Molecular Gene Expression and Genome Wide Profiling in Tamoxifen-Resistant Breast Cancer.

Chit Cheng YEOH

July 2010

PhD report

Supervisor:

Professor Helen Hurst

Cancer Research UK Tumour Biology Unit

Barts and The London School of Medicine and Dentistry

John Vane Science Centre

Charterhouse Square

London EC1M 6BQ

DECLARATION

I declare that the work presented in this thesis is my own, except where stated otherwise in the text and this work has not been submitted for any other degree or professional qualification except as specified. The work was performed between September 2006 and June 2010 in the Centre for Tumour Biology, Institute of Cancer, Barts and The London School of Medicine and Dentistry, Queen Mary University of London.

Chit Cheng YEOH
July 2010

‘Every war is won by surmounting the little battles’

This is my small contribution to our war against Cancer.

This is for you, Amah.

ABSTRACT

Oestrogen receptor positive (ER+) breast cancers (BC) are heterogeneous in both their clinical behaviour and response to therapy. The ER and Progesterone (PgR) are currently the best predictors of response to the anti-oestrogen tamoxifen, yet up to 40% of ER+ breast cancer will relapse despite tamoxifen treatment. New prognostic biomarkers and further biological understanding of tamoxifen resistance (TR) are required. There has been an explosion of greater understanding since the arrival of cutting-edge gene and genomic profiling technology. The two major aims of this research are to develop stable gene signatures that are effective at distinguishing 'prognostic' groups and, when tested directly for response to tamoxifen, a set of 'predictive' markers.

In order to establish cellular pathways responsible for TR, tissue at relapse while on tamoxifen is preferred. However, in practice, this is difficult to obtain. Hence, in this study, I have established TR derivatives of breast cancer cell lines, T47D and ZR75-1, and analysed their gene-expression by microarray. *MAGEA2* and *EGLN3* were 4.0 and 3.8 fold upregulated respectively in TR cell lines. For MAGEA2- and EGLN3-overexpressing lines, the proliferation and growth rates in tamoxifen-containing media were significantly higher (p-value <0.001 and $p<0.05$, respectively) than for control cells. I have investigated possible downstream targets for each protein which may contribute to the mechanism of resistance. Immunohistochemistry validation was performed on a cohort of 196 tamoxifen-treated primary breast tumour tissues: MAGEA2 and EGLN3 were found to be valuable predictive (Positive predictive value of 89%, and 85%, with high sensitivity 38% and 42% respectively) biomarkers for TR in primary breast tumours.

In the human breast tumour arm of this study, 25 frozen samples with known response to tamoxifen were analysed on both SNP6.0 and expression EXON arrays. The integrated analysis suggested that 5 genes (*OPCML, OR10G7, SNF1LK2, PALM and ZBTB-16*) are good predictors of TR, with high negative predictor values (68%, 71%, 59% and 73% respectively for the last 4 genes). Significant regions of copy number variation (CNV) were identified at chromosomes 8q24, 17q21-22 and 11q23-25. The application of this high-resolution approach should lead to a better understanding of the roles of complex genetic alterations in TR.

ACKNOWLEDGEMENTS

To the women in my life: Mum, Ghee and Hui
– Thank-you for showing me what strong women are made of.

To the men of my life: Dad, Anselm, ***Jaap*** and my boys, Alexander and Victor
– Thank-you for bringing out the best in me (what love is capable of).

To Prof Helen Hurst, for her guidance, wisdom, compassion and brilliance.

To Prof Ian Hart, for exemplary leadership, and tenacious work etiquettes.

To Dr Claude Chelala, Dr Amar Ahmad and Dr Tony Wong, for their interest and input into part of my thesis.

To my loyal friends: Jee Wei, Mei Ling & Tim, Pauline, Claudine, Shook Hsia, Ernest & Annabel, Fabian & Rachel, Siang & Chen, Edward & Eunice, Chiat & Jacqueline, Jessie & Lena, Toyin
– Who supported me at all times.

To good colleagues: Chiara Berlato, Ka-yi Chan, Monica Canosa, Angelo Scibetta, Sayka Barry, Chia-Yu Chen, Anne-Marie Young, Ludo Menard, Katie Goodey, Stephen Robinson, Debbie Buckle, Fieke Froeling, George Elia, Ming Yuan, Namaa Audi and Keyur Trivedi.
– For making our workplace so enjoyable.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

APPENDICES

Appendix A - 200 significantly altered genes from microarray expression chip (Affymetrix HU133plus 2.0) in Tamoxifen resistant breast cancer cell lines (T47D TR, ZR 75-1 TR, OD T47D TR and OD ZR TR) after normalising with WT.

Appendix B - 100 significantly altered tamoxifen resistant gene expression for predictive gene from microarray expression chip (EXON 1.0ST human) in 25 human primary breast cancer tissues from patients with Tamoxifen adjuvant therapy.

GLOSSARY OF ABBREVIATIONS

Ab	Antibody
AI	Aromatase Inhibitors
AP-1	Activation Protein-1
ASCL1	Achaete-Scute Complex homolog-Like 1
BSA	Bovine serum albumin
BNIP3	Bcl2 Nineteen kDa Interacting Protein 3
BP	Base pair
cDNA	Complementary DNA
DAPI	4',6-diamidino-2-phenylindole
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulphoxide
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamine tetraacetate
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EGLN	Egl Nine homolog, PHD
ER	Oestrogen Receptors
ERE	Oestrogen Response Element
ERK	Extracellular signal Regulated Kinase
FBS	Fetal Bovine Serum
FCS	Foetal calf serum
FDR	False Discovery Rate
FITC	Fluorescein isothiocyanate
GAPDH	Glucose -3-phosphate dehydrogenase
GLUT1	Glucose Transporter 1
GREB1	Gene Regulated in Breast Cancer 1
GO	Gene Ontology
HER2	Human Epidermal growth factor Receptor-2
HIF	Hypoxia Inducible Factor
HRP	Horseradish Peroxidase
IGFBP3	Insuline-like Growth Factor Binding Protein
IGFR	Insuline-like Growth Factor Receptor
IPA	Ingenuity Pathway
Kb	Kilobase
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MAGEA2	Melanoma Antigen family A-2
MAPK	Mitogen Activated Protein Kinase
mRNA	Messenger RNA
miRNA	MicroRNA
MPPED2	Metallophosphoesterase domain containing 2
NM	Normal Media
OD	Oestrogen Deprived
ODD	Oxygen Dependent Degradation
OD	Optical density

PAGE	Poly Acrylamide Gel Electrophoresis
PCR	Polymerase chain reaction
PBS	Phosphate-buffered saline
PAGE	PolyAcrylamide Gel Electrophoresis
PDZK1	PDZ containing domain 1
pVHL	von Hippel-Lindau Protein
PHD	Prolyl Hydroxylase, EGLN
PI	Propodium iodide
PI3K	Phosphoinositide 3-kinase
PgR	Progesterone Receptor
PS	Phosphatidylserine
PDAC	Pancreatic ductal adenocarcinoma
PSC	Pancreatic stellate cells
qPCR	quantitative (Real Time) Polymerase Chain Reaction
RT	Reverse Transcript
RMA	Robust Multi-array Average
RNA	Ribonucleic acid
RNase	Ribonuclease
SDS	Sodium dodecyl sulphate
SNP	Single Nucleotide Polymorphism
siRNA	Short interfering RNA
SRB	Sulphohodamine assay
TEX14	Testis Expressed 14
TR	Tamoxifen Resistant, TamR
TS	Tamoxifen sensitive
TFF1	Trefoil Factor 1
TR	Tamoxifen resistant
TS	Tamoxifen sensitive
UPD	Uniparent Disomy
VA	Vector alone
VEGF	Vascular Endothelial Growth Factor
WCL	Whole cell lysates
Wt	wild type

UNITS OF CONCENTRATION

M	Molar (moles/litre)
mM	Millimolar (millimoles/litre)
µM	Micromolar (micromoles/litre)
nM	Nanomolar (nanomoles/litre)
pM	Picomolar (picomoles/litre)

UNITS OF LENGTH, AREA, VOLUME, MASS, TIME

m	Metre
cm	Centimetre
mm	Millimetre
µm	Micrometre
nm	Nanometre
ml	Millilitre
µl	microlitre
gr	Gram
µg	Microgram
kg	Kilogram
h	Hour
min	Minute
s	Second

CHAPTER 1: INTRODUCTION

1. Introduction

1.1. Breast Cancer

Breast cancer is the most common cause of cancer in women and the second most common cause of cancer death in women in the United Kingdom. While the majority of new breast cancers are diagnosed as a result of an abnormality seen on a mammogram, a lump or change in consistency of the breast tissue can also be a warning sign of the disease. Heightened awareness of breast cancer risk in the past decades has led to an increase in the number of women undergoing mammography for screening, leading to detection of cancers at earlier stages and a resultant improvement in survival rates. Still, breast cancer is the most common cause of death in women between the ages of 45 and 55 and a woman has a 1 in 9 lifetime risk to be afflicted by breast cancer. Although breast cancer in women is a common form of cancer, male breast cancer does occur and accounts for about 1% of all cancer deaths in men.

Research has yielded much information about the causes of breast cancer, and it is now believed that genetic and/or hormonal factors are the primary risk factors for this disease (Hemminki *et al.*, 2008). Staging systems have been developed to allow doctors to characterize the extent to which a particular cancer has spread and to make decisions concerning treatment options. Breast cancer treatment depends upon many factors, including the type of cancer and the extent to which it has spread. Treatment options for breast cancer may involve surgery (removal of the cancer alone or, in some cases, mastectomy), radiation therapy, hormonal therapy, and chemotherapy (Moulder and Hortobagyi, 2008).

With advances in screening, diagnosis, and treatment, the death rate for breast cancer has declined by about 20% over the past decade, and research is ongoing to develop even more effective screening and treatment programs.

The exact cause of breast cancer is unknown. Science cannot explain why one woman develops breast cancer and another does not. Research has shown that women with certain risk factors are more likely than others to develop this disease.

Studies have found the following risk factors for breast cancer:

1) The chance of getting breast cancer increases as a woman gets older. The average age breast cancer occur in women is 60. This disease is not common before menopause (Scalliet and Kirkove, 2007).

2) A woman who had breast cancer in one breast has an increased risk of getting cancer in her other breast (Meteoglu *et al.*, 2005).

3) A woman's risk of breast cancer is higher if her mother, sister, or daughter had breast cancer. The risk is higher if her family member got breast cancer before age 40. Having other relatives with breast cancer (in either her mother's or father's family) may also increase a woman's risk (Draper et al, 2006).

4) Some women have cells in the breast that look abnormal under a microscope. Having certain types of abnormal cells (atypical hyperplasia and lobular carcinoma in situ [LCIS] increases the risk of breast cancer (Afonso and Bouwman, 2008).

5) Changes in certain genes increase the risk of breast cancer. These genes include BRCA1, BRCA2, and others (Tikhomirova *et al.*, 2007), (Domchek *et al.*), (Troudi *et al.*, 2008). Tests can sometimes show the presence of specific gene changes in families with many women who have had breast cancer. Health care providers may suggest ways to try to reduce the risk of breast cancer, or to improve the detection of this disease in women who have these changes in their genes.

6) Reproductive and menstrual history:

    - If a woman has their first child after 35 years or older, the greater her chance of breast cancer (Afonso and Bouwman, 2008).

    - Women who have early menarche before age 12 are at an increased risk of breast cancer **(Merki-Feld et al, 2008).**

- Women who have late menopause (after age 55) are at an increased risk of breast cancer (Harvey *et al.*, 2008), (Johansson *et al.*, 2008).

- Women who never had children are at an increased risk of breast cancer (Harvey *et al.*, 2008), (Henderson *et al.*, 2008), (Harvey *et al.*, 2008).

- Women who take menopausal hormone therapy with oestrogen plus progestin after menopause also appear to have an increased risk of breast cancer (Barnett *et al.*, 2008).

- Large, well-designed studies have shown no link between abortion or miscarriage and breast cancer **(Brind et al, 2008).**

7) Breast cancer is diagnosed more often in white women than Latina, Asian, or African American women (Markman *et al.*, 2008).

8) Women who had been exposed to radiation therapy to the chest (including breasts) before age 30 are at an increased risk of breast cancer. Studies show that the younger a woman was when she received radiation treatment, the higher her risk of breast cancer later in life.

9) Breast tissue may be dense or fatty. Older women whose mammograms (Harvey *et al.*, 2008) show more dense tissue are at increased risk of breast cancer.

10) Taking Diethylstilbestrol (DES) increases the risk of breast cancer. It was given to some pregnant women in the United States between about 1940 and 1971. Women who took DES during pregnancy may have a slightly increased risk of breast cancer (Titus-Ernstoff *et al.*, 2008).

11) The chance of getting breast cancer after menopause is higher in women who are overweight or obese (Rapp *et al.*, 2008).

12) Women who are physically inactive throughout life may have an increased risk of breast cancer. Being active may help reduce risk by preventing weight gain and obesity (Emaus et al, 2008, 2009).

13) Studies suggest that the more alcohol a woman drinks, the greater her risk of breast cancer (Berstad *et al.*, 2008; Suzuki *et al.*, 2008).

Other possible risk factors are under study. Researchers are studying the effect of diet, physical activity, and genetics on breast cancer risk. They are also examining whether certain substances in the environment can increase the risk of breast cancer. Many risk factors can be avoided. Others, such as family history, cannot be avoided. Women can help protect themselves by staying away from known risk factors whenever possible, but it is also important to keep in mind that most women who have known risk factors do not get breast cancer. Also, most women with breast cancer do not have a family history of the disease. In fact, apart from growing older, most women with breast cancer have no apparent risk to develop this disease.

1.2. Staging of Breast Cancer Research

Breast cancer prognosis and treatment options are generally based on tumour-node-metastasis staging (Greene FL et al, 2002). Hormone receptor status, histologic grade, lymphovascular spread, comorbidities, patient menopausal status and age are also important factors in deciding treatment options for an individual patient.

The TMN Staging is routinely used in the United Kingdom. Table 1.1 illustrates the typical treatment options offered to breast cancer patients based on their staging. Stage 0 is known as carcinoma in-situ. Lobular carcinoma in-situ is usually an incidental finding of abnormal tissue growth in the lobules of the breast. It does not progress but increases the risk of subsequent breast cancer by approximately 7% (Chuba PJ et al, 2005). Local or systemic treatment is not required but patients should be counseled to self-examine the breasts frequently, and annual mammogram plus 6 monthly examinations by a physician. Chemopreventative endocrine treatment (such as selective oestrogen receptor modulator, SERM, like tamoxifen) should be discussed with the patient.

In contrast, ductal carcinoma in situ (DCIS) can progress to invasive breast cancer. The standard treatment for DCIS is breast-conserving surgery (unless in multiple sites, when mastectomy is indicated). Lymph node clearance is usually not performed, as nodal metastsis is rare. Although there are conflicting views, tamoxifen is generally not offered for DCIS.

Stage 1 and 2:

Known as early breast cancer, these are tumours less than 5 cm in size and with no metastasis to the lymph nodes. Treatment option include breast-conserving surgery followed by radiation, this improves cancer-specific survival rate to equivalence with those with mastectomy (Clarke M et al, 2005). Women with stage 1 and II may opt for mastectomy if there is high-risk of local recurrence (Kurtz JM et al, 1990), contraindications to radiation or personal preference.

Stage 3:

Locally advanced breast cancer (LABC) includes tumours of 5cm and larger, with lymph node involvement, or with chest wall and skin involvement, considered inoperable but without distant metastases, and also inflammatory breast cancer. Induction chemotherapy followed by local therapy (surgery, radiation, or both) is becoming standard treatment. Patients have a 55% survival at 5-years (excluding the inflammatory cases) (Giordano SH et al, 2003). The most important prognostic factors are response to induction chemotherapy and lymph node status.

Stage 4:

Many patients who relapsed after early breast cancer treatment, will present with metastatic disease. The 5-year survival is only 23% (Horner MJ et al, 2009), it is therefore important to understand the patients' treatment goals. The option of treatment includes radiation (palliating pain), endocrine therapy or chemotherapy, including targeted monoclonal therapy, like Trastuzumab (Herceptin®). Endocrine therapy is generally more tolerable, but chemotherapy is likely to be used for a timely response. Tailored individual therapy depending on patient tumour type, whether it is ERBB2-positive or EGFR-positive, and patients' preference comes into play as treatment is now aimed to help palliate symptoms rather than be curative.

Table 1.1 A table simplifying typical treatment options for Breast Cancer by Stage.

Table 1.1	Typical	Treament	Options for Breast cancer by Staging		
Cancer stage and type	Primary treatment	Node evaluation	Adjuvant Rx Hormone recp negative	Hormone recp positive	ERBB2 overexpression
Stage 0: insitu					
Lobular Ca in situ	No Rx +/- Tamoxifen	-	-	-	-
Ductal Ca in situ	Sx and DXT	-	-	-	-
Stage 1 & 2: early-stage invasive	Breast Sx and DXT	SLN bx ALN dissection*	Chemotherapy	Chemotherapy & endocrine Rx	Chemotherapy & trastuzumab
Stage 3: Local advanced (Non-inflam)	Induction Chemo, Sx then DXT.	ALN dissection or SLN biopsy.	Induction chemotherapy.	Induction chemotherapy and post op endocrine therapy.	Induction chemo and post-op trastuzumab.
Inflammatory	Induction chemo, Mastectomy & DXT.	ALN dissection.			
Stage 4: Metastatic initial or recurrent.	Address patients Rx goals.	-	Chemotherapy	Endocrine therapy with or without chemotherapy.	Trastuzumab with or without chemotherapy.
Recurrent Local after breast-conser. surgery	Mastectomy	ALN dissection	Chemotherapy	Chemotherapy and endocrine therapy.	Chemotherapy and trastuzumab
Local after mastectomy	Wide excision	ALN dissection			
Local inoperable	Induction chemoRx	ALN dissection			

ALN = axillary lymph node, SLN = sentinal lymph node, Rx = Treatment, Sx = Surgery, DXT = Radiotherapy, recp = Receptor, op = Operation, bx = Biopsy

*- SLN biopsy if clinically negative nodes; otherwise, ALN dissection is recommended.

1.3. Treatment of Breast Cancer

Women with breast cancer have many treatment options. These include surgery, radiation therapy, chemotherapy, hormone therapy, and biological therapy. Many women receive more than one type of treatment.

The choice of treatment depends mainly on the stage of the disease. Treatment options by stage are described below.

Cancer treatment is either local therapy or systemic therapy:

- Local therapy: Surgery and radiation therapy are local treatments. They remove or destroy cancer in the breast. When breast cancer has spread to other parts of the body, local therapy may be used to control the disease in those specific areas.

- Systemic therapy: Chemotherapy, hormone therapy, and biological therapy are systemic treatments. They enter the bloodstream and destroy or control cancer throughout the body. Some women with breast cancer have systemic therapy to shrink the tumour before surgery or radiation. Others have systemic therapy after surgery and/or radiation to prevent the cancer from coming back. Systemic treatments also are used for cancer that has spread.

1.3 Oestrogen Positive Breast Cancer

1.3.1. Oestrogen

1.3.1.1. Breast cancer and Oestrogen

At staging, all breast cancer pathological specimens, i.e. Fine Needle Aspirate (FNA) or tissue biopsy will have their Oestrogen, Progesterone and ErbB2 receptor status confirmed. Breast cancer with ER and PgR positive has 70% chance of responding to Tamoxifen whereas in ER positive and PgR negative or vice versa, only have a 33% chance of responding to Tamoxifen. Sixty five to 70% of all breast cancers are ER positive. Hence it is important to understand the physiological role of oestrogen.

Figure 1.1 illustrates the hormone dependant physiological breast development. Figure 1.2 shows the biochemical pathways involved in the synthesis of endogenous steroids.

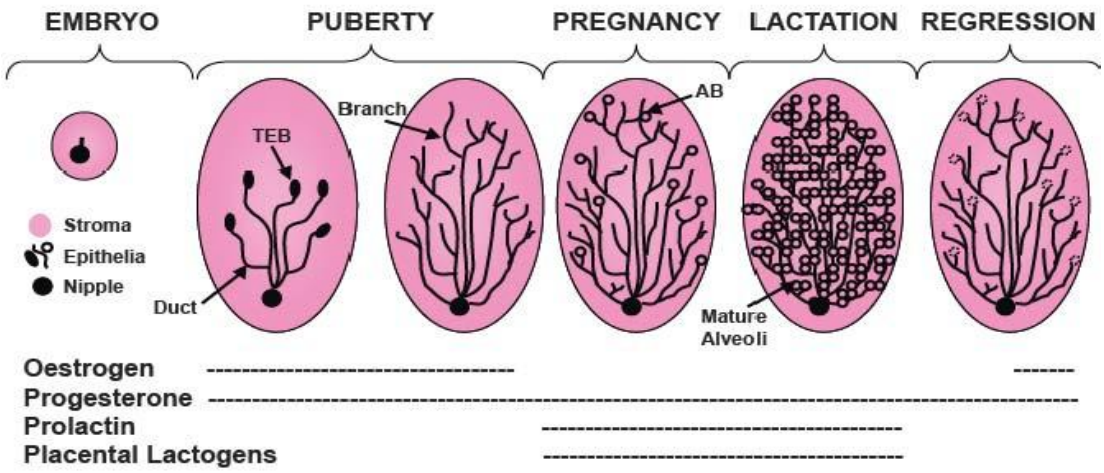


Figure 1.1 Mammary gland development Embryogenesis: A small placode in the ectoderm develops into a bud. At the base of the bud, rapid epithelial proliferation forms a single duct that grows towards a fat pad. Finally, mesenchyme derived structures form the nipple. Puberty: Cyclical ovarian production of oestrogen & progesterone accelerates epithelial duct outgrowth. Terminal end buds (TEB), consisting of an outer layer of cap cells (myoepithelial progenitors) & an inner layer of body cells (luminal progenitors) proliferate rapidly and facilitate ductal outgrowth. In the mature gland, ductal side branches form and disappear with each oestrogen cycle. Pregnancy: Placental lactogens, prolactin, and progesterone stimulate cell proliferation, alveolar bud (AB) formation and alveolar expansion. Lactation: During lactation, luminal cells of mature alveoli synthesise milk, which is transported through the ducts to the nipple. Regression: On cessation of feeding, apoptosis occurs in the secretory epithelium, the surrounding stroma is remodelled to replace apoptosed cells. Finally, cyclical production of ovarian oestrogen and progesterone returns. This illustrates the steps involved in converting cholesterol to oestradiol. Androstenedione, a key intermediary in the pathway is either converted to testosterone, which undergoes aromatisation to oestradiol, or to oestrone and converted to oestradiol by 17βhydroxysteroid reductase. In pre-menopausal women, ovarian granulosa cells produce the majority of oestradiol, with smaller amounts synthesized in the adrenal glands.

Peripherally, precursor hormones such as testosterone are converted by aromatisation to oestradiol; adipose tissue actively converts precursors to oestradiol, this continues post-menopausally. Oestradiol is also produced in the brain and arterial endothelium.

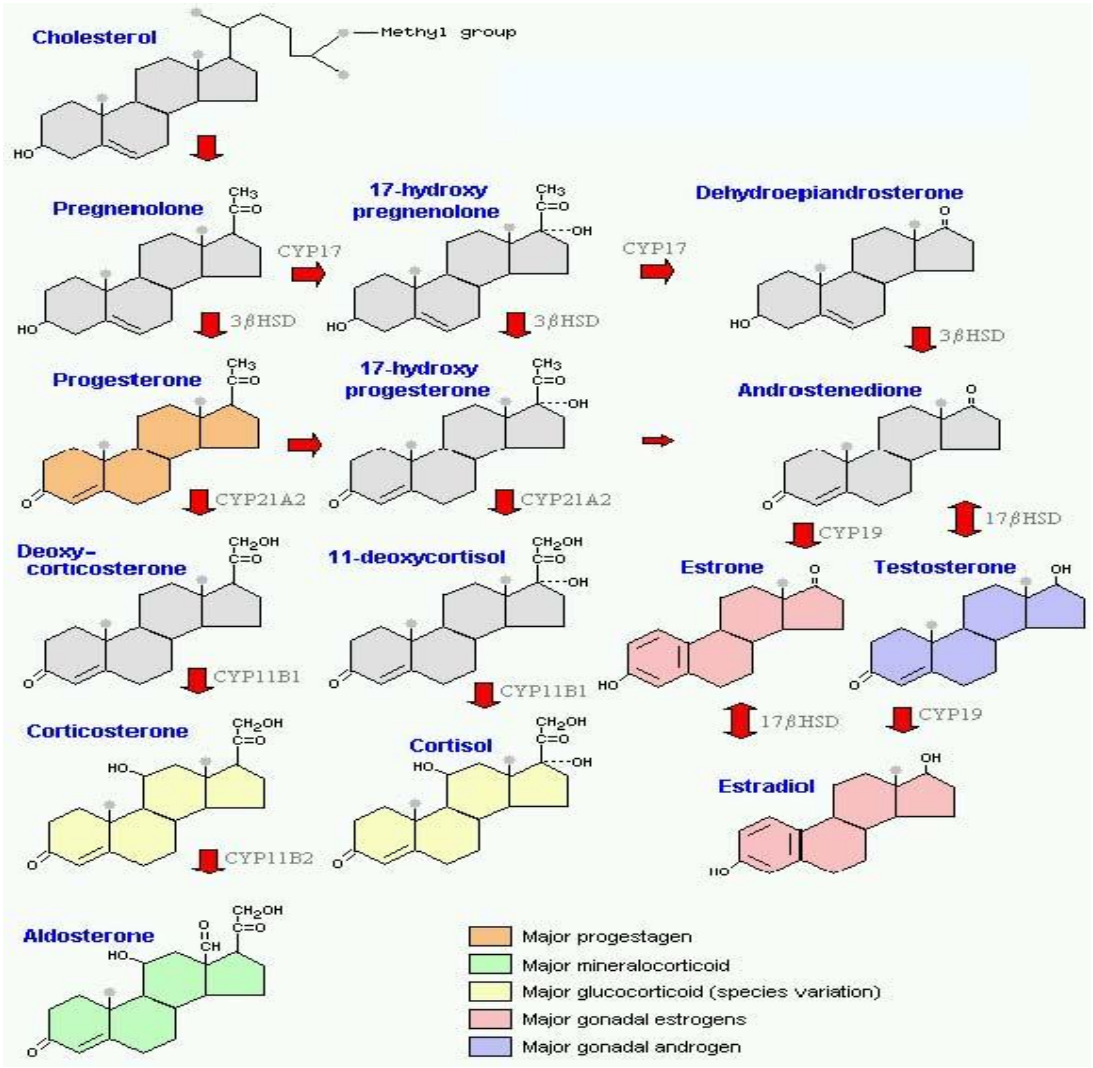


Figure 1.2 Metabolism pathway of steroids. One hypothesis to explain the possibility that oestrogens promote breast cancer through the action of its metabolites, (such as chatecholoestrogens and catechol-quinones) which cause the formation of DNA adducts in experimental models.

