the PR status determined by original screening. **c)** and **d)** PR expression in a grade 2 invasive lobular carcinoma specimen. **c)** PR expression detected by the N-terminal antibody, intense nuclear PR staining (brown) is visible. **d)** PR expression detected by PRA3; nuclear PR staining (brown) can be seen in some cells confirming PR positive status.

Sample 6 was initially scored as PR positive with a Quickscore of 12, however staining with PRA3 characterised this tumour as PR negative (Table 5.1). It is therefore possible that the PR protein present in this tumour may lack the epitope to which PRA3 is targeted. Nuclear staining is clearly visible in analysis of PR expression in sample 6 using the N-terminal antibody (Figure 5.15a), however no staining is observed using PRA3 (Figure 5.15b).

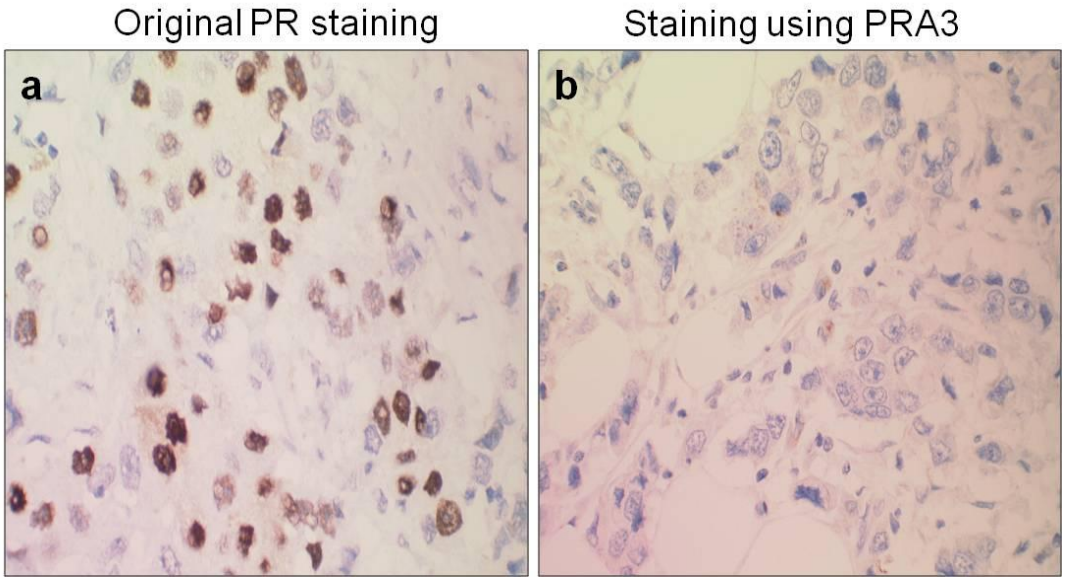
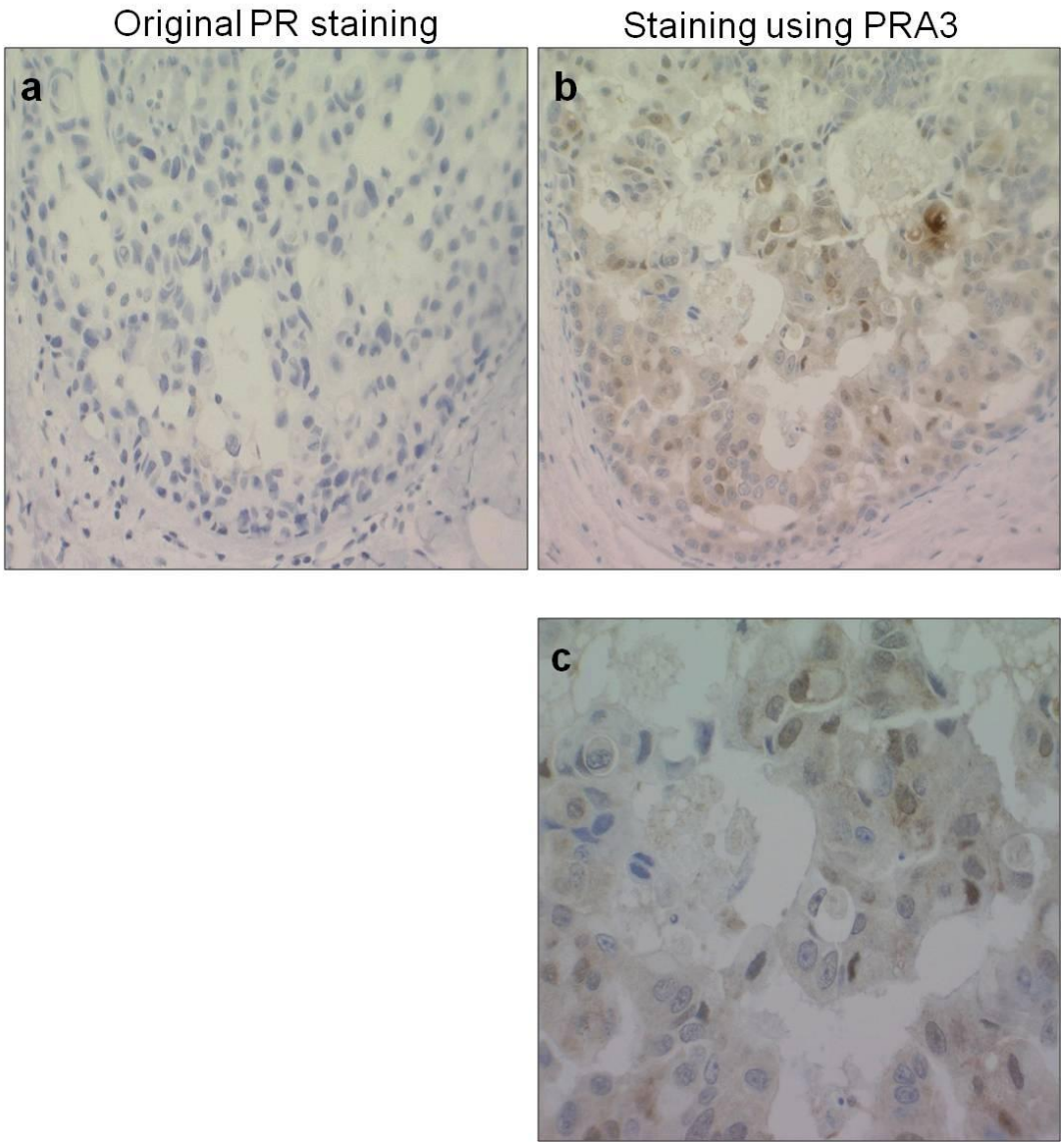


Figure 5.15: Screening with PRA3 indicates a PR negative status in a tumour previously characterised as PR positive.

Sections from a grade 3 invasive ductal carcinoma specimen were analysed for PR expression by immunohistochemistry using an N-terminal antibody and re-assessed using PRA3. Images at x40 magnification. **a)** Original PR staining using the N-terminal antibody demonstrates intense nuclear PR staining (brown). **b)** Staining using PRA3 does not detect nuclear PR expression (lack of brown staining).

Analysis with PRA3 characterised sample 10, a grade 3 invasive ductal carcinoma, as PR positive. However this tumour was originally characterised as PR negative using the N-terminal antibody (Table 5.1). No nuclear staining is visible using the N-terminal PR antibody (Figure 5.16a), however intense nuclear staining is visible using PRA3 (Figure 5.16b). This observed difference in PR status, detected using the different antibodies, suggests that the PR protein expressed in this tumour may be N-terminally truncated. Some cytoplasmic PR staining can also be observed in this sample using PRA3.

Traditional PR screening using N-terminal antibodies focussed on nuclear PR staining which was used to determine the Quickscores detailed in Table 5.1. For direct comparison with the original PR status, re-scoring following staining with PRA3 therefore assessed only nuclear PR expression. However, as described previously in Section 1.1.9, PR can also signal via non-genomic pathways which do not require nuclear localisation. In addition to the nuclear staining detailed in Table 5.1, cytoplasmic PR staining was observed in some tumours. Sample 12 was characterised by both original screening and PRA3 as nuclear PR negative, however cytoplasmic brown staining is visible using PRA3 (Figure 5.17b). A similar pattern of cytoplasmic staining using PRA3 was seen in sample 11 (5.17d), which again was characterised as nuclear PR negative using both the N-terminal antibody and PRA3. This staining suggests that breast tumours may express cytoplasmic PR isoforms which are not detected by N-terminally targeted antibodies and may signal non-genomically.



<u>Figure 5.16:</u> Screening with PRA3 detects PR in a breast tumour originally characterised as PR negative.

PR expression was assessed by immunohistochemistry in a DCIS sample. **a)** Original PR screening using the N-terminal antibody did not detect PR expression, indicated by the lack of brown staining (x20 magnification). **b)** and **c)** Nuclear staining (brown) with PRA3 indicates PR positive status. **b)** x20 magnification, **c)** x40 magnification. Cytoplasmic brown PR staining can also be seen in **b)** and **c)**.

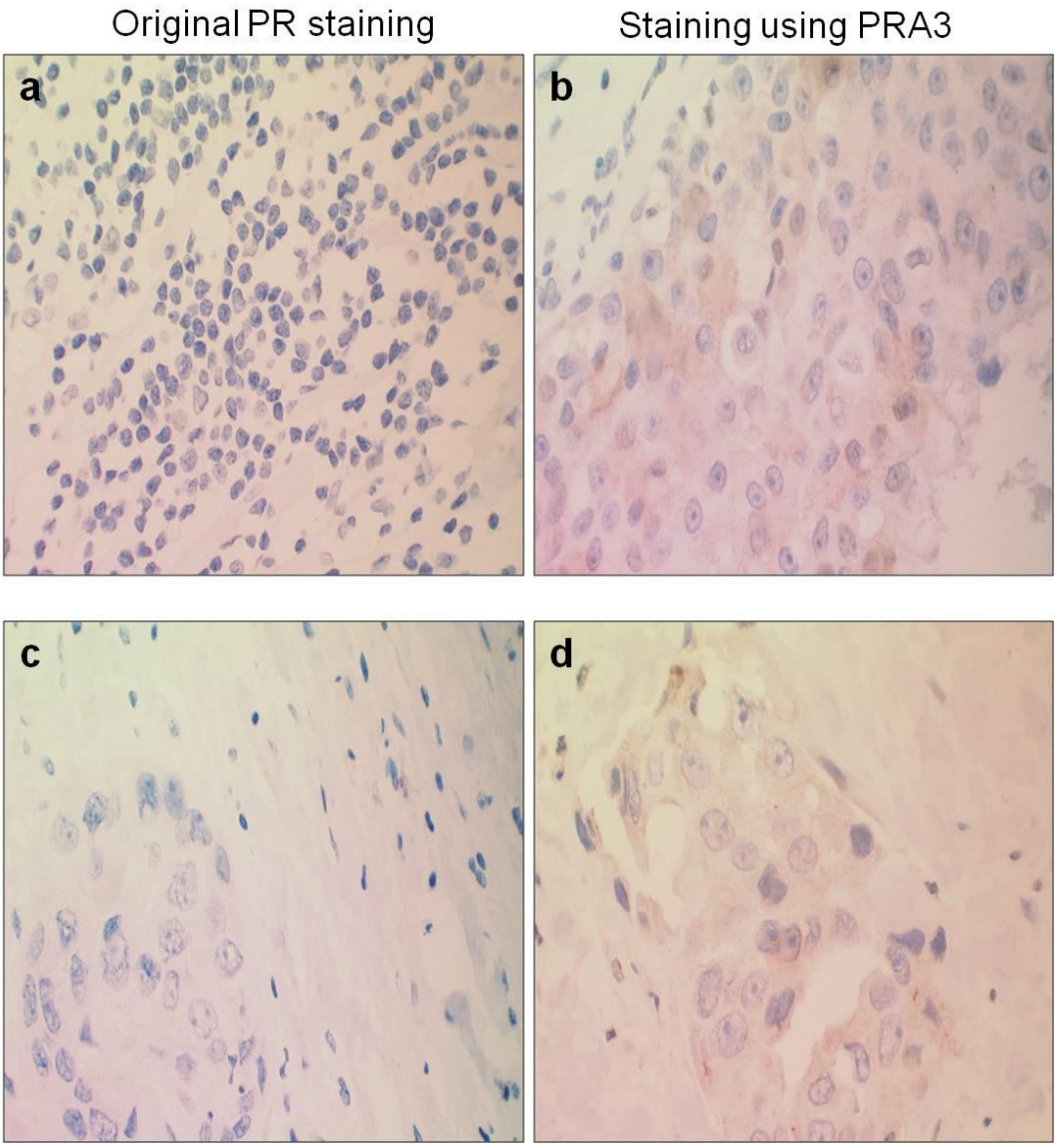


Figure 5.17: Screening with PRA3 indicates cytoplasmic PR expression in tumours characterised as PR negative.

PR expression in breast tumour sections was analysed by immunohistochemistry using the N-terminal screening antibody, and re-assessed using PRA3. **a)** and **b)** PR expression in a DCIS specimen. **a)** Original PR screening using the N-terminal PR antibody did not detect PR expression, indicated by the lack of brown staining (x20 magnification). **b)** Screening with PRA3 reveals cytoplasmic PR expression indicated by brown staining surrounding the nuclei (x40 magnification). **c)** and **d)** PR expression in a grade 3 invasive ductal carcinoma (x40 magnification). **c)** Original PR screening with the N-terminal antibody did not detect PR expression. **d)** PRA3 detects cytoplasmic PR expression.

5.4 Discussion

The analysis of PR mRNA expression in breast tumours presented in this chapter uses a small cohort of frozen breast tumour tissue. Whilst this number is too small to draw firm conclusions regarding differences in patterns of alternative splicing between different groups of tumours (e.g. tumours of different grade), it is sufficient to confirm that alternatively spliced PR mRNA is present in breast tumour tissue, including in samples characterised as PR negative. As described (Section 5.3.1) some RNA degradation may have occurred due to the freezing and storage of these tumour samples. RNA degradation may have affected analysis of long regions by PCR but did not prevent use of the samples in the PR gene walking assay. RNA quality was analysed by NanoDrop spectrophotometry and confirmed by an A260/280 ratio between 1.8 and 2.1.

Fifteen breast tumours were analysed for PR mRNA expression using primers from the PR gene walking assay. For RT-PCR, tumours were ordered according to their hormone receptor status. Information on the individual tumours from the original histology reports from the time of sample collection can be seen in Appendix 2. FFPE sections for fourteen of the tumours used for mRNA analysis were assessed for PR expression by immunohistochemistry using the novel antibody PRA3. As described in Section 5.3.7, sample 15 was not re-assessed by immunohistochemistry due to insufficient tissue being present in the FFPE block.

Analysis of PR alternative splicing in breast tumour tissue using the PR gene walking assay has identified several exon deletion ASEs which occur in breast tumours which are characterised as both PR positive and PR negative. mRNA resulting from retention of the intronic exon M, between exons 3 and 4, which encodes an N-terminally truncated protein was also detected breast tumour samples. Details of the deletions are shown in Table 5.2 and the potential functionality of the putative proteins encoded by each mRNA will be discussed.

Variant	Deletion	In frame	Premature STOP	No. amino acids encoded ΔPR-B/ΔPR-A	Predicted ΔPR-B/ΔPR-A molecular weight (kDa)
full length PR-B/PR-A	n/a	n/a	n/a	933/739	99/82
Δp1,p2	682bp end of exon 1, 27bp start of exon 2	no	exon 2	331/167	34/17
Δp2,3,p4	130bp end of exon 2, exon 3, 72bp start of exon 4	no	exon 4	563/399	57/40
Δ4	exon 4	yes	no	831/667	87/70
$\Delta p4_{126}$	126bp start of exon 4	yes	no	891/727	94/77
Δ5,6	exon 5 and exon 6	no	exon 7	841/677	78/61
Δp6.2	52bp start of exon 6	no	exon 6	803/639	84/67

Table 5.2: Predicted structures of potential PR proteins present in breast tumour tissue.

The mRNA sequence of each sequenced PCR product was entered into the ExPASy Translate Tool to determine the encoded amino acid sequence and identify a change in translational reading frame. The molecular weight of each potential protein was predicted using the ExPASy Compute pI/Mw tool. Details of full length PR-A/B proteins, encoded by the full 8 exon PR mRNA, are given for comparison.

A deletion of 682bp from the end of exon 1 and 27bp from the start of exon 2 (Δp1,p2) was sequenced from sample 9, a grade 1 invasive ductal carcinoma. A PCR product of the same size was also detected in samples 1, 3, 14 and 15 (Figure 5.2) and was sequenced from MDA-MB-231 cells (Section 3.3.1). This mRNA would encode a protein with a truncated N-terminal domain lacking AF1, and lacking the DBD and C-terminal functional domains. Δp2,3,p4 was sequenced from sample 3, a grade 3 invasive ductal carcinoma. This deletion has not previously been reported and would encode a protein lacking the DBD and hinge, and would cause a change in reading frame and truncation within the LBD. Therefore it is unlikely that this protein could function as a PR. The changes in reading frame caused by these deletions, encoding a PTC, may target the mRNA for degradation by NMD.

Exon 4 deleted PR mRNA has previously been reported (Richer et al., 1998; Leygue et al., 1996a) and was detected in the MDA-MB-231 cell line (Section 3.3.1). Deletion of exon 4 would delete the hinge region and the N-terminal portion of the LBD. This deletion was sequenced from sample 1 and sample 9, a grade 2 lobular and grade 1 ductal carcinoma respectively. Δ4 mRNA was detected in the majority of tumour samples studied. Analysis of the relative levels of full length mRNA and Δ4 mRNA generated using an exon 3 sense primer and exon 5 antisense primer (Figure 5.4a) showed that DCIS, samples 10, 12 and 14, appear to express more Δ4 than full length mRNA. In tumours of higher grade, there appears to be a shift towards higher levels of full length mRNA relative to Δ4 (Figure 5.4a). Within the PR exon 4 sequence is a cryptic 3′ splice site which generates the previously reported $\Delta4_{126}$ mRNA (Marshburn et al., 2005; Balleine et al., 1999b). This deletion would truncate 42 amino acids from the hinge region, but the mRNA would remain in the original reading frame, so an encoded protein would possess the normal LBD, DD and AF2. Although the deleted region of exon 4 encodes protein domains essential for nuclear localisation, mutations affecting this region have been demonstrated to only delay, and not to prevent, nuclear localisation and transcriptional response to progestins (Daniel et al., 2010). Therefore, PR-A$\Delta4_{126}$ (77kDa) or PR-B$\Delta4_{126}$ (94kDa) may function as a normal PR protein. PR$\Delta4_{126}$ was sequenced from a PCR product generated using sample 8, sample 11 and sample 15 and similar sized PCR products are visible in a number of other samples (Figure 5.4a and b).

Exon 5 and 6 deleted PR mRNA (Δ5,6) has previously been reported (Misao et al., 2000; Leygue et al., 1996a). This deletion was sequenced from a DCIS specimen, sample 14, from a PCR product generated using primers spanning exon 1 to exon 7. This deletion would change the reading frame and insert a PTC, potentially targeting the mRNA for degradation by NMD. Alternatively it may encode a protein lacking the functional LBD, but possessing DNA binding and phosphorylation domains, as well as the N-terminal Src binding domain, AF1 and AF3. Therefore this protein could function independently of ligand via non-genomic or transcriptional pathways. Similarly to exon 4, the cassette exon 6 contains a cryptic 3′ splice site, deleting 52 bp from the start of the exon. This deletion causes a change in reading frame and truncation within the LBD. Therefore, similarly to Δ5,6 described above, this protein is unlikely to function in response to

hormone but may function through ligand independent pathways. Alternatively this mRNA may be a target for NMD.

PR exon 6 is a cassette exon, meaning deletion does not cause a change in translational reading frame. Whilst an exon 6 deletion (Δ6) was not successfully sequenced from PCR products generated using breast tumour RNA, an appropriately sized deletion was observed in eight of the fifteen tumours using primers spanning exon 5 to exon 7 (Figure 5.5). This deletion would remove a section of the LBD, potentially removing the ligand binding capacity, but would not target the mRNA for NMD. Therefore PRΔ6 protein is more likely to be expressed than the other LBD deletions described above.

PR-M mRNA was also detected in some tumours using primers spanning exon M to exon 5. Analysis of PR-M expression in MDA-MB-231 cells (Section 3.3.1) used primers spanning exon M to exon 4, however PCR using GAPDH primers (data not shown) demonstrated a small amount of gDNA contamination in tumour RNA which was not present in cell line RNA. Therefore a longer gene region was examined by PCR to exclude amplification of this gDNA instead of the cDNA template. PR-M mRNA was detected in tumours of different reported PR status and different tumour grade (Figure 5.6). As described previously, PR-M may be a non-genomic signalling cytoplasmic PR isoform, although thorough characterisation of protein expression and function has not been reported.

Breast tumour tissue expresses alternatively spliced PR mRNA which may encode functionally distinct PR variant proteins, including PR-M. PR-M mRNA may encode a potentially functional N-terminally truncated protein detected in MDA-MB-231 breast cancer cells (discussed in Section 4.4). N-terminally truncated proteins such as PR-M would be undetectable by current breast cancer screening antibodies targeting N-terminal epitopes, leading to 'false negative' tumour statuses. Conversely, non-functional/differently functional PR isoforms may result from exon deletion which would be detected by breast cancer screening, leading to positive status tumours which may not express wild-type PR-A or PR-B. Analysis of the commercially available C-terminal PR antibody C19 demonstrated that non-specific protein binding makes this

antibody unsuitable for immunofluorescent analysis of PR expression in cell lines. Therefore, for the detection of N-terminally truncated PR isoforms in breast cancer, new and more specific antibodies are required. Novel antibodies were therefore developed in this study, targeted to non N-terminal epitopes, and their specificity analysed for use in laboratory techniques and as potential breast cancer screening antibodies. MCF-7 breast cancer cells were used for validation of the novel antibodies since they are known to express high levels of PR protein. Western immunoblotting demonstrated that PRA3, targeted to an internal epitope, detects a range of proteins in MCF-7 breast cancer cells (Figure 5.9). Co-IP experiments further demonstrated that the proteins detected by PRA3 were also detected by the commercial non N-terminally targeted PR antibodies C19 and C20 (Figure 5.12a). Western immunoblotting and Co-IP experiments appeared to demonstrate that the novel antibody PRA4, targeted to a C-terminal epitope, did not detect PR proteins (Figures 5.11 and 5.12b).

To validate that the novel antibody PRA3 was detecting PR protein, and to confirm that PRA4 was not, siRNA knockdown of total PR expression was performed in MCF-7 cells and analysed by Western immunoblotting with each novel antibody (Figure 5.13a and b). Efficient PR knockdown was confirmed by Western immunoblotting using NCL-PgR-B (Figure 5.13c) which was demonstrated in Section 3.3.6 to specifically detect PR-B protein. No knockdown of any protein detected by PRA4 was observed (Figure 5.13b), which may indicate a lack of specificity or may be due to the low quantity of antibody purified being insufficient to adequately detect protein. Knockdown of a protein of approximately 100kDa and another of approximately 40kDa was observed in PR siRNA treated MCF-7 cells by Western immunoblotting using PRA3 (Figure 5.13a), confirming that this antibody detects PR proteins. However, no knockdown of other proteins detected by PRA3 was observed demonstrating some non-specific binding by this antibody.