There is evidence that polymorphisms of the enzymes involved in the formation of oestrogen metabolites such as chatecholoestrogens and catechol-quinones, modulate the risk of breast cancer (Cheng *et al.*, 2005), (Hu *et al.*, 1998), (Lin and Scanlan, 2005), (Lavigne *et al.*, 1997).

There is also evidence to support the theory that oestrogen causes breast cancer via

its pro-proliferative effects on breast tissue; though only 15-25% of normal breast epithelial cells express ER receptors, over 65% of breast cancers are ER-positive and depend on oestrogen for growth. Furthermore, ER expression is higher in the normal breast tissue of women with breast cancer than in women without the disease (Avisar *et al.*, 1998), (Bhandare *et al.*, 2005).

As early as 1896, Beatson demonstrated that depriving breast cancer of endogenous oestrogen, through oopherectomy resulted in control of metastatic disease in ~30% of patients (1986). Today, various pharmacological strategies are used to reduce oestrogen and control the disease, as follows:

i) GnRH analogues, (e.g. goserelin) and GnRH antagonists (eg Cetrorelix) are used in pre-menopausal women to reduce ovarian production of oestrogen.

ii) Aromatase inhibitors prevent the synthesis of oestrogen in breast cancer cells, the ovaries and peripheral tissues; for example: exemestane is a suicide substrate while anastrazole is a competitive inhibitor of aromatase. Aromatase, also known as CYP19 is depicted in the steroidogenesis pathway catalysing the conversion of androstenedione to estrone and the conversion of testosterone to oestradiol (Figure 1.2).

iii) ER antagonists are also widely used; these can be divided into pure anti-oestrogens such as: Faslodex (ICI 182,780) and Selective Estrogen Receptor Modulators, (SERMs) such as tamoxifen and raloxifene.

Tamoxifen and Anastrazole have proven to be effective in the treatment and prevention of breast cancer; they are both widely used in these contexts. (Buzdar et al, 2005); (Baum et al, 2002).

1.3.1.2. The Oestrogen Receptor (ER)

The effects of oestrogen are mediated through the oestrogen receptor ER, which exists in two forms: ERα and ERβ, encoded by separate genes. Oestrogen receptors belong to the nuclear hormone receptor superfamily, a large family of ligand-regulated, zinc finger-containing transcription factors, which share a characteristic structure. ERα and ERβ have distinct tissue expression patterns: (ERα is found in the endometrium, breast, ovaries and pituitary while ERβ occurs in kidney, bone, brain, heart, lungs, intestinal mucosa, prostate and endothelial cells) (Saji *et al.*, 2000); **(Paech et al, 1997**); (Katzenellenbogen *et al.*, 2000a), (Katzenellenbogen and Katzenellenbogen, 2000; Katzenellenbogen *et al.*, 2000b).

1.3.1.3. ER: Mechanisms of action

ERα has been extensively studied and plays a dominant role in the promotion and progression of breast cancer; it is over-expressed in 65-77% of primary breast cancers. By contrast, ERβ may have a tumour suppressor role; expression is reduced in malignancies of the breast, ovary, prostate and colon, and overexpression of ERβ inhibits proliferation and invasion of breast and prostate cancers **(Duong et al, 2006).** Oestrogen is thought to act via several different mechanisms, as follows and summarised in Figure 1.3:

i) In the "classical" model of ligand dependant ER activation; upon oestrogen binding, the receptor undergoes a conformational change, dissociates from chaperone proteins (such as hsp90) it dimerises, and is targeted to the nucleus. In the nucleus, the oestrogen-bound ERα dimer binds DNA at an oestrogen response element (ERE), a palindromic consensus sequence present in the gene regulatory regions of ER-responsive genes, (5'-GGTCAnnnTGACC-3'). Transcription of oestrogen responsive genes is activated through interaction with the basal transcription machinery and recruitment of co-regulatory proteins: co-activators or co-repressors (See Figure 1.4 for summary).

ii) The ligand-bound ER can also interact at "non-classical" sites on DNA by direct protein-protein interaction with other transcription factors (such as AP-1, Sp1, NF-kB) and activate transcription of their target genes.

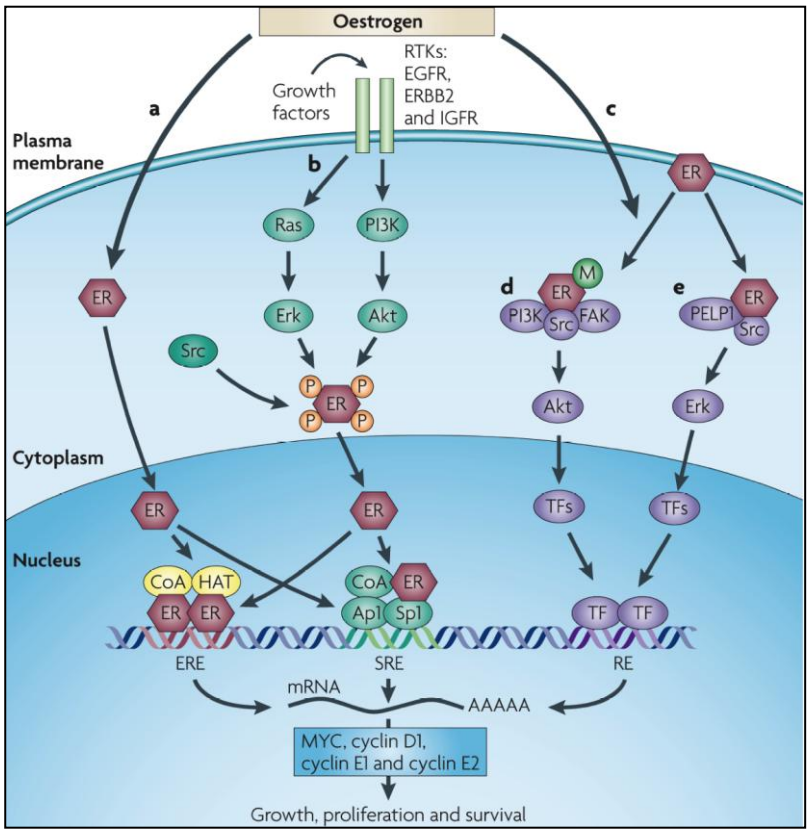


Figure 1.3 Oestrogen action at the molecular level.

(**a**) The ligand-bound oestrogen receptor (ER) activates gene expression by direct dimeric binding to its "classical" DNA response element (ERE), or at "non-classical" sites by interaction with other transcription factors (e.g. AP-1; Sp1), in complexes that include co-activators (CoAs) and histone acetyl transferasae (HATs). **(b**) Receptor tyrosine kinase (RTK) signalling can lead to ligand-independent ER activation via phosphorylation. **(c**) Signalling may also be mediated through "non-genomic" signalling by ER that is localised at the cell membrane or in the cytoplasm. Two recently described pathways are illustrated: (**d**) ligand induced methylation (M) of ER and formation of an ER complex with focal adhesion kinase (FAK) that activates the Akt pathway, and (**e**) activation of Erk signalling by an ER-Src-PELP1 complex. Diagram taken from (Musgrove & Sutherland, 2009).

iii) Ligand-independent activation of ER can occur in the cytoplasm via cross talk with other signalling pathways (e.g. receptor tyrosine kinases such as. EGFR, ErbB2, IGFR). Activation leads to ER phosphorylation at key amino acid residues in the AF-1 activation domain by MAPK/ERK or PI3K/AKT altering the level of activation of ER, resulting in dimerisation and transcription of ER responsive genes (Leo *et al.*, 2005); (Kato et al, 2001). Phosphorylation of co-regulator proteins can also occur, altering their activity and thus modulating the transcriptional activity of ER.

iii) More recently a non genomic pathway has been proposed where cytoplasmic ER becomes activated at the cell membrane leading to increased signalling through alternative pathways that do not result in the binding of ER to the DNA – hence the term "non-genomic".

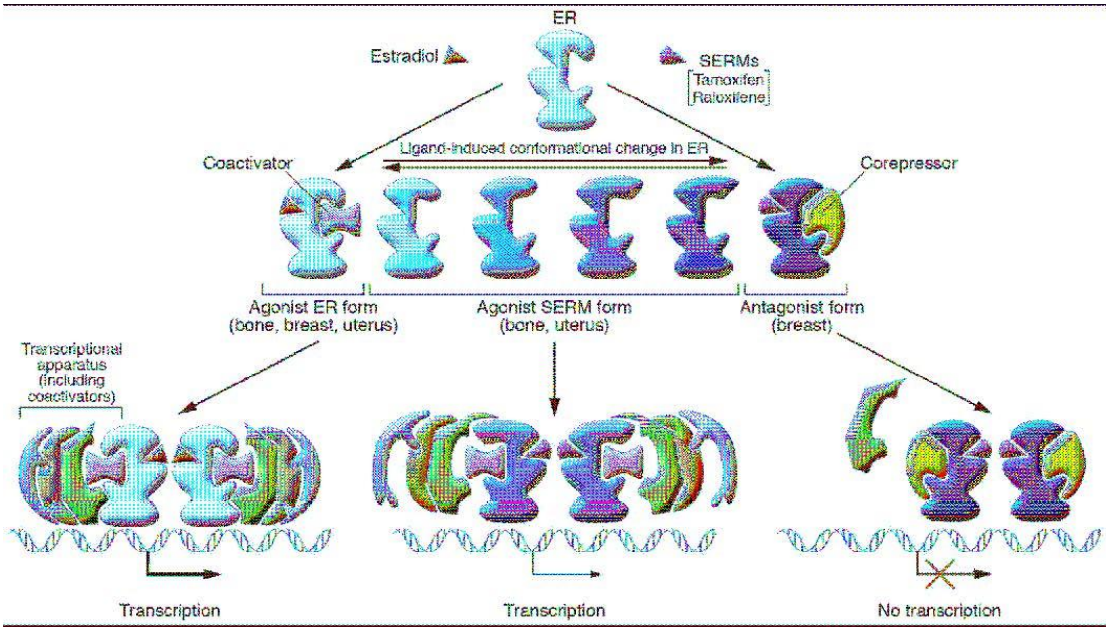


Figure 1.4 ER co-activator recruitment. Upon binding ER ligands, (ER agonists, antagonists, or SERMs) the receptor undergoes a conformational change, which regulates the recruitment of co-regulatory proteins. Co-activators such as SRC1 bind to the active (agonist bound) receptor and activate transcription, while co-repressors interact with the antagonist-bound receptor, inhibiting transcription. Depending on the cell and promoter context, unique and overlapping sets of genes are regulated by the different ligands. (Deroo et al, 2006)

1.3.1.4. Use of tamoxifen and other hormonal therapies

Early clinical trials with tamoxifen established its value in advanced breast cancer (Ward et al, 1973; Ward et al, 1978; Cole et al, 1971; Jordan et al, 1988). Originally, tamoxifen was given to all patients with breast cancer, however it's current use is restricted to hormone receptor positive tumours. For decades, tamoxifen has been the gold standard treatment for ER positive disease in both metastatic and adjuvant settings.

Tamoxifen's safety and efficacy has been established in randomised controlled trials (RCT) involving approximately 30,000 women. Adjuvant treatment with tamoxifen for women with ER positive disease results in a 51% reduction in recurrence rate and 28% reduction in death, standard treatment is for 5 years (Gray et al, 1993; Lewis et al, 2007). Tamoxifen is cost effective and safe. It has been shown to lower serum cholesterol levels and lower the risk of osteoporotic fracture in post-menopausal women, (Mikuls *et al.*, 2005); (Jakesz *et al.*, 2005); (Love *et al.*, 1990). Its less desirable side effects include: hot flushes, an increased risk of uterine cancers, (endometrial carcinoma and uterine sarcoma), thromboembolism and elevated triglycerides (Cuzick *et al.*, 2002); EBCTCG, 1998; (Fisher *et al.*, 2005).

The observation that tamoxifen reduced the incidence of cancer in the contralateral breast by 54% after 5 years prompted further investigation for its use as a chemopreventative. Placebo controlled trials in over 25,000 women showed that tamoxifen reduced breast cancer risk by about 40% and osteoporotic fracture risk by about 32%, IBIS-1, (International Breast Cancer Intervention Study-1), Royal Marsden Hospital Chemoprevention trial, Italian tamoxifen prevention study and NSABP-P1 (National Surgical Adjuvant Breast and Bowel Project P-1) trials.

The Study of Tamoxifen and Raloxifene, (STAR) and NSABP P2 directly compared tamoxifen with raloxifene for chemoprevention. Both studies showed similar risk reduction for invasive breast cancer and osteoporotic fractures with less toxicity for raloxifene. Intriguingly, both trials suggested that tamoxifen had greater activity in the prevention of noninvasive breast cancer (carcinoma in situ – DCIS and LCIS)

(Vogel VG, 2007). The IBIS-1 and Marsden trials/studies has confirmed that the chemoprotective effects for prevention of invasive breast cancer continue long after the end of treatment, while the side effects resolve more quickly, suggesting an increase in the risk benefit profile after the end of active treatment. (Forbes *et al.*, 2008); (Powles *et al.*, 2007).

Over the past 5-10 years, a growing body of evidence has accumulated establishing the clinical superiority of more modern agents, aromatase inhibitors (AIs) in the treatment of hormone responsive breast cancer in postmenopausal women. Meta-analysis of data from 25 trials in the metastatic setting (comprising 8504 patients) demonstrated a significant overall survival advantage for aromatase inhibitors compared with tamoxifen (11% RH reduction, 95% CI = 1% to 19%; P = .03), (Mauri *et al.*, 2006).

However, at the 2008 San Antonio Breast Cancer Symposium, a meta-analysis of Aromatase Inhibitor (AI) trials (Ingle et al, 2008; Jakesz et al, 2008; Mouridsen et al, 2008) showed that it may be premature for oncologists to discard tamoxifen. A meta-analysis of eight large trials using AI alone or with a switch between AI and tamoxifen after 2-3 years, or an extension (defined by 5 years of tamoxifen, followed by 2-3 years of AI) has shown that while patients receiving an AI showed a clear disease-free survival (DFS) advantage, the groups with significant overall survival (OS) were those switched over groups that received tamoxifen first (see Table 1.1). No OS difference, in fact not even a suggested trend towards an improved survival advantage, has been seen in patients receiving an AI alone (Hughes-Davies, et al, 2009). Two possible explanations have been offered for the lack of OS benefit: i) longer followed up is needed; ii) about two-thirds of the patients were node negative, which made it harder to demonstrate a significant absolute mortality difference. The largest AI alone trial (BIG I-98) also shows no statistical difference between letrozole alone and tamoxifen alone (p=0.08). This trial has two cross-over arms but the results of these have not yet been reported. So to date, it is recommended by most hospital trusts to use AI after 2-3 years of tamoxifen to achieve an OS advantage unless there are contraindications to tamoxifen use.

	n	AI	Design	Age	ER+	% n+	FU	DFS HR	Events AI/TA	DFS p-val	OS	Deaths AI/TAM	OS P val	ref
ATAC	5216	A	sub	64	100	35	100	0.85	618/702	0.003	1	472/477	NS	1
BIG 1-98	4922	L	sub	61	98	41	76	0.88	509/565	0.03	0.9	303/343	0.08	2
IES	4602	E	sw	n/s	100	48	56	0.75	339/439	0.0001	0.8	210/251	0.05	3
ABCSG-8	2922	A	sw	n/s	100	25	72	0.79	202/235	0.038	0.8	130/157	0.025	4
ARNO-95	979	A	sw	61	97	25	30	0.66	36/47	0.049	0.5	15/28	0.045	5
ITA/GROCTA	828	A,AG	sw	64	100	85	78		not stated		0.6	48/74	0.007	6
NSABP B33	1598	E	ext	60	100	48	30	0.68	37/52	0.07	1	16/13	NS	7
MA17	5187	L	ext	62	97	46	64	0.68	164/235	0.0001	1	154/155	NS	8

Table 1.1 A summary of the AI trials that have been reported to date.
Abbreviations; sub=substitution, sw=switching. Ext=extension, A=Anastrazole, L=Letrozole, E=Exemestane, AG=Aminogluthemide. Those trials that have a statistically significant mortality benefit are highlighted in pink. This table clearly shows that only the switching trials have been able to show a mortality benefit from the AIs. Refs: 1=ATAC Trialists, 2008; 2=Mouridsen et al, 2008; 3=Coombes et al, 2007; 4=Jakesz et al, 2008; 5=Kaufmann et al, 2007; 6=Boccardo et al, 2007; 7=Mamounas et al, 2008; 8=Ingle et al, 2008. All these analyses are intention to treat (ITT) with no adjustment for crossover and patients are kept in their originally assigned groups even if they crossover from the control arm to the investigational arm. However, for the ATAC and IES data presented in this table, the ER-unknown or ER-negative patients were excluded from analysis. For MA17, a 2005 report showed an OS advantage in the node-positive subgroup; this was not seen in most recent results, probably because of high crossover rate. For ATAC, the figure in the table is for ER-positive subgroup. Reproduced from (Hughes-Davies, et al, 2009).

As for their relative side effect profiles: women taking anastrozole experienced more sexual dysfunction, myalgia and an increased risk of osteoporotic fractures, while tamoxifen was associated with an increased risk of thrombosis and endometrial cancer (Forbes *et al.*, 2008); (Howell *et al.*, 2005); (Baum, 2002); (Buzdar, 2005).

At present, tamoxifen remains in widespread use for numerous reasons: aromatase inhibitors are contraindicated in pre-menopausal women; arguably, clinicians are intrinsically reluctant to change their practice. In the UK, current guidance from the National Institute of Clinical Excellence (NICE) allows patients and their clinicians to choose the appropriate adjuvant hormone therapy from an AI or tamoxifen. No recommendation has been made for one AI over another {National Institute for Health and Clinical Excellence, 2006}. The American Society of Clinical Oncology, ASCO has recommended the use of an AI "at some point" in their treatment for all postmenopausal women with ER positive breast cancer. There are differences in their side effect profile, which means that some women tolerate one or other class of drug more easily. Tamoxifen is still established as an effective treatment for metastatic disease and if patients are given adjuvant AIs, if they relapse they may be given tamoxifen at this point. Therefore one can expect tamoxifen to continue to be in widespread use for some time to come.

However, in the metastatic disease setting, ~50% of patients exhibit *de novo* resistance to tamoxifen, and eventually all patients develop tamoxifen resistance and this clearly limits the use of this drug (Tonetti and Jordan, 1995); (Ali and Coombes, 2002).

1.4. Tamoxifen resistance

The complexity of ER activation and tamoxifen's interaction with ER provides multiple mechanisms by which tamoxifen resistance may occur. Two-thirds of breast cancers express oestrogen receptor-α, which drives breast cancer cell growth. Endocrine therapies are designed to block oestrogen action, but many tumours still exhibit *de novo* or acquired therapeutic resistance. The primary mechanism of de novo resistance to tamoxifen is lack of expression of ERα. However, recently a second intrinsic mechanism has been highlighted, involving the inactivation of cytochrome P450 2D6 (CYP2D6), causing a failure of conversion of tamoxifen to its active metabolite, endoxifen, and consequently reduced response (Hoskin et al, 2009).

In contrast, a number of mechanisms have been proposed to account for acquired resistance. Most of these have been published using results derived from ERα-positive breast cancer cell lines exposed to long term tamoxifen or withdrawal of oestrogen (*in vitro*). Although these studies reflect the range of tamoxifen resistance mechanisms *in vitro*, they are unlikely to describe totally how patients become resistant. This is due to fact that there are relatively few ER+ cell lines which do not represent all *in vivo* phenotypes (see section 1.6) but also because cells in culture cannot reflect the epithelial-stromal or the tumour-host interaction which will modulate resistance *in vivo.*

Here, I will discuss the known mechanisms for endocrine-resistant in breast cancer in two broad classes; a) deregulation of various aspects of oestrogen (ER) signalling, and b) un-related signalling (cross talk) pathways with alternative proliferative and survival stimuli that confer resistance by activating the ER by alternative mechanisms (Table 1.2).

Deregulation of ER signalling:

1) Loss of ERα expression: Since the effects of tamoxifen are mediated through ER, and ERα expression predicts response to tamoxifen, it is logical that loss of ERα expression confers resistance to therapy. However, IHC studies looking at paired tamoxifen sensitive and resistant tumours show that although ERα expression may be lost in some patients who develop tamoxifen resistance, 60-80% continue to express ERα on disease progression (Gutierrez *et al.*, 2005); (Johnston *et al.*, 1995). In addition, ~20% of patients demonstrate a response to further hormone therapy following failure of tamoxifen, suggesting that the ER continues to regulate growth in many tamoxifen-resistant patients (Osborne *et al.*, 2001); (Howell et al, 2001). Cell line models of tamoxifen resistance, such as the ones used in this project continue to respond to β-oestrodiol.

Mutations of the ERα gene may lead to a non-functioning receptor without loss of expression. However, although ERα mutations altering the effects of bound ligand

can be generated in vitro and detected in some resistant cell lines, they are thought to be relatively uncommon clinically (Roodi *et al.*, 1995); (Karnik *et al.*, 1994).

Epigenetic changes may also reduce expression of ER or oestrogen target genes. In a study of DNA methylation status of 148 primary breast tumors, it was found that hypermethylation of the ERα gene, (ESR1) outperformed hormone receptor status as a predictor of clinical response in tamoxifen treated patients. Interesting addition, promoter methylation of CYP1B1, which encodes a tamoxifen and oestradiol-metabolizing cytochrome p450, was also a highly significant predictor of tamoxifen response between tamoxifen-treated and non tamoxifen-treated patients (Widschwendter *et al.*, 2004).

2) Median ERβ mRNA levels measured by RT-PCR were 2-fold higher than ERα in tamoxifen-resistant tumours compared with tamoxifen-sensitive tumours (Speirs *et al.*, 1999) suggesting that aberrations in ERβ levels may contribute to resistance.

3) A common mechanism of drug resistance is through reduced intracellular concentrations of drug as a result of decreased metabolism or increased absorption (pharmacological tolerance). A study analyzing serum and intra-tumoural tamoxifen levels suggested that acquired resistance is associated with reduced intra-tumoural tamoxifen concentrations in the presence of maintained serum levels (Johnston *et al.*, 1993).

Decreased metabolism of tamoxifen to active, agonist metabolites is another potential mechanism of resistance (Osborne *et al.*, 1991). Recent studies have demonstrated the importance of endoxifen, (4-OHN-desmethyltamoxifen) a potent tamoxifen metabolite produced by the action of the cytochrome P450 enzyme, (CYP) 2D6 (Johnson *et al.*, 2004). Polymorphisms in the (CYP) 2D6 gene affect the plasma concentration of endoxifen and clinical outcome of women given hormone therapy. In the NCCTG 89-30-52 study, a retrospective analysis of 256 tamoxifen-treated patients, CYP2D6 genotypes were determined from paraffin-embedded blocks. Women with the CYP2D6 *4/*4 genotype (the less active phenotype comprising ~7% of the European population) had a higher risk of disease relapse and a lower

incidence of hot flashes than women with the wildtype allele (Goetz *et al.*, 2005).

	Molecular aberration	Clinical correlation	Tamoxifen resistance
ER signaling		**Poor**	
Ers and ERRs	ER alpha methylation	Yes	Yes (Clarke R, 2003)
	ER alpha (tuncated)	No	Yes (Shi L, 2009)
	ER alpha phosphorylation	Yes	Yes (Murphy L, 2004)
	ER alpha methylation	ND	ND
	ER beta, ER beta2/beta5	Yes	Yes (Honma N, 2008)
ER-associated	AP1 over-expression	Yes	Yes (Schiff R, 2000)
transcription	NF-kappaB activation	Yes	Yes (Zhou Y, 2005)
factor	NCOA3 amplification	Yes	Yes (Redmond A, 2009)
and co-activators			
	CBP and p300 overexpression	Yes	Yes (Green a, 2008)
	SRC-3 (AIB1)	Yes	Yes (Osborne K, 2003)
Growth factor receptors and cytoplasmic signaling			
Receptors	EGFR overexpression	Yes	Yes (Arpino G, 2004)
	ERBB2 amplification	Yes	Yes (Ali S, 2008)
	PAX2 loss leading to ERBB2 loss	Yes	Yes (Hurtado A, 2008)
	IGF1R overexpression	Yes	ND
	FGFR overexpression	ND	Yes (FGFR4, Meijer D, 2008)
MAPK signaling	Mek and Erk activation	Yes	Yes (Bergqvist J, 2006)
PI3K signaling	Akt activation and overexp	Yes	Yes (Kirkegaard T, 2005)
	PTEN loss	Yes	Yes (Shoman N, 2005)
SRC signaling	SRC activation	ND	ND
	BCAR1	No	Yes (van der Flier, 2000)
	BCAR3 overexpression	No	Associate with good response
			(van Agthoven T, 2009)

Table 1.2 Summary of selected pathways associated with tamoxifen resistance. (Adapted from Musgrove and Sutherland, 2009)

Growth factor receptor and cytoplasmic signalling:

4) Alterations in co-regulatory proteins: SRC-3 (AIB1) is an ER co-activator overexpressed in >50% of breast tumours (Anzick *et al.*, 1997); (Osborne K et al, 2003). In cell line studies, SRC-3, also called AIB1, RAC3, ACTR, and p/CIP, is an ER coactivator that enhances the agonist activity of tamoxifen in resistant cells (Feng *et al.*, 2001). This is particularly seen in tumours expressing epidermal growth factor (EGF) receptor family members leading to activation of MAPKs **(Font de Mora et al, 2000)**. In patients samples (n=316) from those not given adjuvant tamoxifen, high SRC-3 levels were associated with good prognosis, while in tamoxifen-treated patients, high SRC-3 expression was associated with a worse disease free survival (Osborne *et al.*, 2003).

Experimental data suggest that overexpression of other co-activators eg SRC-1 may also enhance the agonist activity of 4-OH tamoxifen (Smith *et al.*, 1997); (Tzukerman *et al.*, 1994). An IHC study examining levels of SRC1 expression in 70 primary breast tumours of known HER2 status (HER2 positive, n = 35) and normal breast tissue found over-expression of SRC-1 was significantly associated with disease recurrence in HER2 positive patients treated with tamoxifen (Meng *et al.*, 2004).

5) Kinase / signal transduction pathways: ERK1/2 expression and activity is increased in several breast cancer cell-line models of endocrine resistance (Lee *et al.*, 2000) (Coutts and Murphy, 1998). Increased ERK 1/2 activity (assessed by phosphorylated MAPK immunostaining) correlated with shorter duration of response to endocrine therapy in clinical breast cancer (Gee *et al.*, 2001). pERK1/ERK2 did not play a role in the phosphorylation of ERα Ser^{118} (Martin L-A et al, 2005).

Two cell-line studies have addressed the possible involvement of the PI3K cell survival pathway with tamoxifen resistance. Transfection of MCF-7 cells with AKT reduced the inhibition of cell growth by tamoxifen, suggesting overexpression of AKT may contribute to tamoxifen resistance (Campbell *et al.*, 2001). PI3KCA mutations have been reported in approximately one third of breast cancers (Bachman et al, 2004). These are reported to effect the downstream signalling (Kang et al, 2005). Clark et al. measured tamoxifen-induced apoptosis with and without the PI3K pathway inhibitor LY294002. Addition of LY294002 significantly increased the pro-apoptotic effects of tamoxifen, particularly in the cell line with the highest endogenous levels of AKT activity (Clark *et al.*, 2002). Stress-activated protein kinase/c-junNH2 terminal kinase (SAPK/JNK). ER can interact with the SAPK/JNK pathway either via binding AP-1 or by direct activation of ER and co-regulators by p38 MAPK.

Activation of the p38 MAPK pathway occurs in response to a number of extracellular stimuli including growth factors, cytokines, physical and chemical stress (Chen *et al.*, 1998). The downstream targets of p38 MAPK include further protein kinases and transcription factors. In cell lines expressing ER; 4-OH

tamoxifen has been shown to activate the p38 MAPK pathway and induce apoptosis. Under these circumstances, inhibition of the p38 signalling pathway greatly reduces the ability of tamoxifen to induce apoptosis. In paired biopsy samples taken pre-treatment and on relapse from patients treated with adjuvant tamoxifen, report (Martin et al, 2005) that elevated ERK1/ERK2/oestrogen receptor cross-talk enhances oestrogen mediated signalling during long-term oestrogen deprivation. Moreover, patients with ER-positive and ErbB2 positive at relapse showed uniformly high expression of p38 MAPK, suggesting that in this subset of patients, activation of ER may have occurred by this route (Dowsett and Ellis, 2003).

In tamoxifen resistant cells, erbB3/erbB2 and erbB3/EGFR heterodimerisation, promoted ERK1/2 and AKT pathway activation and increased cell proliferation and invasion (Hutcheson, 2003; Hutcheson, 2007).

6) Loss of PTEN was found to engage ErbB3 and IGF-IR signalling to promote anti-oestrogen resistance in breast cancer (Miller TW et al, 2009). Tamoxifen treatment inhibited oestrodiol-induced ER transcriptional activity is all shPTEN cell lines but did not abrogate the increased cell proliferation induced by PTEN knockdown. PTEN knockdown increased basal and ligand-induced activation of the IGF-I and ErbB3 receptor tyrosine kinases, and prolonged the association of the p85 PI3K subunit with the IGF-I effector IRS-I with ErbB3, implicating PTEN in the modulation of signalling upstream from PI3K.

7) The most widely studied mechanisms of tamoxifen resistance has been upregulation of ERBB2 expression and activity. ER+ luminal tumours with elevated ERBB2 levels have the poorest prognosis (Kun Y et al, 2003), and about half of ERBB2-positive tumours are also positive for ER (Piccart-Gebhart et al, 2005). Up-regulation of ERBB2 expression is strongly associated with gene amplification, however transcriptional affects have also been documented. In particular, ligand bound ER has been found to repress ERBB2 expression, which is reversed in tamoxifen-containing media in vitro (Bates & Hurst, 1997). A number of transcription factors have been suggested to repress ERBB2 expression, including GATA4 (Hua et al, 2009) and FOXP3 (Zuo et al, 2007) but these have not been directly associated with ER or tamoxifen resistance. In contrast, the PAX2 factor and the ER co-activator SRC-3 have been shown to compete for binding and

regulation of ERBB2 transcription. Increased PAX2 expression (and hence repression of ERBB2) was associated with a better outcome on tamoxifen (Hurtado et al, 2008).

8) Other growth factor receptor families may also contribute to tamoxifen resistance. Increased expression of Fibroblast growth factor receptor 4 (FGFR4) predicted failure on tamoxifen therapy in patients with recurrent breast cancer (Meijer et al, 2008). Although not yet investigated in terms of therapy response, a recent genome-wide association study identified single nucleotide polymorphisms (SNPs) within the second intron of FGFR2 as being associated with a small but highly significant increase in cancer risk in ER+ breast cancer (Easton D et al, 2007).

1.5. Gene expression profiling

One of the conceptually most important discoveries of the past 10 years was the realization that invasive breast cancer is not a single disease with different degrees of ER and HER2 expression and variable histologic features but a collection of several molecularly rather distinct diseases (Bertucci et al., 2004; Dorssers et al., 2001; Rody et al., 2006). Transcriptional profiling revealed large-scale gene expression differences between ER-positive and ER-negative cancers that go far beyond the expression of ER itself. It is plausible that ER-positive and -negative cancers originate from different epithelial precursors, luminal and basal ductal epithelial cells, respectively. Furthermore, among the ER-positive cancers, two distinct subtypes can also be distinguished that show different sensitivities to therapy and have different prognosis. Currently, four different molecular classes of breast cancers can be identified consistently through gene expression profiling. Using the original terminology proposed by Perou et al. these include: (i) "Basal-like" breast cancers that correspond mostly to ER-and HER2-negative, high histologic grade cancers, (ii) "Luminal-A" cancers that are mostly ER-positive and lower grade cancers, (iii) "Luminal-B" cancers that are also mostly ER-positive but often higher grade, and (iv) "HER2-positive" cancers that include most of the HER2 gene amplified cases (**Sorlie et al., 2001; Perou et al. 1999**). It is important to recognize that up to 25-30% breast cancers do not fall into any one of the above robust molecular categories.

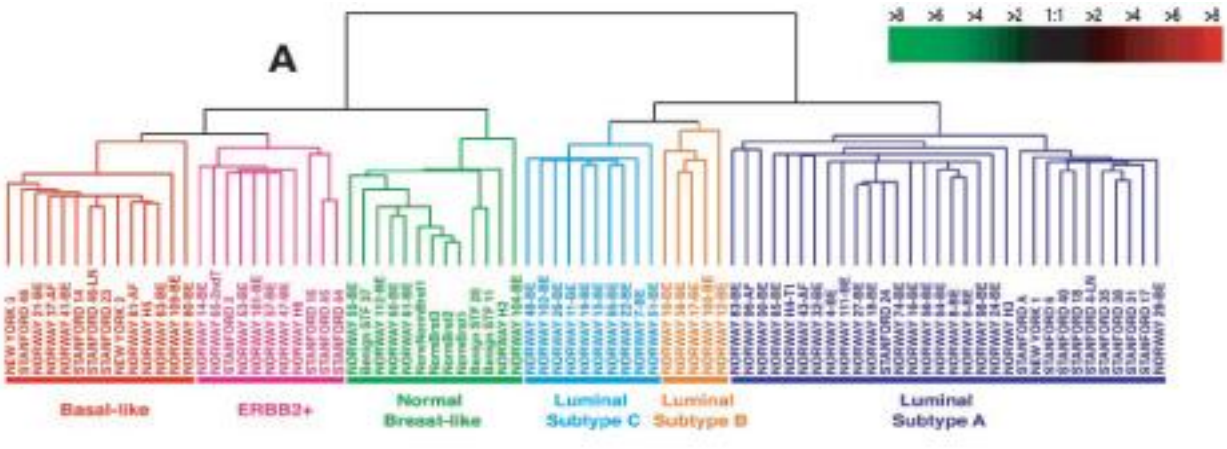


Figure 1.5 Gene Expression patterns of breast carcinoma distinguish tumour subclasses with clinical implications. Perou et al first published this landmark paper in 2002, which showed 5 distinct molecular type of breast cancer, which is associated with histological grade. 'Luminal Subtype A' has the best prognosis while 'basal-like' has the worst prognosis.

The four molecular classes of breast cancer correspond closely, but not perfectly, to well-established clinical phenotypes of breast cancer. This correspondence is reassuring and provides a molecular framework to understand clinical phenotype. It is important to consider that while histological grade can not be targeted with therapies; better understanding of the molecular abnormalities that cause high grade morphologic features may lead to new therapeutic targets. On the other hand, the diagnostic relevance of molecular classification is limited by its close association with ER- and HER2-status and histological grade. Basal-like cancers are almost exclusively ER- and HER2-negative and high grade cancers therefore it is expected that they will have poor prognosis in the absence of adjuvant systemic therapy. This group have higher sensitivity to chemotherapy in general and do not benefit from endocrine treatment. Conversely, Luminal-A cancers that are mostly ER-positive, HER-2 normal and have lower grade will have the highest endocrine sensitivity (but lowest chemotherapy sensitivity) and best prognosis (Hu et al., 2006). To what extent molecular classification provides clinical value beyond routine histopathology parameters remains unknown. However, molecular classification provides a simple summary measure of complex clinical-pathological variables. This is an important potential advantage because considerable variation exists in the assessment of routine histopathological features (i.e. grade and even ER-status) of the same cancer by different pathologists.

1.6. Gene Expression Signatures and Other Markers to Predict Prognosis

Two alternative predictor development strategies exist; the first referred to as the "supervised" marker discovery strategy. This relies on comparing gene expression data from cohorts of cases with known outcome to identify genes that are associated with prognosis or response to therapy and than combine these informative genes into a multivariate prediction model. The second strategy may be called the "hypothesis testing" approach. This starts with defining a hypothesis that particular genes or molecular pathways may influence a clinical outcome of interest and these genes are used to construct a multi-gene predictive signature. Candidate genes may be selected based on existing biological knowledge or can be identified through experiments in vitro.

Regardless of which development strategy is utilized, genomic outcome predictors (i.e. "gene signatures") are conceptually similar to multivariate clinical prognostic prediction models. These prediction tools apply the mathematical principle that individually weak predictive variables, that are at least partly independent of each other, can be combined into prediction models that are more accurate than any single variable alone. The main difference is that genomic predictors combine molecular rather than clinico-pathological variables into a prediction model.

1.7. Prognostic gene signatures

Three distinct gene expression profiling-based prognostic tests were recently developed. One of these, MammaPrint (Agendia Inc., Amsterdam, Netherlands), was recently cleared by the U.S. Food and Drug Administration (FDA) to aid prognostic prediction in node-negative breast cancer. This assay measures the expression of 70 genes and calculates a prognostic score that can be used to categorize patients into "good" or "poor" prognostic risk groups. This test was subsequently evaluated on two separate cohorts of patients that received no systemic adjuvant therapy. The first cohort included 295 patients and showed that those with the good prognosis gene signature had 95% (standard error ±2%) and 85% (±4%) distant metastasis-free survival at 5 and 10 years, respectively. In contrast, the poor

prognostic group had 60% (±4%) and 50% (±4.5%) distant metastasis-free survival at 5 and 10 years, respectively (van de Vijver *et al.*, 2002). A second validation study (n=307) confirmed these findings and showed that patients with the good prognosis signature had 90% (95% confidence interval, 85-96%) distant metastasis-free survival at 10 years, whereas it was 71% (65-78%) in the poor prognosis group (Buyse et al, 2006). Importantly, the MammaPrint signature could re-stratify patients within clinical risk categories defined by the Adjuvant Online (a program used widely for stratifying risk available to the public on the website). Some of the clinically low-risk patients were correctly re-categorized as high-risk based on their gene signature, and some clinically high-risk patients were correctly predicted to be low-risk by the genomic test. However, a recent report also highlighted an important limitation of this test, almost all ER-negative cancers (>90%) are classified as high-risk by † MammaPrint after adjustment for tumor grade, size, and ER status that confirms least partly independent added prognostic value. *MINDACT (Microarray in Node Negative Disease May Avoid Chemotherapy) trial is currently ongoing in Europe to accrual for a prospective study on the predictive value of this gene chip.

A second prognostic signature utilized the "hypothesis testing" discovery strategy. Investigators set out to define the gene expression differences between low and high-histological-grade cancers and assumed that these genes would be able to improve prognostic predictions for morphologically intermediate-grade cancers. Using this approach, a 97-gene genomic grade signature was identified that discriminated between low and high-grade tumours and separated intermediate-grade tumors into two distinct subgroups of lower and higher genomic grade cancers with different prognosis (Sotiriou and Desmedt, 2006). These results were observed across multiple independent data sets generated on different microarray platforms. Not surprisingly, the genomic grade gene index is dominated by genes involved in cell cycle regulation and proliferation (Desmedt and Sotiriou, 2006).

It is important to point out that the various prognostic signatures have very few genes in common. This may seem surprising at first, but it is a common feature of high-dimensional data that contain large numbers of highly correlated variables. Gene expression values are highly correlated with each other and therefore, if the

expression of a particular gene is associated with a particular clinical outcome, all other genes whose expression are closely correlated with that index gene will also correlate with the same clinical outcome. However, the strength of association between any given gene and the clinical outcome varies from training set to training set and therefore, the rank order of the informative genes is unstable when they are ranked by strength of association. Nevertheless, all of these co-expressed genes carry similar information about the outcome of interest and therefore many different statistically equally good predictors can be discovered from the same data set (Ein-Dor et al, 2005). A corollary of this is that different predictors that use information from different genes can predict equally well on a given data set.

A limitation of the current microarray-based prognostic assays is that they only provide moderately precise estimates of risk of recurrence. Also, almost all ER-negative cancers are assigned to high prognostic risk category by the currently available assays. On the other hand, the genomic predictors seem to complement tumour size- and grade-based prognostic models. This is probably driven by the improved ability of the genomic tests to categorize clinically intermediate risk groups (i.e. intermediate grade cancers) into low or high prognostic categories. What constitutes a low enough risk to forgo systemic adjuvant chemotherapy is influenced not only by the absolute risk of relapse but also by the risk of adverse events, the probability of benefit from therapy, and personal preferences. Many patients are willing to accept adjuvant chemotherapy for rather small gains in survival (Ravdin et al, 1998). Molecular prognostic markers may provide little clinical value for these individuals because no predictive test is accurate enough to completely rule out risk of relapse or some potential benefit from adjuvant therapy. However, many other patients are reluctant to accept the toxicities, inconvenience, and costs of chemotherapy for a small and uncertain benefit. For these individuals, more precise prediction of risk of recurrence and sensitivity to adjuvant therapy with genomic tests can assist in making a more informed decision.

1.8. Molecular Predictors of Response to Endocrine Therapy

1.9.1 Oncotype DX assay and other genomic predictors

One of the most important questions for patients with ER-positive breast cancer is whether they should receive adjuvant endocrine therapy alone or also take chemotherapy. Oncotype DX (Genomic Health, Inc., Redwood City, CA), represents a novel and commercially available molecular assay in the United States to assist decision making in this clinical setting. This RT-PCR-based assay represents an important conceptual advance in the diagnosis of ER-positive breast cancers. It measures the expression of 21 genes including ER and HER2 as well as several ER-regulated transcripts and proliferation related-genes including Ki-67. Several of these genes were already known to be associated with outcome and can be assessed with more conventional methods as well. However, an important value of Oncotype DX is that it combines these results into a simple and easily interpretable "recurrence score" (RS). The RS could be used as a continuous variable to estimate the probability of recurrence at 10-years or can be grouped into low-, intermediate- or high-risk categories. Correlation between RS and distant relapse was examined in 668 patients with ER-positive, node-negative cancers treated with tamoxifen who were enrolled in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B14 clinical trial. The 10-year distant recurrence rates were 7% (4-10%), 14% (8-20%), and 30% (24-37%) for the low-, intermediate-, and high-risk categories ($p<0.001$) (Paik et al., 2004). These results suggested that ER-positive patients with high RS may not be treated optimally with 5 years of tamoxifen. Similar results were observed for a community-based patient population (Habel LA, ASCO annual meeting proceeding 2005). The value of the recurrence score for predicting benefit from adjuvant cyclophosphamide, methotrexate, and 5-fluorouracil (CMF or MF) chemotherapy in ER-positive, node-negative breast cancers was also examined. A study that included 651 patients who were enrolled in the NSABP B20 randomized clinical trial, which compared adjuvant tamoxifen with tamoxifen plus CMF (or MF) adjuvant chemotherapy, showed that a higher RS was associated with greater benefit from adjuvant chemotherapy (Paik et al., 2004). The absolute improvement in 10-

year distant recurrence free survival was 28% (60% vs 88%) for patients with RS ≥31, while there was no benefit for patients with RS <18 (test for interaction p=0.038). The hazard ratio for distant recurrence after chemotherapy was 1.31 (0.46-3.78) for patients with RS <18 and 0.26 (0.13-0.53) for patients with scores ≥31. These data indicate that a RS identifies a subset of women with ER-positive and node-negative breast cancers who are at high risk of recurrence despite 5 years of tamoxifen therapy and that this risk can be reduced with adjuvant chemotherapy.

In aggregate, the available data suggests that Oncotype DX can be useful when the decision regarding adjuvant chemotherapy is not straightforward based on routine clinical variables. However, some important caveats must be noted. Oncotype DX is not appropriate for ER-negative patients because they are all categorized as high risk (Badve and Nakshatri, 2009). The predictive performance of this test in patients who receive adjuvant aromatase inhibitor therapy or third generation anthracycline and taxane combination adjuvant chemotherapy remains to be studied. In particular, the magnitude of benefit that patients with low or medium recurrence scores experience when treated with a third-generation adjuvant chemotherapy regimen is unknown. Also, the limited available data suggests that patients with lymph node-positive, ER-positive, low recurrence score disease continue to remain at substantial risk of recurrence if treated with 5 years of tamoxifen therapy only. One study suggested that these patients may have up to 40% risk of local or distant recurrence at 10 years (Albain K, Br C Res Treat 2007; vol 106, abs 10). In contrast, the risk of distant recurrence of patients with the same risk score and lymph node-negative disease was 7%. Oncotype DX is routinely used in the States, and this predictive chip is being formally assessed in Europe by the prospective † TAILORX (Trail Assigning Individual Option for Treatment) trial, which opened in 2007 for accrual.

Several other efforts were made to develop predictors of response to endocrine therapies among ER-positive patients. One group reported that the ratio of HOXB13 and IL17BR genes was predictive of disease-free survival in patients with early-stage, ER-positive, breast cancer, who received treatment with tamoxifen (Ma et al, 2004). Unfortunately, subsequent studies tested this 2-gene ratio in heterogeneously treated patient populations and often used different assay thresholds and

normalization methods; therefore the true prognostic or predictive values of this assay remain uncertain.

1.9. Gene Profiling Platforms

The microarray platforms that were used to develop the current clinical outcome predictors have a clear limitation; they confine their interrogation to the "mRNA world" as it was known 10-15 years ago. In the past few years, our knowledge of the RNA world has evolved rapidly. It is now apparent that a previously unrecognized complex world of small regulatory RNA species exists including microRNA, siRNA, snoRNA and vast regions of non-coding DNA, pseudo-genes and antisense DNA strands are also frequently transcribed. Alternative splice forms of mRNAs are commonly generated from the same gene and can lead to distinct transcripts with different functions. It is almost certain that this extended RNA world contains complementary information not fully captured by measuring the expression of previously know genes. The next generation of DNA arrays (e.g tiling arrays, miRNA arrays) will enable investigators to study the clinical and diagnostic potential of these new RNA species.

By performing our study on concurrent high-throughput genome wide analysis using SNP6.0 from Affymetrix. Copy number variation (CNV) data analysis by the side of EXON gene-level, and 'splice variant' analysis; we aim to overcome some of the limitation listed above. These two new platforms is the result of advancement in array technology with increasingly smaller feature size. The point to appreciate is that with feature densities increasing, more transcribed genome is covered. In fact rather than just measuring the *gene or the genome* expression, the array measures the abundance of RNA or the DNA fragments (*exon or more SNP regions*) in that region, respectively for EXON and SNP6.0 array.

Current molecular models of breast cancer biology are based on interactions between a few hundred molecules. However, gene expression data indicates that at least 5 -10 thousand different mRNA transcripts are expressed in every cancer and most of these have no known function in cancer biology. Many of these genes may prove to be

important novel drug targets and the technology that has lead to their discovery may also serve to select patients for these future therapies.

1.10. MicoRNA in tamoxifen resistance

MicroRNAs are one type of relatively newly identified noncoding RNA. Their mature products are small single-stranded species of around ~22 nucleotides in length (Bartel et al, 2009). MicroRNA play critical roles in silencing genes, often silencing a cluster of about 200 genes, during development. Binding of microRNAs to their target genes occurs by perfect match or mismatch base pairing to complementary sequences within the 3' untranslated region (UTR) of the mRNA. This results either in mRNA degradation (Llave et al, 2002), (Palatnik et al, 2003) or translational repression (Lee et al, 1993), (Esquela-Kerscher et al, 2006). They have a role as onco-miRs or tumour suppressors by regulating apoptosis factors (Cimmino A, Proc Natl Acad Sci). The role of microarray in tamoxifen-resistance has recently been explored by (Zhao et al, 2008), (Maillot et al, 2009), (Miller et al, 2008) and (Pogribny et al, 2007). These studies were mainly on the widespread repression of oestrogen-dependent proteins by a specific miRNA in breast cancer line growth (in-vitro). Only one study (Rodriguez-Gonzalez et al) showed miRNA-30c as an independent predictor of clinical benefit of endocrine therapy in advance oestrogen positive breast cancer.

MicroRNA 221/221 has been shown to be >1.8 fold upregulated in tamoxifen resistant breast cancer cell lines (Miller et al, 2008). The expression of mir-221/222 was also significantly (>2.0) elevated in ERBB2-positive primary breast cancer tissues that are known to be resistant to endocrine therapy. This group elegantly showed that ectopic expression of miR-221/222 render the parental MCF-7 breast cells resistant to tamoxifen. Furthermore, they shown $p27^{KIP1}$, a known target of miR-221/222, and a cell-cycle inhibitor, is reduced by 50% and may be a major role in conferring resistance to tamoxifen in MCF-7 cells.

Zhao JJ et al (Zhao et al, 2008) published simultaneously than miR-221/222 negatively regulates oestrogen receptor-α, at post-translational level, and confer

tamoxifen resistance in breast cancer. They found that the ERα protein but not the mRNA is suppressed in the miR-221/222 carriers, they undertook a search in miRNA Target Scan Database and found two sequence motif of the 3'-UTR of ERα matched miR-221 and miR-222 seed sequences, one of which is highly conserved, between human, mouse and rat. This group also found miR-221/222 to be upregulated in their ERα-negative cancer cell line work and ERα-negative primary breast tumour. Knockdown of miR-221 and miR-222 in MDA-MB-468 (a cell line which has high miR-221/222 and detectable ERα mRNA, but ERα-protein negative), partially restores ERα expression and tamoxifen sensitivity.

In my project, I set out to see if my tamoxifen-resistant primary breast tumour cases and tamoxifen-resistant breast cancer cell lines carried elevated miR-221. And if the positive miR-221 carriers had elevated ERBB2 mRNA expression. As all my primary breast tumour cases were ER-positive patients, we were unable to validate the findings of Zhao JJ et al, regarding miR-221 negatively control ERα protein expression. The aim of this exercise is to validate if our series of primary breast tumour and our breast cell lines (with their TR counterparts), with that of published findings, and I am pleased to say they do.