The novel PR antibody PRA3, targeted to an epitope within the DBD was demonstrated to detect PR in the preliminary cell line validation studies described above. PRA4, targeted to a C-terminal epitope, appeared not to detect PR proteins. Both antibodies were optimised by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital, for use in immunohistochemistry on FFPE breast tumour sections. Epitope

retrieval procedures were optimised for each antibody and optimal staining conditions determined by staining sections of tumours with different reported PR levels using different concentrations of each antibody (data not shown). This optimisation determined the best conditions for use of PRA3 and again demonstrated that PRA4 failed to detect protein.

PRA3 was subsequently used to assess PR expression in fourteen of the breast tumours analysed by RT-PCR. Of the samples analysed, three were DCIS, one grade 1 invasive ductal carcinoma, five grade 2 invasive ductal carcinoma, four grade 3 invasive ductal carcinoma and one grade 2 invasive lobular carcinoma. Sample 15, a grade 2 invasive ductal carcinoma, was not re-assessed due to insufficient tissue being present in the FFPE block as described above. Analysis of PR expression in the remaining fourteen samples using PRA3 was performed by the Histocytopathology Department, Gateshead Queen Elizabeth Hospital. The original PR status of the tumours was assessed using an N-terminal antibody and the Quickscore method at the time of sample collection, and assessed nuclear PR expression. The PR status determined following staining with PRA3, using a modified Quickscore method, was in agreement with the original status for twelve of the fourteen tumours assessed demonstrating that this novel antibody does detect PR (Table 5.1). The modification to the Quickscore method for analysis of PRA3 staining was necessary since the staining was not as intense as commercial antibodies. The modified PRA3 Quickscore therefore considered only the proportion of positively stained nuclei and reported a positive/negative status as described in Section 2.18.3.

The PR status of the other two tumours, sample 6 and sample 10, differed from that originally reported following assessment with the N-terminal antibody. Sample 6 is a grade 3 invasive ductal carcinoma which was originally characterised as PR positive with a Quickscore of 12. Staining with PRA3 did not detect PR and the tumour was scored as PR negative. Sample 10 is a DCIS which was originally characterised as PR negative, however following staining with PRA3 this sample was scored as PR positive. It is interesting to note that the tumour which was originally scored positive but rescored negative is a high grade invasive tumour, whereas the opposite rescoring was seen in a DCIS sample. Further analysis of a greater number of tumours of differing grade is necessary to identify if there is a correlation between tumour grade and the difference in

PR status following assessment with PRA3. As discussed above, PRΔ4 mRNA is more prevalent than full length PR mRNA in DCIS samples (samples 10, 12 and 14). Along with the rescoring of sample 10 as PR positive using PRA3, this data suggests that alternative splicing of PR may be occurring in the early stages of breast cancer development which may affect the PR status as determined using N-terminal antibodies. Whilst N-terminal PR antibodies would detect the protein encoded by an exon 4 deleted PR-A or PR-B variant, it is possible that an N-terminally truncated PR variant may also contain exon deletions. Previous reports studying other hormone receptors have demonstrated that use of differential 5′ exons can lead to differences in the pattern of downstream splicing (Russcher et al., 2007; Wang et al., 2005; Flouriot et al., 2000). For example, the GR 5′ UTR consists of at least nine alternative exons 1 and the first thirteen nucleotides of exon 2 (Bockmuhl et al., 2011; Russcher et al., 2007); alternative exon 1 usage has been linked to alternative 3′ end splicing in human GR (Russcher et al., 2007), and tissue specific variation of mouse GR transcripts resulting from different exons 1 has been reported (Bockmuhl et al., 2011). ERα46 and ERα36 result from splicing of alternative exons from within the 5′ UTR onto exon 2, leading to skipping of exon 1 (Wang et al., 2005; Flouriot et al., 2000). Use of the ERα36 5′ exon results in alternative 3′ end splicing (Wang et al., 2005). If similar, currently unidentified, exons exist in the PR 5′ UTR or within intron 1, similar to exons S, T and M within intron 3, then splicing of these exons onto exon 2 may favour different patterns of downstream splicing including exon 4 skipping. This alternative splicing would encode an N-terminally truncated PR protein with DNA and ligand binding capacity which could be detected by PRA3 but not the N-terminal antibody and may be expressed preferentially in different tissue types or disease states, such as DCIS.

Cytoplasmic PR staining was detected in several breast tumour samples (Figure 5.17). Whilst no apparent correlation can be seen in this small cohort of samples between the presence of cytoplasmic staining and the ASEs detected by RT-PCR, this staining potentially indicates the expression of PR isoforms which may lack the NLS encoded by exon 4 and function by cross-talk with cytoplasmic signalling pathways. Interaction of cytoplasmic PR with c-Src, EGFR and Stat transcription factors have been previously reported (Faivre et al., 2008; Proietti et al., 2005; Boonyaratanakornkit et al., 2001). However, this staining is only evident using PRA3 so the cytoplasmic PR proteins detected may lack the NTD which contains the domain responsible for interaction with

c-Src (Boonyaratanakornkit et al., 2001). Alternatively, the cytoplasmic staining may be the result of non-specific binding, although it was not visible in all samples.

PR status has traditionally been considered as a surrogate marker of a functional ER signalling pathway (Osborne, 1998), however tumours which are ER negative, but PR positive, such as sample 6 in this study, suggests that ER is not essential for PR expression. Furthermore, the value of PR status as both a predictor of endocrine therapeutic outcome and prognosis (Anderson et al., 2011; Cui et al., 2005; Bardou et al., 2003) highlight the importance of accurate assessment of PR status in breast cancer. A tumour characterised as PR negative would be considered less likely to respond to endocrine therapies. However, as exemplified by sample 10 (Table 5.1) assessment with N-terminal antibodies may be unable to detect all PR proteins. Therefore the original PR status may give an inaccurate impression of the likely response to endocrine therapies. Breast cancer therapeutics targeting PR, such as SPRMs, are being developed (Im and Appleman, 2010). In order to target these therapies to the correct patients, accurate assessment of PR status would be vital, including the detection of N-terminally truncated variants which are currently not detected by breast cancer PR screening. Further development of non N-terminal PR antibodies such as PRA3 and PRA4 could contribute to improved detection of PR variants in breast cancer, thus contributing to accurate therapeutic targeting.

PRA3 appears in the Western immunoblotting, Co-IP, siRNA and immunohistochemistry presented in this chapter to detect PR proteins. However, Western immunoblotting following siRNA knockdown demonstrates that non PR proteins are also detected by PRA3. Binding to these proteins does not appear to occur in immunohistochemistry; if PRA3 bound ubiquitously expressed non PR proteins in the immunohistochemical screening, then all sections would stain positive. Despite this apparent non-specific binding observed in cell line studies, the PR status of the majority of tumours agreed with the original PR status, including PR negative. It is possible therefore that the non-specific binding requires protein linearisation or that the epitope retrieval conditions used prior to immunohistochemistry may not expose the epitope in non PR proteins, so PRA3 does not detect the protein in its native form by

immunohistochemistry. Alternatively, it is possible that the non-PR proteins detected by PRA3 are cell line specific and not expressed in breast tumour tissue.

In summary, this study of a small cohort of frozen breast tumour data validates the hypothesis that alternatively spliced PR mRNA, including that encoding the N-terminally truncated PR-M isoform, is present in breast tumour tissue. Novel antibodies PRA3 and PRA4 have been developed in an attempt to detect N-terminally truncated as well as full length PR proteins. PRA4 does not appear to detect PR protein and therefore re-purification and validation of this C-terminally targeted antibody is necessary. PRA3 detects PR proteins and confirms the PR status of the majority of tumours examined. Two tumours were re-scored using PRA3, confirming that the expression of variant PR proteins may affect the PR status determined using PR antibodies targeting different epitopes. This data provides a foundation for further comprehensive analysis of PR expression in breast tumour tissue, comparison with non-tumour breast tissue and development of novel PR antibodies for breast cancer screening.

6 Regulation of PR alternative splicing events in MCF-7 breast cancer cells

6.1 Introduction

As described in Section 1.4.2, splicing involves the recognition of a 5′ splice site by U1 snRNP, the branch point adenosine by U2 snRNP and assembly of the mature spliceosome by recruitment of the U4/U6-U5 tri-snRNP. The spliceosome then catalyses the trans-esterification of the 5′ GU by the branch point adenosine, and the trans-esterification of the 3′ AG by the 3′ hydroxyl group of the upstream exon (Staley and Guthrie, 1998). The recognition of the 5′ and 3′ splice sites is thought to usually occur across the relatively short distance of an exon, a theory termed 'exon definition' (Berget, 1995). Recruitment of snRNPs is regulated by interaction with *trans*-acting splicing factors bound to *cis*-acting splicing regulatory elements which can be either exonic or intronic and either enhance or silence splicing; i.e. exon splice enhancers/silencers (ESE/ESS) or intron splice enhancers/silencers (ISE/ISS) (Hertel, 2008).

SR proteins are a family of splicing factors which commonly bind to ESEs and enhance splicing (Long and Caceres, 2009). The structure and function of SR proteins is described in Section 1.4.3. SR proteins contain RRMs which mediate binding to specific RNA sequences and the function of SR proteins depends on the context in which they are bound, i.e. within an exon or within an intron, and the proximity of the SR protein binding site to other regulatory elements (Long and Caceres, 2009). Antagonistic functions have been described between different SR proteins as well as between SR proteins and hnRNP splicing factors (Graveley et al., 2001).

Alternative splicing of PR pre-mRNA was characterised in this study in the reportedly PR negative MDA-MB-231 breast cancer cell line in Chapter 3 and in a small cohort of breast tumour samples in Chapter 5. Amongst the ASEs identified in both MDA-MB-231 cells and breast tumour tissue was deletion of exon 4, which had previously been reported in breast tumour tissue and other human tissues (Hodges et al., 1999; Misao et al., 1998; Leygue et al., 1996a). Furthermore, two different deletions of parts of exon 4

were observed; a novel 76bp deletion was seen within exon 4 in MDA-MB-231 cells, whereas in breast tumour tissue a 126bp deletion was observed from the start of exon 4. This latter deletion had previously been detected in breast tumour and normal endometrial tissue (Marshburn et al., 2005; Balleine et al., 1999b) and is associated with a cryptic 3′ AG splice site within exon 4. Further analysis of the PR exon 4 sequence identifies a 5′ GU splice site with a poor match to the consensus GUA/GAGU sequence, a 3′AG, a short PPT and a consensus BPS of CTGAC containing the conserved adenosine in the correct position. PR exon 4 is 306bp long, a greater length than is necessary for efficient 'exon definition' (Berget, 1995). Therefore, PR exon 4 is an ideal exon for enhancer dependent alternative splicing and was examined for the presence of potential RBP binding sites and preliminary investigation of its splicing regulation was performed.

Analysis of PR alternative splicing in MDA-MB-231 cells did not identify deletion of the whole of exon 6 which had previously been identified in breast tumour and other tissues as well as breast cancer cell lines including MCF-7 (Hodges et al., 1999; Misao et al., 1998; Leygue et al., 1996a). In MDA-MB-231 breast cancer cells an in-frame deletion of 96bp was detected within exon 6, and in breast tumour tissue a previously reported 52bp deletion from the start of exon 6 was observed. This mRNA had previously been reported at a greater frequency in breast tumour than normal breast tissue (Nagao et al., 2003) and is associated with a cryptic 3′ AG splice site within exon 6. Because exon 6 alternative splicing had previously been reported, and a cryptic splice site exists within the exon, preliminary investigation of the regulation of exon 6 alternative splicing was also undertaken.

Consensus binding motifs for the SR proteins SRSF1, SRSF2, SRSF5 and SRSF6 have been identified using a functional systematic evolution of ligands by exponential enrichment (SELEX) method (Liu et al., 2000; Liu et al., 1998). Briefly, ESEs were identified by the ability of oligonucleotide sequences to promote efficient splicing of pre-mRNA from a mini-gene. Pre-mRNA splicing of sequences containing ESE oligonucleotides was then performed in the presence of individual splicing factors (SRSF1, SRSF2, SRSF5 or SRSF6). The oligonucleotide sequences which promote splicing in the presence of each splicing factor were analysed to determine consensus

binding motifs (Liu et al., 2000; Liu et al., 1998). The ESEfinder program analyses an input sequence for the presence of the binding motifs and provides a score for each putative binding site based on how closely it matches the consensus and how frequently a nucleotide was seen at a particular position in SELEX experiments (Cartegni et al., 2003). The sequences of PR exons 4 and 6 were entered into ESEfinder to identify potential SR protein binding sites.

Splicing Rainbow searches the input exon sequence for the presence of binding motifs for a range of splicing factors using information gathered from literature and stored in the Alternative Splicing Database AEdb Motif database (Stamm et al., 2006). Results are presented in colour coded form to indicate the consensus match for each splicing factor and have been tabulated for presentation in this thesis.

Both Splicing Rainbow and ESEfinder use the traditional nomenclature for SR proteins which has recently been replaced by a new convention (Manley and Krainer, 2010) which is used throughout this thesis and described in Section 1.4.3. Briefly the old and new nomenclature are as follows for the genes/proteins studied herein; SF2/ASF refers to SRSF1, SC35 to SRSF2, SRp40 to SRSF5 and SRp55 to SRSF6.

6.2 Aims

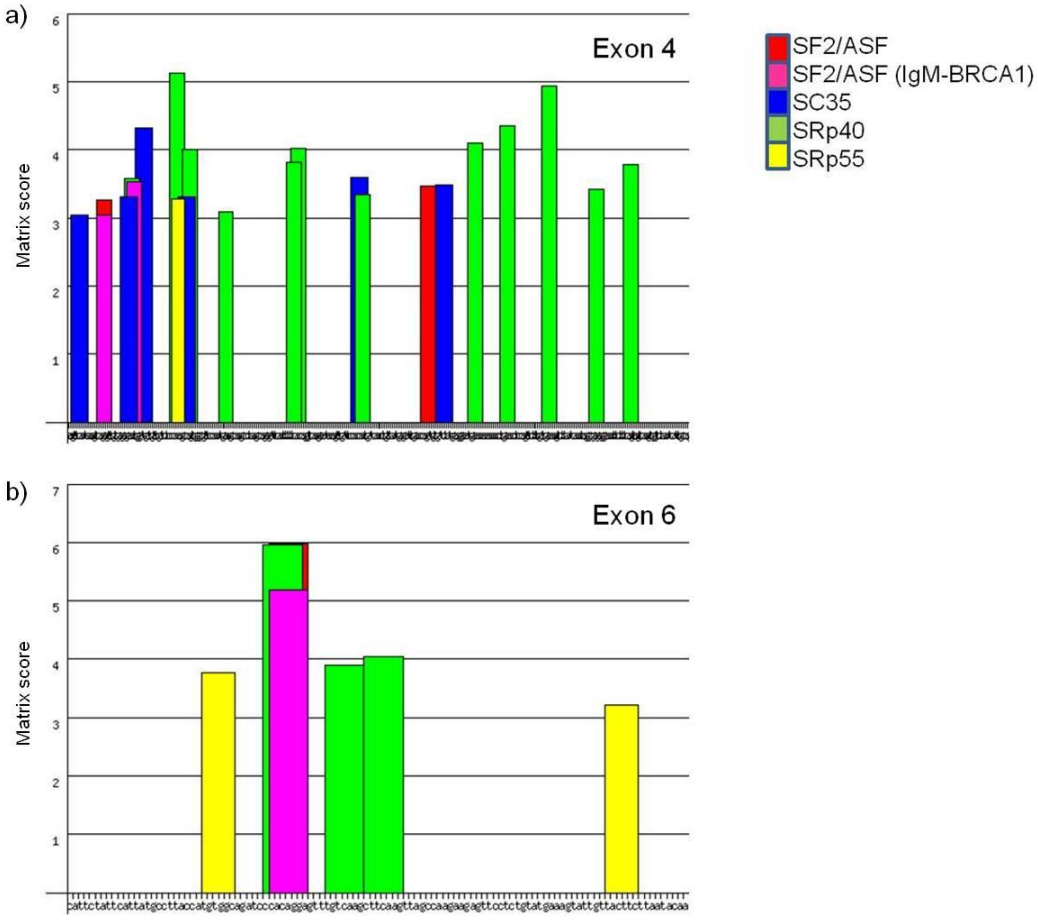
The aims of this chapter were to:

- Predict the potential binding sites for a range of splicing factors present within the PR exon 4 and exon 6 sequences.
- Knockdown the expression of SRSF1, SRSF2, SRSF5 and SRSF6 in MCF-7 cells.
- Perform RT-PCR experiments for a preliminary examination into the effect of splicing factor knockdown on the alternative splicing events involving PR exons 4 and 6.

6.3 Results

6.3.1 Identifying SR protein binding sites in PR exons 4 and 6

Analysis of PR exon 4 using ESEfinder demonstrated that this exon contains potential binding sites for SRSF1, SRSF2, SRSF5 and SRSF6. Eleven potential SRSF5 and six potential SRSF2 binding sites are seen throughout the exon 4 sequence. Interestingly, four of the six SRSF2 sites are located close to the 5′ end of the exon (Figure 6.1a) which is rich in potential RBP binding sites, with overlapping sites for SRSF1 and SRSF2, SRSF2 and SRSF5, and SRSF5 and SRSF6. Exon 6 contains far fewer potential binding sites; three potential SRSF5 binding sites are present, one of which overlapped with a high scoring potential SRSF1 sites. Two potential SRSF6 sites are present in exon 6, and no SRSF5 binding sites were detected (Figure 6.1b).



<u>Figure 6.1:</u> Histograms representing the presence of potential SR protein binding sites within PR exons 4 and 6.

PR exons 4 and 6 (x-axis) were screened for the presence of potential SR protein binding motifs by the ESEfinder database. Sequence similarity to a consensus matrix motif was identified by ESEfinder and the potential binding motif was scored to indicate binding probability (y-axis). **a)** Potential SR protein binding motifs within PR exon 4. **b)** Potential SR protein binding motifs present in PR exon 6.

The presence of potential binding sites for SR proteins within the alternatively spliced PR exons 4 and 6 was also analysed using Splicing Rainbow. The sequence of each exon was entered into the program and analysed for the presence of consensus binding sequences for each RBP. Binding sites for the SRSF1, SRSF2, SRSF5 and SRSF6 proteins analysed by ESEfinder were again observed in exon 4 using Splicing Rainbow (Table 6.1). High scoring potential binding sequences for other SR proteins are also detected in PR exon 4, especially for SRSF3 (SRp20) towards the 3′ end of the sequence (Table 6.2). Binding sites for other RBPs known to have roles in activating constitutive or alternative splicing were also detected within exon 4 (Table 6.3).

Factor	Consensus type	Score	Nuc start	Nuc end
SF2/ASF	matrix 10-mer	5.472	25	34
	matrix 10-mer	4.776	34	43
	matrix 10-mer	5.761	101	110
	matrix 10-mer	4.135	125	134
	matrix 10-mer	4.23	130	139
	matrix 10-mer	6.01	197	206
	matrix 10-mer	6.394	202	211
	matrix 10-mer	6.179	207	216
	matrix 9-mer	3.452	194	202
	matrix 7-mer	3.692	29	35
	matrix 7-mer	4.163	43	49
	matrix 7-mer	3.564	181	187
	matrix 7-mer	3.077	237	243
SC35	matrix 11-mer B	6.159	143	153
	matrix 10-mer C	8.041	45	54
	matrix 10-mer D	7.359	77	86
	matrix 10-mer D	8.075	176	185
	matrix 9-mer F	6.715	257	265
	matrix 7-mer G	9.785	50	56
	matrix 7-mer G	5.755	235	241
	matrix 7-mer H	4.712	179	185
	matrix 7-mer H	6.115	197	203
	matrix 7-mer H	5.407	257	263
	matrix 8-mer I	3.279	40	47
	matrix 8-mer I	4.293	47	54
	matrix 8-mer I	3.181	67	74
	matrix 8-mer I	3.604	148	155
	matrix 8-mer I	3.233	188	195
SRp40	matrix 6-mer	4.049	64	69
	matrix 6-mer	3.415	221	226
	matrix 6-mer	3.353	241	246
	matrix 6-mer	4.109	293	298
	matrix 5-mer	4.591	44	48
	matrix 5-mer	5.593	65	69
	matrix 5-mer	3.171	137	141
	matrix 5-mer	3.145	200	204
	matrix 5-mer	3.564	205	209
	matrix 5-mer	3.941	220	224
	matrix 5-mer	3.145	240	244
SRp55	matrix 7-mer	6.113	42	48
	matrix 7-mer	5.366	57	63
	matrix 7-mer	6.074	63	69
	matrix 7-mer	3.363	69	75
	matrix 7-mer	3.116	98	104
	matrix 7-mer	5.032	118	124
	matrix 7-mer	4.688	120	126
	matrix 7-mer	3.729	126	132
	matrix 7-mer	3.954	145	151
	matrix 7-mer	4.476	147	153
	matrix 7-mer	3.593	150	156
	matrix 7-mer	3.229	198	204
	matrix 7-mer	4.762	203	209
	matrix 7-mer	3.124	208	214
	matrix 7-mer	3.525	210	216
	matrix 7-mer	5.211	218	224
	matrix 7-mer	3.276	223	229
	matrix 7-mer	4.78	238	244
	matrix 7-mer	3.225	252	258
	matrix 7-mer	3.597	276	282
	matrix 7-mer	3.41	297	303

Table 6.1: Confirmation of the presence SR protein binding sites within PR exon 4.

The PR exon 4 sequence was screened for the presence of SRSF1 (SF2/ASF), SRSF2 (SC35), SRSF5 (SRp40) and SRSF6 (SRp55) binding sequences using Splicing Rainbow. The presence of binding sites previously detected by ESEfinder was confirmed.

Factor	Consensus type	Score	Nuc start	Nuc end
SRp20	matrix 9-mer	5.631	248	256
	matrix 8-mer	5.004	116	123
	matrix 8-mer	5.217	244	251
	matrix 8-mer	5.96	268	275
	matrix 8-mer	5.301	293	300
	matrix 7-mer	5.244	155	161
	matrix 7-mer	5.476	203	209
9G8	matrix 9-mer	4.164	85	93
	matrix 9-mer	3.506	130	138
	matrix 9-mer	5.8	259	267
Tra2B	AAGAAGAA	6 matches	261	268

Table 6.2: Binding motifs for other SR and SR related proteins are present within PR exon 4.

Screening of PR exon 4 using splicing rainbow also revealed the presence of potential binding sites for other SR proteins not analysed by ESEfinder.

Factor	Consensus type	Score	Nuc start	Nuc end
hnRNP A2/B1	2x poly-T	6 T's (in 9)	111	120
	poly-T	6 T's (in 9)	230	238
	ATTTA	Exact match	8	11
hnRNP C1/C2	TTTTT	Exact match	116	120
	ATTTA	Exact match	8	12
hnRNP E1/E2	2x poly-C	6 C's (in 9)	58	70
hnRNP F	poly-G	6 G's (in 9)	258	266
hnRNP G	3x poly-A	7 A's (in 10)	5	16
	3x poly-A	7 A's (in 10)	205	216
hnRNP H	poly-G	6 G's (in 9)	258	266
hnRNP I (PTB)	7x poly-Y	8 pyrimidines (in 10)	52	67
	7x poly-Y	8 pyrimidines (in 10)	110	125
	poly-Y	8 pyrimidines (in 10)	145	154
	6x poly-Y	8 pyrimidines (in 10)	223	239
	5x poly-Y	8 pyrimidines (in 10)	267	280
	TCTT	Exact match	231	234
	TCTT	Exact match	245	248
	TCTT	Exact match	272	275
hnRNP K	matrix 6-mer	5.798	61	66
	matrix 6-mer	4.256	79	84
	matrix 6-mer	4.256	80	85
	matrix 6-mer	6.038	96	101
hnRNP U	2x poly-G	7.5 G's (in 10)	31	42
	poly-G	7.5 G's (in 10)	257	266
hnRNP A0	ATTTA	Exact match	8	12
HuR	poly-T	7 T's (in 10)	111	120
	(T/A)TTT(T/A)	Exact match	8	12
	(T/A)TTT(T/A)	Exact match	116	120

Table 6.3: Binding sequences for RNA binding proteins within PR exon 4.