Then, I set off to investigate for new novel miRNAs from our combined integrated microarray analyses, and validated them for correlation with TR cases, in breast cell lines and human primary tissue. In the future, if time permit, I intend to carry out in-vitro study of the most promising miRNA from our study.

1.11. Melanoma Associated Antigens (MAGE)

The melanoma antigen (MAGE) genes were initially isolated from melanomas based on their almost exclusive tumour-specific expression pattern. Subsequently a large number of human genes encoding tumour-specific antigens were isolated, including melanoma antigen families MAGE, BAGE, GAGE and LAGE (Chen et al, 1998; De Backer et al, 1999; Zendman et al, 2002). The MAGE family has since been divided into two sub-families termed Class I and Class II MAGE genes. The only conserved

domain found in all members of this family is the MAGE homology domain (MHD), a stretch of ~200 amino acids located towards the carboxy-terminus of the proteins (Chomez et al, 2001). Within Class I, the MAGE-A (12 genes), MAGE-B (6 genes) and MAGE-C (3 genes) subfamilies are highly (50-80%) homologous and positioned on chromosome X. Their lack of expression in somatic adult tissues but frequent aberrant expression in tumours defines them as cancer-testis (CT-X) antigens, proteins encoded on the X chromosome which are normally only expressed by gametes and trophoblast cells (Simpson et al, 2005). The Class II MAGE antigens in contrast are encoded by genes found at a variety of chromosomal locations, expressed in adult tissues and have not been found to be upregulated in tumours (Forslund & Nordqvist, 2001).

The mechanism behind the activation of Class I *MAGE* genes in cancer has generally been assumed to be due to genome-wide DNA hypometylation (Simpson et al, 2005), which is a frequently observed epigenetic event during carcinogenesis and is directly associated with induction of tumours in mice (Ehrlich et al, 2009; Gama-Sosa et al, 1983). However, it has been recently reported that the multifunctional DNA binding protein, BORIS, itself a CT-X antigen, is able to induce epigenetic reprogramming (Loukinov et al, 2006) and was shown to act as a potent activator for the expression of several *MAGEA* genes (Vatolin et al, 2005). In this context, deregulated BORIS (found in numerous human cancers) could contribute to the induction of *MAGE* expression in tumours.

The precise role of MAGE antigens remains unclear, although Necdin, one of the Class II proteins, is thought to regulate growth and differentiation is certain cell types (Chapman & Knowles, 2009). The MHD is believed to act as a protein-protein interaction domain and MAGE antigens may act as scaffolding proteins to regulate the activity of key cellular proteins, including p53. In a study examining the involvement of *MAGEA2* expression in the acquisition of resistance to etoposide, it was suggested that MAGEA2 protein acted to suppress wild-type p53 activity thereby protecting chemoresistant cells from apoptosis. MAGEA2 was shown to directly complex with p53 and recruit HDAC3 to repress p53 activity as a transcription factor (Monte et al, 2006). In a separate study, a number of class I

MAGE antigens (from the A, B and C subfamilies) were found to be able to interact indirectly with p53 via another scaffolding protein, the transcriptional co-repressor protein KAP1, again resulting in a suppression of apoptosis (Yang et al, 2007).

The data above suggest that MAGE protein expression in tumours may therefore be associated with treatment failure. These more recent findings have added weight to the concept of using MAGE as targets for cancer immunotherapy. The absence of expression in somatic tissues, but induction in a variety of tumour types including melanoma, small cell lung carcinoma, germ cell tumours and also breast cancer (Caballero & Chen, 2009; Grigoriadis et al, 2009) makes MAGE proteins ideal tumour-specific antigens. However, the ability of MAGE antigens to induce spontaneous cytotoxic T-lymphocyte (CTL)-dependent immune responses in cancer patients particularly marks them out as useful targets for immunotherapy (Van den Eynde et al, 1995). Cancer vaccine phase II trials based on recombinant MAGEA3 antigen are currently in progress for lung cancer and melanoma and show promise (Atanackovic et al, 2008; Caballero & Chen, 2009). Due to their high sequence homology, it is difficult to generate antibodies that can distinguish between MAGE subfamily members, therefore it is likely that tumours that express a range of *MAGEA* genes may potentially be targeted using this therapy.

1.12. EGLN3 (Elg Nine Homolog 3)

Hypoxia Inducible Factor (HIF), is the master transcriptional regulator of hypoxia-induced gene expression and consists of a labile α subunit and a stable β subunit (also known as HIFβ or ARNT). In the presence of oxygen, HIFα family members (ubiquitous HIF-1α or cell type-specific HIF-2α) are hydroxylated at one of two conserved prolyl residues by members of the egg-laying-defective nine (EGLN)

family (also termed prolyl hydroxylase domain (PHD) proteins) which act as intracellular oxygen sensors. Prolyl hydroxylation generates a binding site for a ubiquitin ligase complex containing the von Hippel-Lindau (VHL) tumour suppressor protein, and results in HIFα degradation via the proteasome. In addition, in the presence of oxygen, HIFα function is modulated by asparagine hydroxylation by FIH (factor-inhibiting HIF), which inhibits HIFα recruitment of the transcriptional coactivators p300 and CBP, and hence its function as a transcriptional activator. During hypoxia, EGLN/PHD activity decreases such that HIF degradation is blocked leading to the activation of its ~100 target genes including *VEGF* and *GLUT1* (see Figure 1.6, reviewed Loboda et al, 2010).

QuickTime™ and a
decompressor
are needed to see this picture.

There are three known members in the EGLN/PHD family, EGLN1, EGLN2 and EGLN3 (Taylor et al, 2001), in human and mouse. The minimal HIF-derived peptides efficiently hydroxylated by these enzymes are typically long (~19 mers), in contrast to the X-Pro-Gly peptides acted on by collagen prolyl hydroxylases (Hirsila et al, 2003). All three proteins have been reported to hydroxylate HIFα and to have similar dependence on oxygen and the co-factors Fe (II) and 2-oxo-gutarate. EGLN1 (PHD2) is considered to be the primary HIFα prolyl hydroxylase under normal conditions (Ivan et al, 2008). It is possible that EGLN2 (PHD1) and EGLN3 (PHD3) regulate HIFα under different conditions. EGLN3 is itself upregulated during hypoxia and is considered to be a HIF target gene. EGLN2 is primarily nuclear while EGLN3 is seen in both the cytoplasm and the nucleus (Metzen et al, 2007; Marxsen et al, 2004).

Regulation of EGLN is associated with mitochondrial generated species (Chandel et al, 2010), nitric oxide (Metzen et al, 2003), (Sandau et al, 2001; Sandau et al, 2001), oncogenes v-src, activated ras and PI3K/AKT (Chan et al, 2002) and of course its own expression and stability. Of the three family members, less is understood about the role of EGLN3; it has been suggested to regulate apoptosis in neural cells and this may involve protein aggregation (Rantanen et al, 2008). In addition, a non-HIF hydroxylation target, ATF4, has been reported for EGLN3 (Koditz et al, 2007).

The ability to withstand prolonged hypoxia is one of the hallmarks of tumours. Numerous studies that have reported aberrant activation of HIF in many tumour types and several key HIF target genes are associated with the induction of tumour angiogenesis. Consequently, strategies to target HIF in tumours are actively being explored (reviewed, Semenza, 2010). The expression of the EGLN proteins has also been associated with cancer, in particular EGLN2 has been shown to be oestrogen regulated in breast cancer and associate with tamoxifen resistance (Seth et al, 2002). More recently, *EGLN2* was found to promote proliferation in breast and other tumour cell lines in a cyclin D1-dependent, but HIF-independent manner, thereby suggesting that small molecule inhibitors of these enzymes currently in development may have a role in cancer therapy (Zhang et al, 2009).

1.13. Aims of Project

The project is aimed at elucidating the mechanism and genes responsible for tamoxifen resistance (TR). The study will be done on *in vitro* TR breast cell lines and human TR breast tissue.

Cell line study:

1) Generate in-house TR and oestrogen-deprived (OD) versions of ER+ve cell lines T47D and ZR75-1.
2) Perform gene profiling on Affymetrix HU133 plus 2.0 gene chips on RNA from T47D, ZR75-1 and the OD T47D, ZR75-1 cell lines that have been transformed into TR.
3) Compare data with that previously obtained using WT and TR MCF-7 cells.
4) Validate genes with significantly differentiated expression from profiling using real-time PCR (qPCR). Immuno-histochemistry (IHC).
5) Functional study on biologically relevant/interesting genes by stable overexpression in tamoxifen sensitive breast lines to determine their effect (+/- tamoxifen) on proliferation, migration, invasion *in vitro* and if possible in xenograft studies.
6) Determine the relationship of functionally validated gene(s) with response to tamoxifen in tumour samples from ER+ve tumours of known outcome on tamoxifen monotherapy (see point 3 below).

Human TR Breast Cancer Tissue:

1) Our study aims to determine a set of genetic markers that is predictive of the response tamoxifen therapy using EXON expression arrays and genome-wide SNP6.0 Affymetrix chips.
2) The study will be done on a small (n=25 specimens) but unique series of fresh frozen samples (training set) from Guys and St Thomas / King's College London (GSTFT/KCL) Breast Tissue Bank). Cases were selected from patients participating in the European Oncology Research Trial Consortium (EORTC) 10850 & 10851 studies under the LREC Ref 06/Q0603/25. The samples are from patients with a known response to tamoxifen and for whom complete clinical follow-up is available.
3) The selected predictive gene-sets will be validated using qPCR and IHC on 129 validation cases from the same EORTC trial and an independent TMA from Bart's and the London hospital, for which we have the complete clinical follow up data.

In conclusion, the study aim to find predictive markers which are dependent and independent on Oestrogen Receptor (ESR1) co-expressed genes for the response to tamoxifen. From the TR cell line study, I hope this study will reveal the escape mechanisms from the effect of tamoxifen, which may shed light on pathways which account for tamoxifen resistance.

CHAPTER 2: MATERIALS AND METHODS

2. Cell Culture

2.1. Mammary Lines

Cell lines were cultured as described in Table 2.1. All cells were regularly passaged to maintain exponential growth. Mammary cell lines T47D and ZR-75 were grown in DMEM with 10% charcoal stripped Foetal Calf Serum at 10% CO_2. These were grown for six months and labelled Oestrogen Deprieved (OD) T47D and OD ZR-75. All cells were regularly screened for *Mycoplasma sp.* contamination.

Cell Line	Description	Media
MCF-7 (WT) Tenovus	Mammary Carcinoma	**DMEM, 10% Foetal Calf Serum.**
ZR75-1	Mammary Carcinoma	RPMI, 10% Foetal Calf Serum.
T47D	Mammary Carcinoma	DMEM, 10% Foetal Calf Serum.
OD ZR75-1	Mammary Carcinoma	RPMI, 10% charcoal-stripped Foetal Calf Serum.
OD T47D	Mammary Carcinoma	DMEM, 10% charcoal-stripped Foetal Calf Serum.

Table 2.1 List of cell lines and cell culture conditions. MCF-7 and T47D were grown in 10% CO_2 at 37^{o}C. ZR75-1 were grown in 5% CO_2 at 37^{o}C.

2.1.1. Tamoxifen Resistant Breast Cell lines

The above cell lines were generated by growing in media containing 10^{-7} M hydroxyl-tamoxifen. These were grown for six months with regular media changes and labelled Tamoxifen Resistant (TR).

Cell Line	Description	Media
MCF-7 (Tenovus) TR	Mammary Carcinoma	**DMEM, 10% Foetal Calf Serum and 10^{-7} M hydroxy-tamoxifen.**
ZR TR	Mammary Carcinoma	RPMI, 10% Foetal Calf Serum and 10^{-7} M hydroxy-tamoxifen.
T47D TR	Mammary Carcinoma	DMEM, 10% Foetal Calf Serum and 10^{-7} M hydroxy-tamoxifen.
OD ZR TR	Mammary Carcinoma	RPMI, 10%charcoal-stripped Foetal Calf Serum and 10^{-7} M hydroxy-tamoxifen.
OD T47D TR	Mammary Carcinoma	DMEM, 10% charcoal-stripped Foetal Calf Serum and 10^{-7} M hydroxy-tamoxifen.

2.2. RNA Interference

2.2.1. siRNA Selection

Initially sequences specific to MageA2 were selected for the generation of siRNA oligonucleotides based on published suggestions (Yang *et al.*, 2007b). Also used was pool of four siRNAs, based on highly conserved MageA2 sequence (SMARTpool; Dharmacon Research). These oligonucleotides were able to knock down all tested MageA genes (MageA1,-A2,-A3, and –A6; data not shown). BLAST searches were performed to verify the MageA2 specificity. A non-specific random sequence, Allstar® negative control (Qiagen) was used for a non-silencing control. These siRNA target sequences are detailed in Table 2.2 and 2.3.

Target	Name	Target Sequence 5' - 3'
Mage-A2	Mage-A2_1	AUUCGUUCACAAUAUAGGCUU
	Mage-A2_2	UCUCCACCGAUCUUUAGUGUU
	Mage-A2_3	GUCCUGGCAAUUUCUGAGGUU
	Mage-A2_4	UAUCACACGAGGCAGUGGAUU
Mage-A2B	Mage-A2B_1	AUUCGUUCACAAUAUAGGCUU
	Mage-A2B_2	UCUCCACCGAUCUUUAGUGUU
	Mage-A2B_3	CCUCAGAAAUUGCCAGGACUU
	Mage-A2B_4	UAUCACACGAGGCAGUGGAUU
Non Silencing Control siRNA	All star negative control	AATTCTCCGAACGTGTCACGT

Table 2.2 siRNA Target Sequences. The non-specific random sequence (Qiagen) was used for a non-silencing control for transient transfection.

Catalogue Number	Name	GenBank code
J-006350-09, -10, -11, -12	Human MAGEA2	NM_175742
J-019148-09, -10, -11, -12	Human MAGEA2B	NM_153488

Table 2.3 MageA2 siRNA targeting sequences from Dhamacon. The sequence is identical to that shown in Table 2.2.

2.2.2. Transient Transfection of siRNA Oligonucleotides

T47D TR and OD T47D TR cells at 60 % confluency were transfected in six well plates using Oligofectamine as per the manufacturer's instructions (Invitrogen). Briefly the siRNA was diluted in DMEM to a volume of 185µl. At the same time 3µl of Oligofectamine reagent was diluted in 12 µl of DMEM, and allowed to equilibrate for 5-10 minutes. The diluted transfection reagent was then added the diluted siRNA, mixed by inversion and incubated at room temperature for 20 minutes. Following two washes of cells in PBS, 800µl of DMEM was added per well. The transfection complexes were then added to the cells and incubated at 37 ^{o}C. Four hours later 500µl of 3x Complete Medium (DMEM + 10% FCS) was added per well, and Mage-A2 knock down was assayed at time points thereafter. All the siRNA concentrations referred to in this report represent the final siRNA concentration in 1ml, the volume in which transfection complexes were incubated on the cells for 4 hours.

2.2.3. Transfection via Nucleophoresis using Amaxa Nucleofector

The same siRNA from Dharmacon were also transfected via the Amaxa Nucleofection method. The optimised protocol can be found on the Amaxa company website (Amaxa, url). The Amaxa nucleofection kit came with Nucleofection Equipment with the appropriate software installed, (Ver V2.4 for Nucleofector 1 Device), was available and a kit containing the Nucleofector solution and suppliment, pmax GFP (as a positive control) and curvettes was purchased.

1 X 10^6 T47D TR cells were needed for each well of the 6-well plates used. These cells are harvested, trypsinised and gently spun down at 1000 G for 1 minute in a 15ml falcon tube. The media was removed by vacuum suction and the pellet left undisturbed. Meanwhile, the media and nucleofector solution is equilibrated to room temperature. The cells are resuspended in 100µl of Nucleofector solution and supplements followed by the addition of 0.5-3µg of siRNA, in this case 3µg of MageA2 siRNA was found to be optimal. The nucleofector sample is transfered into

an Amaxa certified curvette placed in the curvette holder and the program is started. T47D cell lines require the X-05 or the X-005 program. The cells were then immediately removed and 900µl of pre-warmed culture medium was added to the curvette and the total volume of approximately 1000µl is now seeded into each well of a 6-well plate containing 1ml of medium per well. The final volume per well is 2 ml. Cells are incubated at 37ºC and 10% CO_2. The cells were than incubated for 24 hours before lysis and whole cell extraction.

2.2.4. Generation of Stable MageA2 Expression Clones

The *MAGEA2* (Clone ID: 8327628) coding sequence was obtained in the pCR4-TOPO vector from the IMAGE Consortium. The insert was excised using an EcoRI restriction digest and cloned into the pcDNA3.1 expression vector, which carries a strong mammalian promoter, CMV (cytomegalovirus) and also the Neomycin (G418) selection marker. Prior to this, an optimisation experiment of ascending doses (200mcg, 300mcg, 400 mcg, 500 mcg, 600 mcg/ml of media) of G418 concentration was undertaken to determine the dose just below the kill-dose of specific breast cancer lines. Cells were left in G418 containing media for 48 hours.

An outline of the strategy used to achieve this is shown in Figure 2.1 and the map of the final clone is shown in Figure 2.2.

2.2.4.1. Enzyme digestion

All restriction enzymes (RE) used for DNA digestion and related buffers and solutions used were supplied by New England Biolabs (NEB). In most cases, 1µg of DNA was digested with 1Unit of restriction enzyme in 1× enzyme buffer in a total volume of 20µl for 1h at 37ºC. Small aliquots were run on an agarose gel to confirm digestion.

2.2.4.2. Agarose Gel Electrophoresis

Enzyme digests were analyzed by agarose gel electrophoresis. 20µl of the samples were loaded in wells alongside DNA ladder (Hyperladder 1, Bioline) and the gel was run for 1h at 100V for optimal separation. In order to make the gel, agarose powder

(1%) was added to 1X TBE (108g Tris base/55g Boric Acid/9.3 EDTA in 1L of H_2O), and dissolved by boiling. 0.5µg/ml of ethidium bromide was added to the gel to facilitate visualization of DNA fragments under UV light. Fragments of appropriate size (MAGEA2, 1kb, see Results) were excised from the gel and purified using Zymoclean Gel DNA Recovery kit.

2.2.4.3. Ligation

Linearized expression vector pcDNA3.1 (previously dephosphorylated using Shrimp Alkaline Phosphatase in 1× USB RX SAP buffer) was incubated overnight at 16ºC with T4 DNA ligase (1µl), insert and T4 DNA Ligase buffer in a total volume of 20µl. A control reaction with no insert was also prepared. The relative ratio of insert to vector was 3:1.This ratio was calculated as follows: amount of vector used (ng) was multiplied by the size of the insert (Kb) and the product was divided by the size of the vector (Kb)**.**

2.2.4.4. Transformation

2 µl of the ligation reaction was transformed into competent E.coli. The protocol is as follows: Add 20µl ligation reaction to 50µl competant cells. Incubate on ice for 5-30 min. Heat shock for 30 sec at 42ºC, incubate on ice for 2 min. Add 250 µl SOC buffer and incubate at 37ºC for 1 hour. 20µl from each transformation was plated on a prewarmed agar plate (LB medium +Amplicilin) and incubated overnight at 37 to select for positive clones. Next day individual colonies from agar plates were picked and placed into separate tubes containing 5ml LB + Ampicilin to grow up.

2.2.4.5. Maxiprep and Miniprep

Isolation of small or large quantities of plasmid DNA was achieved by using Miniprep or Maxiprep Qiagen kit, respectively, by following the manufacturer's instructions.

2.2.4.6. In vitro translation

1µg of designed construct, pcDNA3.1/MAGEA2, and 2µg of the control plasmid, encoding luciferase (Promega), was used for mRNA synthesis, and the translation

reactions were carried out according to the manufacturer's instructions using the TnT Coupled Reticulocyte Lysate System (Promega).

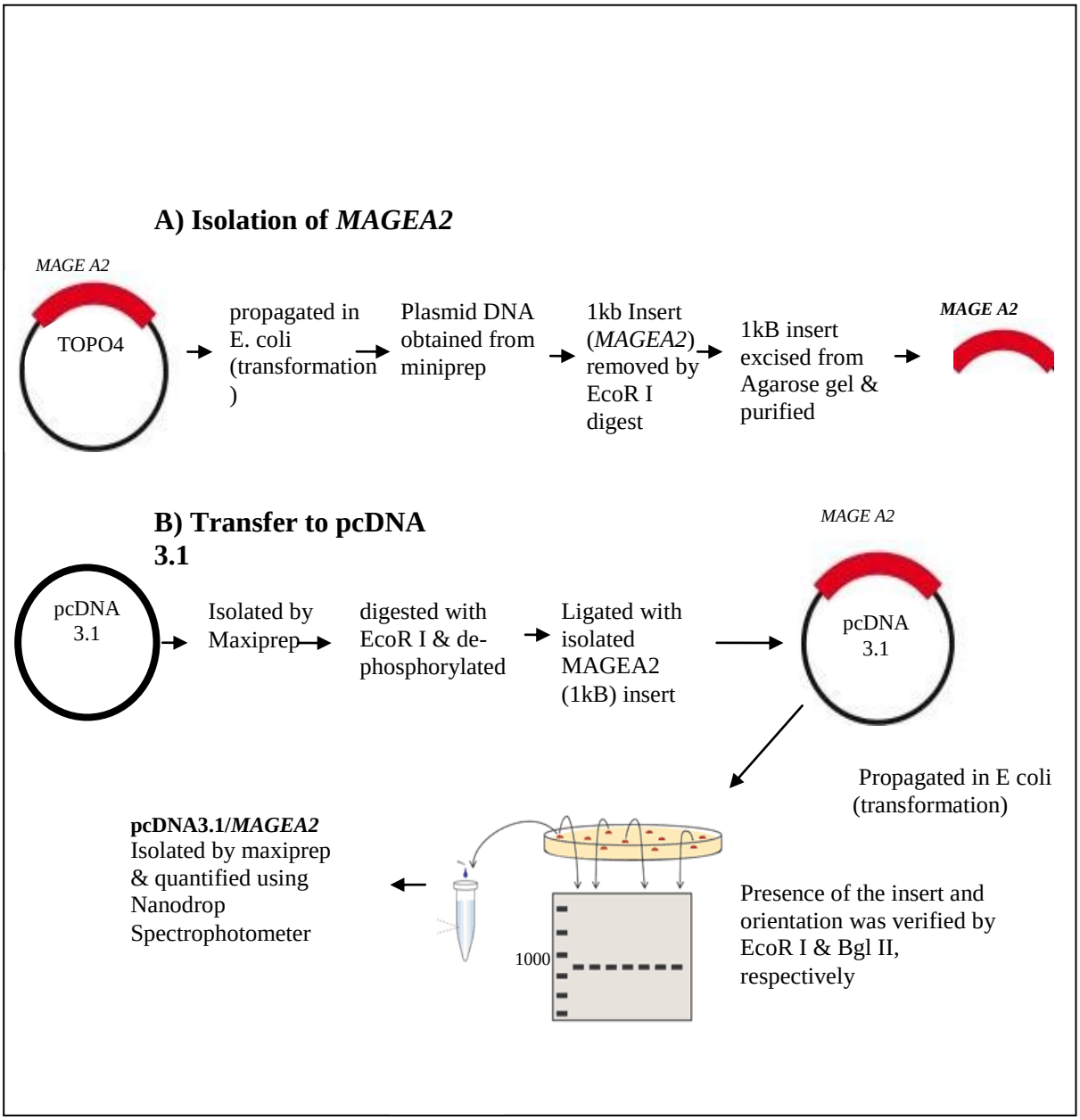


Figure 2.1 Cloning strategy (A) Sub-cloning stages where MAGEA2 on TOPO4 vector (from the IMAGE consortium) is propagated then amplified by miniprep. The MAGEA2 insert is then removed after digested by EcoR1. The construct MAGEA2 is purified. (B) MAGEA2 is then transferred into vector pcDNA3.1 after pcDNA3.1 is digested and ligated with MAGEA2. Finally the presence and orientation of the insert was verified by EcoR1 and Bgl-11.

The orientation of the *MAGEA2* insert was determined by a Bgl II digest. This enzyme cuts once within the insert (see Figure 8B) and once in the vector. Clones in the correct orientation should give a band of 1.7 Kb as shown in Figure 3A for the clones in lanes 1, 3, 5 and 7.One of these was selected and prepared for further experiments.

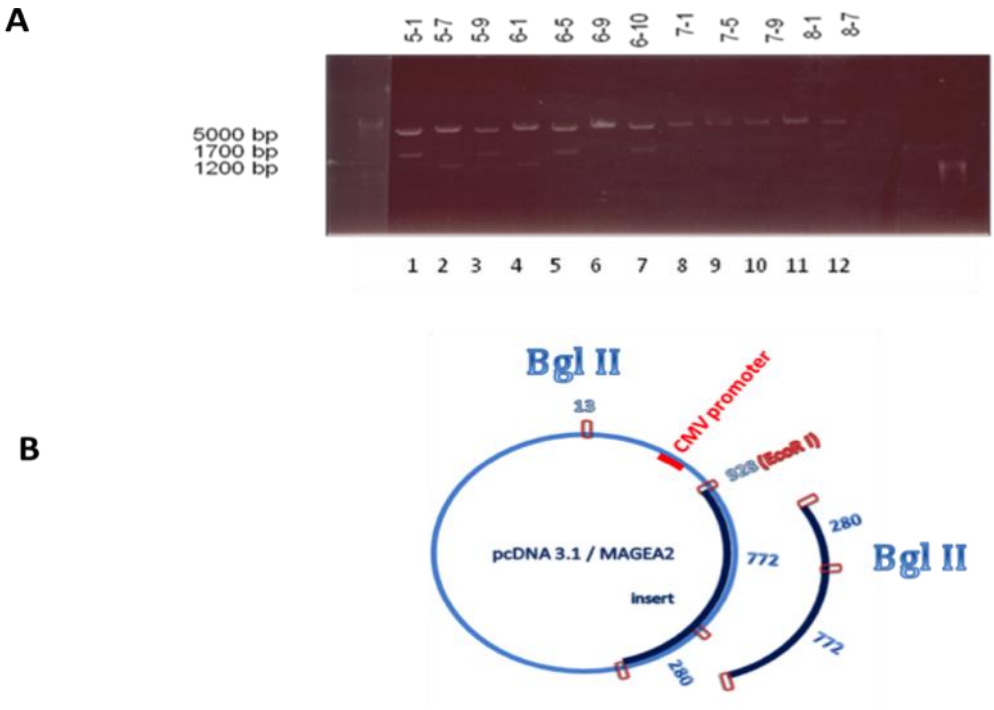


Figure 2.2 Characterisation and structure of the *MAGEA2* expression plasmid. **A)** Bgl II restriction mapping of 12 miniprep test clones of pcDNA3.1/*MAGEA2*. The samples were separated on a 1% agarose gel. **B)** Map of the final expression construct with *MAGEA2* cloned into EcoRI site at position 926 in pcDNA3.1. Bgl II cuts pcDNA3.1at 13bp and also cuts asymmetrically within the 1025bp EcoRI *MAGEA2* insert as shown. Thus, pcDNA3.1/*MAGEA2* cut with Bgl II will give a fragment of 1.7Kb or 1.2Kb depending on the orientation of the *MAGEA2* insert. The plasmid constructs that have the correct orientation are those with a 1.7Kb Bgl II fragment.

2.3. Western Blotting

2.3.1. Whole Cell Extracts

Cells were lysed using extract buffer (8M Urea, 1M Thiourea, 0.5% CHAPS, 24mM Spermine, 50mM DTT). Extracts were collected using cell scrapers. The approximate protein content of each extract was determined using the Bradford Assay (Biorad) and a BSA standard curve ranging from 0 – 20µg/ul. Typically 5µg of whole cell extracts were used in SDS PAGE.

2.3.2. SDS PAGE and Western Blotting

Samples were separated on 6-12% SDS PAGE gels prepared with the BioRad Mini Protean system. Full range molecular weight Rainbow Markers (Amersham) were loaded to allow size determination of detected proteins. Separated proteins were transferred onto a nitrocellulose membrane (Amersham) using semi-dry blotting apparatus (Biorad). The membrane was blocked (5% Marvel 0.1% Tween-20 in PBS) for an hour at room temperature. The primary antibody was diluted in blocking solution and incubated with the membrane for an hour at room temperature. The membranes were washed in blocking solution twice and a further 3 times in PBS / 0.1% Tween-20 to remove excess secondary antibody. Supersignal WestFemto reagent (Pierce) was used for chemiluminescence and the membrane was exposed on autoradiograph film. A summary of the antibodies used is shown in Table 2.4.

Antigen	Animal origin	Source	Dilution
Mage-A (pan-A)	Mouse	Zymed	1:500
Mage-A2 (3ES)	Mouse	Santa Cruz	1:200
Mage-A2	Rabbit	Abcam	1:100
PgR	Mouse	Santa Cruz	1:200
PgR	Mouse	Dako	1:100
ErbB2	Rabbit	Cell Signalling	1:1000
Egln3	Mouse	Gift from Peter Ratcliffe	1:20
KAP1	Rabbit	Abcam	1:500
Egln3	Mouse	Peter Ratcliff, Oxford	1:30
HDAC3 (B-12)	Mouse	Santa Cruz	1:500
p53 (DO-1)	Mouse	Santa Cruz	1:500
Acetyl-p53	Rabbit	Cell Signalling	1:1000
Phospho-p53 (Ser 15)	Rabbit	Cell Signalling	1:1000
Phospho-p53 (Ser 20)	Rabbit	Cell Signalling	1:1000
Phospho-p53 (Ser 46)	Rabbit	Cell Signalling	1:1000
Phospho-p53 (thr 18)	Rabbit	Cell Signalling	1:1000
Phospho-p53 (Ser 37)	Rabbit	Cell Signalling	1:1000
PARP-1	Mouse	Santa Cruz	1:250
THRAP5	Rabbit	Abcam	1:100
v-Myb	Mouse	Abcam	1:500
VGLL1	Rabbit	Abcam	1:50
PDZK1	Mouse	Abcam	1:100
HnRNPA2B1	Mouse	Abcam	1:100
HDAC3	Mouse	Santa Cruz	1:100
ERα (HC-20)	Rabbit	Santa Cruz	1:500
Inhibin βA	Mouse	R&D systems	1:10
Ku-70 (C19)	Goat	Santa Cruz	1:1000
Myc (9E10)	Mouse	Santa Cruz	1:500
p21 (DCS60)	Mouse	Cell Signalling	1:1000
p27 (# 2552)	Rabbit	Cell Signalling	1:1000
p53 (DO1)	Mouse	Santa Cruz	1:1000
p300 (N-15)	Rabbit	Santa Cruz	1:500
GAPDH	Mouse	Santa Cruz	1:100
PCNA	Rabbit	Santa Cruz	1:1000

Table 2.4 List of antibodies and their dilutions for Western blotting.

2.4. Fresh frozen tissue for microarray study

Fresh frozen tissues from 25 patients for the microarray studies came from the EORTC 10850 & 10851 trial and generously donated to us by Guys and St Thomas / King's College London (GSTFT/KCL) Breast Tissue Bank the Tissue Bank for the study of tamoxifen resistance in breast cancer (Ethnic approval number LREC Ref 06/Q0603/25). All cases were Oestrogen-receptor positive (ER-positive). They were all treated with tamoxifen after surgery. These cases are unique for the reasons below:

1) The patients, which we have defined as tamoxifen Sensitive, were patients who had positive surgical margins (i.e. incomplete resection) after their primary breast surgery. These patients chose not to return to have a repeat resection. The treatment with tamoxifen led to long-term survival. They are 'true' sensitive cases.

2) The patients, which we have defined as tamoxifen Resistant, were patients who had primary surgery and then adjuvant tamoxifen treatment and relapsed/had a recurrent within 2 years.

3) All their clinical follow-up (1984-1991) data is complete. They were randomised to receive either lumpectomy or radical mastectomy followed by tamoxifen 20mg daily. No further treatment for their diagnosis of breast cancer.

4) In the 25 samples, 3 patients had paired samples; where frozen tissues from primary surgery and relapsed stage were obtained. The relapsed fresh tissues will give informative microarray data, which will represent the tamoxifen-resistance mechanism better.

Ten normal breast tissues were donated by Prof Louise Jones for normalisation (LREC Ref 05Q403/199).

2.5. RNA extraction

2.5.1. Extraction of total RNA from tissue samples

Total RNA was isolated from tissues with TRIzol Reagent (Invitrogen Life Technologies) according to the manufacturer's protocol. Briefly, 50 mg of each frozen tissue was homogenized in 1 ml of TRIzol Reagent (Invitrogen) using a power homogenizer (IKA ULTRA-TURRAX, T25 basic). After centrifugation at 12,000 g for 10 min at 4^{o}C the supernatant was transferred to a fresh tube and the content was sheared 10-20 times using a 20gauge needle/syringe in order to fragment genomic DNA. 0.2 ml of chloroform was then added per 1 ml of TRIzol reagent and mixed by vigorous shaking for 15 seconds. Following centrifugation at 12,000 g for 15 minutes at 4^{o}C, the colourless upper aqueous phase was collected. Subsequently, 0.5 ml of isopropyl alcohol was added per 1 ml of TRIzol Reagent and incubated at -20^{o}C for 20 minutes. After spinning at 12,000 g at 4^{o}C for 10 minutes the supernatant was removed and the RNA pellet was washed with 1 ml of cold 75% ethanol and spun at 7,500 g for 5 minutes at 4^{o}C. The supernatant was carefully removed and the pellet was air-dried for 5-10 minutes then dissolved in diethyl pyrocarbonate (DEPC)-treated water.

2.5.2. Extraction of total RNA from cell lines

Cells were washed twice with ice cold PBS. In each well of a six well plate 250µl TRIzol reagent (Invitrogen) was added to the cells and incubated for 5 minutes at room temperature. TRIzol extracts were collected using cell scrapers. 50µl chloroform was added per 250µl TRIzol, vortexed for 15 seconds, and then incubated for 2-3 minutes at room temperature. Samples were then centrifuged at 12,000g for 15 minutes at 4˚C and the aqueous upper phase was transferred to a new microcentrifuge tube. An equal volume of 70% ethanol was added and mixed by pipetting. The mixture was applied to an RNeasy Mini column (QIAGEN) then centrifuged for 15 seconds at 8000g. RNA clean-up was followed using the manufactures instructions, including the DNAse I digestion step. The RNA concentration was estimated by spectrophotometer and formaldehyde-MOPS

denaturing agarose gel electrophoresis was then used to assess RNA quality. All extracted RNA were quantified and quality checke by BioAgilent NanoCHIP 2100 Bioanalyzer. (Agilent, url)

2.6. DNA extraction

High-quality DNA was extracted from small amount 10mg of breast tissue using the DNA easy Micro Kit (Qiagen). DNA was quantified and qualified by NanoDrop ND-1000 Spectrometer (NanoDrop Technologies). DNA was also checked via agarose gel electrophoresis.

2.7. Gene Expression Microarrays

2.7.1. The Affymetrix GeneChip

All experiments were performed using Human Genome U133 Plus 2.0 high-density oligonucleotide arrays (Affymetrix, *url*). Oligonucleotides of 25 base pairs in length are used to probe message levels in the samples. Each gene of interest is represented by a set of oligonucleotides comprised of 11 probe pairs. Each probe pair is composed of a perfect match (PM) probe against a section of the mRNA molecule of interest, and a mismatch (MM) probe that is created by changing the middle (13th) base of the PM with the intention of measuring non-specific binding. The HG-U133-Plus2 array contains 54,000 probe sets, representing an estimated 47,000 human transcripts.

2.7.2. Target Preparation, Microarray Hybridisation, Staining and Scanning

High quality RNA from each sample was used to prepare biotinylated target RNA, according to the manufacturer's recommendations (Affymetrix, *url*). An overview of this procedure is shown in Figure 2.3. Briefly, 5µg of total RNA was used to generate first-strand cDNA by using a T7-linked oligo(dT) primer. After second-strand synthesis, *in vitro* transcription was performed with biotinylated UTP and CTP, resulting in approximately 100-fold amplification of RNA. This labelled cRNA target was quantified and then fragmented before preparation of the

hybridisation cocktail. Spike in controls were added to the fragmented cRNA (15µg per HG-U133Plus2 array), before overnight hybridisation. Arrays were then washed and stained with streptavidin-phycoerythrin, before being scanned on an Affymetrix GeneChip 3000 scanner. After scanning, array images were assessed by eye to confirm scanner alignment, the absence of significant bubbles or scratches on the chip surface, and the absence of slides with very high background (scanning and image analysis was performed by Tracy Chaplin, Institute of Cancer, Charterhouse Square).

2.7.3. Data Analysis

2.7.3.1. Quality Control

Quality Control of the probes prepared were analysed by Affymetrix Console® (Publicly made available software by Affymetrix launched in 2007). In order to ensure the arrays fulfilled a series of Affymetrix recommended quality control metrics (described in detail in the Data Analysis Fundamentals Manual, Affymetrix, *url*), the raw data were pre-processed using the "simpleaffy" BioConductor package. Simpleaffy uses the Affymetrix MAS 5.0 array processing algorithms (described in detail below in section 2.7.3) to provide access to these metrics.

2.7.3.2. Data Analysis

Data analysis was performed using BioConductor (Gentleman *et al.*, 2004) and Partek® Genomic Suite (Version 4.0) software. BioConductor is an open development software project, providing access to a wide range of statistical and graphical approaches for the analysis of genomic data. It works through the R open source programming language (Ihaka and Gentleman, 1996). Data was processed from the *.CEL file format which contains information on background values and perfect match and mismatch intensities. Data processing was done using "affy" (Gautier *et al.*, 2004) and "simpleaffy" (Wilson & Miller, 2005) packages in BioConductor.

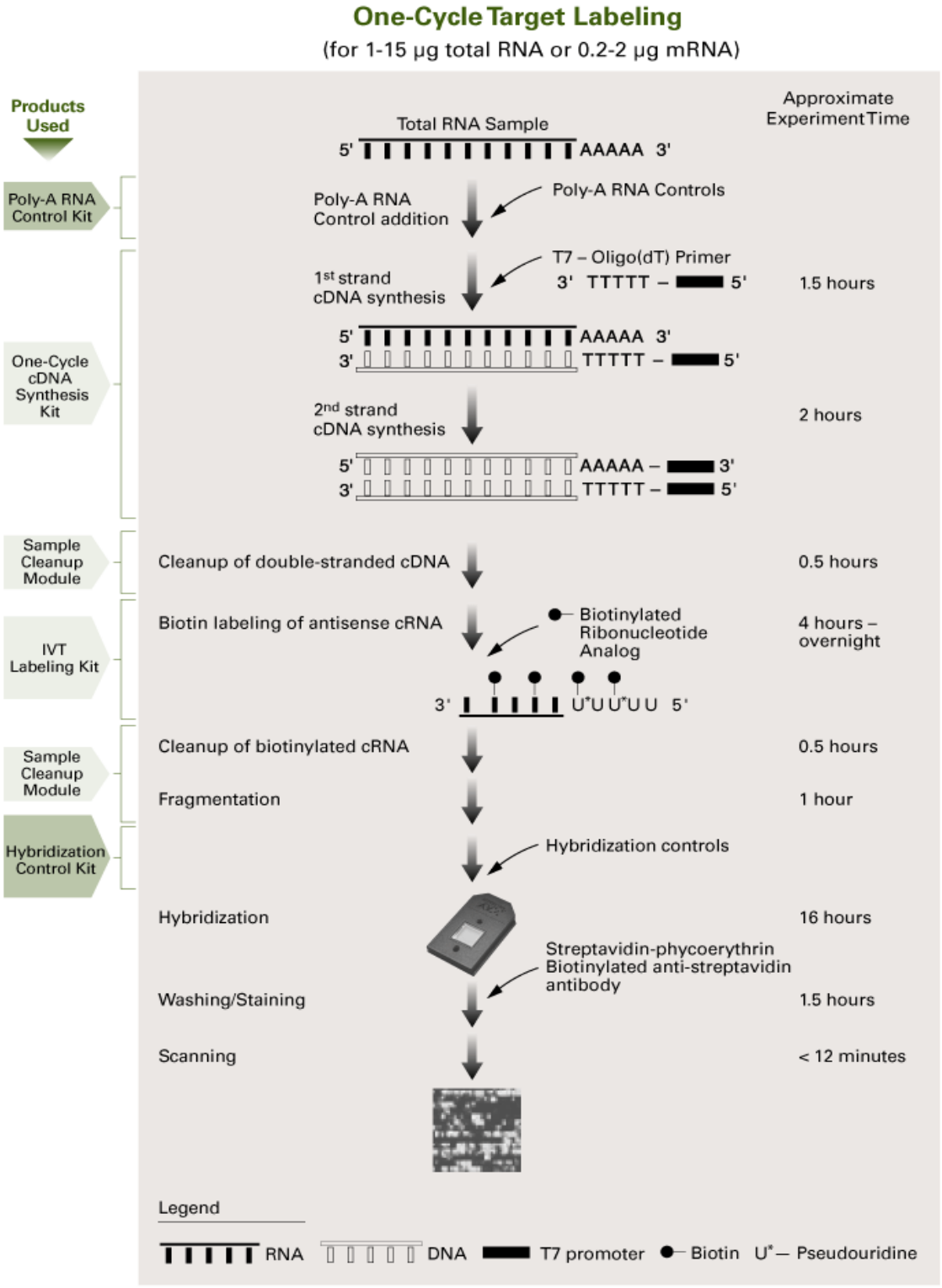


Figure 2.3 Affymetrix GeneChip Eukaryotic Sample and Array Processing. A more detailed description of this procedure can be found in Section 3: Eukaryotic Sample and Array Processing of the GeneChip Expression Analysis Technical Manual (Affymetrix, *url*).

2.7.3.2.1. Average background

The average background is the level of signal detected by the scanner surrounding the signal from the specific features. No Affymetrix guidelines are suggested for average background only that the values are similar and typically in the range of 20 to 100. Any arrays with unusually high background would be discarded from this analysis following the visual inspection.

2.7.3.2.2. Scale factor

Before the data from different arrays can be compared, a global normalisation method needs to be conducted in order to minimise variation between arrays. Variation is caused by biological and experimental factors throughout the microarray protocol, starting from the sample preparation and ending with data acquisition. The scaling or normalisation factors should be comparable among arrays. A large discrepancy scale factor (3 fold or greater) may indicate significant assay variability or sample degradation leading to poor quality data.

2.7.3.2.3. 3' to 5' ratios for β-actin and GAPDH

β-actin and GAPDH are used to assess RNA sample and assay quality. Specifically, the intensities of the 3' probe sets for β-actin and GAPDH are compared to the intensities of the corresponding 5' probe sets. The ratio of the 3' probe set to the 5' probe set should generally be no more than 3 for the targets prepared using the one cycle target labelling procedure described in Figure 2.3. A high 3′ to 5′ ratio indicates degraded RNA or inefficient transcription of cDNA or biotinylated cRNA.

In addition to the "simpleaffy" quality control measures described above, target cRNA quality was also assessed using the array-by-array cRNA digestion plot produced using the "affy" BioConductor *plotAffyRNAdeg* function. This averages individual probe intensities by their location in each probe set and following a scaling transformation an average can then be taken over all probe sets on the array. A side-by-side plot of these averages, then illustrates any global patterns of 5' to 3' probe intensity. Any abnormally low levels of 3' intensity would illustrate a degraded RNA or inefficient transcription of cDNA or biotinylated cRNA.

2.7.3.3. Data Correction, Normalisation and Transformation

After careful consideration of the quality control measures described above the correction, normalisation and transformation of raw array data took place, following a four-step process of Background Correction, Normalisation, Perfect Match (PM)/ Mismatch (MM) Correction and Summarisation. The data analysis was performed using the "affy" BioConductor package using the *expresso* function. In order to establish the most appropriate method for the microarray experiments performed in this study, the Affymetrix recommended *mas5* method (Affymetrix, *url*) was assessed along with variations of the Robust Multi-array Average (RMA) process (Irizarry *et al.*, 2003), specifically altering Background Correction and PM/MM Correction steps. The detail of each method used is described below.

Background correction

Background correction is the process of correcting probe intensities on an array by subtracting the background level of signal detected by the scanner. The following approaches were assessed.

- ***none*** – no background correction.
- ***rma*** - Developed by Irizarry and colleagues, this correction uses a model that assumes observed intensity is the sum of an exponential signal component and a linear noise component. PM probe intensities are corrected using a global model for the distribution of probe intensities (Irizarry *et al.*, 2003).
- ***mas5*** – This is the method recommended by Affymetrix (Affymetrix, *url*), where a chip is broken into subgrids, and background is calculated for each region based on the lowest 2% of probe intensities. For each region, a weighted average background value is calculated using the distances of the probe location and the areas surrounding the probes of the different regions. Individual probe intensity is then adjusted based upon the average background for each region.
- ***GC Robust Multichip Analysis (GC RMA)*** – Developed and further improvised version of the RMA for exploration, normalisation and summarisation of high-density oligonucleotide probes level data.

- ***Model based Expression Index (MBEL, dCHIP)*** – Developed by Li and Wang (PNAS, 2001, 98(1); 31-36) to calculate expression index computation and outlier detection.

Normalisation Methods

Normalisation is the process of removing non-biological variability between arrays. All global normalisation methods work on the assumption that there is no variability between different microarrays. Here, we used ***mas5*** - developed by Affymetrix which uses a global scaling factor in order to normalise the data between each microarray. A scaling factor is calculated based on the average of all the intensities, after removing the intensities in the lowest 2% and highest 2%. This factor is then used to correct the intensities across all the probe sets on all the arrays.

Perfect Match / Mismatch Correction

PM correction is the process of adjusting PM intensities based on information from the MM intensity values.

pmonly - No PM/MM correction was performed and only PM values were used for analyses. It is widely reported that MM probesets may be detecting signal as well as non-specific binding and therefore including the MM parameter will contribute to the overall noise in the data analysis (Naef *et al.,* 2002; Irizarry *et al.,* 2003).

mas5 – Recommended by Affymetrix, in this method an ideal MM value is subtracted from the PM intensity value, always leaving a positive value. An "artificial" mismatch value is computed when the MM intensity is greater than or equal to the PM and results in a PM-MM that is close to zero.

Summarisation Method

In order to combine the pre-processed probe intensities together in order to compute a single expression measure for each probe set on the array, a summarisation method was employed

medianpolish – Described by Irizarry and colleagues (Irizarry *et al.,* 2003), median polish uses a multi chip linear model fitted to the data from each probe set and the result value is in log2 scale.

mas5 – Recommended by Affymetrix uses a robust average using 1-step Tukey bi-weight on log2 scale.

2.8. Exon Array

2.8.1. The Affymetrix GeneChip®Whole Transcript (WT) Sense Target Labeling Assay

All experiments were performed using Human Gene®Chip Exon 1.0 ST Array designed to generate amplified and biotinylated sense-strand DNA targets from the whole transcript without bias (Affymetrix, *url*). The WT Assay is not compatible with Gene®Chip arrays designed to focus on the 3' ends of transcriptions, such as the Human Gene®Chip HU 133 plus 2.0.

Human U 133 plus 2.0	**Human Gene®Chip Exon 1.0 ST**
1.3 Million probes	5.3 Million probes
54,000 probe sets	1.4 Million probe sets (284,000 core, 523,000 extended and 580,000 full)
11 Perfect Match (PM)/Mismatch (MM) probe pairs	4 Prefect Match (PM) probes per probe set
Interrogated strand is ANTISENSE	SENSE
3' end of the mRNA	Whole Transcript level (Random Hexamer primers)
Hybridising intensity = (PM)- (MM) targeting the 3' end	Detection above background (DABG) Which is comparing the PM with the background probes
Few different algorithms used; Robust Multiarray Average (RMA), Microarray Suite (MAS 5.0)	Only Probe Logarithmic Intensity error (PLIER) Is used to minimised error at low and high abundance
11micron per feature size	5 micron per feature size

Table 2.5 Table comparing the Affymetrix HU 133 plus 2.0 chip with Affymetrix Human Gene Chip Exon 1.0 ST.

2.8.2. Sense Target Preparation, Microarray Hybridisation, Staining and Scanning

The Affymetrix GeneChip® Whole Transript (WT) Sense Target Labelling assay is designed to generate amplified and biotinylated sense-strand DNA targets from the entire expressed genome without bias. The 'Sense Target' prepared with this assay, and the probes on the arrays have been selected to distribute throughout the entire length of each transcript.

In my study, I have used the 100ng protocol for the reasons that our main interest is in the gene-level analysis and that the frozen tissue specimens we have were of small quantity.

For more information regarding the performances of the two protocols on Gene 1.0 ST Array refer to the Whole Transcript Sense Target labelling Assay Performance white paper on the Affymetrix website.

As outlined in Figure 2.3, the two protocols merged where double-stranded cDNA is synthesized with random hexamers tagged with T7 promoter sequence. The double-stranded cDNA is subsequently used as a template and amplified by T7 RNA polymerase producing many copies of antisense cRNA. In the second cycle of cDNA synthesis, random hexamers are used to prime reverse transcription of the cRNA from the first cycle to produce single-stranded DNA in the sense orientation.

In order to reproducibly fragment the single-stranded DNA and improve the robustness of the assay, a novel approach is utilised where dUTP is incorporated in the DNA during the second-cycle, first-strand reverse transcription reaction. This single-stranded DNA sample is then treated with combination of uracil DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1(APE 1) that specifically recognises the unnatural dUTP residues and breaks the DNA strand. DNA is labelled by terminal deoxynucleotidyl transferase (TdT) with the Affymetrix ® proprietary DNA Labeling Reagent that is covalently linked to biotin.