Splicing rainbow revealed the presence of binding sites for a number of RBPs with known roles in splcing activation or silencing.

Within the sequence of PR exon 6 the presence of SRSF1, SRSF5 and SRSF6 binding sites detected by ESEfinder is confirmed by Splicing Rainbow (Table 6.4). In addition Splicing Rainbow identifies potential SRSF2 binding sites within PR exon 6 (Table 6.4) as well as high scoring binding sequences for SRSF3 and other SR proteins (Table 6.5). PR exon 6 also contains potential binding sites for a number of other RBPs which were not analysed by ESEfinder (Table 6.6).

Factor	Consensus type	Score	Nuc start	Nuc end
SF2/ASF	matrix 10-mer	4.436	8	17
	matrix 10-mer	4.427	46	55
	matrix 9-mer	5.056	88	96
	matrix 7-mer	4.797	2	8
	matrix 7-mer	3.484	5	11
	matrix 7-mer	6.378	57	63
SC35	matrix 11-mer B	7.117	73	83
	matrix 10-mer D	6.998	45	54
	matrix 9-mer F	5.044	60	68
	matrix 9-mer F	8.331	88	96
	matrix 7-mer H	4.127	3	9
	matrix 7-mer H	6.022	60	66
	matrix 7-mer H	4.127	88	94
	matrix 8-mer I	3.084	95	102
SRp40	matrix 6-mer	4.157	117	122
	matrix 5-mer	5.593	1	5
	matrix 5-mer	5.174	58	62
SRp55	matrix 7-mer	6.092	56	62
	matrix 7-mer	3.562	67	73
	matrix 7-mer	4.408	74	80
	matrix 7-mer	3.744	116	122
	matrix 7-mer	3.721	123	129
	matrix 7-mer	4.203	126	132
	matrix 6-mer	4.087	3	8

Table 6.4: SR protein binding sites within exon 6 are confirmed by Splicing Rainbow.

Splicing Rainbow confirms the presence of SRSF1, SRSF5 and SRSF6 binding sites which had previously been identified by ESEfinder in PR exon 6. Additional SRSF2 motifs are identified.

Factor	Consensus type	Score	Nuc start	Nuc end
SRp20	matrix 9-mer	5.37	17	25
	matrix 8-mer	7.723	16	23
	matrix 8-mer	6.427	25	32
	matrix 8-mer	5.489	28	35
	matrix 8-mer	5.822	72	79
	matrix 8-mer	6.473	117	124
	matrix 8-mer	6.286	120	127
	matrix 7-mer	5.947	74	80
9G8	matrix 9-mer	5.511	8	16
Tra2B	AAGAAGAA	6 matches	8	15
	AAGAAGAA	6 matches	9	16
	AAGAAGAA	6 matches	12	19
	AAGAAGAA	7.5 matches	87	94

Table 6.5: SR protein binding sites within PR exon 6.

The presence of binding sequences for SR proteins not analysed by ESEfinder are detected by Splicing rainbow within PR exon 6.

Factor	Consensus type	Score	Nuc start	Nuc end
hnRNP A2/B1	2x poly-T	6 T's (in 9)	113	124
	GTTTG	Exact match	64	67
	TTAG	Exact match	81	84
hnRNP C1/C2	TTAG	Exact match	81	84
hnRNP E1/E2	TTAG	Exact match	81	84
hnRNP H	TGTGG	Exact match	45	49
hnRNP I (PTB)	5x poly-Y	8 pyrimidines (in 10)	17	33
	5x poly-Y	8 pyrimidines (in 10)	94	105
	4x poly-Y	8 pyrimidines (in 10)	113	125
	TCTT	Exact match	121	124
hnRNP K	matrix 6-mer	5.798	54	59

Table 6.6: RNA binding protein binding sequences within PR exon 6.

Splicing Rainbow detected the presence of binding sites for RBPs with roles in splicing regulation within PR exon 6 which are not analysed by ESEfinder.

6.3.2 Knockdown of SR protein expression in MCF-7 cells

Potential binding sites for SRSF1, SRSF2, SRSF5 and SRSF6 were identified within PR exons 4 and 6 by the ESEfinder and Splicing Rainbow databases. Therefore, the effect of knocking down expression of these splicing factors on ASEs involving PR exons 4 and 6 was examined experimentally. MCF-7 breast cancer cells were treated with siRNA targeting each of these genes and cultured for 48 hours before harvesting and RNA extraction of RNA. RT-PCR was then performed using primers to detect each splicing factor mRNA and confirm successful knockdown (Figure 6.2). The siRNA transfections were performed, and protein knockdown confirmed by Western immunoblotting for each SR protein, by Dr H. Gautrey (unpublished data). Knockdown of *SRSF* mRNA and the effect on PR splicing are described below.

A substantial reduction of each splicing factor was observed in siRNA treated cells (Figure 6.2). For each gene, cells were treated with 2 different siRNAs targeting different regions of the gene to ensure efficient knockdown. siRNA knockdown of each of the genes appears to give a reduction in the level of SRSF1 mRNA detected (Figure 6.2a, lanes 1-8) compared to control samples; however the greatest effect was seen with the SRSF1 specific siRNA (Figure 6.2a, lanes 1 and 2). Only one of the SRSF2 siRNAs appears to give significant knockdown in the level of SRSF2 mRNA detected (Figure 6.2b, lane 4). No off target knockdowns of SFRS2 were observed. A slight reduction in SRSF5 mRNA was observed after SRSF5 siRNA treatment (Figure 6.2c, lane 5), and a reduction of SRSF6 was seen after SRSF6 siRNA treatment (Figure 6.2d, lanes 7 and 8). However, detection of SRSF5 mRNA was also greatly reduced when cells were treated with SFRS1 siRNA (Figure 6.2d, lanes 1 and 2). No reduction of any of the splicing factor genes was observed in cells treated with siRNA to knockdown GAPDH, an untargeted negative control siRNA, a vehicle only control or in untreated cells. GAPDH knockdown was seen in cells treated with GAPDH siRNA (Figure 6.2e lane 9).

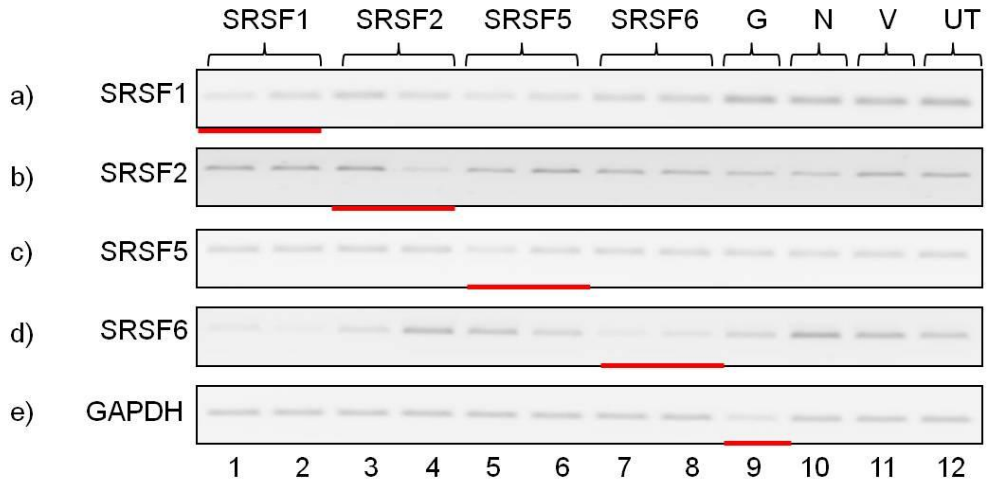


Figure 6.2: siRNA knockdown of splicing factor mRNA in MCF-7 cells.

MCF-7 cells in 12 well culture plates were treated with siRNA to knockdown expression of the splicing factors SRSF1 (lanes 1 and 2), SRSF2 (lanes 3 and 4), SRSF5 (lanes 5 and 6) and SRSF6 (lanes 7 and 8). Cells were also treated with control siRNAs and a vehicle only control. Untreated cells were included for comparison. G= positive transfection control GAPDH siRNA (knockdown confirmed by PCR using GAPDH primers, highlighted by red underlining), lane 9; N= Negative control untargeted siRNA, lane 10; V= Vehicle (transfection agent) only control, lane 11; U= Untreated cells, lane 12. The splicing factor targeted by the siRNA is indicated along the top and the mRNA detected by the PCR primers used is indicated beside each panel. Knockdown of each gene by the appropriate siRNA is highlighted by the red underlining of each panel. a**)** RT-PCR was performed using RNA from these cells and primers to detect SRSF1. **b)** RT-PCR to detect SRSF2. **c)** RT-PCR to detect SRSF5. **d)** RT-PCR to detect SRSF6. **e)** RT-PCR to detect GAPDH.

6.3.3 The effect of SR protein knockdown on alternative splicing of PR exon 4

The effect of depleting cells of SR proteins by siRNA knockdown on PR exon 4 ASEs was examined by PCR using primers spanning exons 3 to 5 and cDNA from MCF-7 cells in which knockdown of the splicing factors SRSF1, SRSF2, SRSF5 and SRSF6 had previously been demonstrated.

RT-PCR was performed using RNA from cells 48 hours post siRNA treatment with primers spanning exons 3 to 5. Levels of full length PCR product (Figure 6.3a) and exon 4 deleted (Δ4) PCR product (Figure 6.3b) were examined in cells in which the SR proteins had been knocked down and compared to control siRNA treated or untreated cells. Knockdown of GAPDH was observed in the positive transfection control cells, treated with GAPDH siRNA (Figure 6.3c, lane 9). Knockdown of SRSF1 (Figure 6.3a, lane 2) or SRSF5 (Figure 6.3a, lane 5) appeared to give a slightly increased level of full length PCR product compared to control samples. No effects on the levels of PCR product detected were observed after SRSF2 or SRSF6 knockdown (Figure 6.3a and b, lanes 3, 4 and 7, 8).

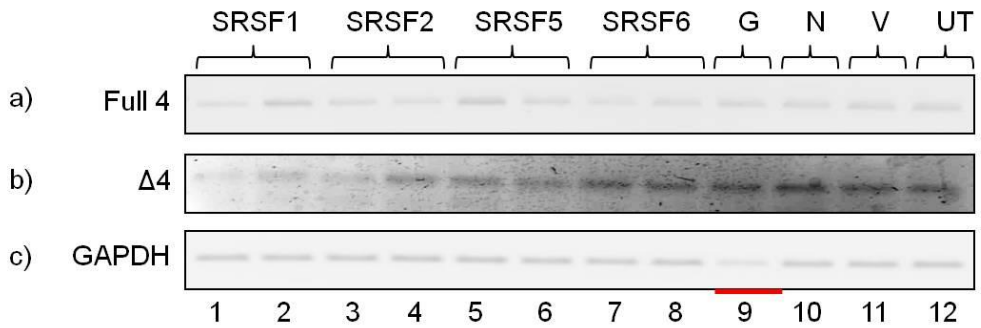


Figure 6.3: Effect of SR protein knockdown on PR exon 4 splicing after 48 hours.

Expression of SRSF1 (lanes 1 and 2), SRSF2 (lanes 3 and 4), SRSF5 (lanes 5 and 6) and SRSF6 (lanes 7 and 8) was knocked down by siRNA treatment in MCF-7 cells. Cells were harvested 48 hours after siRNA treatments. The effect of knockdown on the alternative splicing of PR exon 4 was examined by RT-PCR using primers spanning exons 3 to 5. The siRNA or control treatment is indicated along the top, and PCR product detected at the side of each panel. **(a)** Full length PCR product. **(b)** Exon 4 deleted PCR product. Levels of each PCR product in splicing factor knockdown cells were compared to control treated or untreated cells. G= GAPDH siRNA (knockdown confirmed by PCR using GAPDH primers **(c)**, underlined red), lane 9; N= Negative control siRNA, lane 10; V= Vehicle (transfection agent) only treated cells, lane 11; U= Untreated cells, lane 12.

Since little effect of splicing factor knockdown on PR exon 4 alternative splicing was observed 48 hours post siRNA treatment, siRNA treatments were repeated and cells cultured for 72 hours to allow downstream effects of the loss of SR protein expression to be detected. SRSF1 and SRSF5 knockdown did not change the levels of PCR products detected after 72 hours in MCF-7 cells by this assay (Figure 6.4a and b, lanes 1, 2 and 5, 6). In cells treated with the second SRSF2 siRNA, which produced the greatest reduction of SRSF2 mRNA after 48 hours (Figure 6.2b, lane 4), a reduced level of full length PR PCR product (Figure 6.4a, lane 4), and an increased level of PRΔ4 PCR product are detected (Figure 6.4b, lane 4). This indicates that SRSF2 may promote the inclusion of PR exon 4. Knockdown of SRSF6 appears to reduce the level of full length PR mRNA (Figure 6.4a, lanes 7 and 8), but no effect on PRΔ4 mRNA is observed in this assay (Figure 6.4b, lanes 7 and 8).

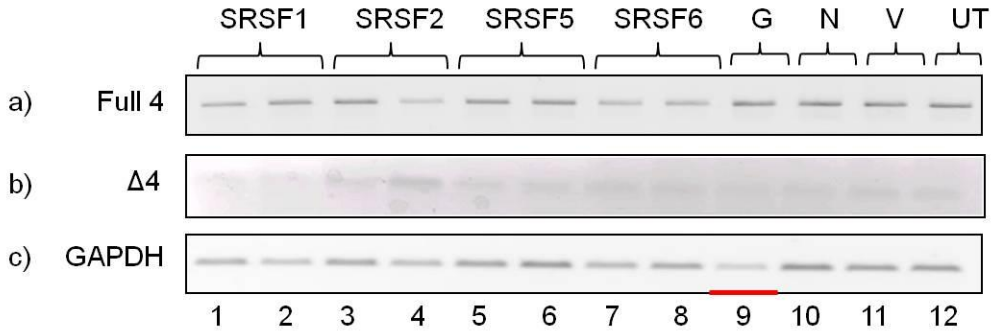


Figure 6.4: Effect of SR protein knockdown on PR exon 4 splicing after 72 hours.

Expression of SRSF1 (lanes 1 and 2), SRSF2 (lanes 3 and 4), SRSF5 (lanes 5 and 6) and SRSF6 (lanes 7 and 8) was knocked down by siRNA treatment in MCF-7 cells. Cells were harvested 72 hours after siRNA treatments. The effect of knockdown on the alternative splicing of PR exon 4 was examined by RT-PCR using primers spanning exons 3 to 5. The siRNA or control treatment is indicated along the top, and PCR product detected at the side of each panel. **(a)** Full length PCR product. **(b)** Exon 4 deleted PCR product. Levels of each PCR product in splicing factor knockdown cells were compared to control treated or untreated cells. G= GAPDH siRNA (knockdown confirmed by PCR using GAPDH primers **(c)**, underlined red), lane 9; N= Negative control siRNA, lane 10; V= Vehicle (transfection agent) only treated cells, lane 11; U= Untreated cells, lane 12.

6.3.4 The effect of SR protein knockdown on alternative splicing of PR exon 6

The effect of SR protein knockdown on PR exon 6 alternative splicing was examined by PCR using primers spanning exons 5 to 7 and MCF-7 cell cDNA 48 hours after siRNA treatments. SRSF1 knockdown appeared to slightly reduce the level of exon 6 deleted mRNA (PRΔ6) detected in this system (Figure 6.5b, lanes 1 and 2). SRSF2 knockdown with the second siRNA reduced full length PR mRNA (Figure 6.5a, lane 4). SRSF6 knockdown appeared to slightly increase the level of PRΔ6 mRNA (Figure 6.5b, lanes 7 and 8). Similarly to exon 4 splicing, no effect was seen with SRSF5 knockdown relative to control samples.

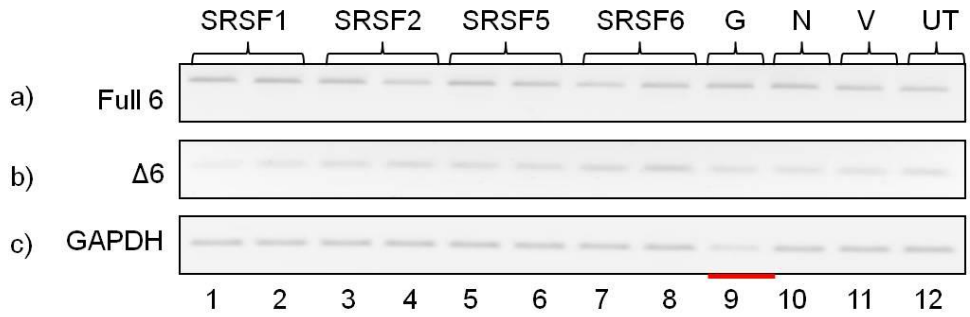


Figure 6.5: Effect of SR protein knockdown on PR exon 6 splicing after 48 hours.

Expression of SRSF1 (lanes 1 and 2), SRSF2 (lanes 3 and 4), SRSF5 (lanes 5 and 6) and SRSF6 (lanes 7 and 8) was knocked down by siRNA treatment in MCF-7 cells. Cells were harvested 48 hours after siRNA treatments. The effect of knockdown on the alternative splicing of PR exon 6 was examined by RT-PCR using primers spanning exons 5 to 7. The siRNA or control treatment is indicated along the top, and PCR product detected at the side of each panel. **(a)** Full length PCR product. **(b)** Exon 6 deleted PCR product. Levels of each PCR product in splicing factor knockdown cells were compared to control treated or untreated cells. G= GAPDH siRNA (knockdown confirmed by PCR using GAPDH primers **(c)**, underlined red), lane 9; N= Negative control siRNA, lane 10; V= Vehicle (transfection agent) only treated cells, lane 11; U= Untreated cells, lane 12.

Since a greater effect of splicing factor knockdown on PR exon 4 splicing was observed after 72 hours, PCR was performed to further examine the effect of SR protein knockdown on exon 6 splicing using the primers spanning exons 5 to 7 and MCF-7 cDNA from cells cultured for 72 hours post siRNA treatment. Reduced levels of Δ6 PCR product were detected in MCF-7 cells after SRSF1 knockdown (Figure 6.6b, lanes 1 and 2). Knockdown of SRSF2, SRSF5 and SRSF6 appeared to lead to increased detection of Δ6 PCR product (Figure 6.6b lanes 3-8) relative to control samples. This increase was particularly evident after SRSF6 knockdown (Figure 6.6b, lane 7) and this SRSF6 knockdown also appeared to reduce detection of full length PCR product (Figure 6.6a, lane 7) compared to control samples and samples treated with other SRSF siRNAs.

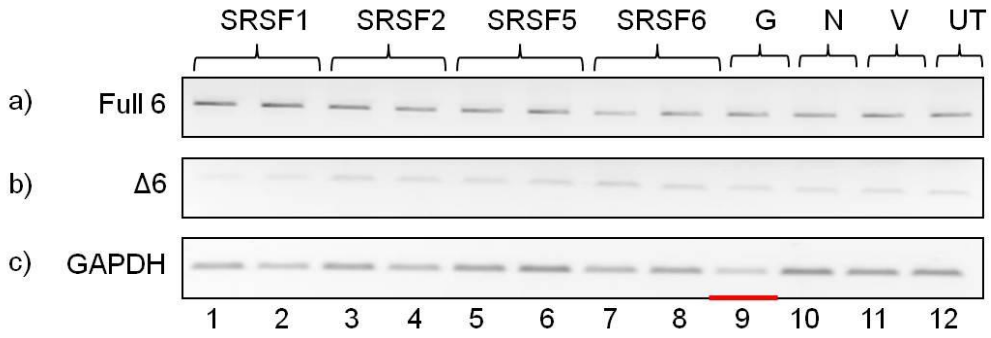


Figure 6.6: Effect of SR protein knockdown on PR exon 6 splicing after 72 hours.

Expression of SRSF1 (lanes 1 and 2), SRSF2 (lanes 3 and 4), SRSF5 (lanes 5 and 6) and SRSF6 (lanes 7 and 8) was knocked down by siRNA treatment in MCF-7 cells. Cells were harvested 48 hours after siRNA treatments. The effect of knockdown on the alternative splicing of PR exon 6 was examined by RT-PCR using primers spanning exons 5 to 7. The siRNA or control treatment is indicated along the top, and PCR product detected at the side of each panel. **(a)** Full length PCR product. **(b)** Exon 6 deleted PCR product. Levels of each PCR product in splicing factor knockdown cells were compared to control treated or untreated cells. G= GAPDH siRNA (knockdown confirmed by PCR using GAPDH primers **(c)**, underlined red), lane 9; N= Negative control siRNA, lane 10; V= Vehicle (transfection agent) only treated cells, lane 11; U= Untreated cells, lane 12.

6.4 Discussion

Analysis of the PR exon 4 sequence identified a putative 5′ splice site GU, 3′ AG, weak PPT and a consensus BP in addition to a cryptic 3′ AG within the exon, 126 nucleotides upstream of the 3′ splice site of intron 3 (Figure 6.7). Binding motifs for SRSF1, SRSF2, SRSF5 and SRSF6 were identified within PR exon 4. A high number of putative ESE sequences were identified using ESEfinder close to the 5′ end of exon 4, potentially promoting alternative splicing of this exon (Figure 6.1a). Potential binding motifs for SRSF2 and to a lesser extent SRSF1 were particularly prevalent at the 5′ end of exon 4. Potential SRSF6 binding motifs were located throughout the sequence and were prevalent in proximity to the 3′ end of the exon. Additional motifs for these SR proteins were identified using Splicing Rainbow (Table 6.1). Discrepancies between the two analyses may be attributed to the different methods used by the two programs to identify motifs and also to the different thresholds for reporting a sequence match to a binding motif. The threshold in ESEfinder was set to report only strong consensus matches whereas Splicing Rainbow reports all potential motifs.

gtcg aaaatttaaa aagttcaata aagtcagagt tgtgagagca ctggatgctg ttgctctccc
acagccagtg ggcgttccaa atgaaagcca agccctaagc cagagattca ctttttcacc aggtcaagac
atacagttga ttccaccact gatcaacctg ttaatgagca ttgaaccaga tgtgatctat gcaggacatg
acaacacaaa acctgacacc tccagttctt tgctgacaag tcttaatcaa ctaggcgaga ggcaacttct
ttcagtagtc aagtggtcta aatcattgcc ag

Figure 6.7: PR exon 4 cDNA sequence with potential splice site sequences identified.

5′ splice site GT and 3′ splice site AG are highlighted in red, cryptic 3′ splice site highlighted by the yellow background. PPT highlighted in blue font. BPS underlined with conserved adenosine in bold font.

Exon 4 deleted PR mRNA was detected in MDA-MB-231 cells (Section 3.3.1) and in MCF-7 cells (data not shown). Expression of SR proteins in MCF-7 and MDA-MB-231 cells has previously been profiled in our laboratory: SRSF1, SRSF5 and SRSF6 were all expressed in both cell lines (E. Hong unpublished data). Since both cell lines exhibited a similar profile of SR protein expression and both express PRΔ4 mRNA, MCF-7 cells which also express PRΔ6 mRNA were used for analysing the effect of splicing factor knockdown on PR alternative splicing. MCF-7 cells also express a higher level of PR mRNA and have been demonstrated in this thesis to be easily manipulated by transfection with siRNA (Section 3.3.6).

siRNA knockdown of SRSF1, SRSF2, SRSF5 and SRSF6 was achieved in the MCF-7 breast cancer cell line (Figure 6.2). Knockdown was assessed after 48 hours demonstrating reduced levels of each SRSF gene following targeted siRNA treatment. In addition to the specific knockdown targeted to each SRSF gene, each siRNA, especially that which targeted SRSF5, appeared to reduce expression of SRSF1 mRNA. Knockdown of SRSF1 additionally appeared to reduce levels of SRSF6 mRNA. The sequence identity between different SR proteins (Shepard and Hertel, 2009), and thus the similarity in gene sequence may explain the off target effects of siRNAs targeted to individual SRSF genes.

The effect of SR protein knockdown on PR exon 4 alternative splicing was assessed after 48 hours (Figure 6.3) and 72 hours (Figure 6.4) by RT-PCR using primers spanning exons 3 to 5. A greater effect of SRSF knockdown was observed after 72 hours, attributable to the need for degradation of existing proteins to see an effect from reduced gene expression. After 48 hours SRSF1 and SRSF5 siRNA appeared to result in a slight increase in full length PR mRNA (Figure 6.3). SRSF5 siRNA appeared to give the smallest specific knockdown and resulted in a much greater level of SRSF1 knockdown than SRSF5 relative to controls (Figure 6.2). Therefore the effects of both sets of siRNA could be due to the reduction in SRSF1 expression.

After 72 hours, knockdown of SRSF2 using the second siRNA appeared to reduce levels of full length PR (Figure 6.4a, lane 4) and increase levels of exon 4 deleted

(PRΔ4) mRNA (Figure 6.4b, lane 4). Figure 4.2b demonstrates that efficient knockdown of SRSF2 was only achieved using the second siRNA (lane 4). Therefore, this preliminary data suggests that SRSF2 may promote the inclusion of PR exon 4. A high number of potential SRSF2 binding motifs were identified by ESEfinder and Splicing Rainbow at the 5′ end of exon 4, suggesting that SRSF2 binding in this region may promote the use of the 3′ splice site of intron 3 upstream of exon 4. Similarly SRSF6 knockdown reduced full length PR mRNA levels, implicating this splicing factor in exon 4 inclusion and again numerous potential binding motifs for this splicing factor were identified by ESEfinder or Splicing Rainbow at the 5′ end of exon 4.

The majority of the high scoring SR protein binding motifs detected by ESEfinder in PR exon 6 were for SRSF6 (Figure 6.1) and knockdown of this splicing factor appeared to both decrease levels of full length mRNA detected by RT-PCR using primers spanning exons 5 to 7, and increase the levels of the truncated exon 6 deleted (PRΔ6) mRNA (Figure 4.6). High scoring potential binding motifs for SRSF1 or SRSF2 were identified by either ESEfinder or Splicing Rainbow (Figure 6.1 and Table 6.4). SRSF1 knockdown appeared to slightly decrease the level of PRΔ6 mRNA detected suggesting that this splicing factor promotes exon skipping, while SRSF2 knockdown decreased full length PR mRNA levels suggesting a role in promoting exon inclusion (Figure 6.6).

The knockdown of SRSF mRNA was assessed by semi-quantitative RT-PCR whereby PCR reactions are performed for a cycle number which is optimised to allow comparison of PCR products in the logarithmic phase of amplification. This method allows identification of a reduction in mRNA levels following siRNA treatment compared to controls, although absolute quantification is not possible. Absolute quantification of the level of knockdown could be achieved by qRT-PCR but is not necessary in this instance. Analysis of the effect of SRSF knockdown on PR mRNA and alternative splicing was again assessed by semi-quantitative RT-PCR. Whilst the cycle numbers used for this PCR appear high, the relatively low level of PR mRNA in MDA-MB-231 cells means that the amplifications are still in the logarithmic phase. The use of RT-PCR rather than alternative methods for analysis of exon skipping is described in Section 3.4. Splice specific primers to detect just the Δ4 or Δ6 mRNA could be used to quantify the changes in expression following SRSF knockdown and is an experiment

which could be performed when this preliminary study is expanded. However, SR proteins other roles in addition to functioning as splicing factors, including an involvement in preventing RNA polymerase stalling and promoting transcriptional elongation (Long and Caceres, 2009; Moore and Proudfoot, 2009). Knockdown of SRSF6 led to a reduction in full length mRNA detected using primers spanning exon 3-exon 5 (Figure 6.4a, lanes 7 and 8). However, no parallel increase in Δ4 mRNA was detected (Figure 6.4b, lanes 7 and 8) suggesting that this could be a transcriptional effect, rather than the splicing effect discussed above. This effect would not have been identified using splice specific primers for qRT-PCR, further justifying the use of RT-PCR and agarose gel electrophoresis in this chapter.

Potentially antagonistic roles for SR proteins in the regulation of the alternatively spliced cassette exons 4 and 6 of PR have been described. SRSF1 appears to promote exon 4 skipping whereas SRSF2 may promote exon 4 inclusion, and SRSF1 appears to promote exon 6 skipping whereas SRSF2 and SRSF6 may promote exon inclusion. Antagonistic functions of SR proteins have previously been reported (Solis et al., 2008; Gallego et al., 1997) and thus the balance of SRSF1 and SRSF2 expression, as well as SRSF6 may influence the alternative splicing of PR exon 6. Whilst SR proteins are usually associated with exon inclusion, roles for this family of proteins in exon skipping have been reported; for example, SRSF1 promotes the skipping of SRSF3 exon 4 (Jumaa and Nielsen, 1997). Furthermore, the presence of overlapping binding sites for antagonistic SR proteins and other splicing factors in both exons (Figure 6.1 and Tables 6.1-6.6) suggests that the balance of expression levels of different splicing factors competing for the same sites could regulate PR exon inclusion. Splicing regulatory elements consisting of overlapping and competing binding motifs have also previously been described. For example composite exonic regulatory elements of splicing (CERES) have been reported, consisting of overlapping enhancer and silencer elements which bind different splicing factors (Pagani et al., 2003b). The CERES element within *CFTR* exon 12 is capable of binding both SR proteins and hnRNPs, and mutations of the element alter the pattern of exon inclusion depending on the splicing factor binding which is disrupted (Haque et al., 2010).

The potential impact of the off target knockdown of SRSF1 by SRSF5 siRNA has been discussed above. The differing effect on PR exon 4 and exon 6 alternative splicing of other siRNAs targeting individual SR proteins suggests that the off target knockdown may not be sufficient to affect the overall pattern of splicing. For example knockdown of SRSF1 also appeared to reduce SRSF6 mRNA levels, however knockdown of SRSF1 appears to have the opposite effect on PR exon 6 splicing to knockdown of SRSF6 (Figure 6.6). Off target knockdown may reduce the effect observed with each siRNA but does not appear to completely mask the effects in most cases.

The results presented in this chapter represent a preliminary investigation into the regulation of alternative splicing of the cassette exons 4 and 6 of PR, implicating different combinations of SR proteins in the regulation of each ASE. A more comprehensive study of the splicing regulation will also investigate the roles of other splicing factors, such as hnRNPs which are known to exhibit functions antagonistic to SR proteins (Caceres et al., 1994). The presence of splicing regulatory elements within the intronic regions flanking the exons, as well as the ability of SR proteins and other splicing factors to bind to specific regulatory sequences identified within the alternatively spliced PR exons, must also be considered. Details of future investigations into the regulation of PR alternative splicing will be described in Section 7.2.

7 Summary and future work

7.1 Summary

In Chapter 3 of this thesis an RT-PCR based PR gene walking assay was used to characterise PR expression in breast cancer cell lines. PR mRNA was detected both in MCF-7 cells and in reportedly PR negative MDA-MB-231 cells. In MDA-MB-231 cells PR alternative splicing events were detected which result in deletions from within exon 1 and spanning exon 1 to exon 2, leading to a change in translational reading frame. A variant mRNA, termed Δp1c, was detected with an in frame deletion within exon 1, deleting AF1 encoded by this region from a potential protein. Deletions of exon 2 or exon 3 were also detected; each of these exons encodes one zinc finger of the DBD and deletion of exon 3 would encode a protein lacking just this portion of the DBD, whilst exon 2 deletion would cause a change in reading frame, generating a PTC within exon 3. A deletion of part of exon 4, encoding the hinge and part of the LBD was observed in MDA-MB-231 cells which would cause a change in reading frame and truncation within the LBD. A deletion of the entire cassette exon 4 was also observed which would generate a protein with DNA binding ability but lacking the NLS and part of the LBD. Several deletions affecting regions within exons 4 to 6 were also observed which would cause a change in reading frame and disruption of the LBD. mRNA containing the intronic exon M was also detected in MDA-MB-231 cells, which potentially encodes an N-terminally truncated protein also lacking the DBD. mRNA for the putative N-terminally truncated isoforms PR-S and PR-T were additionally detected in MCF-7 cells, along with PR-i45 mRNA resulting from retention of intronic exons between exons 4 and 5 which truncate the C-terminal end of the protein.

MDA-MB-231 cells appear PR negative by Western immunoblotting using N-terminal PR antibodies. However, using the C-terminal antibody C19 a range of proteins are detected in this cell line by Western immunoblotting, suggesting expression of N-terminally truncated alternatively spliced variants in this cell line. siRNA knockdown of PR expression in MCF-7 cells, followed by Western immunoblotting using C19 confirms that this antibody does detect truncated PR proteins; reduced expression of proteins of 80kDa, 55kDa and a doublet between 35-40kDa were seen following siRNA transfection. It was further demonstrated that these proteins are not the products of

proteolytic degradation of full length PR and that the range of protein sizes detected is not due to differential phosphorylation of a single isoform. However, due to non-specific binding by C19, this antibody does not appear suitable for analysis of PR protein expression by immunofluorescent staining. Therefore, novel and more specific PR antibodies which detect N-terminally truncated PR spliced variants are required.

In Chapter 4, a series of functional assays were used to investigate the potential of truncated proteins detected by C19 to function as PR. Progesterone binding, dimerisation, interaction with the PR nuclear co-factor PSF and DNA binding were investigated. Potentially functional proteins were detected in MDA-MB-231 cells; a 35kDa nuclear protein could bind progesterone, PSF and PRE, an 80kDa protein could bind progesterone and PSF, and a 100kDa protein could bind progesterone and PRE. Cytoplasmic proteins were also detected in MDA-MB-231 cells with functional capabilities similar to PR-A/PR-B; a protein of approximately 50-55kDa could bind progesterone and PRE, whilst an approximately 42kDa progesterone binding protein was also detected by C19. High molecular weight protein bands were detected by C19 following Co-IP with the PR antibody C20, representing potential denaturation resistant PR heterodimers, suggesting that the truncated proteins detected by C19 may be capable of dimerisation. The detection of N-terminally truncated proteins by these assays suggest that the ASEs characterised in MDA-MB-231 cells may encode functional proteins. The molecular weights of the functional PR proteins identified in this chapter correspond to the molecular weights of PR proteins identified by C19 following siRNA knockdown of total PR expression in MCF-7 cells in Chapter 3.

Chapter 5 provides preliminary validation of the detection of alternatively spliced PR mRNA in breast tumour tissue using the PR gene walking assay. Fifteen frozen breast tumour samples were analysed by RT-PCR and alternatively spliced PR mRNA was detected in breast tumours which were originally characterised as both PR positive and PR negative. A deletion of the end of exon 1 and the beginning of exon 2, Δp1,p2 which was observed in MDA-MB-231 cells, was seen in a number of breast tumours. Exon 4 deleted PR mRNA was also detected in several breast tumours with an apparent correlation between the level of Δ4 mRNA relative to full length mRNA and tumour grade; in DCIS more Δ4 was present than full length, whereas in high grade invasive

tumours less Δ4 mRNA was present relative to full length. Two partial exon deletions were seen in breast tumour tissue which were not detected in MDA-MB-231 cells, resulting from use of cryptic 3′ splice sites within exon 4 ($\Delta p4_{126}$) and exon 6 (Δp6.2). $\Delta p4_{126}$ would lead to deletion of the hinge and NLS but leave the LBD intact, whilst Δp6.2 would change the translational reading frame and truncate the LBD. Deletion of the whole of exon 6 was also observed in breast tumour tissue, a deletion which would remove part of the LBD but not change the reading frame.

The data presented in Chapter 3 of this thesis supports existing literature (Samalecos and Gellersen, 2008; Madsen et al., 2007) and validates the hypothesis that commercial non N-terminal PR antibodies detect non-PR proteins and are therefore unsuitable for PR screening by immunohistochemistry. Therefore, novel PR antibodies PRA3, targeted to the DBD, and PRA4, targeted to the C-terminal end of the LBD, were generated and validated in Chapter 5 of this study. PRA3 appeared in preliminary Western immunoblotting, Co-IP and siRNA experiments to detect PR proteins whereas in these initial experiments PRA4 did not. Further purification and validation of PRA4 is therefore necessary. The novel antibodies were optimised for immunohistochemical analysis of PR expression in FFPE breast tumour sections; again PRA4 did not satisfactorily detect PR. PRA3 was used to re-score nuclear PR expression in fourteen breast tumour samples which had been assessed using an N-terminal antibody, confirming the original PR status in twelve of the fourteen samples analysed, confirming the specificity of this novel antibody. One sample originally scored PR negative but appeared PR positive after staining using PRA3, validating the hypothesis that N-terminal antibodies may be unable to detect some PR proteins expressed in breast tumour tissue. The final tumour was originally scored PR positive but appeared PR negative using PRA3, demonstrating that whilst non N-terminal PR antibodies may detect additional proteins, internal exon deletions may make it difficult to target an epitope which is present in all isoforms. Further analysis of PR alternative splicing in a large cohort of breast tumour specimens would be necessary to identify cancer specific alternative splicing events which could be used to target antibodies to specific isoforms.

Chapter 6 of this thesis presents a preliminary investigation into the regulation of PR alternative splicing, focussing on exon 4 which was alternatively spliced in both MDA-

MB-231 cells and in breast tumour tissue, and exon 6, alternative splicing of which was observed specifically in breast tumour tissue in this study. siRNA targeting genes encoding the SR proteins SRSF1, SRSF2, SRSF5 and SRSF6 were used to investigate the role of these splicing factors in PR exon 4 and exon 6 inclusion. SRSF1 appears to promote the skipping of both exon 4 and exon 6, whereas SRSF2 which is known to function antagonistically to SRSF1 appears to promote inclusion of both exons. Similarly SRSF6 may promote the inclusion of PR exon 6.

7.2 Future work

Future work based on the findings from this study will focus on:

- Investigating a role for truncated PR proteins in rapid signalling pathways.
- Defining the structure of the functional low molecular weight PR proteins detected in this study.
- Re-purification and further validation of novel PR antisera generated in this study with the aim of developing novel monoclonal antibodies for laboratory use and clinical PR screening.
- Analysis of PR expression in a large cohort of breast tumour and normal breast tissue specimens to identify any cancer specific PR ASEs which may provide targets for novel antibodies to detect cancer specific protein isoforms.

The investigation into PR alternative splicing regulation will be extended to examine other ASEs identified in this study. Future work will include:

- An investigation into the role of splicing factors in regulating the PR ASEs identified in this study using gene knockdown, protein:RNA binding studies and mutation analyses.
- Analysis of the function of exon deleted PR proteins overexpressed in a PR negative cell line, investigating the effects on gene expression and rapid signalling.

8 References

Abdel-Hafiz, H., Dudevoir, M. L. and Horwitz, K. B. (2009) 'Mechanisms underlying the control of progesterone receptor transcriptional activity by SUMOylation', *J Biol Chem,* 284, (14), pp. 9099-108.

Abdel-Hafiz, H., Takimoto, G. S., Tung, L. and Horwitz, K. B. (2002) 'The inhibitory function in human progesterone receptor N termini binds SUMO-1 protein to regulate autoinhibition and transrepression', *J Biol Chem,* 277, (37), pp. 33950-6.

Albert, J. M., Gonzalez-Angulo, A. M., Guray, M., Sahin, A., Tereffe, W., Woodward, W. A., Strom, E. A., Hunt, K. K., Tucker, S. L. and Buchholz, T. A. (2011) 'Patients with only 1 positive hormone receptor have increased locoregional recurrence compared with patients with estrogen receptor-positive progesterone receptor-positive disease in very early stage breast cancer', *Cancer,* 117, (8), pp. 1595-601.

Alvarez, R. H. (2010) 'Present and future evolution of advanced breast cancer therapy', *Breast Cancer Res,* 12 Suppl 2, pp. S1.

Anderson, H., Hills, M., Zabaglo, L., A'Hern, R., Leary, A. F., Haynes, B. P., Smith, I. E. and Dowsett, M. (2011) 'Relationship between estrogen receptor, progesterone receptor, HER-2 and Ki67 expression and efficacy of aromatase inhibitors in advanced breast cancer', *Ann Oncol*.

Arpino, G., Weiss, H., Lee, A. V., Schiff, R., De Placido, S., Osborne, C. K. and Elledge, R. M. (2005) 'Estrogen receptor-positive, progesterone receptor-negative breast cancer: association with growth factor receptor expression and tamoxifen resistance', *J Natl Cancer Inst,* 97, (17), pp. 1254-61.

Aupperlee, M. D. and Haslam, S. Z. (2007) 'Differential hormonal regulation and function of progesterone receptor isoforms in normal adult mouse mammary gland', *Endocrinology,* 148, (5), pp. 2290-300.

Aupperlee, M. D., Smith, K. T., Kariagina, A. and Haslam, S. Z. (2005) 'Progesterone receptor isoforms A and B: temporal and spatial differences in expression during murine mammary gland development', *Endocrinology,* 146, (8), pp. 3577-88.

Bachman, K. E., Argani, P., Samuels, Y., Silliman, N., Ptak, J., Szabo, S., Konishi, H., Karakas, B., Blair, B. G., Lin, C., Peters, B. A., Velculescu, V. E. and Park, B. H. (2004) 'The PIK3CA gene is mutated with high frequency in human breast cancers', *Cancer Biol Ther,* 3, (8), pp. 772-5.

Bain, D. L., Franden, M. A., McManaman, J. L., Takimoto, G. S. and Horwitz, K. B. (2000) 'The N-terminal region of the human progesterone A-receptor. Structural analysis and the influence of the DNA binding domain', *J Biol Chem,* 275, (10), pp. 7313-20.

Bain, D. L., Franden, M. A., McManaman, J. L., Takimoto, G. S. and Horwitz, K. B. (2001) 'The N-terminal region of human progesterone B-receptors: biophysical and biochemical comparison to A-receptors', *J Biol Chem,* 276, (26), pp. 23825-31.

Balleine, R. L., Earl, M. J., Greenberg, M. L. and Clarke, C. L. (1999a) 'Absence of progesterone receptor associated with secondary breast cancer in postmenopausal women', *Br J Cancer,* 79, (9-10), pp. 1564-71.

Balleine, R. L., Hunt, S. M. and Clarke, C. L. (1999b) 'Coexpression of alternatively spliced estrogen and progesterone receptor transcripts in human breast cancer', *J Clin Endocrinol Metab,* 84, (4), pp. 1370-7.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J. and Frey, B. J. (2010) 'Deciphering the splicing code', *Nature,* 465, (7294), pp. 53-9.

Bardou, V. J., Arpino, G., Elledge, R. M., Osborne, C. K. and Clark, G. M. (2003) 'Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases', *J Clin Oncol,* 21, (10), pp. 1973-9.

Bateman, A. C. (2006) 'Pathology of breast cancer', *Women's Health Medicine,* 3, (1), pp. 18-21.

Bates, D. O., Cui, T. G., Doughty, J. M., Winkler, M., Sugiono, M., Shields, J. D., Peat, D., Gillatt, D. and Harper, S. J. (2002) 'VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma', *Cancer Res,* 62, (14), pp. 4123-31.

Beatson, G. T. (1983) 'On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment, with illustrative cases', *CA: A Cancer Journal for Clinicans,* 33, (2), pp. 108-121.

Beleut, M., Rajaram, R. D., Caikovski, M., Ayyanan, A., Germano, D., Choi, Y., Schneider, P. and Brisken, C. (2010) 'Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland', *Proc Natl Acad Sci U S A,* 107, (7), pp. 2989-94.

Bemmo, A., Dias, C., Rose, A. A., Russo, C., Siegel, P. and Majewski, J. (2010) 'Exon-level transcriptome profiling in murine breast cancer reveals splicing changes specific to tumors with different metastatic abilities', *PLoS One,* 5, (8), pp. e11981.

Beral, V. (2003) 'Breast cancer and hormone-replacement therapy in the Million Women Study', *Lancet,* 362, (9382), pp. 419-27.

Berget, S. M. (1995) 'Exon recognition in vertebrate splicing', *J Biol Chem,* 270, (6), pp. 2411-4.

Berglund, J. A., Abovich, N. and Rosbash, M. (1998) 'A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition', *Genes Dev,* 12, (6), pp. 858-67.

Berry, D. A., Cirrincione, C., Henderson, I. C., Citron, M. L., Budman, D. R., Goldstein, L. J., Martino, S., Perez, E. A., Muss, H. B., Norton, L., Hudis, C. and Winer, E. P. (2006) 'Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer', *JAMA*, 295, (14), pp. 1658-67.

Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. and Luhrmann, R. (2008) 'Isolation of an active step I spliceosome and composition of its RNP core', *Nature*, 452, (7189), pp. 846-50.

Blakely, C. M., Stoddard, A. J., Belka, G. K., Dugan, K. D., Notarfrancesco, K. L., Moody, S. E., D'Cruz, C. M. and Chodosh, L. A. (2006) 'Hormone-induced protection against mammary tumorigenesis is conserved in multiple rat strains and identifies a core gene expression signature induced by pregnancy', *Cancer Res*, 66, (12), pp. 6421-31.

Blencowe, B. J., Ahmad, S. and Lee, L. J. (2009) 'Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes', *Genes Dev*, 23, (12), pp. 1379-86.

Bockmuhl, Y., Murgatroyd, C. A., Kuczynska, A., Adcock, I. M., Almeida, O. F. and Spengler, D. (2011) 'Differential Regulation and Function of 5'-Untranslated GR-Exon 1 Transcripts', *Mol Endocrinol*, 25, (7), pp. 1100-10.

Boise, L. H., Gonzalez-Garcia, M., Postema, C. E., Ding, L., Lindsten, T., Turka, L. A., Mao, X., Nunez, G. and Thompson, C. B. (1993) 'bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death', *Cell*, 74, (4), pp. 597-608.

Boonyaratanakornkit, V., McGowan, E., Sherman, L., Mancini, M. A., Cheskis, B. J. and Edwards, D. P. (2007) 'The role of extranuclear signaling actions of progesterone receptor in mediating progesterone regulation of gene expression and the cell cycle', *Mol Endocrinol*, 21, (2), pp. 359-75.

Boonyaratanakornkit, V., Scott, M. P., Ribon, V., Sherman, L., Anderson, S. M., Maller, J. L., Miller, W. T. and Edwards, D. P. (2001) 'Progesterone receptor contains a proline-rich motif that directly interacts with SH3 domains and activates c-Src family tyrosine kinases', *Mol Cell*, 8, (2), pp. 269-80.