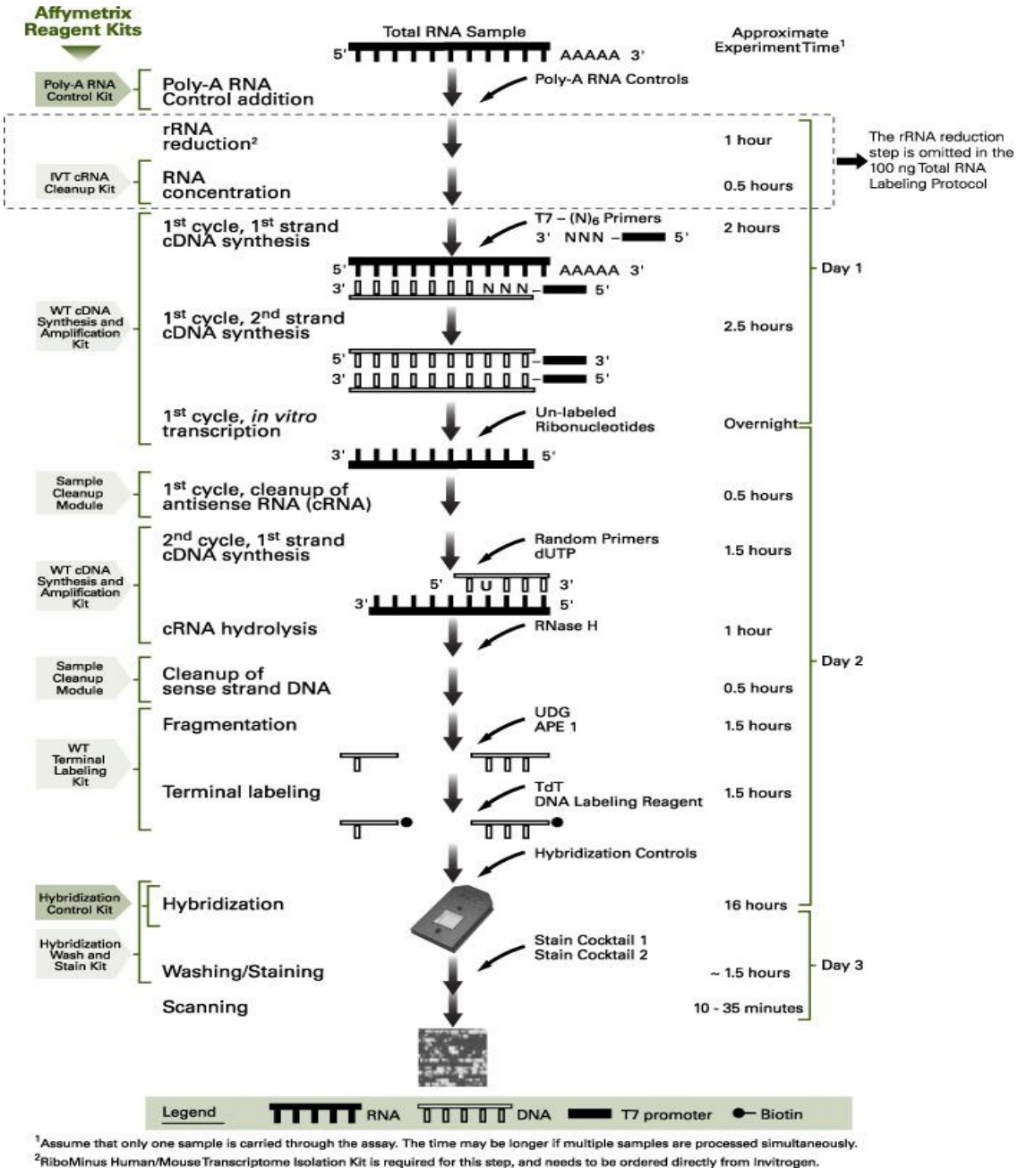


Figure 2.4 Affymetrix Exon Gene®Chip (WT) Sense Targeting 1.0 Eukaryotic Sample and Array Processing. A more detailed description of this procedure can be found in Section 4: Eukaryotic Sample and Array Processing of the Exon Gene®Chip Expression Analysis Technical Manual (Affymetrix, *url*).

2.8.3. Differential Gene Expression Analysis for both HU133 plus 2.0 and Exon GeneChip® WT ST1.0 Arrays

Filtering

Expression profiling experiments used to identify genes that change their expression between two groups. Therefore, it is important to first filter the data to include only those probe sets that change in expression. A commonly used method is to filter genes based on the fold change between the *test* and reference groups. However, when filtering on fold change there is a risk of ignoring genes that change significantly but are below the arbitrary fold change threshold. A more biologically sensitive method is to order genes from low to high standard deviation, in order to identify the most variable genes across the conditions analysed. The top 2500 (for HU 133 plus2.0) and 10000 (for Human Exon 1.0 ST) probe sets with highest variance can then be used in clustering and statistical testing.

Hierarchical Clustering

Hierarchical clustering analysis allows the monitoring of overall patterns of gene expression between the normalised arrays, and uses standard statistical algorithms to arrange the genes according to a similarity in gene expression patters. The hierarchical clustering was performed using Partek® Genomic Suite, based on the approach used by Eisen and colleagues (Eisen *et al.*, 1998). The hierarchical clustering analysis aimed to produce a map of results where probe sets were grouped together based on similarities in their patterns of normalised expression across all of the microarrays. The similarity or dissimilarity between a pair of objects in the data set was found by evaluating a distance measure and assuming a normal distribution of gene expression values, it is appropriate to use the Pearson correlation coefficient, which calculates the similarity measure based on a linear model. The objects are then grouped into a hierarchical cluster tree (dendrogram) by linking newly formed clusters. The same algorithms can then be applied to cluster the experimental samples for similarities in the overall patterns of gene expression. Hierarchical clustering analysis was preformed on the normalised and filtered gene expression data.

Statistical Analyses

This process involves the identification of diferentially-expressed genes between different experimental conditions. The Welch's t-test was applied on filtered data using the one-way ANOVA setting in Partek® Genomic Suite. This is a parametric test that works on the assumption that log intensity of microarray data are normally distributed. Another common assumption of parametric statistical analyses, such as Student's t-test, is that the variability of a gene is constant across treatment types. This is difficult to assess for microarray data, so it is safest to assume that variance may differ between treatment and control. Welch's t-test corrects for difference in variability, and does not detect it, therefore is more suitable for microarray data differential gene expression analysis. In a conjunction with this test a False Discovery Rate (FDR) multiple test correction was applied across the significant genes using GeneSpring. Multiple testing corrections adjust p-values to correct for occurrence of false positives. False positives are genes that are identified as significant changes following statistical tests, when their true state is unchanged. A False Discovery Rate of 5% (p-value <0.05) on an array of 54000 reporters would mean that on any size gene list, 2700 genes would be expect to be false leads. The Benjamini and Hochberg FDR correction was applied across the significant genes (Benjamini & Hochberg, 1995). This test reduces the number of false positives without enriching the number of 'false negatives', which can be the case for other types of correction (Benjamini & Hochberg, 1995). Briefly, the p-values are ranked from the smallest to the largest. The largest p-value remains as standard. The second largest p-value is multiplied by the total number of genes in the gene list of differentially expressed genes divided by its rank. The same approach is repeated with the second largest p-value and so on, until no gene is found to be significant. The resulting FDR corrected values mean that a FDR of 5% (FDR corrected p-value <0.05) on a gene list of 500 would expect 25 to be false leads, regardless of the number of reporters on the array.

2.8.4. Ingenuity Pathway Analysis (IPA)

2.8.4.1. Introduction to IPA

I have used Ingenuity Pathway Analysis (IPA version 4.0), a web-based application (www.Ingenuity.com) that enables building signaling networks from gene/protein expression data. It is based on IPKB (Knowledge Base database), which is currently the largest curated database containing millions of computable relationships between genes, proteins, drugs and diseases.

A data set containing Affymetrix probe identifiers and their corresponding 'fold change' values was uploaded in Excel. Each probe identifier was mapped to its corresponding object in the IPK base. To build a network, IPA searches the IPKB for interactions between focus genes/proteins and all other gene objects stored in the Knowledge base ('Focus genes' show direct interaction with other genes in the knowledge base). It then generates a set of networks with a maximum of 35 genes, and computes a score for each network. The score shows the likelihood that a gene is placed in a network due to random chance (for example, a score of 2 gives a 99% confidence that the focus genes are not being generated by random chance). In addition, IPA's Global Functional Analysis feature provides an overview of biological functions associated with a set of dysregulated genes/proteins, with functions displaying a p-value <0.05 being significant. The significance values for these analyses are calculated using the right-tailed Fisher's Exact test. Similarly, the Canonical Pathway Analysis feature shows which of the known signalling and metabolic pathways are altered in the user input data. IPA's canonical pathways are based on its own curation as well as on KEGG (Kyoto Encyclopedia of Genes and Genomes).

2.8.5. Affymetrix Human Genome-Wide SNP 6.0 Arrays

Nsp/Sty 5.0/6.0 Assay protocol, Washing, Staining and Scanning

The Human Genome-Wide SNP 6.0 Arrays were purchased from Affymetrix (Santa Clara, CA, USA), containing 906,600 SNPs in a single chip with a physical distance

of on average 23.6kb. SNP array experiments were performed according to the manufacturer's instructions. Briefly, 250 ng of tumor DNA was digested by either Nsp1 or Sty1 restriction enzyme (New England Biolabs). After ligation to an appropriate adaptor for each enzyme, a PCR reaction was carried out using a generic primer that recognizes the adapter sequence. The PCR products from four reactions were pooled, concentrated, fragmentated by DNase I and subsequently labeled with biotin. Hybridization was performed at 45°C for 16h in a hybridization machine (Affymetrix). After washing and staining the arrays, the signal intensities were measured on an Affymetrix GeneChip Scanner 3000 and the raw images were analyzed using the GCOS (Ver1.4) and the GTYPE (Ver4.1) software that implements a new genotyping algorithm, BRLMM.

SNP 6.0 Data Analysis

To assess DNA copy number variations (CNV) and Loss of Heterozygosity (LOH), the Copy Number Analysis from the Partek (Partek, Version 4.0) software was used. In the analysis, the inferred copy numbers at each SNP locus was estimated by applying the hidden Markov model (HMM) and the segmentation algorithms. With the GCH software, nine algorithms were implemented in the "CGHweb" software (http://compbio.med.harvard.edu/CGHweb/) {Lai, 2008 #1}; Forward-Backward Fragment-Annealing Segmentation, Gaussian Model with Adaptive Penalty, Locally weighted scatterplot smoother, Quantile Smoothing, Circular Binary Segmentation, Fused Lasso (cghFLasso), GLAD, Wavelet smoothing and Running Average. The HMM parameters were set up based on comparison between our reference data (Tamoxifen Sensitive samples were used as reference) and the Tamoxifen Resistant data. The analysis described above is implemented in Partek (see "Supplementary Methods" in ref. 32 for details.). "Automatic analysis" mode was selected in which the software performed pair-wise tests for all of the references. Genetic gains (DCN≥3) and losses (DCN≦1) were defined according to the working criteria of the Partek software. High-level amplifications and homozygous deletions were determined to be CN gains ≥ 5 and CN deletion = 0, respectively. The LOH output from Partek was verified by the Affymetrix CNAT (Ver3.0) software, in which a set of 110 built-in reference files are available to calculate the probability of LOH at each probe. (The non-BRLMM data were used for this analysis because the BRLMM

data cannot be processed in CNAT Ver3.0.) The great majority of LOH calls from Partek agreed with CNAT. To avoid detecting false-positive changes due to random noise in allele intensity at individual SNPs, we set a minimum physical length of at least five consecutive SNPs for putative genetic alterations. The physical position of all SNPs (n=116,204) on the arrays were mapped according to the UCSC Genome Browser on Human May 2004 Assembly. The gene annotation was computationally determined after combining the information available in RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) and Ensembl (http://www.ensembl.org) databases. Taking structural variation in the human genome into account, recurrent regions of copy number variations (CNVs) were also excluded from the analysis (http://projects.tcag.ca/variation/) (Iafrate *et al.*, 2004), (Sebat *et al.*, 2004), (Redon *et al.*, 2006), (Freeman *et al.*, 2006).

2.9. Quantitative PCR

2.9.1. Reverse Transcription Reaction

A microgram of total RNA was used to generate cDNA. Reverse transcription reactions were performed using sterile plasticware throughout and aerosol filter tips to reduce contamination. All materials were sourced from Applied Biosystems. The reactions were prepared on ice as follows; 1μg total RNA, 5.5mM $MgCl_2$, 2.5mM dNTP mix, 2.5μM Random Hexamers, RNase Inhibitor (0.4U/μl), MultiScribe Reverse Transcriptase (1.25U/μl) and RNAse free water to make the reaction volume up to 50μl. Samples were then transferred to a thermocycler and reactions incubated at 25^{o}C for 10 minutes, 48^{o}C for 30 minutes followed by 95^{o}C for 5 minutes to inactivate the enzyme. After the RT reaction it can be assumed that 1μg of total RNA corresponds to 1μg of cDNA.

2.9.2. Quantitative “Real Time” PCR reaction (qPCR)

Pre-designed transcript specific primer-probe sets for use in for qPCR reactions were purchased from Applied Biosystems, the details of these are outlined in Table 2.5. Briefly, the probe is supplied dye-labelled at the 5’ end. The fluorescence of this dye

is controlled by a quencher at the 3'end. The probe binds to the DNA between the two primers. As the reaction proceeds, the dye is cleaved from the probe and thus released from the quencher. Therefore as the reaction progresses the fluorescent emission from the dye increases allowing accurate quantification of the target sequence. Reaction mixes (25µl) were prepared in triplicate on a 96-well plate using 15ng cDNA per reaction and Universal PCR master mix (Applied Biosystems) at 1X. Approximate controls were included. In order to assess the efficiency of the PCR reaction and to allow relative quantification, a standard curve was run alongside the samples. The standard curve consisted of four separate dilutions of cDNA per reaction and was prepared in triplicate; 25ng, 6.25ng, 1.5625ng and 0.39ng. The PCR reaction was carried out on the 7700 Sequence Detection System (Applied Biosystems) using the following program as standard: 50^{o}C for 2 minutes (AmpErase UNG step), 95^{o}C for 10 minutes to activate the AmpliTaq Gold, then 40 cycles of 95^{o}C for 15 seconds, 60^{o}C for 1 minute.

Transcript	**Assay ID**	**Probe Dye Layer**
GAPDH	4319413E	VIC
VGLL1	Hs00212387_m1	FAM
AKR1C3	Hs00366267_m1	FAM
MAGEA2	Hs00606323_s1	FAM
MED16	Hs00193899_m1	FAM
EGLN3	Hs00222966_m1	FAM
GREB1	Hs00536409_m1	FAM
PDZK1	Hs00536409_m1	FAM
MYBL1	Hs00277143_m1	FAM
HNRNPA2B1	Hs00242600_m1	FAM

Table 2.6 Details of primer-probe sets used in qPCR analyses. All PCR reactions were analysed by agarose gel electrophoresis to ensure the presence of a single product under standard PCR conditions.

Results were initially analysed using the Sequence Detection software version 1.9.1. (Applied Biosystems). The amplification plots were observed in both linear and semi-log plots, with the background corrected and the threshold cycles determined. Following this the standard curve was plotted showing the slope (PCR efficiency) and the correlation coefficient. A slope of –3.3 relates a 100% efficient PCR reaction, a ten-fold increase in PCR product every 3.3 cycles. The PCR efficiency was considered satisfactory above 98% and only if all the samples fell within the points of the standard curve. Data was analysed according to the Standard Curve Method for relative quantification (Applied Biosystems, *url*). Standard curves were prepared for both the target (e.g. Mage-A2) and the endogenous reference (e.g. *GAPDH*). For each experimental sample, the relative quantity of target and endogenous reference levels was determined from the appropriate standard curve. The target amount was then divided by the endogenous reference amount to obtain a normalised target value. The transfection control sample was used as a calibrator and each of the normalised target values were divided by the calibrator normalised target value to generate the relative expression levels. Triplicate samples were used to generate standard errors.

2.10. Co-ImmunoPrecipitate Assays

All of the Co-IP assays presented in this thesis were performed in collaboration with Tony Wong, who performed the Co-IP and IP work on MAGEA2 project.

2.10.1. Immunoprecipitation

Immunoprecipitation experiment was performed as follows. First, cell pellets were lysed with IPH buffer (50nM Tris-HCI pH8.0, 150mM NaCI, 5mM EDTA, 0.5% NP40, 0.1mM PMSF, 5µM Trichostatin A (TSA) and protease inhibitor cocktail) on ice for 30 minutes. 1 mg of cell lysate was incubated with 50µl of Dynabeads® (Invitrogen) and either 2µg mouse anti-p53, 4µg anti-MAGEA2 (Santa Cruz) or control IgG antibody overnight at 4°C. The next day, immunoprecipitates were washed three times with IPH buffer for 5 minutes at 4°C, and resuspended in 45µl of 2X SDS western loading buffer. The samples were resolved via SDS-PAGE and transferred onto PVDF membranes. Protein detection was achieved by western blot analysis.

2.11. In vitro functional analysis

2.11.1. Proliferation assay

Cell number was determined by counting cells using a Z1 Coulter particle counter (Coulter Electronics). 70-80% confluent cells were plated in 2 well plates at a density of 2.5 x 10^4cells/well and allowed to adhere overnight at 37^{o}C. The cells were treated with their normal media or media with 10^{-7}M Tamoxifen (Sigma cat no. H7904) for 8 days. After treatment, cells were trypsinized, 80 µl of cells was mixed with 20 ml of Isoton Coulter balanced electrolyte and loaded in a Coulter counter. Each experiment was carried out in triplicate. Cells were counted every 24 hours for 8 days. Cells were split every third day during cell growth assay.

2.11.2. Sulphorhodamine (SRB) assay

In addition to cell counting, a non-mitochondrial cytotoxicity assay, suphorhodamine assay (SRB) was used to evaluate the effect of the clones on cell proliferation in normal media and tamoxifen media. This method relies on the uptake of the negatively charged pink aminoxanthine dye, SRB by basic amino acids in the cells. The greater the number of cells, the greater the amount of dye is taken up, and after fixing, when the cells are lysed, the released dye will give a more intense colour and greater absorbance. Cells were seeded at a density of 5 x 10^3 cells/well in 100 µl in a 96-well plate, nine wells per cell line. After 24 h, the cells were treated with the appropriate media, normal or the tamoxifen (10^{-7} M) containing. At a given time point, the cells were fixed with ice-cold TCA for 1 hour at 4°C. Cells were then washed 5 times with distilled water. Cells were then dried at 56°C for no more than 5mins. Fifty µl of SRB (0.4% in 1% acetic acid) was added into each well for 30mins in room temperature. The cells are then washed quickly with 1% acetic acid five times. Finally, 100µl of 10mM Tris Base was added, on a rocker for 5mins, and read at 492nM wave-length.

2.11.3. Annexin V assay

The annexin V binds to negatively charged phospholipid, like phosphatidylserine. During apoptosis the cells react to annexin V as soon as chromatin condenses but before the plasma membrane loses its ability to exclude PI. Hence by staining cells with a combination of fluorecenated annexin V and PI it is possible to detect nonapoptotic live cells, early apoptotic cells and late apoptotic or necrotic cells. Cells from the exponentially growing were collected at the indicated time and added to the floating cells and analysed together. Aliquots of cells (>0.5 X 10^6) were centrifuged at 1000 rpm for 5 mins and washed with PBS. The cell pellet was resuspended in 100 µl of labelling solution (Annexin-V-fluos, Boehringer Mannheim) containin 2 µl annexin V labelling reagent and 0.1 µg propidium iodine (Calbiochem. La Jolla, CA) and incubated for 10-15 min, as per manufacturer instructions. Immediately after adding 0.4ml of incubation buffer (10mM HEPES.NaOH, 140 mm NaCl, 5 mM $CaCl_2$) analysis of red (annexin V) and white (PI uptake) fluorescence of individual cells was measured with FACScan flow cytometer (Becton Dickinson,

Erembodegem, Belgium). The data were analysed using the Prism software package supplied by Graphpad Software Inc. Comparisons among treatments were performed using one-way analysis of variance (ANOVA). When significance was observed ($P \leq 0.05$), Tukey's multiple comparison test was performed to determine which means differed from the control by a significant margin. All results are expressed as the mean ± S.D. of triplicate treatments. Results shown are from single experiments, representative of a minimum of three. Where appropriate on figures significance is indicated as ***$P \leq 0.001$; **$P \leq 0.01$, *$P \leq 0.05$.

2.11.4. Transwell Migration assay

Transwell plate which is especially designed to have a total of 24-wells, with an inner well cradle in them (24-well size with an 8μm pore size filter; Costar). I had used 6 wells for MCF-7 VA cells and 6 wells for MCF-7 MAGEA2 expressing clone (C24) for this experiment. The bottom (outer membrane) of the inner well is coated with 0.5% of BSA in PBS to completely block the membrane. The bottom well is then filled with media with 0.5% BSA, using BSA as the cell attractant. The top well is then filled with 100000 cells, which has been washed clean of any serum (twice washed with PBS), and in serum free RPMI. The plates are left in 37°C for 18 hours in their normal incubator.

Cells are trypsinised and counted (Model TTC, CASY 1, Scharfe system Gmbh, Rentlinen, Germany) after 18hours. The percentage of cells that are found in the bottom well relative to the total cells is calculated and presented as histogram graphs by exel software.

2.12. Immunofluorescence (ICC)

Coverslips (round 13-mm) were placed in a 24-well plate and sterilised with 70% ethanol for 20 minutes followed by three washes with PBS. Cells (at 5×10^4) were plated onto the coverslips and grown overnight at 37^oC. After three washes in PBS cells were fixed in 4% paraformaldehyde for 15 minutes at room temperature. Following three washes in PBS cells were permeabilized (when necessary) with 0.1% Triton X-100 in PBS for 5 min at 4^oC. Cells were then washed three times with PBS and blocked in 1% BSA for 30 min at room temperature. Subsequently, cells

were incubated with primary antibodies diluted in BSA for 1 hour at room temperature. After three 10-minutes washes with PBS, cells were incubated with secondary antibodies diluted in BSA and incubated for 30 minutes at room temperature in the dark. Secondary antibodies were: DAKO (1:200) polyclonal Rabbit anti-mouse FITC IgG, and Alexa Fluor 546–conjugated goat anti-rabbit IgG (1:200; Molecular Probes, Invitrogen). After three 10-minutes washes with PBS, 50 µL DAPI-containing mounting agent gel (Prolong® Gold antifade reagent DAPI; Molecular Probes) was added on to glass slides with the coverslips with the stained cells placed upside down on the glass slides and left in the dark for 1h to solidify. The stained slides were analyzed using the laser scanning confocal microscope Zeiss LSM 510.

2.13. Independent Validation Cohort

The independent validation cohort consisted of 76 cases from Guy's and St Thomas' (GSTH) Hospital (EORTC 10850 & 10851), 42 cases with full clinical information from The Royal London Hospital (RLH), 7 paired-cases (primary and relapsed tissue, kindly donated by Dr Simak Ali) from Charing Cross hospital and 71 TMA from Leeds (kindly donated by Dr Valerie Speirs). Of which all were paraffin slides sliced at 0.4µm thick, except the 42 cases from RLH and 71 cases from Leeds, which were Tissue Microarray (TMA) made from core punch at tumour cell regions selected by Prof Louise Jones (qualified pathologist from RLH). These were identified as cases of breast cancer from 1984-2005 with complete post-primary surgery follow up data on adjuvant tamoxifen therapy. A third of the cases with no recurrence after 10 years were classed as tamoxifen sensitive, and two-thirds of the cases were considered to be tamoxifen resistant as the patients relapsed within 2 years of adjuvant tamoxifen. The latter cohorts, patients from Caring Cross and Leeds had other treatment for their breast cancer apart from tamoxifen. In addition, in the Leeds cohort, a third of the patients were ER-negative. The 76 cases from GSTH were patients where the patients who had tamoxifen monotherapy, with no subsequent chemotherapy. This made the cohort unique as their survival analysis is purely as the consequent of the effect of tamoxifen on the course of ER-positive breast cancer.

A summary of our patient characteristics in our 196-validation cohort is shown in Table 2.7. For all the genes validated in immunohistochemistry staining, the cohort consisted of the 118 cases from GSTH and BTLH. For the two-genes which I have also done in-vitro study; MAGEA2 and EGLN3, I have validated them across GSTH, BTLH, Charing Cross hospital (paired for primary and relapsed tumour paraffin slides) and also from the Leeds cohort, which total up to 196 patients. To construct tissue microarrays we have used a Tissue Arrayer (Beecher Instruments, MD, USA). This invaluable resource is now ready for my candidate validation step.

	GSTH	BTLH	Charing Cros	Leeds
	n=76	n=42	n=7	n=71
Grade				
1	38	6	0	36
2	36	18	0	30
3	2	18	7	5
Unknown	0	0	0	0
Type				
Lobular	8	8	0	9
Ductal	68	34	7	52
Others	0	0	0	10
Nodal status				
0	11	20	0	16
1 to 3	48	14	0	6
>4	4	8	0	2
Unknown	13	0	7	47
Size				
T1 (<20mm)	32	26	0	26
T2 (20-50mm)	38	14	0	13
T3 (>50mm)	6	0	0	3
Unknown	0	2	7	29
Age				
<50 y	0	0	0	16
>50 y	76	42	7	55
Oestrogen receptor status				
Positive	75	41	7	53
Negative	0	1	0	18
Unknown	1	0	0	0
Progesterone receptor status				
Positive	76	8	7	2
Negative	0	26	0	2
Unknown	0	8	0	67
Median Follow up	77 mo (range 11-229, SD 49.7)			

Table 2.7 Table with the 196 patients and their characteristics. Most patients were ER positive and had Tamoxifen treatment. All tumour grades were known. Lymph nodes status was known in most cases.