Bourdon, J. C., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., Xirodimas, D. P., Saville, M. K. and Lane, D. P. (2005) 'p53 isoforms can regulate p53 transcriptional activity', *Genes Dev*, 19, (18), pp. 2122-37.

breastcancer.org *What is Breast Cancer?* Available at: http://www.breastcancer.org/symptoms/understand_bc/what_is_bc.jsp (Accessed: 04/07/2011).

breastcancer.org(b) *Diagnosis of DCIS*. Available at: http://www.breastcancer.org/symptoms/types/dcis/diagnosis.jsp (Accessed: 04/07/2011).

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S. K., McMahon, J. A., McMahon, A. P. and Weinberg, R. A. (2000) 'Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling', *Genes Dev,* 14, (6), pp. 650-4.

Brisken, C., Park, S., Vass, T., Lydon, J. P., O'Malley, B. W. and Weinberg, R. A. (1998) 'A paracrine role for the epithelial progesterone receptor in mammary gland development', *Proc Natl Acad Sci U S A,* 95, (9), pp. 5076-81.

Buckanovich, R. J. and Darnell, R. B. (1997) 'The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo', *Mol Cell Biol,* 17, (6), pp. 3194-201.

Burstein, H. J., Keshaviah, A., Baron, A. D., Hart, R. D., Lambert-Falls, R., Marcom, P. K., Gelman, R. and Winer, E. P. (2007) 'Trastuzumab plus vinorelbine or taxane chemotherapy for HER2-overexpressing metastatic breast cancer: the trastuzumab and vinorelbine or taxane study', *Cancer,* 110, (5), pp. 965-72.

Caceres, J. F. and Kornblihtt, A. R. (2002) 'Alternative splicing: multiple control mechanisms and involvement in human disease', *Trends Genet,* 18, (4), pp. 186-93.

Caceres, J. F., Stamm, S., Helfman, D. M. and Krainer, A. R. (1994) 'Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors', *Science,* 265, (5179), pp. 1706-9.

Caldon, C. E., Sutherland, R. L. and Musgrove, E. (2010) 'Cell cycle proteins in epithelial cell differentiation: implications for breast cancer', *Cell Cycle,* 9, (10), pp. 1918-28.

Calza, S., Hall, P., Auer, G., Bjohle, J., Klaar, S., Kronenwett, U., Liu, E. T., Miller, L., Ploner, A., Smeds, J., Bergh, J. and Pawitan, Y. (2006) 'Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients', *Breast Cancer Res,* 8, (4), pp. R34.

Cargnello, M. and Roux, P. P. (2011) 'Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases', *Microbiol Mol Biol Rev,* 75, (1), pp. 50-83.

Carnevale, R. P., Proietti, C. J., Salatino, M., Urtreger, A., Peluffo, G., Edwards, D. P., Boonyaratanakornkit, V., Charreau, E. H., Bal de Kier Joffe, E., Schillaci, R. and Elizalde, P. V. (2007) 'Progestin effects on breast cancer cell proliferation, proteases activation, and in vivo development of metastatic phenotype all depend on progesterone receptor capacity to activate cytoplasmic signaling pathways', *Mol Endocrinol,* 21, (6), pp. 1335-58.

Cartegni, L. and Krainer, A. R. (2003) 'Correction of disease-associated exon skipping by synthetic exon-specific activators', *Nat Struct Biol,* 10, (2), pp. 120-5.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. and Krainer, A. R. (2003) 'ESEfinder: A web resource to identify exonic splicing enhancers', *Nucleic Acids Res,* 31, (13), pp. 3568-71.

Cerliani, J. P., Guillardoy, T., Giulianelli, S., Vaque, J. P., Gutkind, J. S., Vanzulli, S. I., Martins, R., Zeitlin, E., Lamb, C. A. and Lanari, C. (2011) 'Interaction between FGFR-2, STAT5, and Progesterone Receptors in Breast Cancer', *Cancer Res,* 71, (10), pp. 3720-31.

Chatterton, R. T., Jr., Lydon, J. P., Mehta, R. G., Mateo, E. T., Pletz, A. and Jordan, V. C. (2002) 'Role of the progesterone receptor (PR) in susceptibility of mouse mammary gland to 7,12-dimethylbenz[a]anthracene-induced hormone-independent preneoplastic lesions in vitro', *Cancer Lett,* 188, (1-2), pp. 47-52.

Chen, L. L., Sabripour, M., Wu, E. F., Prieto, V. G., Fuller, G. N. and Frazier, M. L. (2005) 'A mutation-created novel intra-exonic pre-mRNA splice site causes constitutive activation of KIT in human gastrointestinal stromal tumors', *Oncogene,* 24, (26), pp. 4271-80.

Cheng, C., Yaffe, M. B. and Sharp, P. A. (2006) 'A positive feedback loop couples Ras activation and CD44 alternative splicing', *Genes Dev,* 20, (13), pp. 1715-20.

Cheung, J. and Smith, D. F. (2000) 'Molecular chaperone interactions with steroid receptors: an update', *Mol Endocrinol,* 14, (7), pp. 939-46.

Chlebowski, R. T., Anderson, G. L., Gass, M., Lane, D. S., Aragaki, A. K., Kuller, L. H., Manson, J. E., Stefanick, M. L., Ockene, J., Sarto, G. E., Johnson, K. C., Wactawski-Wende, J., Ravdin, P. M., Schenken, R., Hendrix, S. L., Rajkovic, A., Rohan, T. E., Yasmeen, S. and Prentice, R. L. (2010) 'Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women', *JAMA,* 304, (15), pp. 1684-92.

Chlebowski, R. T., Kuller, L. H., Prentice, R. L., Stefanick, M. L., Manson, J. E., Gass, M., Aragaki, A. K., Ockene, J. K., Lane, D. S., Sarto, G. E., Rajkovic, A., Schenken, R., Hendrix, S. L., Ravdin, P. M., Rohan, T. E., Yasmeen, S. and Anderson, G. (2009) 'Breast cancer after use of estrogen plus progestin in postmenopausal women', *N Engl J Med,* 360, (6), pp. 573-87.

Chou, Y. C., Uehara, N., Lowry, J. R. and Shyamala, G. (2003) 'Mammary epithelial cells of PR-A transgenic mice exhibit distinct alterations in gene expression and growth potential associated with transformation', *Carcinogenesis,* 24, (3), pp. 403-9.

Clarke, R. B. (2005) 'Isolation and characterization of human mammary stem cells', *Cell Prolif,* 38, (6), pp. 375-86.

Clarke, R. B., Howell, A., Potten, C. S. and Anderson, E. (1997) 'Dissociation between steroid receptor expression and cell proliferation in the human breast', *Cancer Res,* 57, (22), pp. 4987-91.

Clemm, D. L., Sherman, L., Boonyaratanakornkit, V., Schrader, W. T., Weigel, N. L. and Edwards, D. P. (2000) 'Differential hormone-dependent phosphorylation of progesterone receptor A and B forms revealed by a phosphoserine site-specific monoclonal antibody', *Mol Endocrinol,* 14, (1), pp. 52-65.

Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T. and Chabot, B. (2008) 'Antagonistic effects of the SRp30c protein and cryptic 5' splice sites on the alternative splicing of the apoptotic regulator Bcl-x', *J Biol Chem,* 283, (31), pp. 21315-24.

Condon, J. C., Hardy, D. B., Kovaric, K. and Mendelson, C. R. (2006) 'Up-regulation of the progesterone receptor (PR)-C isoform in laboring myometrium by activation of nuclear factor-kappaB may contribute to the onset of labor through inhibition of PR function', *Mol Endocrinol,* 20, (4), pp. 764-75.

Conzen, S. D. (2008) 'Minireview: nuclear receptors and breast cancer', *Mol Endocrinol,* 22, (10), pp. 2215-28.

Cork, D. M., Lennard, T. W. and Tyson-Capper, A. J. (2008) 'Alternative splicing and the progesterone receptor in breast cancer', *Breast Cancer Res,* 10, (3), pp. 207.

Corless, C. L., McGreevey, L., Town, A., Schroeder, A., Bainbridge, T., Harrell, P., Fletcher, J. A. and Heinrich, M. C. (2004) 'KIT gene deletions at the intron 10-exon 11 boundary in GI stromal tumors', *J Mol Diagn,* 6, (4), pp. 366-70.

Cowper, A. E., Caceres, J. F., Mayeda, A. and Screaton, G. R. (2001) 'Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing', *J Biol Chem,* 276, (52), pp. 48908-14.

Coyle, Y. M., Xie, X. J., Hardy, D. B., Ashfaq, R. and Mendelson, C. R. (2007) 'Progesterone receptor expression is a marker for early stage breast cancer: implications for progesterone receptor as a therapeutic tool and target', *Cancer Lett,* 258, (2), pp. 253-61.

Creighton, C. J., Kent Osborne, C., van de Vijver, M. J., Foekens, J. A., Klijn, J. G., Horlings, H. M., Nuyten, D., Wang, Y., Zhang, Y., Chamness, G. C., Hilsenbeck, S. G., Lee, A. V. and Schiff, R. (2009) 'Molecular profiles of progesterone receptor loss in human breast tumors', *Breast Cancer Res Treat,* 114, (2), pp. 287-99.

Croce, C. M. (2008) 'Oncogenes and cancer', *N Engl J Med,* 358, (5), pp. 502-11.

CRUK *Breast Cancer Statistics*. Available at: http://info.cancerresearchuk.org/cancerstats/types/breast/ (Accessed: 04/07/2011).

Cui, X., Schiff, R., Arpino, G., Osborne, C. K. and Lee, A. V. (2005) 'Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy', *J Clin Oncol,* 23, (30), pp. 7721-35.

Daniel, A. R., Gaviglio, A. L., Czaplicki, L. M., Hillard, C. J., Housa, D. and Lange, C. A. (2010) 'The progesterone receptor hinge region regulates the kinetics of transcriptional responses through acetylation, phosphorylation, and nuclear retention', *Mol Endocrinol,* 24, (11), pp. 2126-38.

Daniel, A. R., Qiu, M., Faivre, E. J., Ostrander, J. H., Skildum, A. and Lange, C. A. (2007) 'Linkage of progestin and epidermal growth factor signaling: phosphorylation of progesterone receptors mediates transcriptional hypersensitivity and increased ligand-independent breast cancer cell growth', *Steroids,* 72, (2), pp. 188-201.

David, C. J. and Manley, J. L. (2010) 'Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged', *Genes Dev,* 24, (21), pp. 2343-64.

Davies, S., Dai, D., Wolf, D. M. and Leslie, K. K. (2004) 'Immunomodulatory and transcriptional effects of progesterone through progesterone A and B receptors in Hec50co poorly differentiated endometrial cancer cells', *J Soc Gynecol Investig,* 11, (7), pp. 494-9.

Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L. and Tindall, D. J. (2008) 'Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance', *Cancer Res,* 68, (13), pp. 5469-77.

Desai, A. J., Luqmani, Y. A., Walters, J. E., Coope, R. C., Dagg, B., Gomm, J. J., Pace, P. E., Rees, C. N., Thirunavukkarasu, V., Shousha, S., Groome, N. P., Coombes, R. and Ali, S. (1997) 'Presence of exon 5-deleted oestrogen receptor in human breast cancer: functional analysis and clinical significance', *Br J Cancer,* 75, (8), pp. 1173-84.

Dillon, R. L., White, D. E. and Muller, W. J. (2007) 'The phosphatidyl inositol 3-kinase signaling network: implications for human breast cancer', *Oncogene,* 26, (9), pp. 1338-45.

Doherty, J. K., Bond, C., Jardim, A., Adelman, J. P. and Clinton, G. M. (1999) 'The HER-2/neu receptor tyrosine kinase gene encodes a secreted autoinhibitor', *Proc Natl Acad Sci U S A,* 96, (19), pp. 10869-74.

Dowsett, M., Cuzick, J., Wale, C., Howell, T., Houghton, J. and Baum, M. (2005) 'Retrospective analysis of time to recurrence in the ATAC trial according to hormone receptor status: an hypothesis-generating study', *J Clin Oncol,* 23, (30), pp. 7512-7.

Dunbier, A. K., Anderson, H., Ghazoui, Z., Salter, J., Parker, J. S., Perou, C. M., Smith, I. E. and Dowsett, M. (2011) 'Association between breast cancer subtypes and response to neoadjuvant anastrozole', *Steroids,* 76, (8), pp. 736-40.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001) 'Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells', *Nature,* 411, (6836), pp. 494-8.

Elston, C. W. (1998) *The Breast.*Third ed: Churchill Livingstone.

Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G. and Krainer, A. R. (2000) 'Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1', *Mol Cell Biol,* 20, (22), pp. 8303-18.

Evans, R. M. (1988) 'The steroid and thyroid hormone receptor superfamily', *Science,* 240, (4854), pp. 889-95.

Ezponda, T., Pajares, M. J., Agorreta, J., Echeveste, J. I., Lopez-Picazo, J. M., Torre, W., Pio, R. and Montuenga, L. M. (2010) 'The oncoprotein SF2/ASF promotes non-small cell lung cancer survival by enhancing survivin expression', *Clin Cancer Res,* 16, (16), pp. 4113-25.

Faivre, E. J., Daniel, A. R., Hillard, C. J. and Lange, C. A. (2008) 'Progesterone receptor rapid signaling mediates serine 345 phosphorylation and tethering to specificity protein 1 transcription factors', *Mol Endocrinol,* 22, (4), pp. 823-37.

Faivre, E. J. and Lange, C. A. (2007) 'Progesterone receptors upregulate Wnt-1 to induce epidermal growth factor receptor transactivation and c-Src-dependent sustained activation of Erk1/2 mitogen-activated protein kinase in breast cancer cells', *Mol Cell Biol,* 27, (2), pp. 466-80.

Flouriot, G., Brand, H., Denger, S., Metivier, R., Kos, M., Reid, G., Sonntag-Buck, V. and Gannon, F. (2000) 'Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1', *EMBO J,* 19, (17), pp. 4688-700.

Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B. A., Gayther, S. A., Zelada-Hedman, M. and et al. (1998) 'Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium', *Am J Hum Genet,* 62, (3), pp. 676-89.

Fu, X. D. (1993) 'Specific commitment of different pre-mRNAs to splicing by single SR proteins', *Nature,* 365, (6441), pp. 82-5.

Gaildrat, P., Krieger, S., Thery, J. C., Killian, A., Rousselin, A., Berthet, P., Frebourg, T., Hardouin, A., Martins, A. and Tosi, M. (2010) 'The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements', *J Med Genet,* 47, (6), pp. 398-403.

Galea, M. H., Blamey, R. W., Elston, C. E. and Ellis, I. O. (1992) 'The Nottingham Prognostic Index in primary breast cancer', *Breast Cancer Res Treat,* 22, (3), pp. 207-19.

Gallego, M. E., Gattoni, R., Stevenin, J., Marie, J. and Expert-Bezancon, A. (1997) 'The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A', *EMBO J,* 16, (7), pp. 1772-84.

Gardner, L. B. (2010) 'Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis', *Mol Cancer Res,* 8, (3), pp. 295-308.

Garneau, D., Revil, T., Fisette, J. F. and Chabot, B. (2005) 'Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x', *J Biol Chem,* 280, (24), pp. 22641-50.

Gellersen, B., Fernandes, M. S. and Brosens, J. J. (2009) 'Non-genomic progesterone actions in female reproduction', *Hum Reprod Update,* 15, (1), pp. 119-38.

Giangrande, P. H., Kimbrel, E. A., Edwards, D. P. and McDonnell, D. P. (2000) 'The opposing transcriptional activities of the two isoforms of the human progesterone receptor are due to differential cofactor binding', *Mol Cell Biol,* 20, (9), pp. 3102-15.

Giangrande, P. H., Pollio, G. and McDonnell, D. P. (1997) 'Mapping and characterization of the functional domains responsible for the differential activity of the A and B isoforms of the human progesterone receptor', *J Biol Chem,* 272, (52), pp. 32889-900.

Graham, J. D. and Clarke, C. L. (1997) 'Physiological action of progesterone in target tissues', *Endocr Rev,* 18, (4), pp. 502-19.

Graham, J. D., Yager, M. L., Hill, H. D., Byth, K., O'Neill, G. M. and Clarke, C. L. (2005) 'Altered progesterone receptor isoform expression remodels progestin responsiveness of breast cancer cells', *Mol Endocrinol,* 19, (11), pp. 2713-35.

Graham, J. D., Yeates, C., Balleine, R. L., Harvey, S. S., Milliken, J. S., Bilous, A. M. and Clarke, C. L. (1995) 'Characterization of progesterone receptor A and B expression in human breast cancer', *Cancer Res,* 55, (21), pp. 5063-8.

Graham, J. D., Yeates, C., Balleine, R. L., Harvey, S. S., Milliken, J. S., Bilous, A. M. and Clarke, C. L. (1996) 'Progesterone receptor A and B protein expression in human breast cancer', *J Steroid Biochem Mol Biol,* 56, (1-6 Spec No), pp. 93-8.

Graupner, V., Schulze-Osthoff, K., Essmann, F. and Janicke, R. U. (2009) 'Functional characterization of p53beta and p53gamma, two isoforms of the tumor suppressor p53', *Cell Cycle,* 8, (8), pp. 1238-48.

Graveley, B. R. (2000) 'Sorting out the complexity of SR protein functions', *RNA,* 6, (9), pp. 1197-211.

Graveley, B. R., Hertel, K. J. and Maniatis, T. (2001) 'The role of U2AF35 and U2AF65 in enhancer-dependent splicing', *RNA,* 7, (6), pp. 806-18.

Gromoll, J., Lahrmann, L., Godmann, M., Muller, T., Michel, C., Stamm, S. and Simoni, M. (2007) 'Genomic checkpoints for exon 10 usage in the luteinizing hormone receptor type 1 and type 2', *Mol Endocrinol,* 21, (8), pp. 1984-96.

Gross, G. E., Clark, G. M., Chamness, G. C. and McGuire, W. L. (1984) 'Multiple progesterone receptor assays in human breast cancer', *Cancer Res,* 44, (2), pp. 836-40.

Guiochon-Mantel, A., Loosfelt, H., Lescop, P., Christin-Maitre, S., Perrot-Applanat, M. and Milgrom, E. (1992) 'Mechanisms of nuclear localization of the progesterone receptor', *J Steroid Biochem Mol Biol,* 41, (3-8), pp. 209-15.

Guo, X., Chen, Q. R., Song, Y. K., Wei, J. S. and Khan, J. (2011) 'Exon array analysis reveals neuroblastoma tumors have distinct alternative splicing patterns according to stage and MYCN amplification status', *BMC Med Genomics,* 4, pp. 35.

Hagan, C. R., Regan, T. M., Dressing, G. E. and Lange, C. A. (2011) 'ck2-Dependent Phosphorylation of Progesterone Receptors (PR) on Ser81 Regulates PR-B Isoform-Specific Target Gene Expression in Breast Cancer Cells', *Mol Cell Biol,* 31, (12), pp. 2439-52.

Hamilton, A. J. and Baulcombe, D. C. (1999) 'A species of small antisense RNA in posttranscriptional gene silencing in plants', *Science,* 286, (5441), pp. 950-2.

Hanahan, D. and Weinberg, R. A. (2000) 'The hallmarks of cancer', *Cell,* 100, (1), pp. 57-70.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of cancer: the next generation', *Cell,* 144, (5), pp. 646-74.

Hanley, K. Z., Siddiqui, M. T., Lawson, D., Cohen, C. and Nassar, A. (2009) 'Evaluation of new monoclonal antibodies in detection of estrogen receptor, progesterone receptor, and Her2 protein expression in breast carcinoma cell block sections using conventional microscopy and quantitative image analysis', *Diagn Cytopathol,* 37, (4), pp. 251-7.

Haque, A., Buratti, E. and Baralle, F. E. (2010) 'Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12', *Nucleic Acids Res,* 38, (2), pp. 647-59.

He, X., Arslan, A. D., Pool, M. D., Ho, T. T., Darcy, K. M., Coon, J. S. and Beck, W. T. (2011) 'Knockdown of splicing factor SRp20 causes apoptosis in ovarian cancer cells and its expression is associated with malignancy of epithelial ovarian cancer', *Oncogene,* 30, (3), pp. 356-65.

Hennessy, B. T., Gonzalez-Angulo, A. M., Stemke-Hale, K., Gilcrease, M. Z., Krishnamurthy, S., Lee, J. S., Fridlyand, J., Sahin, A., Agarwal, R., Joy, C., Liu, W., Stivers, D., Baggerly, K., Carey, M., Lluch, A., Monteagudo, C., He, X., Weigman, V., Fan, C., Palazzo, J., Hortobagyi, G. N., Nolden, L. K., Wang, N. J., Valero, V., Gray, J. W., Perou, C. M. and Mills, G. B. (2009) 'Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics', *Cancer Res,* 69, (10), pp. 4116-24.

Hertel, K. J. (2008) 'Combinatorial control of exon recognition', *J Biol Chem,* 283, (3), pp. 1211-5.

Hertel, K. J. and Maniatis, T. (1999) 'Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing', *Proc Natl Acad Sci U S A,* 96, (6), pp. 2651-5.

Herynk, M. H. and Fuqua, S. A. (2004) 'Estrogen receptor mutations in human disease', *Endocr Rev,* 25, (6), pp. 869-98.

Hinck, L. and Silberstein, G. B. (2005) 'Key stages in mammary gland development: the mammary end bud as a motile organ', *Breast Cancer Res,* 7, (6), pp. 245-51.

Hirata, S., Shoda, T., Kato, J. and Hoshi, K. (2000) 'The novel isoform of the progesterone receptor cDNA in the human testis and detection of its mRNA in the human uterine endometrium', *Oncology,* 59 Suppl 1, pp. 39-44.

Hirata, S., Shoda, T., Kato, J. and Hoshi, K. (2002) 'The novel exon, exon T, of the human progesterone receptor gene and the genomic organization of the gene', *J Steroid Biochem Mol Biol,* 80, (3), pp. 365-7.

Hirata, S., Shoda, T., Kato, J. and Hoshi, K. (2003) 'Isoform/variant mRNAs for sex steroid hormone receptors in humans', *Trends Endocrinol Metab,* 14, (3), pp. 124-9.

Hisatomi, H., Kohno, N., Wakita, K., Nagao, K., Hirata, H., Hikiji, K. and Harada, S. (2003) 'Novel alternatively spliced variant with a deletion of 52 BP in exon 6 of the progesterone receptor gene is observed frequently in breast cancer tissues', *Int J Cancer,* 105, (2), pp. 182-5.

Hodges, Y. K., Richer, J. K., Horwitz, K. B. and Horwitz, L. D. (1999) 'Variant estrogen and progesterone receptor messages in human vascular smooth muscle', *Circulation,* 99, (20), pp. 2688-93.

Hofseth, L. J., Raafat, A. M., Osuch, J. R., Pathak, D. R., Slomski, C. A. and Haslam, S. Z. (1999) 'Hormone replacement therapy with estrogen or estrogen plus medroxyprogesterone acetate is associated with increased epithelial proliferation in the normal postmenopausal breast', *J Clin Endocrinol Metab,* 84, (12), pp. 4559-65.

Hofstetter, G., Berger, A., Fiegl, H., Slade, N., Zoric, A., Holzer, B., Schuster, E., Mobus, V. J., Reimer, D., Daxenbichler, G., Marth, C., Zeimet, A. G., Concin, N. and Zeillinger, R. (2010) 'Alternative splicing of p53 and p73: the novel p53 splice variant p53delta is an independent prognostic marker in ovarian cancer', *Oncogene,* 29, (13), pp. 1997-2004.

Hopp, T. A., Weiss, H. L., Hilsenbeck, S. G., Cui, Y., Allred, D. C., Horwitz, K. B. and Fuqua, S. A. (2004) 'Breast cancer patients with progesterone receptor PR-A-rich tumors have poorer disease-free survival rates', *Clin Cancer Res,* 10, (8), pp. 2751-60.

Hovland, A. R., Powell, R. L., Takimoto, G. S., Tung, L. and Horwitz, K. B. (1998) 'An N-terminal inhibitory function, IF, suppresses transcription by the A-isoform but not the B-isoform of human progesterone receptors', *J Biol Chem,* 273, (10), pp. 5455-60.

Hsiao, W. C., Cho, W. C., Lin, P. W., Lin, S. L., Lee, W. Y. and Young, K. C. (2006) 'Quantitative profile of estrogen receptor variants/isoforms in Taiwanese women with breast cancer', *Eur J Surg Oncol*, 32, (5), pp. 492-7.

http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home *ESEfinder*. Available at: http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home (Accessed: 05/07/2011).

http://web.expasy.org/compute_pi/ *Compute pI/Mw*. Available at: http://web.expasy.org/compute_pi/ (Accessed: 04/07/2011).

http://web.expasy.org/translate/ *Translate*. Available at: http://web.expasy.org/translate/ (Accessed: 04/07/2011).

http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8 *Splicing Rainbow*. Available at: http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8 (Accessed: 05/07/2011).

http://xylian.igh.cnrs.fr/bin/align-guess.cgi *ALIGN*. Available at: http://xylian.igh.cnrs.fr/bin/align-guess.cgi (Accessed: 01/03/2011).

Hu, R., Dunn, T. A., Wei, S., Isharwal, S., Veltri, R. W., Humphreys, E., Han, M., Partin, A. W., Vessella, R. L., Isaacs, W. B., Bova, G. S. and Luo, J. (2009) 'Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer', *Cancer Res*, 69, (1), pp. 16-22.

Ibrahim, E. C., Schaal, T. D., Hertel, K. J., Reed, R. and Maniatis, T. (2005) 'Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers', *Proc Natl Acad Sci U S A*, 102, (14), pp. 5002-7.

Im, A. and Appleman, L. J. (2010) 'Mifepristone: pharmacology and clinical impact in reproductive medicine, endocrinology and oncology', *Expert Opin Pharmacother*, 11, (3), pp. 481-8.

Ing, N. H., Beekman, J. M., Tsai, S. Y., Tsai, M. J. and O'Malley, B. W. (1992) 'Members of the steroid hormone receptor superfamily interact with TFIIB (S300-II)', *J Biol Chem*, 267, (25), pp. 17617-23.

Ishii, H., Shoda, Y., Yomogida, K., Hamada, T. and Sakuma, Y. (2011) 'Identification of C-terminally and N-terminally truncated estrogen receptor alpha variants in the mouse', *J Steroid Biochem Mol Biol*, 124, (1-2), pp. 38-46.

Ismail, P. M., Li, J., DeMayo, F. J., O'Malley, B. W. and Lydon, J. P. (2002) 'A novel LacZ reporter mouse reveals complex regulation of the progesterone receptor promoter during mammary gland development', *Mol Endocrinol*, 16, (11), pp. 2475-89.

Iverson, A. A., Gillett, C., Cane, P., Santini, C. D., Vess, T. M., Kam-Morgan, L., Wang, A., Eisenberg, M., Rowland, C. M., Hessling, J. J., Broder, S. E., Sninsky, J. J., Tutt, A., Anderson, S. and Chang, S. Y. (2009) 'A single-tube quantitative assay for mRNA levels of hormonal and growth factor receptors in breast cancer specimens', *J Mol Diagn*, 11, (2), pp. 117-30.

Jacobsen, B. M., Richer, J. K., Schittone, S. A. and Horwitz, K. B. (2002) 'New human breast cancer cells to study progesterone receptor isoform ratio effects and ligand-independent gene regulation', *J Biol Chem*, 277, (31), pp. 27793-800.

Jacobsen, B. M., Schittone, S. A., Richer, J. K. and Horwitz, K. B. (2005) 'Progesterone-independent effects of human progesterone receptors (PRs) in estrogen receptor-positive breast cancer: PR isoform-specific gene regulation and tumor biology', *Mol Endocrinol*, 19, (3), pp. 574-87.

Jang, S. and Yi, L. S. (2005) 'Identification of a 71 kDa protein as a putative non-genomic membrane progesterone receptor in boar spermatozoa', *J Endocrinol*, 184, (2), pp. 417-25.

Jia, R., Li, C., McCoy, J. P., Deng, C. X. and Zheng, Z. M. (2010) 'SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance', *Int J Biol Sci*, 6, (7), pp. 806-26.

Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R. and Shoemaker, D. D. (2003) 'Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays', *Science*, 302, (5653), pp. 2141-4.

Jumaa, H. and Nielsen, P. J. (1997) 'The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation', *EMBO J*, 16, (16), pp. 5077-85.

Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. and Moore, M. J. (2002) 'Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis', *RNA*, 8, (4), pp. 426-39.

Jurica, M. S. and Moore, M. J. (2003) 'Pre-mRNA splicing: awash in a sea of proteins', *Mol Cell*, 12, (1), pp. 5-14.

Kakar, M., Kanwal, C., Davis, J. R., Li, H. and Lim, C. S. (2006) 'Geldanamycin, an inhibitor of Hsp90, blocks cytoplasmic retention of progesterone receptors and glucocorticoid receptors via their respective ligand binding domains', *AAPS J*, 8, (4), pp. E718-28.

Kang, L., Zhang, X., Xie, Y., Tu, Y., Wang, D., Liu, Z. and Wang, Z. Y. (2010) 'Involvement of estrogen receptor variant ER-alpha36, not GPR30, in nongenomic estrogen signaling', *Mol Endocrinol*, 24, (4), pp. 709-21.

Kanopka, A., Muhlemann, O. and Akusjarvi, G. (1996) 'Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA', *Nature*, 381, (6582), pp. 535-8.

Kariagina, A., Aupperlee, M. D. and Haslam, S. Z. (2007) 'Progesterone receptor isoforms and proliferation in the rat mammary gland during development', *Endocrinology*, 148, (6), pp. 2723-36.

Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D. and Krainer, A. R. (2007) 'The gene encoding the splicing factor SF2/ASF is a proto-oncogene', *Nat Struct Mol Biol*, 14, (3), pp. 185-93.

Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H. and Chambon, P. (1990) 'Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B', *EMBO J,* 9, (5), pp. 1603-14.

Katiyar, P., Ma, Y., Fan, S., Pestell, R. G., Furth, P. A. and Rosen, E. M. (2006) 'Regulation of progesterone receptor signaling by BRCA1 in mammary cancer', *Nucl Recept Signal,* 4, pp. e006.

Katiyar, P., Ma, Y., Riegel, A., Fan, S. and Rosen, E. M. (2009) 'Mechanism of BRCA1-mediated inhibition of progesterone receptor transcriptional activity', *Mol Endocrinol,* 23, (8), pp. 1135-46.

Kawano, T., Fujita, M. and Sakamoto, H. (2000) 'Unique and redundant functions of SR proteins, a conserved family of splicing factors, in Caenorhabditis elegans development', *Mech Dev,* 95, (1-2), pp. 67-76.

Kim, E., Goren, A. and Ast, G. (2008) 'Insights into the connection between cancer and alternative splicing', *Trends Genet,* 24, (1), pp. 7-10.

Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Gervais-Bird, J., Madden, R., Paquet, E. R., Koh, C., Venables, J. P., Prinos, P., Jilaveanu-Pelmus, M., Wellinger, R., Rancourt, C., Chabot, B. and Abou Elela, S. (2008) 'Multiple alternative splicing markers for ovarian cancer', *Cancer Res,* 68, (3), pp. 657-63.

Koduri, S., Goldhar, A. S. and Vonderhaar, B. K. (2006) 'Activation of vascular endothelial growth factor (VEGF) by the ER-alpha variant, ERDelta3', *Breast Cancer Res Treat,* 95, (1), pp. 37-43.

Koletsa, T., Kostopoulos, I., Charalambous, E., Christoforidou, B., Nenopoulou, E. and Kotoula, V. (2008) 'A splice variant of HER2 corresponding to Herstatin is expressed in the noncancerous breast and in breast carcinomas', *Neoplasia,* 10, (7), pp. 687-96.

Lai, Y. L., Mau, B. L., Cheng, W. H., Chen, H. M., Chiu, H. H. and Tzen, C. Y. (2008) 'PIK3CA exon 20 mutation is independently associated with a poor prognosis in breast cancer patients', *Ann Surg Oncol,* 15, (4), pp. 1064-9.

Lamb, C. A., Helguero, L. A., Giulianelli, S., Soldati, R., Vanzulli, S. I., Molinolo, A. and Lanari, C. (2005) 'Antisense oligonucleotides targeting the progesterone receptor inhibit hormone-independent breast cancer growth in mice', *Breast Cancer Res,* 7, (6), pp. R1111-21.

Lange, C. A., Shen, T. and Horwitz, K. B. (2000) 'Phosphorylation of human progesterone receptors at serine-294 by mitogen-activated protein kinase signals their degradation by the 26S proteasome', *Proc Natl Acad Sci U S A,* 97, (3), pp. 1032-7.

Langer, W., Sohler, F., Leder, G., Beckmann, G., Seidel, H., Grone, J., Hummel, M. and Sommer, A. (2010) 'Exon array analysis using re-defined probe sets results in reliable identification of alternatively spliced genes in non-small cell lung cancer', *BMC Genomics,* 11, pp. 676.

Lapidus, R. G., Ferguson, A. T., Ottaviano, Y. L., Parl, F. F., Smith, H. S., Weitzman, S. A., Baylin, S. B., Issa, J. P. and Davidson, N. E. (1996) 'Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors', *Clin Cancer Res,* 2, (5), pp. 805-10.

Lapuk, A., Marr, H., Jakkula, L., Pedro, H., Bhattacharya, S., Purdom, E., Hu, Z., Simpson, K., Pachter, L., Durinck, S., Wang, N., Parvin, B., Fontenay, G., Speed, T., Garbe, J., Stampfer, M., Bayandorian, H., Dorton, S., Clark, T. A., Schweitzer, A., Wyrobek, A., Feiler, H., Spellman, P., Conboy, J. and Gray, J. W. (2010) 'Exon-level microarray analyses identify alternative splicing programs in breast cancer', *Mol Cancer Res,* 8, (7), pp. 961-74.

Launonen, V., Stenback, F., Puistola, U., Bloigu, R., Huusko, P., Kytola, S., Kauppila, A. and Winqvist, R. (1998) 'Chromosome 11q22.3-q25 LOH in ovarian cancer: association with a more aggressive disease course and involved subregions', *Gynecol Oncol,* 71, (2), pp. 299-304.

Lee, E. Y. and Muller, W. J. (2010) 'Oncogenes and tumor suppressor genes', *Cold Spring Harb Perspect Biol,* 2, (10), pp. a003236.

Lewis, B. P., Green, R. E. and Brenner, S. E. (2003) 'Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans', *Proc Natl Acad Sci U S A,* 100, (1), pp. 189-92.

Leygue, E., Dotzlaw, H., Watson, P. H. and Murphy, L. C. (1996a) 'Identification of novel exon-deleted progesterone receptor variant mRNAs in human breast tissue', *Biochem Biophys Res Commun,* 228, (1), pp. 63-8.

Leygue, E. R., Watson, P. H. and Murphy, L. C. (1996b) 'Estrogen receptor variants in normal human mammary tissue', *J Natl Cancer Inst,* 88, (5), pp. 284-90.

Li, X., Lonard, D. M. and O'Malley, B. W. (2004) 'A contemporary understanding of progesterone receptor function', *Mech Ageing Dev,* 125, (10-11), pp. 669-78.

Liao, N., Zhang, G. C., Liu, Y. H., Li, X. R., Yao, M., Xu, F. P., Li, L. and Wu, Y. L. (2011) 'HER2-positive status is an independent predictor for coexisting invasion of ductal carcinoma in situ of the breast presenting extensive DCIS component', *Pathol Res Pract,* 207, (1), pp. 1-7.

Lin, S. L., Yan, L. Y., Zhang, X. T., Yuan, J., Li, M., Qiao, J., Wang, Z. Y. and Sun, Q. Y. (2010) 'ER-alpha36, a variant of ER-alpha, promotes tamoxifen agonist action in endometrial cancer cells via the MAPK/ERK and PI3K/Akt pathways', *PLoS One,* 5, (2), pp. e9013.

Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q. and Krainer, A. R. (2000) 'Exonic splicing enhancer motif recognized by human SC35 under splicing conditions', *Mol Cell Biol,* 20, (3), pp. 1063-71.

Liu, H. X., Zhang, M. and Krainer, A. R. (1998) 'Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins', *Genes Dev,* 12, (13), pp. 1998-2012.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L. and Hannon, G. J. (2004) 'Argonaute2 is the catalytic engine of mammalian RNAi', *Science,* 305, (5689), pp. 1437-41.

Liu, Z., Auboeuf, D., Wong, J., Chen, J. D., Tsai, S. Y., Tsai, M. J. and O'Malley, B. W. (2002) 'Coactivator/corepressor ratios modulate PR-mediated transcription by the selective receptor modulator RU486', *Proc Natl Acad Sci U S A,* 99, (12), pp. 7940-4.

Liu, Z., Wong, J., Tsai, S. Y., Tsai, M. J. and O'Malley, B. W. (2001) 'Sequential recruitment of steroid receptor coactivator-1 (SRC-1) and p300 enhances progesterone receptor-dependent initiation and reinitiation of transcription from chromatin', *Proc Natl Acad Sci U S A,* 98, (22), pp. 12426-31.

Loeb, L. and Hesselberg, C. (1917) 'The Cyclic Changes in the Mammary Gland under Normal and Pathological Conditions : Ii. The Changes in the Pregnant Guinea Pig, the Effect of Lutein Injections, and the Correlation between the Cycle of the Uterus and Ovaries and the Cycle of the Mammary Gland', *J Exp Med,* 25, (2), pp. 305-21.

Long, J. C. and Caceres, J. F. (2009) 'The SR protein family of splicing factors: master regulators of gene expression', *Biochem J,* 417, (1), pp. 15-27.

Lydon, J. P., DeMayo, F. J., Funk, C. R., Mani, S. K., Hughes, A. R., Montgomery, C. A., Jr., Shyamala, G., Conneely, O. M. and O'Malley, B. W. (1995) 'Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities', *Genes Dev,* 9, (18), pp. 2266-78.

Lydon, J. P., Ge, G., Kittrell, F. S., Medina, D. and O'Malley, B. W. (1999) 'Murine mammary gland carcinogenesis is critically dependent on progesterone receptor function', *Cancer Res,* 59, (17), pp. 4276-84.

Ma, Y., Katiyar, P., Jones, L. P., Fan, S., Zhang, Y., Furth, P. A. and Rosen, E. M. (2006) 'The breast cancer susceptibility gene BRCA1 regulates progesterone receptor signaling in mammary epithelial cells', *Mol Endocrinol,* 20, (1), pp. 14-34.

Madsen, G., Macintyre, D. A., Mesiano, S. and Smith, R. (2007) 'Progesterone receptor or cytoskeletal protein?', *Reprod Sci,* 14, (3), pp. 217-22.

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. and Luhrmann, R. (2002) 'Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome', *Science,* 298, (5601), pp. 2205-8.

Makhoul, I. and Kiwan, E. (2011) 'Neoadjuvant systemic treatment of breast cancer', *J Surg Oncol,* 103, (4), pp. 348-57.

Manley, J. L. and Krainer, A. R. (2010) 'A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)', *Genes Dev,* 24, (11), pp. 1073-4.

Marshburn, P. B., Zhang, J., Bahrani-Mostafavi, Z., Matthews, M. L., White, J. and Hurst, B. S. (2005) 'Variant progesterone receptor mRNAs are co-expressed with the wild-type progesterone receptor mRNA in human endometrium during all phases of the menstrual cycle', *Mol Hum Reprod,* 11, (11), pp. 809-15.

Matlin, A. J., Clark, F. and Smith, C. W. (2005) 'Understanding alternative splicing: towards a cellular code', *Nat Rev Mol Cell Biol,* 6, (5), pp. 386-98.

Matter, N., Herrlich, P. and Konig, H. (2002) 'Signal-dependent regulation of splicing via phosphorylation of Sam68', *Nature,* 420, (6916), pp. 691-5.

Mazoyer, S., Puget, N., Perrin-Vidoz, L., Lynch, H. T., Serova-Sinilnikova, O. M. and Lenoir, G. M. (1998) 'A BRCA1 nonsense mutation causes exon skipping', *Am J Hum Genet,* 62, (3), pp. 713-5.

Mc Cormack, O., Chung, W. Y., Fitzpatrick, P., Cooke, F., Flynn, B., Harrison, M., Fox, E., Gallagher, E., McGoldrick, A., Dervan, P. A., McCann, A. and Kerin, M. J. (2008) 'Progesterone receptor B (PRB) promoter hypermethylation in sporadic breast cancer: progesterone receptor B hypermethylation in breast cancer', *Breast Cancer Res Treat,* 111, (1), pp. 45-53.

McDonnell, D. P., Shahbaz, M. M., Vegeto, E. and Goldman, M. E. (1994) 'The human progesterone receptor A-form functions as a transcriptional modulator of mineralocorticoid receptor transcriptional activity', *The Journal of Steroid Biochemistry and Molecular Biology,* 48, pp. 425-432.

McGowan, E. M., Russell, A. J., Boonyaratanakornkit, V., Saunders, D. N., Lehrbach, G. M., Sergio, C. M., Musgrove, E. A., Edwards, D. P. and Sutherland, R. L. (2007) 'Progestins reinitiate cell cycle progression in antiestrogen-arrested breast cancer cells through the B-isoform of progesterone receptor', *Cancer Res,* 67, (18), pp. 8942-51.

McKenna, N. J., Lanz, R. B. and O'Malley, B. W. (1999) 'Nuclear receptor coregulators: cellular and molecular biology', *Endocr Rev,* 20, (3), pp. 321-44.

McManus, C. J. and Graveley, B. R. (2011) 'RNA structure and the mechanisms of alternative splicing', *Curr Opin Genet Dev.*

Menard, S., Pupa, S. M., Campiglio, M. and Tagliabue, E. (2003) 'Biologic and therapeutic role of HER2 in cancer', *Oncogene,* 22, (42), pp. 6570-8.

Mersin, H., Yildirim, E., Berberoglu, U. and Gulben, K. (2008) 'The prognostic importance of triple negative breast carcinoma', *Breast,* 17, (4), pp. 341-6.

Meyer, M. E., Pornon, A., Ji, J. W., Bocquel, M. T., Chambon, P. and Gronemeyer, H. (1990) 'Agonistic and antagonistic activities of RU486 on the functions of the human progesterone receptor', *EMBO J,* 9, (12), pp. 3923-32.

Migliaccio, A., Piccolo, D., Castoria, G., Di Domenico, M., Bilancio, A., Lombardi, M., Gong, W., Beato, M. and Auricchio, F. (1998) 'Activation of the Src/p21ras/Erk pathway by progesterone receptor via cross-talk with estrogen receptor', *EMBO J,* 17, (7), pp. 2008-18.

Mirza, S., Sharma, G., Prasad, C. P., Parshad, R., Srivastava, A., Gupta, S. D. and Ralhan, R. (2007) 'Promoter hypermethylation of TMS1, BRCA1, ERalpha and PRB in serum and tumor DNA of invasive ductal breast carcinoma patients', *Life Sci*, 81, (4), pp. 280-7.

Misao, R., Nakanishi, Y., Sun, W., Iwagaki, S., Fujimoto, J. and Tamaya, T. (2000) 'Identification of exon-deleted progesterone receptor mRNAs in human uterine endometrial cancers', *Oncology*, 58, (1), pp. 60-5.

Misao, R., Sun, W. S., Iwagaki, S., Fujimoto, J. and Tamaya, T. (1998) 'Identification of various exon-deleted progesterone receptor mRNAs in human endometrium and ovarian endometriosis', *Biochem Biophys Res Commun*, 252, (2), pp. 302-6.

Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlansky, F., Guiochon-Mantel, A., Galibert, F. and Milgrom, E. (1987) 'Complete amino acid sequence of the human progesterone receptor deduced from cloned cDNA', *Biochem Biophys Res Commun*, 143, (2), pp. 740-8.

Misrahi, M., Venencie, P. Y., Saugier-Veber, P., Sar, S., Dessen, P. and Milgrom, E. (1993) 'Structure of the human progesterone receptor gene', *Biochim Biophys Acta*, 1216, (2), pp. 289-92.

Mohamed, M. K., Tung, L., Takimoto, G. S. and Horwitz, K. B. (1994) 'The leucine zippers of c-fos and c-jun for progesterone receptor dimerization: A-dominance in the A/B heterodimer', *J Steroid Biochem Mol Biol*, 51, (5-6), pp. 241-50.

Moore, M. J. and Proudfoot, N. J. (2009) 'Pre-mRNA processing reaches back to transcription and ahead to translation', *Cell*, 136, (4), pp. 688-700.

Moroy, T. and Heyd, F. (2007) 'The impact of alternative splicing in vivo: mouse models show the way', *RNA*, 13, (8), pp. 1155-71.

Mote, P. A., Bartow, S., Tran, N. and Clarke, C. L. (2002) 'Loss of co-ordinate expression of progesterone receptors A and B is an early event in breast carcinogenesis', *Breast Cancer Res Treat*, 72, (2), pp. 163-72.

Mote, P. A., Leary, J. A., Avery, K. A., Sandelin, K., Chenevix-Trench, G., Kirk, J. A. and Clarke, C. L. (2004) 'Germ-line mutations in BRCA1 or BRCA2 in the normal breast are associated with altered expression of estrogen-responsive proteins and the predominance of progesterone receptor A', *Genes Chromosomes Cancer*, 39, (3), pp. 236-48.

Mukherjee, A., Soyal, S. M., Li, J., Ying, Y., He, B., DeMayo, F. J. and Lydon, J. P. (2010) 'Targeting RANKL to a specific subset of murine mammary epithelial cells induces ordered branching morphogenesis and alveologenesis in the absence of progesterone receptor expression', *FASEB J*, 24, (11), pp. 4408-19.

Mulac-Jericevic, B., Lydon, J. P., DeMayo, F. J. and Conneely, O. M. (2003) 'Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform', *Proc Natl Acad Sci U S A*, 100, (17), pp. 9744-9.

Mulac-Jericevic, B., Mullinax, R. A., DeMayo, F. J., Lydon, J. P. and Conneely, O. M. (2000) 'Subgroup of reproductive functions of progesterone mediated by progesterone receptor-B isoform', *Science,* 289, (5485), pp. 1751-4.

Nagao, K., Kohno, N., Wakita, K., Hikiji, K., Yamamoto, S., Hirata, H. and Hisatomi, H. (2003) 'Expression of a novel splicing variant deleting exons 4 and 6 of the progesterone receptor gene is a rare event in breast cancer', *Oncol Rep,* 10, (2), pp. 305-8.

Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A. and Gray, J. W. (2006) 'A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes', *Cancer Cell,* 10, (6), pp. 515-27.