2.14. Immunohistochemistry (IHC)

To confirm expression of selected differentially expressed genes at the protein level, immunohistochemical (IHC) analysis was performed using the Ventana Discovery™ System (Ventana Discovery™ System, Illkirch, France) following the manufacturer's protocol. Paraffin-embedded tissue blocks and tissue microarrays were cut into 4 -µm-thin sections and stained with Haematoxylin and Eosin (H&E). For IHC, all slides were deparaffinised and processed for antigen retrieval with SSC (Standard saline citrate buffer). After blocking, the slides were incubated with the respective antibodies. For the negative control, primary antibody was omitted from the reaction. Staining for pan-MAGEA, EGLN3 and the human predictive genes (EPHA7, PALM2, SNF1LK2, ZBTB16, OR10G7, OPCML, SP2, RUNX1T1 and ENPP2) was performed using the 3-3' diaminobenzidine (DAB) detection kit (Ventana); the slides were counterstained with haematoxylin. A reddish-brown precipitate indicated positive immunoreactivity. The sources and dilutions of antibodies are shown in table 2.8.

Antibody	Microarray change	Species raised	Clonality	Source	IHC
MAGEA2	3.83	Mouse	Monoclonal	Zymed	1 in 50
EGLN3	3.65	Rabbit	Monoclonal	Abcam	1 in300
EPHA7	3.94	Rabbit	Polyclonal	Santa Cruz	1 in 100
PALM2	-4.12	Mouse	Monoclonal	Abnova	1 in 200
SNF1K2	-4.02	Rabbit	Polyclonal	Abcam	1 in 50
ZBTB16	-4.99	Rabbit	Polyclonal	Atlas	1 in 50
OR10G7	3.99	Goat	Polyclonal	Santa Cruz	1 in 50
OPCML	3.92	Goat	Polyclonal	R&D sytems	1 in 25
MAF	-4.26	Rabbit			
Sp2	-4.26	Rabbit	Polyclonal	Santa Cruz	1 in 10
RUNX1T1	-4.57	Mouse	Monoclonal	Abgent	1 in 50
ENPP2	-3.92	Rabbit	Polyclonal	Cosmo Bio Co	1 in 25

Table 2.8 Antibody details and dilutions for Immunohistochemistry (IHC) analyses.

2.14.1. Evaluation of staining

The stained tissue sections were scored by Dr Yaohe Wang (trained pathologist) and I on the basis of both the extent and the intensity of the immunoreactivity and average scores were taken. The staining intensity was graded on a 0-3 scale 0 (negative/ no staining), 1 (weak immunoreactivity), 2 (moderate immunoreactivity) and 3 (strong immunoreactivity). The extent of immunoreactivity was scored according to the percentage of stained cells in relation to the entire section as (0 points for no staining, 1 point for less than 20%, 2 points for 20-50% and 3 points for more than 50% of the cells). The product of the intensity and extent scores was used as the final staining score. Negative or weakly positive cases had a score of 0-3 points, moderately positive had a score of 4-6, and strongly positive cases had a final score greater than 6. The mean was then use as a cut off, with all scored above the mean, labelled positive and all below the mean being negative.

2.15. Survival analysis

All analyses were done using Prism (Version 6.0). A p value of <0.05 defined statistical significance. Kaplan-Meier curves were plotted from data of disease free survival (DFS) and overall survival (OS). Data from patients who were lost (<1%) from followed up were treated as censored data. Comparison of survival curves were analysed using both Log-rank (Mantel-Cox) test, and Gehan-Breslow-Wilcoxon test. As the results were similar, we have shown only the p-value from Gehan-Breslow-Wilcoxon test. Hazard ratios and 95% Confidence Intervals (95% CI) were also given alongside the p-value calculated.

2.16. MicroRNA extraction using the miRVana PARIS kit, preparation of cDNA and real-time PCR with microRNA primers.

Cells (102-107 cultured cells) or fresh frozen tissues (0.5-250mg) can be used for extraction of miRNA, small RNA and small nuclear RNA (snRNA) with mirVana miRNA isolation kit. The mirVana PARIS kit can effectively extract miRNA and small RNA simultaneously with protein extraction (whole cell extract). The protocol is easy to follow, and the method involved the principle of organic extraction followed by the purification using a silicate matrix (spin columns). The enriched extraction protocol efficiently purified all RNA larger than 10nt, and up to 200nt.

cDNA were made from miRNA using the high-capasity miRNA reverse transcriptase kit (part number: 4366596) from Applied Biosystem® (ABI). For qPCR, Tagman assay (see below table for the assay ID) for miRNA were purchased, and housekeeping miRNA, such as hsa-mir-19a, hsa-mir-106a and RNU24 were used as recommended by ABI for breast cancer cell lines and human tissues respectively.

Name of miRNA	Assay ID
has-miR-221	4373077
has-miR-657	4380922
has-miR-19a	4427975
has-miR-RNU24	4427975

Table 2.9 Part number for the miRNA Tagman probe and primers from Applied Biosystems.

CHAPTER 3: RESULTS OF TAMOXIFEN RESISTANT CELL LINES

3. Oestrogen, Progesterone and ErbB2 Expression in cultured breast cancer (TR) Cell lines

The work on TR is largely related with the hormone receptor status on the breast cancer cells. All breast cancer in the clinical setting is typed for hormonal status. We have a series of breast cancer cell lines, which we use for our laboratory work. Oestrogen receptor (ER), progesterone receptor (PgR) and ErbB2 receptor (HER2) status from all the parental cell lines were typed using immunohistochemistry. Prof. Loiuse Jones provided this service (Table 3.1).

Cell Line	ER	PgR	Her2/ErbB2	p21/WAF1	p53
MCF-7 (ATCC)	30%	60%	0	Pos	WT
ZR75-1	80%	90%	3+	Pos	WT
MCF-7 WT (Tenovus)	60%	60%	1+	Pos	WT
MCF-7 RE	80%	30%	2+	Pos	WT
MCF-7 RL	70%	25%	2+	Pos	WT
Ca151	Neg	Neg	0	NK	NK
T47D	60%	90%	0	Pos	Mutant
MDA-MB-453	Neg	Neg	3+	NK	WT
MDA-MB 436	Neg	Neg	0	NK	WT
MDA-MB-361	Pos	Neg	3+	Pos	WT
HCC1500	Pos	Pos	0	NK	WT
HCC38	Neg	Neg	0	NK	WT
SKBR3	Neg	Neg	3+	Pos	Mutant

Table 3.1 ER, PgR, HER2, p21(WAF1) and p53 status across a panel of parental breast cancer cell lines. Tam RE=Tamoxifen resistant early, Tam RL= tamoxifen resistant late. Pos=positive, Neg=Negative, NK=Not known, WT=wild type, MuT=Mutant. The immunohistochemistry staining is expressed as percentages (%) of cells positive for ER or PgR in the nucleus. 1+ and 2+=low HER2 expression; 3+=membrane staining. The MCF-7 derived cell lines from the Tenovus Institute in Cardiff, which I have labelled as MCF-7 (Tenovus), which were maintained in Tamoxifen containing media for 3 and 6 months (RE=resistant early and RL=resistant late) respectively.

3.1. Generation and characterisation of Oestrogen Deprived and Tamoxifen resistant T47D and ZR75-1 lines

As part of our study, we generated Oestrogen deprived cell lines with the aim of studying breast cancer cell lines, which are independent of the oestrogen pathway for their growth. Oestrogen deprived (OD) T47D and ZR75-1 were generated by growing parental cells in medium containing charcoal-stripped serum for duration of 6 months. The resulting lines were labelled as OD T47D and OD ZR respectively. These cells kept the same morphological features and were tested intermittently for Mycoplasma contamination. The OD cells grew relatively slowly, and only required passaging every 7 days as compared to parental cell lines which required splitting every 4th or 5th day.

Tamoxifen Resistant (TR) cell lines were generated by growing T47D, ZR75-1, OD T47D and OD ZR in Tamoxifen containing media. Hydroxy-4-OH-Tamoxifen at 10^{-7} M was added to specific media (See Chapter 2.1.2.) and maintained for 6 months' duration. Initially cells arrested growth but eventually re-entered the cell cycle and could be expanded. At this point they were considered as a separate cell line. The lines were labelled T47D TR, ZR TR, OD T47D TR and OD ZR TR. The cells were tested intermittently for Mycoplasma contamination. TR cells were slow growing and required splitting once every 10-12 days.

Whole cell extracts (WCE) from MCF-7 (CRUK), MCF-7 WT (Tenovus), T47D, ZR75-1 and their OD and Tamoxifen resistant (TR) counterparts were assessed for ERα, ErbB2 and AP-2γ levels using Western blotting (Figure 3.1). In keeping with the literature (Orso *et al.*, 2004), ERα levels were acutely increased in cell lines exposed to tamoxifen containing media for 24 to 72 hours. The late TR cell lines expressed similar levels of ERα as the parental cell lines. The OD TR cell lines showed almost complete loss of ERα protein expression of. ErbB2 protein expression was increased in Tamoxifen late resistant cell lines. This is in keeping with the many publications that report an inverse relationship between ErbB2 and Progesterone (PgR) (Ponzone *et al.*, 2006). Progesterone mRNA expression was significantly down-regulated in TR cell lines in our microarray study (Table 3.3.4).

AP-2γ expression showed minimal any change in TR lines when compared with parental cells. This is contrary to published findings (Orso *et al.*, 2004).

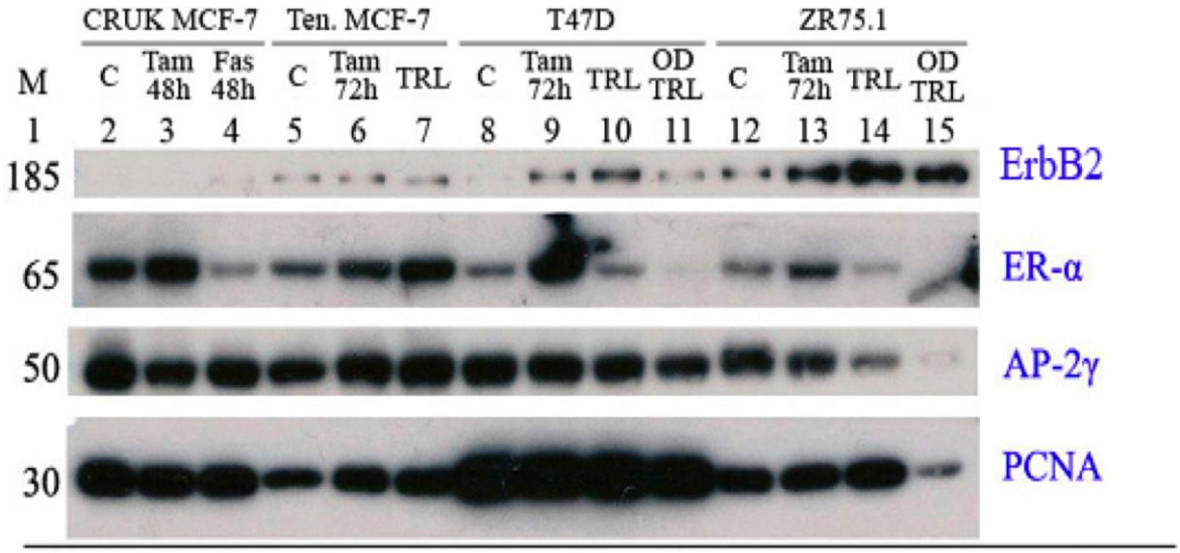


Figure 3.1 Expression of key breast markers in in-vitro models of tamoxifen resistance. WB of WCE (10μg) from MCF-7 (CRUK), MCF-7 WT (Tenovus), T47D, ZR75-1 and their TR counterparts. C=Control/parental lines grown in normal oestrogenic media. Tam 48h=exposure of the cells to 48h of Tamoxifen (10^{-7}M) containing media. Fas 48h=exposure of the cells to 48h of Faslodex (10^{-7}M) containing media. Tam 72h=exposure of the cells to 72h of tamoxifen containing media. TRL=Tamoxifen Resistant Late, grown for 6 months in tamoxifen containing media. ErbB2 is Her2 receptor protein, ER∝ is the oestrogen receptor, and AP-2γ is a transcription factor related to mammary gland genesis and development. PCNA was used as a loading control.

3.2. Expression profiling of T47DTR, ODT47DTR, ZRTR and ODZRTR

Good quality RNA was extracted (refer to Material and methods 2.6.) from T47D TR, OD T47D TR, ZR TR and OD ZR TR cell lines. Probes were prepared according to the Affymetrix protocol for the HU133 2.0 plus microarray chip. The probes were tested for quality using TEST Chips prior to hybridising on HU133 2.0 plus. The results were analysed using 'Affymetrix Console' for quality control (QC).

The analysis of the data was undertaken by a bio-informatics team, Dr Claude Chelala. The analysis normalised the TR cell line expression data with parental cell line (wild-type) data, obtained from the Gene Expression Omnibus (GEO) Datasets. These were T47D (Accession number: GSM70667) and ZR75-1 (Accession number: GSM70668), which were also arrayed on HU133 2.0 plus. Our TR cell line expression data have been submitted in the Gene Expression Omnibus (GEO) under the Accession number GSE22664). Figure 3.2. illustrates two Venn diagrams which demonstrate the data set comparisons made to identify common genes deregulated in both ZR75-1 and T47D resistant lines that might be considered to confer the TR phenotype. There are 428 genes, which are in both subsets of Oestrogen-deprived T47D and oestrogen-deprived ZRTR, which may be responsible for the Tamoxifen-resistance. These 428 genes are within the 555 genes, which is the common altered genes between the T47DTR and ZRTR.

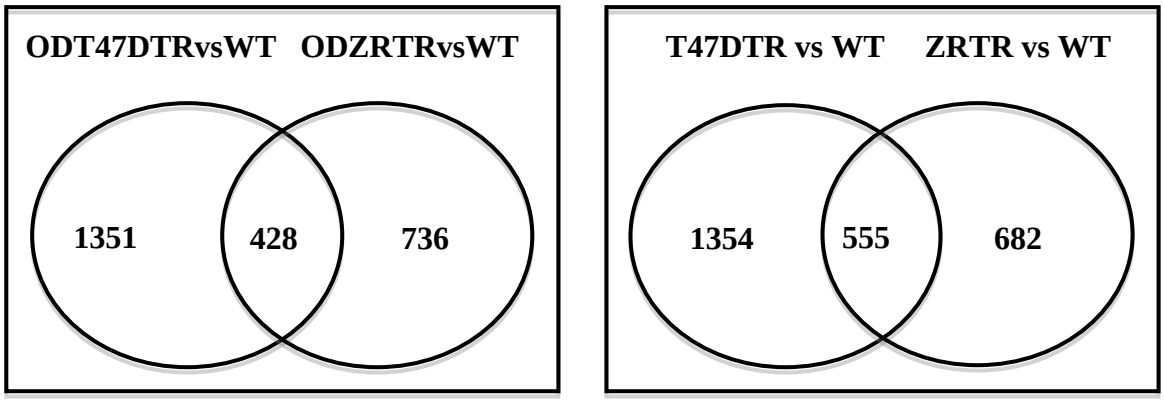


Figure 3.2 Venn diagrams showing overlap between significantly expressed genes from the ODT47DTR over the T47D (WT), ODZRTR over ZR (WT), T47DTR over T47D (WT) and the ZRTR over ZR (WT) respectively. The subset of genes, which overlapped, is seen as the overlap regions between 2 circles. There were 428 significantly altered genes in common between ODT47D TR and ODZR TR compared to WT cells (LH panel). There were 555 significantly altered genes in common between T47D TR and ZR TR compared to WT cells (RH panel).

3.3. Hierarchical Clustering Analysis

Bioconductor normalised data were imported into the limma package (R from Bioconductor) in order to allow a hierarchical clustering analysis to monitor overall patterns of gene expression between the normalised arrays (Materials and methods 2.7.3; Eisen *et al.*, 1998). Briefly, a hierarchical clustering analysis produces a map of results where probe sets were grouped together based on similarities in their patterns of normalised expression across the six arrays. As the processed data follows a normal distribution it is appropriate to use the Pearson correlation coefficient, which calculates the similarity measure based on a linear model. As we are interested in genes that change their expression between the *reference* and *test* groups, data were filtered to include the 10,000 probe sets with the highest variance across the six arrays. The filter cut-off of 10,000 probe sets were used instead of the the conventional 2,500 probe sets for higher stringency and to compensate for the fact that our reference were single experiment rather than duplicates or triplicates. The same algorithm was also applied to cluster the experimental samples for similarities in their overall patterns of gene expression. The resulting dendrogram is shown in Figure 3.3. The *reference* samples formed their own cluster, indicating that there is enough distinction in expression in this variable subset of probe sets from the TR breast cancer cell lines, T47D TR, OD T47D TR, ZR TR and OD ZR TR. Arrays hybridised were not replicates but formed their own cluster separate from the *reference* samples, i.e.T47D (WT) and ZR75-1 (WT) from the clusters of TR breast cancer cell lines. The hybridised arrays of the TR breast cancer cell lines formed their own clusters for each breast cancer cell linage, i.e. T47D TR and OD T47D TR were separated from ZR TR and OD ZR TR as illustrated in Figure 3.3. Given that both the cancer cell lines, T47D and ZR75-1 are 90% PgR-positive but 80% and 60% ER-positive respectively (see Table 3.1), it is likely that their ER-positivity differences played a larger part in their expression profile differences than the TR factor.

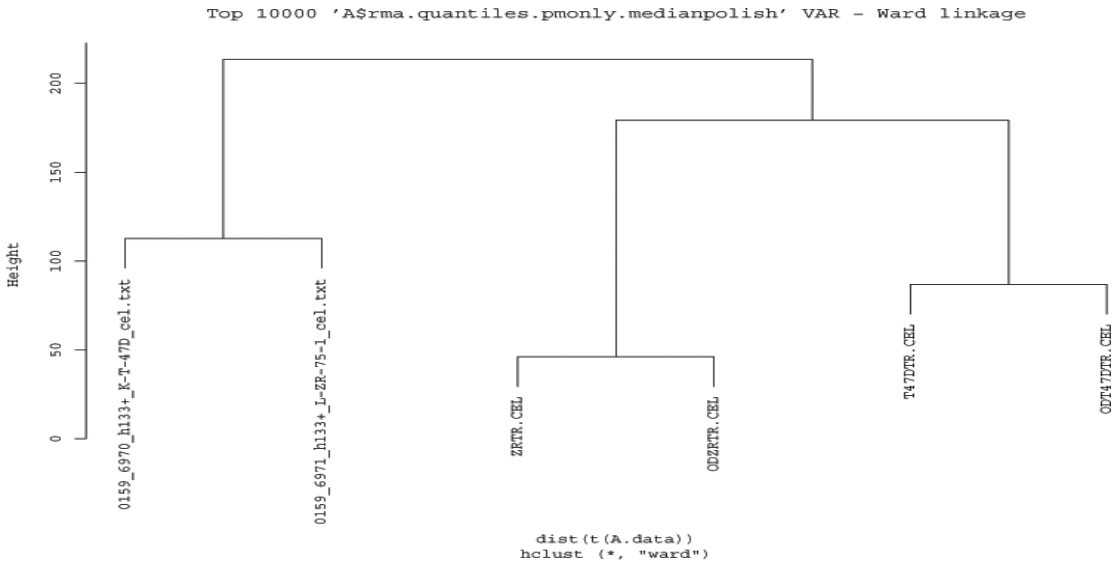


Figure 3.3 Hierarchical cluster tree generated with R software after normalisation (RMA) using APT tools. 0159_6970_h133+_K-T-47D_cel.txt and 0159_6371_h133+_L-ZR-75-1_cel.txt were the GEO datasets for T47D(WT) and ZR75-1(WT) respectively. ZRTR, OD ZRTR, T47DTR and ODT47DTR (Accession number on GEO repository: GSE22664) were abbreviations for ZR75-1 Tamoxifen resistant, oestrogen deprived ZR75-1 Tamoxifen resistant; T47D tamoxifen resistant and oestrogen deprived T47D Tamoxifen resistant.

3.3.1. Quality diagnostics using PLM

For summarisation of quality assessment of Affymetrix Genechip data, we have used the Probe-Level Model (PLM), which allowed us to examine the QC statistic described in chapter 2.7.3.3. Normalised unscaled summarised expression (NUSE) is a method to estimate summarised expression at Chip level. Each chip will have NUSE for each probe, which can be summarised by the median. Specifically, NUSE values are computed using:

$$\text{NUSE}(\theta_{gi}) = \frac{SE(\theta_{gi})}{\text{med}_i\,(SE(\theta_{gi}))}$$

where θ_{gi} = log scale estimates of expression for each gene g on each array i. SE is standard error and med=median. This provides a useful summary of the residual and can be use to judge quality relative to other chip. Median NUSE is a number that fluctuates around the value 1.0 - 'high' values, such as 1.05, indicate 'worse' (unusual) chips. Another similar method, Relative Log Expression (RLE) summarises the relative log intensity of the signals. This has a narrower range of variance, which makes the obvious outlier, if there is one, easier to spot. Figure 3.4 illustrates NUSE and RLE quality assessment at chip level of our TR breast cell lines cohort. There is some degree of correlation exists between NUSE and RLE summaries.

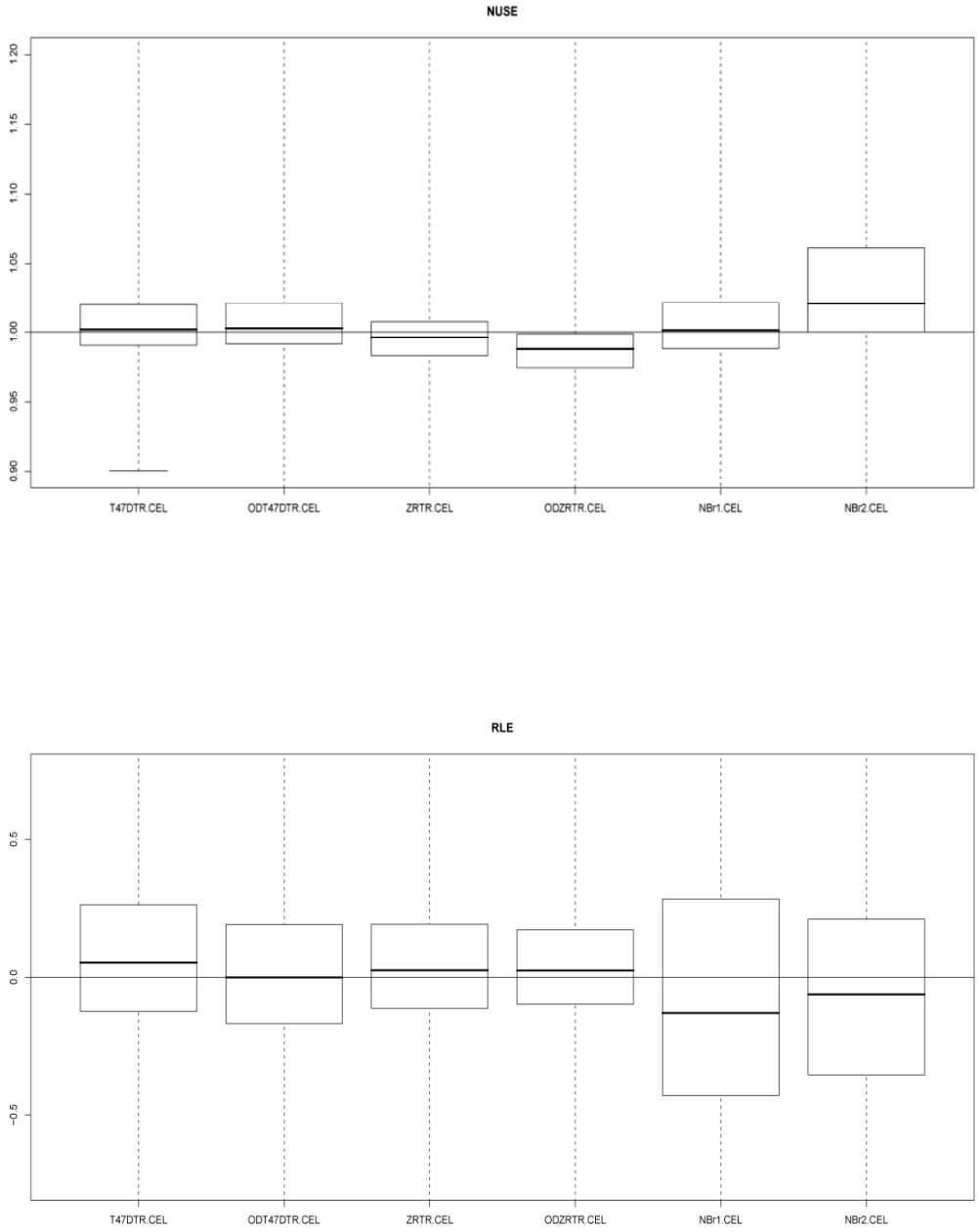


Figure 3.4 Quality control assessments on the TR expression array datasets. The NUSE (Normalised Unscaled Summarised Expression) and RLE (Relative Log Expression) at chip level. This is a method of quality control.

3.4. Statistical Analysis of Tamoxifen Resistant Differential Gene Expression

Summarisation is the process of combining the multiple probe intensities for each probe-set to produce an expression value. RMA (Irizarry *et al.*, 2003), is the expression measure used in our study. Bioconductor normalised data were imported into Limma in R software in order to allow statistical analyses to be carried out. As well as fitting a normal distribution, another assumption of statistical analyses, such as the widely used Student's t-test, is that the variability of a gene is constant across treatment types. However, in the absence of accurate diagnosis methods, it is safest to assume that variability may differ between our *reference* and *test* groups. As we have confirmed that our data fits a normal distribution, it is appropriate to conduct a Welch's t-test, this is a parametric analysis that corrects for difference in variability.