Nicholson, P. and Muhlemann, O. (2010) 'Cutting the nonsense: the degradation of PTC-containing mRNAs', *Biochem Soc Trans,* 38, (6), pp. 1615-20.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N. and Muhlemann, O. (2010) 'Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors', *Cell Mol Life Sci,* 67, (5), pp. 677-700.

Nilsen, T. W. and Graveley, B. R. (2010) 'Expansion of the eukaryotic proteome by alternative splicing', *Nature,* 463, (7280), pp. 457-63.

Nowak, D. G., Amin, E. M., Rennel, E. S., Hoareau-Aveilla, C., Gammons, M., Damodoran, G., Hagiwara, M., Harper, S. J., Woolard, J., Ladomery, M. R. and Bates, D. O. (2010) 'Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis', *J Biol Chem,* 285, (8), pp. 5532-40.

Nowak, D. G., Woolard, J., Amin, E. M., Konopatskaya, O., Saleem, M. A., Churchill, A. J., Ladomery, M. R., Harper, S. J. and Bates, D. O. (2008) 'Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors', *J Cell Sci,* 121, (Pt 20), pp. 3487-95.

Olopade, O. I., Adeyanju, M. O., Safa, A. R., Hagos, F., Mick, R., Thompson, C. B. and Recant, W. M. (1997) 'Overexpression of BCL-x protein in primary breast cancer is associated with high tumor grade and nodal metastases', *Cancer J Sci Am,* 3, (4), pp. 230-7.

Onate, S. A., Boonyaratanakornkit, V., Spencer, T. E., Tsai, S. Y., Tsai, M. J., Edwards, D. P. and O'Malley, B. W. (1998) 'The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors', *J Biol Chem,* 273, (20), pp. 12101-8.

Onate, S. A., Tsai, S. Y., Tsai, M. J. and O'Malley, B. W. (1995) 'Sequence and characterization of a coactivator for the steroid hormone receptor superfamily', *Science,* 270, (5240), pp. 1354-7.

Osborne, C. K. (1998) 'Steroid hormone receptors in breast cancer management', *Breast Cancer Res Treat,* 51, (3), pp. 227-38.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P. A. (1986) 'Splicing of messenger RNA precursors', *Annu Rev Biochem,* 55, pp. 1119-50.

Pagani, F. and Baralle, F. E. (2004) 'Genomic variants in exons and introns: identifying the splicing spoilers', *Nat Rev Genet,* 5, (5), pp. 389-96.

Pagani, F., Buratti, E., Stuani, C. and Baralle, F. E. (2003a) 'Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9', *J Biol Chem,* 278, (29), pp. 26580-8.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T. and Baralle, F. E. (2003b) 'New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12', *Hum Mol Genet,* 12, (10), pp. 1111-20.

Pajares, M. J., Ezponda, T., Catena, R., Calvo, A., Pio, R. and Montuenga, L. M. (2007) 'Alternative splicing: an emerging topic in molecular and clinical oncology', *Lancet Oncol,* 8, (4), pp. 349-57.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J. and Blencowe, B. J. (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet,* 40, (12), pp. 1413-5.

Pang, Y. and Thomas, P. (2011) 'Progesterone signals through membrane progesterone receptors (mPRs) in MDA-MB-468 and mPR-transfected MDA-MB-231 breast cancer cells which lack full-length and N-terminally truncated isoforms of the nuclear progesterone receptor', *Steroids*.

Payne, S. J., Bowen, R. L., Jones, J. L. and Wells, C. A. (2008) 'Predictive markers in breast cancer--the present', *Histopathology,* 52, (1), pp. 82-90.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000) 'Molecular portraits of human breast tumours', *Nature,* 406, (6797), pp. 747-52.

Persson, I., Weiderpass, E., Bergkvist, L., Bergstrom, R. and Schairer, C. (1999) 'Risks of breast and endometrial cancer after estrogen and estrogen-progestin replacement', *Cancer Causes Control,* 10, (4), pp. 253-60.

Pettigrew, C., Wayte, N., Lovelock, P. K., Tavtigian, S. V., Chenevix-Trench, G., Spurdle, A. B. and Brown, M. A. (2005) 'Evolutionary conservation analysis increases the colocalization of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms', *Breast Cancer Res,* 7, (6), pp. R929-39.

Petz, L. N., Ziegler, Y. S., Schultz, J. R., Kim, H., Kemper, J. K. and Nardulli, A. M. (2004) 'Differential regulation of the human progesterone receptor gene through an estrogen response element half site and Sp1 sites', *J Steroid Biochem Mol Biol,* 88, (2), pp. 113-22.

Piekielko-Witkowska, A., Wiszomirska, H., Wojcicka, A., Poplawski, P., Boguslawska, J., Tanski, Z. and Nauman, A. (2010) 'Disturbed expression of splicing factors in renal cancer affects alternative splicing of apoptosis regulators, oncogenes, and tumor suppressors', *PLoS One,* 5, (10), pp. e13690.

Pierson-Mullany, L. K. and Lange, C. A. (2004) 'Phosphorylation of progesterone receptor serine 400 mediates ligand-independent transcriptional activity in response to activation of cyclin-dependent protein kinase 2', *Mol Cell Biol,* 24, (24), pp. 10542-57.

Pollard, A. J., Krainer, A. R., Robson, S. C. and Europe-Finner, G. N. (2002) 'Alternative splicing of the adenylyl cyclase stimulatory G-protein G alpha(s) is regulated by SF2/ASF and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and involves the use of an unusual TG 3'-splice Site', *J Biol Chem,* 277, (18), pp. 15241-51.

Pollard, A. J., Sparey, C., Robson, S. C., Krainer, A. R. and Europe-Finner, G. N. (2000) 'Spatio-temporal expression of the trans-acting splicing factors SF2/ASF and heterogeneous ribonuclear proteins A1/A1B in the myometrium of the pregnant human uterus: a molecular mechanism for regulating regional protein isoform expression in vivo', *J Clin Endocrinol Metab,* 85, (5), pp. 1928-36.

Poole, A. J., Li, Y., Kim, Y., Lin, S. C., Lee, W. H. and Lee, E. Y. (2006) 'Prevention of Brca1-mediated mammary tumorigenesis in mice by a progesterone antagonist', *Science,* 314, (5804), pp. 1467-70.

Price, T. M., Hansen, E. L. and Oliver, T. N. (2005) 'Immunofluorescent localization of a novel progesterone receptor(s) in a T47D-Y breast cancer cell line lacking genomic progesterone receptor expression', *J Soc Gynecol Investig,* 12, (8), pp. 610-6.

Proietti, C., Salatino, M., Rosemblit, C., Carnevale, R., Pecci, A., Kornblihtt, A. R., Molinolo, A. A., Frahm, I., Charreau, E. H., Schillaci, R. and Elizalde, P. V. (2005) 'Progestins induce transcriptional activation of signal transducer and activator of transcription 3 (Stat3) via a Jak- and Src-dependent mechanism in breast cancer cells', *Mol Cell Biol,* 25, (12), pp. 4826-40.

Proietti, C. J., Rosemblit, C., Beguelin, W., Rivas, M. A., Diaz Flaque, M. C., Charreau, E. H., Schillaci, R. and Elizalde, P. V. (2009) 'Activation of Stat3 by heregulin/ErbB-2 through the co-option of progesterone receptor signaling drives breast cancer growth', *Mol Cell Biol,* 29, (5), pp. 1249-65.

Raponi, M., Kralovicova, J., Copson, E., Divina, P., Eccles, D., Johnson, P., Baralle, D. and Vorechovsky, I. (2011) 'Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6', *Hum Mutat,* 32, (4), pp. 436-44.

Revil, T., Toutant, J., Shkreta, L., Garneau, D., Cloutier, P. and Chabot, B. (2007) 'Protein kinase C-dependent control of Bcl-x alternative splicing', *Mol Cell Biol,* 27, (24), pp. 8431-41.

Rhodes, A. and Jasani, B. (2009) 'The oestrogen receptor-negative/progesterone receptor-positive breast tumour: a biological entity or a technical artefact?', *J Clin Pathol,* 62, (1), pp. 95-6.

Richer, J. K., Jacobsen, B. M., Manning, N. G., Abel, M. G., Wolf, D. M. and Horwitz, K. B. (2002) 'Differential gene regulation by the two progesterone receptor isoforms in human breast cancer cells', *J Biol Chem,* 277, (7), pp. 5209-18.

Richer, J. K., Lange, C. A., Wierman, A. M., Brooks, K. M., Tung, L., Takimoto, G. S. and Horwitz, K. B. (1998) 'Progesterone receptor variants found in breast cells repress transcription by wild-type receptors', *Breast Cancer Res Treat,* 48, (3), pp. 231-41.

Roscigno, R. F. and Garcia-Blanco, M. A. (1995) 'SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome', *RNA,* 1, (7), pp. 692-706.

Rose-Hellekant, T. A., Schroeder, M. D., Brockman, J. L., Zhdankin, O., Bolstad, R., Chen, K. S., Gould, M. N., Schuler, L. A. and Sandgren, E. P. (2007) 'Estrogen receptor-positive mammary tumorigenesis in TGFalpha transgenic mice progresses with progesterone receptor loss', *Oncogene,* 26, (36), pp. 5238-46.

Rosen, P. P. (2001) *Rosen's Breast Pathology.*2nd ed Philadelphia, PA. : Lippincott Williams & Wilkins.

Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A., Howard, B. V., Johnson, K. C., Kotchen, J. M. and Ockene, J. (2002) 'Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial', *JAMA,* 288, (3), pp. 321-33.

Russcher, H., Dalm, V. A., de Jong, F. H., Brinkmann, A. O., Hofland, L. J., Lamberts, S. W. and Koper, J. W. (2007) 'Associations between promoter usage and alternative splicing of the glucocorticoid receptor gene', *J Mol Endocrinol,* 38, (1-2), pp. 91-8.

Russo, J., Ao, X., Grill, C. and Russo, I. H. (1999) 'Pattern of distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland', *Breast Cancer Res Treat,* 53, (3), pp. 217-27.

Russo, J., Mailo, D., Hu, Y. F., Balogh, G., Sheriff, F. and Russo, I. H. (2005) 'Breast differentiation and its implication in cancer prevention', *Clin Cancer Res,* 11, (2 Pt 2), pp. 931s-6s.

Russo, J. and Russo, I. H. (2004) 'Development of the human breast', *Maturitas,* 49, (1), pp. 2-15.

Rybak, J. N., Roesli, C., Kaspar, M., Villa, A. and Neri, D. (2007) 'The extra-domain A of fibronectin is a vascular marker of solid tumors and metastases', *Cancer Res,* 67, (22), pp. 10948-57.

Samalecos, A. and Gellersen, B. (2008) 'Systematic expression analysis and antibody screening do not support the existence of naturally occurring progesterone receptor (PR)-C, PR-M, or other truncated PR isoforms', *Endocrinology,* 149, (11), pp. 5872-87.

Saner, K. J., Welter, B. H., Zhang, F., Hansen, E., Dupont, B., Wei, Y. and Price, T. M. (2003) 'Cloning and expression of a novel, truncated, progesterone receptor', *Mol Cell Endocrinol,* 200, (1-2), pp. 155-63.

Sartorius, C. A., Melville, M. Y., Hovland, A. R., Tung, L., Takimoto, G. S. and Horwitz, K. B. (1994) 'A third transactivation function (AF3) of human progesterone receptors located in the unique N-terminal segment of the B-isoform', *Mol Endocrinol,* 8, (10), pp. 1347-60.

Shav-Tal, Y. and Zipori, D. (2002) 'PSF and p54(nrb)/NonO--multi-functional nuclear proteins', *FEBS Lett,* 531, (2), pp. 109-14.

Shen, H. and Green, M. R. (2004) 'A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly', *Mol Cell,* 16, (3), pp. 363-73.

Shen, H., Kan, J. L. and Green, M. R. (2004) 'Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly', *Mol Cell,* 13, (3), pp. 367-76.

Shen, T., Horwitz, K. B. and Lange, C. A. (2001) 'Transcriptional hyperactivity of human progesterone receptors is coupled to their ligand-dependent down-regulation by mitogen-activated protein kinase-dependent phosphorylation of serine 294', *Mol Cell Biol,* 21, (18), pp. 6122-31.

Shepard, P. J. and Hertel, K. J. (2009) 'The SR protein family', *Genome Biol,* 10, (10), pp. 242.

Shkreta, L., Michelle, L., Toutant, J., Tremblay, M. L. and Chabot, B. (2011) 'The DNA damage response pathway regulates the alternative splicing of the apoptotic mediator Bcl-x', *J Biol Chem,* 286, (1), pp. 331-40.

Shyamala, G., Louie, S. G., Camarillo, I. G. and Talamantes, F. (1999) 'The progesterone receptor and its isoforms in mammary development', *Mol Genet Metab*, 68, (2), pp. 182-90.

Shyamala, G., Yang, X., Silberstein, G., Barcellos-Hoff, M. H. and Dale, E. (1998) 'Transgenic mice carrying an imbalance in the native ratio of A to B forms of progesterone receptor exhibit developmental abnormalities in mammary glands', *Proc Natl Acad Sci U S A*, 95, (2), pp. 696-701.

Silberstein, G. B., Van Horn, K., Shyamala, G. and Daniel, C. W. (1996) 'Progesterone receptors in the mouse mammary duct: distribution and developmental regulation', *Cell Growth Differ*, 7, (7), pp. 945-52.

Silverman, B. G., Siegelmann-Danieli, N., Braunstein, R. and Kokia, E. S. (2011) 'Trends in breast cancer incidence associated with reductions in the use of hormone replacement therapy', *Cancer Epidemiol*, 35, (1), pp. 11-6.

Simian, M., Bissell, M. J., Barcellos-Hoff, M. H. and Shyamala, G. (2009) 'Estrogen and progesterone receptors have distinct roles in the establishment of the hyperplastic phenotype in PR-A transgenic mice', *Breast Cancer Res*, 11, (5), pp. R72.

Sivaraman, L., Conneely, O. M., Medina, D. and O'Malley, B. W. (2001a) 'p53 is a potential mediator of pregnancy and hormone-induced resistance to mammary carcinogenesis', *Proc Natl Acad Sci U S A*, 98, (22), pp. 12379-84.

Sivaraman, L., Hilsenbeck, S. G., Zhong, L., Gay, J., Conneely, O. M., Medina, D. and O'Malley, B. W. (2001b) 'Early exposure of the rat mammary gland to estrogen and progesterone blocks co-localization of estrogen receptor expression and proliferation', *J Endocrinol*, 171, (1), pp. 75-83.

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A. and McGuire, W. L. (1987) 'Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene', *Science*, 235, (4785), pp. 177-82.

Smith, C. L., Onate, S. A., Tsai, M. J. and O'Malley, B. W. (1996) 'CREB binding protein acts synergistically with steroid receptor coactivator-1 to enhance steroid receptor-dependent transcription', *Proc Natl Acad Sci U S A*, 93, (17), pp. 8884-8.

Smith, D. F. (1993) 'Dynamics of heat shock protein 90-progesterone receptor binding and the disactivation loop model for steroid receptor complexes', *Mol Endocrinol*, 7, (11), pp. 1418-29.

Solis, A. S., Peng, R., Crawford, J. B., Phillips, J. A., 3rd and Patton, J. G. (2008) 'Growth hormone deficiency and splicing fidelity: two serine/arginine-rich proteins, ASF/SF2 and SC35, act antagonistically', *J Biol Chem*, 283, (35), pp. 23619-26.

Song, X., Sui, A. and Garen, A. (2004) 'Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: effects on steroidogenesis and oncogenesis', *Proc Natl Acad Sci U S A*, 101, (2), pp. 621-6.

Sorlie, T. (2004) 'Molecular portraits of breast cancer: tumour subtypes as distinct disease entities', *Eur J Cancer,* 40, (18), pp. 2667-75.

Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A. L. (2001) 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc Natl Acad Sci U S A,* 98, (19), pp. 10869-74.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. and Botstein, D. (2003) 'Repeated observation of breast tumor subtypes in independent gene expression data sets', *Proc Natl Acad Sci U S A,* 100, (14), pp. 8418-23.

Springwald, A., Lattrich, C., Skrzypczak, M., Goerse, R., Ortmann, O. and Treeck, O. (2010) 'Identification of novel transcript variants of estrogen receptor alpha, beta and progesterone receptor gene in human endometrium', *Endocrine,* 37, (3), pp. 415-24.

Srebrow, A. and Kornblihtt, A. R. (2006) 'The connection between splicing and cancer', *J Cell Sci,* 119, (Pt 13), pp. 2635-41.

Staley, J. P. and Guthrie, C. (1998) 'Mechanical devices of the spliceosome: motors, clocks, springs, and things', *Cell,* 92, (3), pp. 315-26.

Stamm, S. (2002) 'Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome', *Hum Mol Genet,* 11, (20), pp. 2409-16.

Stamm, S., Riethoven, J. J., Le Texier, V., Gopalakrishnan, C., Kumanduri, V., Tang, Y., Barbosa-Morais, N. L. and Thanaraj, T. A. (2006) 'ASD: a bioinformatics resource on alternative splicing', *Nucleic Acids Res,* 34, (Database issue), pp. D46-55.

Takimoto, G. S., Hovland, A. R., Tasset, D. M., Melville, M. Y., Tung, L. and Horwitz, K. B. (1996) 'Role of phosphorylation on DNA binding and transcriptional functions of human progesterone receptors', *J Biol Chem,* 271, (23), pp. 13308-16.

Taylor, A. H., McParland, P. C., Taylor, D. J. and Bell, S. C. (2009) 'The cytoplasmic 60 kDa progesterone receptor isoform predominates in the human amniochorion and placenta at term', *Reprod Biol Endocrinol,* 7, pp. 22.

Taylor, S. E., Patel, II, Singh, P. B., Nicholson, C. M., Stringfellow, H. F., Gopala Krishna, R. K., Matanhelia, S. S., Martin-Hirsch, P. L. and Martin, F. L. (2010) 'Elevated oestrogen receptor splice variant ERalphaDelta5 expression in tumour-adjacent hormone-responsive tissue', *Int J Environ Res Public Health,* 7, (11), pp. 3871-89.

Tejedor, J. R. and Valcarcel, J. (2010) 'Gene regulation: Breaking the second genetic code', *Nature,* 465, (7294), pp. 45-6.

Tetel, M. J., Giangrande, P. H., Leonhardt, S. A., McDonnell, D. P. and Edwards, D. P. (1999) 'Hormone-dependent interaction between the amino- and carboxyl-terminal domains of progesterone receptor in vitro and in vivo', *Mol Endocrinol*, 13, (6), pp. 910-24.

Tetel, M. J., Jung, S., Carbajo, P., Ladtkow, T., Skafar, D. F. and Edwards, D. P. (1997) 'Hinge and amino-terminal sequences contribute to solution dimerization of human progesterone receptor', *Mol Endocrinol*, 11, (8), pp. 1114-28.

Thomas, P. (2008) 'Characteristics of membrane progestin receptor alpha (mPRalpha) and progesterone membrane receptor component 1 (PGMRC1) and their roles in mediating rapid progestin actions', *Front Neuroendocrinol*, 29, (2), pp. 292-312.

Tomlinson, I. P., Nicolai, H., Solomon, E. and Bodmer, W. F. (1996) 'The frequency and mechanism of loss of heterozygosity on chromosome 11q in breast cancer', *J Pathol*, 180, (1), pp. 38-43.

Topper, Y. J. and Freeman, C. S. (1980) 'Multiple hormone interactions in the developmental biology of the mammary gland', *Physiol Rev*, 60, (4), pp. 1049-106.

Tsai, M. J. and O'Malley, B. W. (1994) 'Molecular mechanisms of action of steroid/thyroid receptor superfamily members', *Annu Rev Biochem*, 63, pp. 451-86.

Tung, L., Abdel-Hafiz, H., Shen, T., Harvell, D. M., Nitao, L. K., Richer, J. K., Sartorius, C. A., Takimoto, G. S. and Horwitz, K. B. (2006) 'Progesterone receptors (PR)-B and -A regulate transcription by different mechanisms: AF-3 exerts regulatory control over coactivator binding to PR-B', *Mol Endocrinol*, 20, (11), pp. 2656-70.

Tung, L., Shen, T., Abel, M. G., Powell, R. L., Takimoto, G. S., Sartorius, C. A. and Horwitz, K. B. (2001) 'Mapping the unique activation function 3 in the progesterone B-receptor upstream segment. Two LXXLL motifs and a tryptophan residue are required for activity', *J Biol Chem*, 276, (43), pp. 39843-51.

Tyson-Capper, A. J., Bailey, J., Krainer, A. R., Robson, S. C. and Europe-Finner, G. N. (2005) 'The switch in alternative splicing of cyclic AMP-response element modulator protein CREM{tau}2{alpha} (activator) to CREM{alpha} (repressor) in human myometrial cells is mediated by SRp40', *J Biol Chem*, 280, (41), pp. 34521-9.

Tyson-Capper, A. J., Shiells, E. A. and Robson, S. C. (2009) 'Interplay between polypyrimidine tract binding protein-associated splicing factor and human myometrial progesterone receptors', *J Mol Endocrinol*, 43, (1), pp. 29-41.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J. and Darnell, R. B. (2006) 'An RNA map predicting Nova-dependent splicing regulation', *Nature*, 444, (7119), pp. 580-6.

van Alphen, R. J., Wiemer, E. A., Burger, H. and Eskens, F. A. (2009) 'The spliceosome as target for anticancer treatment', *Br J Cancer,* 100, (2), pp. 228-32.

van Dijk, M. A., Hart, A. A. and van 't Veer, L. J. (2000) 'Differences in estrogen receptor alpha variant messenger RNAs between normal human breast tissue and primary breast carcinomas', *Cancer Res,* 60, (3), pp. 530-3.

Vegeto, E., Shahbaz, M. M., Wen, D. X., Goldman, M. E., O'Malley, B. W. and McDonnell, D. P. (1993) 'Human progesterone receptor A form is a cell- and promoter-specific repressor of human progesterone receptor B function', *Mol Endocrinol,* 7, (10), pp. 1244-55.

Venables, J. P. (2004) 'Aberrant and alternative splicing in cancer', *Cancer Res,* 64, (21), pp. 7647-54.

Venables, J. P. (2007) 'Downstream intronic splicing enhancers', *FEBS Lett,* 581, (22), pp. 4127-31.

Venables, J. P., Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Koh, C., Gervais-Bird, J., Lapointe, E., Froehlich, U., Durand, M., Gendron, D., Brosseau, J. P., Thibault, P., Lucier, J. F., Tremblay, K., Prinos, P., Wellinger, R. J., Chabot, B., Rancourt, C. and Elela, S. A. (2008) 'Identification of alternative splicing markers for breast cancer', *Cancer Res,* 68, (22), pp. 9525-31.

Venables, J. P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., Lucier, J. F., Thibault, P., Rancourt, C., Tremblay, K., Prinos, P., Chabot, B. and Elela, S. A. (2009) 'Cancer-associated regulation of alternative splicing', *Nat Struct Mol Biol,* 16, (6), pp. 670-6.

Venkitaraman, A. R. (2002) 'Cancer susceptibility and the functions of BRCA1 and BRCA2', *Cell,* 108, (2), pp. 171-82.

Vezain, M., Saugier-Veber, P., Goina, E., Touraine, R., Manel, V., Toutain, A., Fehrenbach, S., Frebourg, T., Pagani, F., Tosi, M. and Martins, A. (2010) 'A rare SMN2 variant in a previously unrecognized composite splicing regulatory element induces exon 7 inclusion and reduces the clinical severity of spinal muscular atrophy', *Hum Mutat,* 31, (1), pp. E1110-25.

Vranic, S., Gatalica, Z., Deng, H., Frkovic-Grazio, S., Lee, L. M., Gurjeva, O. and Wang, Z. Y. (2011) 'ER-alpha36, a novel isoform of ER-alpha66, is commonly over-expressed in apocrine and adenoid cystic carcinomas of the breast', *J Clin Pathol,* 64, (1), pp. 54-7.

Wan, J., Sazani, P. and Kole, R. (2009) 'Modification of HER2 pre-mRNA alternative splicing and its effects on breast cancer cells', *Int J Cancer,* 124, (4), pp. 772-7.

Wan, Y., Coxe, K. K., Thackray, V. G., Housley, P. R. and Nordeen, S. K. (2001) 'Separable features of the ligand-binding domain determine the differential subcellular localization and ligand-binding specificity of glucocorticoid receptor and progesterone receptor', *Mol Endocrinol,* 15, (1), pp. 17-31.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. and Burge, C. B. (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature,* 456, (7221), pp. 470-6.

Wang, Z., Zhang, X., Shen, P., Loggie, B. W., Chang, Y. and Deuel, T. F. (2005) 'Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66', *Biochem Biophys Res Commun,* 336, (4), pp. 1023-7.

Wang, Z., Zhang, X., Shen, P., Loggie, B. W., Chang, Y. and Deuel, T. F. (2006) 'A variant of estrogen receptor-{alpha}, hER-{alpha}36: transduction of estrogen- and antiestrogen-dependent membrane-initiated mitogenic signaling', *Proc Natl Acad Sci U S A,* 103, (24), pp. 9063-8.

Ward, A. J. and Cooper, T. A. (2010) 'The pathobiology of splicing', *J Pathol,* 220, (2), pp. 152-63.

Ward, R. D. and Weigel, N. L. (2009) 'Steroid receptor phosphorylation: Assigning function to site-specific phosphorylation', *Biofactors,* 35, (6), pp. 528-36.

Wardell, S. E., Boonyaratanakornkit, V., Adelman, J. S., Aronheim, A. and Edwards, D. P. (2002) 'Jun dimerization protein 2 functions as a progesterone receptor N-terminal domain coactivator', *Mol Cell Biol,* 22, (15), pp. 5451-66.

Warnmark, A., Treuter, E., Wright, A. P. and Gustafsson, J. A. (2003) 'Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation', *Mol Endocrinol,* 17, (10), pp. 1901-9.

Wei, L. L., Gonzalez-Aller, C., Wood, W. M., Miller, L. A. and Horwitz, K. B. (1990) '5'-Heterogeneity in human progesterone receptor transcripts predicts a new amino-terminal truncated "C"-receptor and unique A-receptor messages', *Mol Endocrinol,* 4, (12), pp. 1833-40.

Wei, L. L., Hawkins, P., Baker, C., Norris, B., Sheridan, P. L. and Quinn, P. G. (1996) 'An amino-terminal truncated progesterone receptor isoform, PRc, enhances progestin-induced transcriptional activity', *Mol Endocrinol,* 10, (11), pp. 1379-87.

Wei, L. L. and Miner, R. (1994) 'Evidence for the existence of a third progesterone receptor protein in human breast cancer cell line T47D', *Cancer Res,* 54, (2), pp. 340-3.

Wei, L. L., Norris, B. M. and Baker, C. J. (1997) 'An N-terminally truncated third progesterone receptor protein, PR(C), forms heterodimers with PR(B) but interferes in PR(B)-DNA binding', *J Steroid Biochem Mol Biol,* 62, (4), pp. 287-97.

Wiehle, R. D., Christov, K. and Mehta, R. (2007) 'Anti-progestins suppress the growth of established tumors induced by 7,12-dimethylbenz(a)anthracene: comparison between RU486 and a new 21-substituted-19-nor-progestin', *Oncol Rep,* 18, (1), pp. 167-74.

Wu, J. Y. and Maniatis, T. (1993) 'Specific interactions between proteins implicated in splice site selection and regulated alternative splicing', *Cell,* 75, (6), pp. 1061-70.

Yamanaka, T., Hirata, S., Shoda, T. and Hoshi, K. (2002) 'Progesterone receptor mRNA variant containing novel exon insertions between exon 4 and exon 5 in human uterine endometrium', *Endocr J,* 49, (4), pp. 473-82.

Yarden, Y. (2001) 'The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities', *Eur J Cancer,* 37 Suppl 4, pp. S3-8.

Yeates, C., Hunt, S. M., Balleine, R. L. and Clarke, C. L. (1998) 'Characterization of a truncated progesterone receptor protein in breast tumors', *J Clin Endocrinol Metab,* 83, (2), pp. 460-7.

Yudt, M. R., Berrodin, T. J., Jelinsky, S. A., Hanna, L. A., Brown, E. L., Chippari, S., Bhat, R. A., Winneker, R. C. and Zhang, Z. (2006) 'Selective and opposing actions of progesterone receptor isoforms in human endometrial stromal cells', *Mol Cell Endocrinol,* 247, (1-2), pp. 116-26.

Zahler, A. M., Neugebauer, K. M., Lane, W. S. and Roth, M. B. (1993) 'Distinct functions of SR proteins in alternative pre-mRNA splicing', *Science,* 260, (5105), pp. 219-22.

Zhang, X., Moor, A. N., Merkler, K. A., Liu, Q. and McLean, M. P. (2007) 'Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors', *Endocrinology,* 148, (11), pp. 5295-304.

Zhang, X. T., Kang, L. G., Ding, L., Vranic, S., Gatalica, Z. and Wang, Z. Y. (2011) 'A positive feedback loop of ER-alpha36/EGFR promotes malignant growth of ER-negative breast cancer cells', *Oncogene,* 30, (7), pp. 770-80.

Zhou, Z., Licklider, L. J., Gygi, S. P. and Reed, R. (2002) 'Comprehensive proteomic analysis of the human spliceosome', *Nature,* 419, (6903), pp. 182-5.

Zhu, J., Mayeda, A. and Krainer, A. R. (2001) 'Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins', *Mol Cell,* 8, (6), pp. 1351-61.

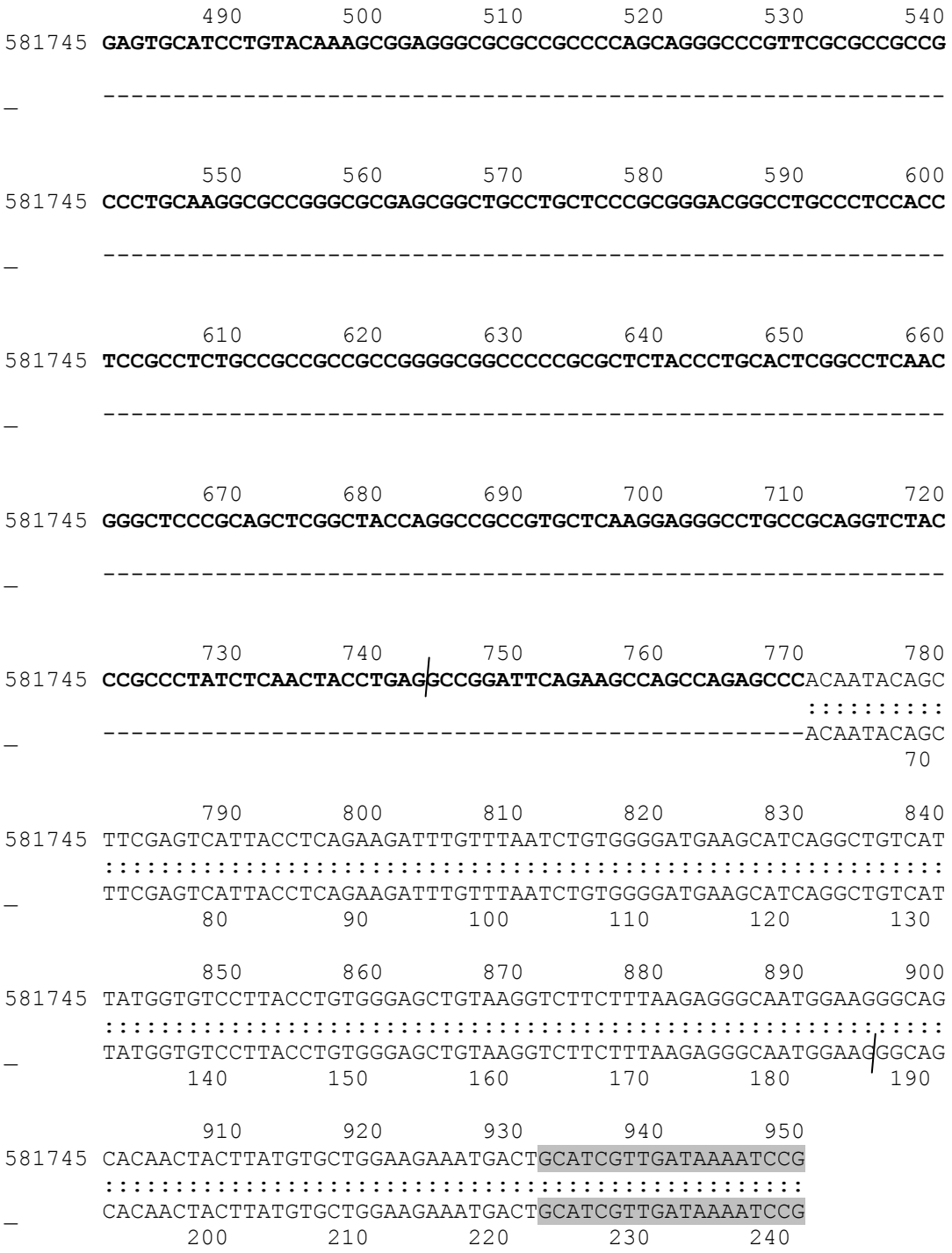
Zhu, N., Eghbali, M., Helguera, G., Song, M., Stefani, E. and Toro, L. (2005) 'Alternative splicing of Slo channel gene programmed by estrogen, progesterone and pregnancy', *FEBS Lett,* 579, (21), pp. 4856-60.

Appendix 1

Examples of DNA sequence analysis using ALIGN. Examples shown illustrate ASEs identified in both MDA-MB-231 cells and in breast tumour tissue.

Δp1,p2 S4/A6

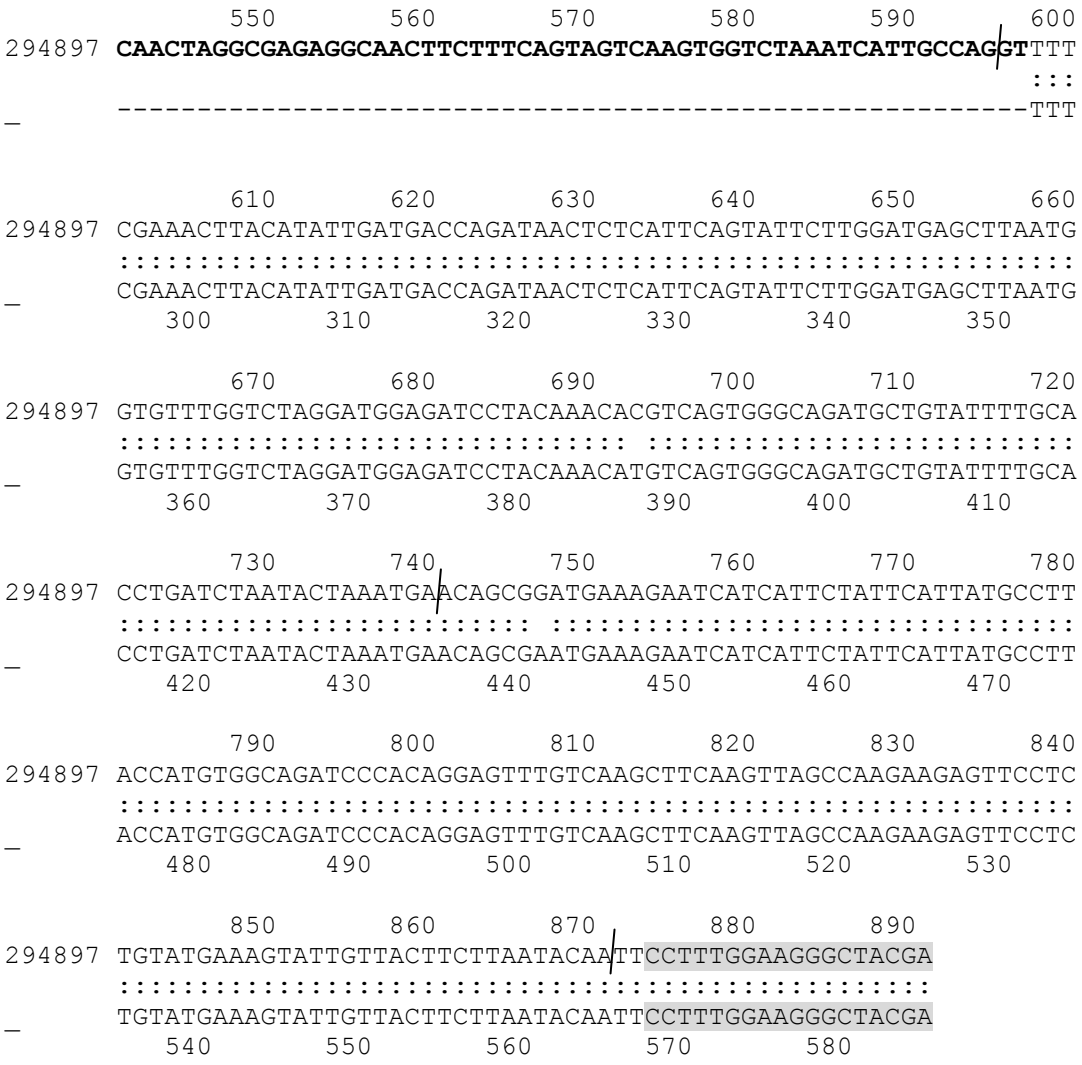
```
>_                                                        950 nt vs.
>_                                                        241 nt
scoring matrix: , gap penalties: -12/-2
25.4% identity;        Global alignment score: -464

               10        20        30        40        50        60
581745 ACGGTGATGGATTTCATCCACGTGCCTATCCTGCCTCTCAATCACGCCTTATTGGCAGCC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
_      ACGGTGATGGATTTCATCCACGTGCCTATCCTGCCTCTCAATCACGCCTTATTGGCAGCC
               10        20        30        40        50        60

               70        80        90       100       110       120
581745 CGCACTCGGCAGCTGCTGGAAGACGAAAGTTACGACGGCGGGGCCGGGGCTGCCAGCGCC
       :
_      C-------------------------------------------------------------

              130       140       150       160       170       180
581745 TTTGCCCCGCCGCGGAGTTCACCCTGTGCCTCGTCCACCCCGGTCGCTGTAGGCGACTTC

_      -------------------------------------------------------------

              190       200       210       220       230       240
581745 CCCGACTGCGCGTACCCGCCCGACGCCGAGCCCAAGGACGACGCGTACCCTCTCTATAGC

_      -------------------------------------------------------------

              250       260       270       280       290       300
581745 GACTTCCAGCCGCCCGCTCTAAAGATAAAGGAGGAGGAGGAAGGCGCGGAGGCCTCCGCG

_      -------------------------------------------------------------

              310       320       330       340       350       360
581745 CGCTCCCCGCGTTCCTACCTTGTGGCCGGTGCCAACCCCGCAGCCTTCCCGGATTTCCCG

_      -------------------------------------------------------------


              370       380       390       400       410       420
581745 TTGGGGCCACCGCCCCCGCTGCCGCCGCGAGCGACCCCATCCAGACCCGGGGAAGCGGCG

_      -------------------------------------------------------------


              430       440       450       460       470       480
581745 GTGACGGCCGCACCCGCCAGTGCCTCAGTCTCGTCTGCGTCCTCCTCGGGGTCGACCCTG

_      -------------------------------------------------------------
```



Primer sequences shaded grey, deleted region in bold font; 682bp 3′ end of exon 1 and 27bp 5′ end of exon 2 deleted. Exon 1/2 and 2/3 boundaries indicated (|).

Δ4 SA/A2

```
>                                                      891 nt vs.
>_                                                     585 nt
scoring matrix: , gap penalties: -12/-2
65.3% identity;        Global alignment score: 1697

               10        20        30        40        50        60
294897 CCCTATCTCAACTACCTGAGGCCGGATTCAGAAGCCAGCCAGAGCCCACAATACAGCTTC
       :::::::::::::::::::::::::::::::::::::::::::::::::::: :::::
_      CCCTATCTCAACTACCTGAGGCCGGATTCAGAAGCCAGCCAGAGCCCACAATACGGCTTC
               10        20        30        40        50        60

               70        80        90       100       110       120
294897 GAGTCATTACCTCAGAAGATTTGTTTAATCTGTGGGGATGAAGCATCAGGCTGTCATTAT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
_      GAGTCATTACCTCAGAAGATTTGTTTAATCTGTGGGGATGAAGCATCAGGCTGTCATTAT
               70        80        90       100       110       120

              130       140       150       160       170       180
294897 GGTGTCCTTACCTGTGGGAGCTGTAAGGTCTTCTTTAAGAGGGCAATGGAAGGGCAGCAC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
_      GGTGTCCTTACCTGTGGGAGCTGTAAGGTCTTCTTTAAGAGGGCAATGGAAGGGCAGCAC
              130       140       150       160       170       180

              190       200       210       220       230       240
294897 AACTACTTATGTGCTGGAAGAAATGACTGCATCGTTGATAAAATCCGCAGAAAAAACTGC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
_      AACTACTTATGTGCTGGAAGAAATGACTGCATCGTTGATAAAATCCGCAGAAAAAACTGC
              190       200       210       220       230       240

              250       260       270       280       290       300
294897 CCAGCATGTCGCCTTAGAAAGTGCTGTCAGGCTGGCATGGTCCTTGGAGGTCGAAAATTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::
_      CCAGCATGTCGCCTTAGAAAGTGCTGTCAGGCTGGCATGGTCCTTGGAGGT---------
              250       260       270       280       290

              310       320       330       340       350       360
294897 AAAAAGTTCAATAAAGTCAGAGTTGTGAGAGCACTGGATGCTGTTGCTCTCCCACAGCCA

_      ------------------------------------------------------------

              370       380       390       400       410       420
294897 GTGGGCGTTCCAAATGAAAGCCAAGCCCTAAGCCAGAGATTCACTTTTTCACCAGGTCAA

_      ------------------------------------------------------------

              430       440       450       460       470       480
294897 GACATACAGTTGATTCCACCACTGATCAACCTGTTAATGAGCATTGAACCAGATGTGATC

_      ------------------------------------------------------------

              490       500       510       520       530       540
294897 TATGCAGGACATGACAACACAAAACCTGACACCTCCAGTTCTTTGCTGACAAGTCTTAAT

_      ------------------------------------------------------------
```



Primer sequences shaded grey, deleted region in bold font; entire 306bp exon 4 sequence. Exon 1/2, 2/3, 3/4, 4/5, 5/6 and 6/7 boundaries indicated (|).

Appendix 2

Sample No.	ER	PR	HER2	Grade	Nodes	NPI	Patient age	Other information from histology report
1	+	18	n/a	Lobular G2	1/5	4.5	65	
2	+	6	+	Ductal G2	1/4	4.36	59	
3	+	15	n/a	Ductal G2	3/9	5.6	70	
4	+	+	n/a	Ductal G2	4/10	5.48	58	PR assessed on previous biopsy
5	+	15	-	Ductal G3	11/11	6.5	50	
6	-	12	+	Ductal G3	10/15	7.8	46	
7	+	0	n/a	Ductal G1/2	0	3.4	66	Multiple foci some G1, some G2
8	+	0	n/a	Ductal G2	3/8	4.66	51	
9	+	0	n/a	Ductal G1	n/a	n/a	81	
10	+	0	n/a	DCIS	0	n/a	72	
11	+	0	-	Ductal G3	n/a	n/a	52	
12	-	-	n/a	DCIS	n/a	n/a	45	PR not mentioned
13	-	0	+	Ductal G4	0/10	n/a	63	
14	-	-	n/a	DCIS	n/a	n/a	56	PR scored from previous biopsy, nodes and NPI not mentioned
15	-	0	+	Ductal G2	2/8	n/a	68	Residual post chemotherapy, therefore no NPI

Information extracted from histology reports for clinical material used in this study.

+/- status indicated when PR Quickscore not recorded. n/a= not available; HER2 not assessed for all samples, node involvement and NPI not routinely recorded.

NPI = Nottingham prognostic index

NPI = [0.2 x S] + N + G

S = size of the index lesion in cm

N = number of lymph nodes involved: 0 nodes =1, 1-3 nodes = 2, >3 nodes = 3

G = grade of tumour: Grade I =1, Grade II =2, Grade III =3

≥2.0 to ≤2.4, 5yr survival 93%;

>2.4 to ≤3.4, 5yr survival 85%;

>3.4 to ≤5.4, 5yr survival 70%;

≥5.4, 5yr survival 50%.

Appendix 3

Publications, presentations and prizes arising from this study

Publications

Cork, D.M., Lennard, T.W. and Tyson-Capper, A.J. 'Alternative splicing generates functional truncated progesterone receptor proteins in breast cancer cells' (Submitted).

Cork, D.M., Lennard, T.W. and Tyson-Capper, A.J. (2010) 'Alternatively spliced progesterone receptor expression in breast cancer' (http://www.isge2010.com.)

Cork, D.M., Lennard, T.W. and Tyson-Capper, A.J. (2009) 'Progesterone Receptor Variants in Breast Cancer', *Reproductive Sciences,* 16, (3) (Supplement) A268.

Cork, D. M., Lennard, T. W. and Tyson-Capper, A. J. (2008) 'Alternative splicing and the progesterone receptor in breast cancer', *Breast Cancer Res,* 10, (3), pp. 207 (Included at the end of Appendix 3).

Presentations

23/06/2010	Non-genomic progesterone receptor expression in breast cancer. (Oral presentation, ICM Research Day, Newcastle University)
05/03/2010	Alternatively spliced progesterone receptor expression in breast cancer. (Oral presentation, 14th World Congress of Gynecological Endocrinology, Florence, Italy)
11/01/2010	Alternative splicing and the progesterone receptor in breast cancer. (Oral presentation, ICM Science Forum, Newcastle University)
12/05/2009	Progesterone Receptor Variants in Breast Cancer. (Oral presentation, RNA Interest Group, Newcastle University)
17/03/2009	Progesterone Receptor Variants in Breast Cancer. (Poster presentation, SGI, 56th Annual Meeting, Glasgow)

23/01/2009 Progesterone Receptor Variants in Breast Cancer Cell Lines. (Oral presentation, EURASNET UK RNA Splicing Workshop, Rydal Hall, Cumbria)

28/07/2008 Alternative Splicing and the Progesterone Receptor in Breast Cancer. (Poster presentation, RNA2008, The RNA Society (USA) Annual Congress, Berlin)

30/06/2008 Characterisation of Progesterone Receptors in Breast Cancer: Potential for Novel Diagnostic and Prognostic Markers. (Oral presentation, School of Surgical & Reproductive Sciences 'Science Forum', Newcastle University).

Prizes

23/06/2010 'Speaker's Prize' for best student oral presentation at the ICM Research Day, Newcastle University.

04/03/2010 Young Scientist Travel and Conference Registration Award, International Society for Gynecological Endocrinology Abstract Competition for the 14th World Congress of Gynecological Endocrinology.

27/11/2008 2nd Prize, ICM Poster Competition, Newcastle University.