As we are interested in identifying genes that change their expression between the *reference* and *test* groups, it was important first to filter the data to include only those probe sets that change in expression. We chose to order the list from low to high standard deviation. This would enable us to identify the genes that are changed the most in expression across the six arrays, and remove probe sets from the analysis that change very little. It was felt that this is more informative than filtering on fold change, for which there is a risk of ignoring genes that change significantly but are below the arbitrary fold change threshold.

Limma software from R was used to calculate a Welch's t-test on the 2500 probe sets with highest variance across the six arrays. A Benjamini and Hochberg False Discovery Rate (FDR) test (Benjamini *et al.*, 2001) was also applied in order to identify the percentage of false positive genes that could have been included in the final lists of differentially expressed genes. Appendix 1 displays the 578 probe sets that changed their expression significantly (FDR corrected, $p<0.05$) between the *reference (*T47D (WT) and ZR75-1 (WT)) and *test* (Tamoxifen Resistant cell line) groups.

3.5. TR Gene Expression Profile Initial Observations and validation

Significant changes in gene expression were observed when TR cell lines were analysed against T47D (WT) and ZR75-1 (WT). The significantly up-regulated and significantly down-regulation genes are summarised in Table 3.2. The full gene lists from the analysis is listed in Appendix A. A smaller subset of probe sets (180 genes) regulated at a $p<0.01$ probability was used for subsequent Ingenuity Pathway Analysis. In the results presented in this thesis, probe sets were assigned to gene symbols and gene descriptions based on the January 2007 release of the HG-U133_Plus2 Affymetrix NetAffx Annotation files (Liu et al, 2003; www.affymetrix.com).

The resulting list of TR genes from this study was compared with results from a previous study in our lab (Charlotte Moss, PhD thesis 2009), which compared expression profiles of MCF-7 and MCF-7 TRL cell lines. This comparison is found in Appendix C. The platform used for the MCF-7 study was HU133A Chips, which have fewer probe-sets compared to HU133 2.0 plus, but despite this, there were still distinct similarities between our studies. The common TR genes found in both studies are AKR1C3, AKR1C2, PGR, DKK1, MARCKS and GPNMB.

We chose to validate 5 up-regulated genes, MAGEA2, AKR1C3, THRAP5, VGLL1 and EGLN3, and 4 down-regulated genes, GREB1, PDZK1, hnRNP2A1B and MYBL1 based on the relevance of these genes to breast cancer in published literature, the reproducibility when compared with other breast cell lines, i.e. MCF7 TR, gene involvement in proposed pathways in relation to resistance to treatment and the availability of commercial antibodies. These chosen genes were validated using quantitative PCR (qPCR) and immunohistochemistry.

MAP	symbol	NAME	Fold change
13q33	EFNB2	ephrin-B2	3.713047555
11q13	ALDH3B2	aldehyde dehydrogenase 3 family, B2	4.539575528
22q13.33	SCO2	SCO cytochrome oxidase deficient homo 2	3.477794836
Xq26.3	VGLL1	vestigial like 1 (Drosophila)	4.119736333
10p15-p14	AKR1C3	aldo-keto reductase family 1, member C3	4.081036655
Xq28	MAGEA12	melanoma antigen family A, 12	4.647332681
12p12.2-p12.1	LDHB	lactate dehydrogenase B	5.573450211
22q13	SERHL2	serine hydrolase-like 2	6.461724005
Xq28	MAGEA2B	melanoma antigen family A, 2B	3.84036576
Xq26.3	VGLL1	vestigial like 1 (Drosophila)	6.279577017
Xq28	CSAG2	CSAG family, member 2	3.580457482
20q12	ITGB4BP	integrin beta 4 binding protein	3.539804845
14q13.1	EGLN3	egl nine homolog 3 (C. elegans)	3.655909188
19p13.3	ALKBH7	alkB, alkylation repair homolog 7 (E. coli)	3.521589408
NA	NA	NA	3.875260601
19q13.42	CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5	4.254616896
20p11.21	ABHD12	abhydrolase domain containing 12	3.519955768
19p13.3	THRAP5	thyroid hormone receptor associated protein 5	4.257983055

MAP	symbol	NAME	Fold change
2p25.1	GREB1	GREB1 protein	-6.90466979
1q21	PDZK1	PDZ domain containing 1	-6.57908464
NA	NA	NA	-5.89388337
11q22-q23	PGR	progesterone receptor	-4.93601507
7p12	EGFR	epidermal growth factor receptor	-4.86841307
9q34.3	OLFM1	olfactomedin 1	-4.76255748
8q22	MYBL1	v-myb myeloblast viral oncogene homolog	-4.69112967
10q22-q23	RPS24	ribosomal protein S24	-4.14395185
8q24.12	TRPS1	trichorhinophalangeal syndrome I	-4.03544262
7p15	HNRPA2B1	heterogeneous nucl. ribonucleopro A2/B1	-3.99102865
12q24.21	THRAP2	thyroid hormone receptor ass protein 2	-3.95992594
11q13.1	TncRNA	trophoblast-derived noncoding RNA	-3.94963705
16p12.2	LOC23117	KIAA0220-like protein	-3.94621846

Table 3.2 Significant changes in gene expression in TR cell lines. Changes observed in probe sets between the *reference* and *test* (TR cell lines) groups at the indicated False Discovery Rate (FDR=0.05). The genes were then exported to Ingenuity Pathway Analysis software and the top ten up regulated are shown at the ***top table***, while the top ten down-regulated are in the ***bottom table***. Fold changes are seen on the right margin.

3.6. Validation of array results

These chosen genes were validated using quantitative real-time PCR (see Material and Methods 2.8.) using RNA prepared from parental and OD and TR versions of MCF-7, T47D and ZR75-1 cells. In addition, RNA from the MDAMB231 line, which is ER-negative and PgR-negative and therefore inherently Tamoxifen resistant was also tested as a control. Where appropriate antibodies were available, genes where also validated with immunohistochemical staining of cell pellets embedded in paraffin (Appendix D).

I have simplified the results in the table below (Table 3.3), which showed that the genes were positively correlated with microarray results when validated with qPCR. Commercially available antibodies for the genes/protein were purchased and used on breast cancer cell line and the tamoxifen-resistant cell line pellet, to assess if antibodies are good biomarker. We used the best antibodies for IHC-P staining for further validation across an independent cohort of breast cancer patients. MAGEA2 and EGLN3 antibodies were used with the results shown in Chapter 3.7.7 and Chapter 3.8.6.

Map	Symbol		qPCR	IHC-P in human
Xq28	MAGEA2	melanoma antigen family A, 2	pos	see Chapter 3.7.7
Xq26.3	VGLL1	Vestigial like 1 (drosophilia)	pos	good antibody
14q13.1	EGLN3	egl nine homolog 3 (C.elegans)	pos	see Chapter 3.8.6
10p15-p14	AKR1C3	Aldoketoreductase enz 1C3	pos	good antibody
19p13.3	THRAP5	Thyroid hormone receptor ass protein 5	pos	good antibody
2p25.1	GREB1	GREB1 protein	pos	good antibody
1q21	PDZK1	PDZ domain containing 1	pos	good antibody
11q22-q23	PGR	Progesterone receptor	pos	ND
8q22	MYBL1	v-myb myeloblast viral oncogene homolog	pos	good antibody
7p15	HNRPA2B1	heterogeneous nucl. Ribonucleopro A2/B1	pos	good antibody

Table 3.3 Positive validation of the genes/proteins from the tamoxifen-resistant microarray study. MCF-7, T47D and ZR75-1 and their Tamoxifen-resistant (TR) counterparts were grown and made into cell pellet and set into a paraffin block. These were then made into paraffin slides, used for staining with the commercially available antibodies for our genes of interest. These cells were also extracted for RNA, which we made into cDNA. The cDNA (1:3 dilution) was used with Tagman primers for the genes to access for mRNA expression in the three breast cancer cell lines and its TR counterparts.

3.7. Functional study (in-vitro) of MAGEA2 in TR

3.7.1. MageA2

MAGEA2 was chosen to study further as it was consistently up regulated in T47DTR, ODT47DTR, ZRTR and ODZRTR. In addition it was not detected in MDAMB231 (our negative control), which does not express negative hormone receptors. As described in the Introduction (see Chapter 1.10), the *MAGEA* gene family is not expressed in most adult tissues but is frequently up regulated in several tumour types including malignant melanoma, germ cell tumours and, to a lesser extent, breast cancer. Little is known about *MAGEA* function but there are recent suggestions of involvement in chemotherapy resistance, and regulation of apoptosis via the p53 pathway (Monte *et al.*, 2006). It was suggested that MageA2 could form a complex with p53 and thereby reduce its activity as a transcription factor. I set out to examine if *MAGEA2* over-expression may also have a role in the development of Tamoxifen resistance in breast cancer cells.

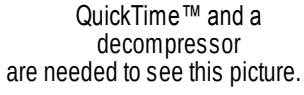
3.7.2. MAGEA2 is over-expressed in a panel of Tamoxifen-Resistant cell lines

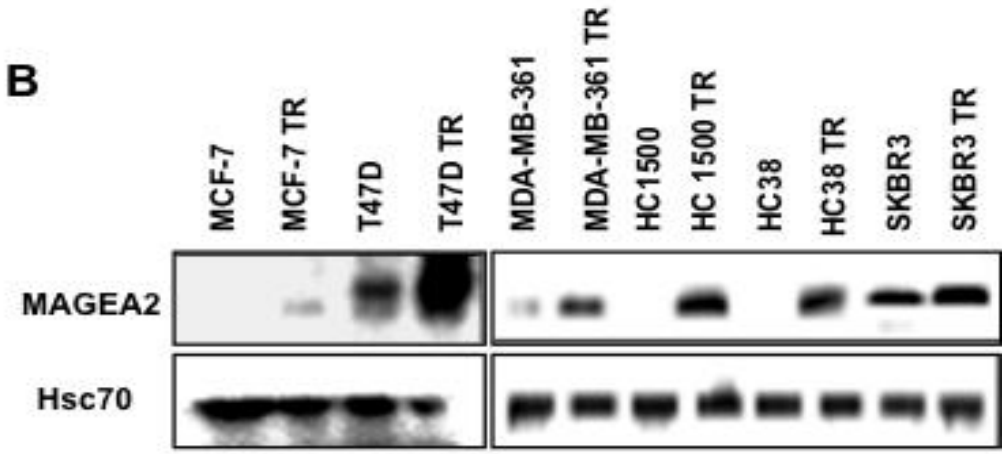
To aid our study into gene expression changes in Tamoxifen resistant breast cell lines, additional lines (both ER positive and negative) were generated by maintaining the cells in Tamoxifen-containing media (10^{-7}M) for at least three months (see Chapter 2.1.1). Cell lysates were prepared from the panel of TR cells and their wt counterparts and analysed for MageA2 expression by immunoblotting (see Figure 3.5B). All the ER-positive cell lines, MCF-7, T47D, MDA-MB361 and HCC1500, showed no or very low MageA2 protein expression in the wt lines with significant induction expression in the TR lines. The ER-negative SKBR3 line is considered to be inherently Tamoxifen resistant since it carries an amplification of the *ERBB2* gene (see Table 3.1). Interestingly, the wt SKBR3 cells already expressed elevated levels of MageA2, which were maintained when the cells were grown in Tamoxifen media.

MAGEA2 expression was also examined at the mRNA levels using qPCR with broadly similar results (Figure 3.5C).

The mRNA expression resonated the same result as seen in protein immuno-blotting, apart from the data from MCF-7 and MCF-7 TR. We did not succeed in detecting MAGEA2 by real-time PCR, despite using different probes and primer, or various different starting cDNA amounts. We analysed clones of tamoxifen-resistant MCF-7 cell lines to test for mRNA expression with no avail. When using the generic MAGEA probe and primer from ABI, we detected mRNA and an up-regulation of MAGEA. (A possible explanation is that MCF-7 has a very low level of MAGEA2 mRNA, due to its low turn-over of the MageA2 protein).

A


QuickTime™ and a
decompressor
are needed to see this picture.


B
MCF-7
MCF-7 TR
T47D
T47D TR
MDA-MB-361
MDA-MB-361 TR
HC1500
HC 1500 TR
HC38
HC38 TR
SKBR3
SKBR3 TR
MAGEA2
Hsc70

3.7.3. Generation of MAGEA2 overexpressing cell lines

Since induction of *MAGEA2* overexpression appears to be a characteristic of several tamoxifen resistant cell lines, I decided to investigate if the encoded protein can play a functional role in resistance to tamoxifen. To achieve this, I generated stable *MAGEA2* overexpressing lines in wt, tamoxifen sensitive MCF-7 and T47D breast tumour lines. As a first step, a *MAGEA2* mammalian expression construct was made using a cDNA clone obtained from the I.M.A.G.E Consortium. As detailed in the Materials and Methods (Chapter 2.2.4), the insert was excised and cloned into the pcDNA3.1 expression vector which carries a strong mammalian promoter, CMV (cytomegalovirus) and also the Neomycin selection marker.

T47D and MCF-7 cells were transfected with the pcDNA3.1/*MAGEA2* plasmid or "empty" pcDNA3.1 (vector alone, VA, negative control) by nucleoefector technique. Conditions were optimized for T47D and MCF-7 cell lines as suggested by the manufacturer (Amaxa, see Chapter 2.2.4.6.). Separate transfection experiments used either circular or linearised plasmids. In the latter case plasmids were digested using the restriction enzyme Kas I within the ampicillin resistance cassette. In order to optimise transfection conditions, different concentrations of cells were plated and two different concentrations of the construct added, 1µg or 2µg. The selection drug (G418) was added to the media the day after transfection at a previously optimised concentration (see Chapter 2.2.4). For the vector alone controls, a pool of G418-resistant colonies for each cell line was made and maintained as separate cell lines, named MCF-7/VA and T47D/VA, grown in standard media supplemented with G418. For the MAGEA2 transfected cells, individual colonies were picked after 16 days. The colonies were expanded and levels of *MAGEA2* expression were determined using qPCR and Western blotting. Two high-expressing, positive clones (c18 and c24 for MCF-7, and c30 and c34 for T47D) were used in subsequent experiments (see Figure 3.7A & Figure 3.8A).

3.7.3. Growth of MAGEA2 expressing clones in Tamoxifen-containing media

To test if expression of *MAGEA2* is able to confer resistance to Tamoxifen, cell count assays were performed in triplicate for each cell line comparing wild-type (wt) cells, vector alone (VA), and two MAGEA2-expressing clones in Tamoxifen-containing media. For the MCF-7 derived cells (see Figure 3.6B), there was a highly significant difference between growth of the clones, which continued to proliferate in Tamoxifen-containing media, and their VA and wt counterparts which showed growth arrest (p-value of 0.00185 for MCF-7 c18 versus VA averaged from day 1 to day 8, Students t-test). Broadly similar results were also obtained when the T47D derived cells were analysed (shown in Figure 3.6D) with the MAGEA2 expressing clones again able to sustain growth in Tamoxifen–containing media (p-value of 0.00435 for T47D c30 versus VA averaged from day 1 to day 8, Students t-test). This experiment was repeated again in triplicates a month later with the similar results.

To investigate further the growth differences between control and MAGEA2-expressing lines, I analysed if the MAGEA2 clones showed a proliferation advantage in normal media over parental / VA control cells using a BrdU incorporation assay to measure the proportion of cells in S-phase and sub-G1. However, there was no significant difference found between control and MAGEA2-expressing cells suggesting that MageA2 expression does not confer an inherent cell cycle proliferation advantage to cells. (data not shown). I next examined if MageA2 expression can protect cells from apoptosis, particularly in the presence of Tamoxifen. The MCF-7 control (wt and VA) cells and the two MageA2 clonal lines were grown in the presence and absence of Tamoxifen-containing media for 48 hours and then assayed for Annexin V binding, a recognised hallmark of early apoptotic cells. The samples were additionally stained with propidium iodide (PI) to differentiate between intact cells (AnnV-PI-), early apoptotic (AnnV+PI-) and late apoptotic/necrotic cells (AnnV+PI+) using FACS analysis. As expected, the control cells (wt and VA) showed increased percentages of dead and dying cells when Tamoxifen was added to the media, but both c18 and c24 had reduced levels of cell

death in Tamoxifen (pvalue=0.027* and 0.043* respectively for VA and C18, and VA and C24 in Tamoxifen media, see Figure 3.7). Similar results were also found for the T47D derived lines (Figure 3.8).

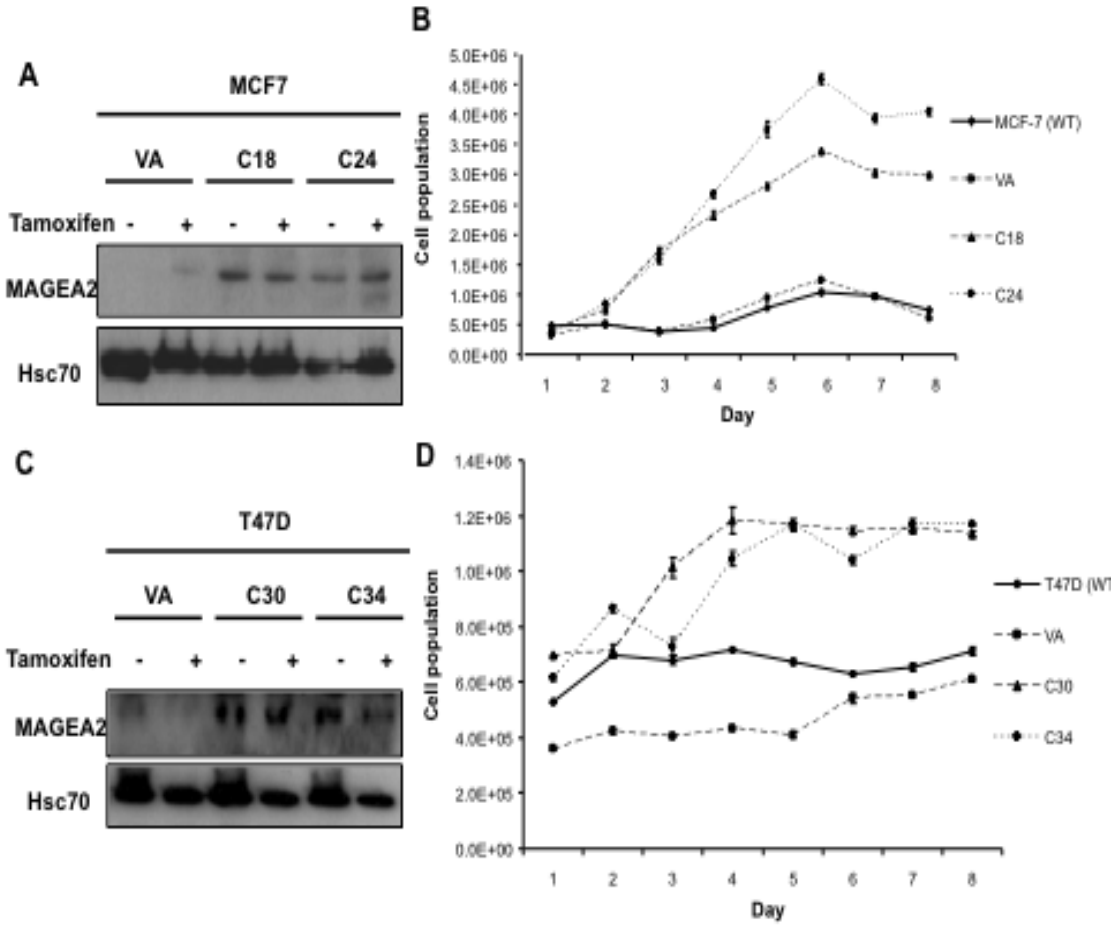


Figure 3.6 MageA2 stable over-expressing clones have a proliferation advantage in MCF-7 and T47D breast cancer cell lines in tamoxifen containing media. (A, C) Western blot showing stable MAGEA2 overexpression in MCF-7 and T47D. (B, D) The cell count study showed that there was an increase in cell population compared with non-transfected and mock-transfected MCF-7 or T47D in tamoxifen-containing media compared clones with VA. This analysis was obtained from two individual experiments with triplicates for each group. (Asterisk (*,**) $p<0.001$ relevant to vector alone control, student t' test)

Cell counting for two individual MAGEA2-expressing stable MCF-7 lines (Clone 18 and 24) and T47D lines (Clone 30 and C34) compared to wt cells and those transfected with vector alone (VA). 75,000 cells/well were seeded into 6-well plates in media supplemented with 10^{-7}M tamoxifen. Cells were counted from triplicate wells (Coulter cell counter) daily for eight days. This analysis was obtained from one of two individual experiments which both gave very similar results. Growth rates at later time points declined due to cell confluence.

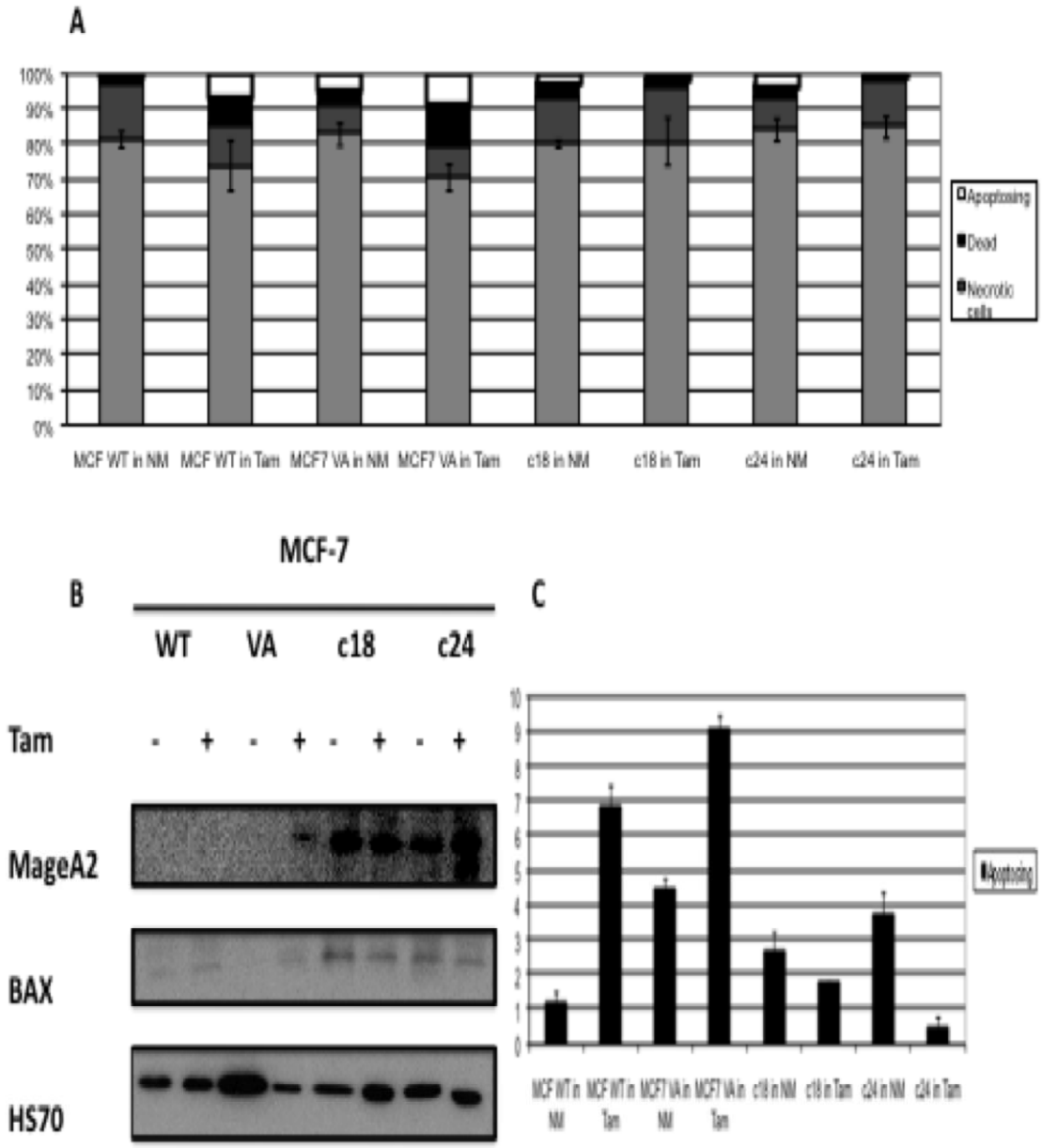


Figure 3.7 ***MAGEA2*-expressing clones show reduced apoptosis in Tamoxifen media compared to controls.** For each line (wt MCF-7, VA, c18, c24) 100,000 cells/well were plated in 6-well plates, and grown in normal media either without or with 10^{-7}M Tamoxifen, using triplicate wells for each line and condition. Cells were harvested at 24 hours and were processed using ApopNexin Annexin V FITC Apoptosis Kit (Millipore) according to the manufacture's protocol and analysed on a FACScalibur Flow Cytometer. Quadrant statistics were used to identify the percentage of apoptotic (propidium iodide-negative, annexin V-positive) cells – seen on graph (A) as white box. There are reduced apoptosis cells in the clones (C18, and C24) when compared with WT and VA when exposed to Tamoxifen media. (for methods, see Chapter 2.9.1)

(B) Immunoblot showed the MageA2 protein is present in C18 and C24 as seen in the last four lanes. Tamoxifen when added into the media and left for 24 hours are indicated as + when

present and – when absent. BAX protein is elevated in all the Tamoxifen containing experiments, and more distinct in the C18 and C24, even when they are in their normal media without the Tamoxifen.

(C) Graph showing the percentage of apoptosing cells in the Annexin V and PI study in their normal media and Tamoxifen containing media. The percentage of apoptosing cells is higher in the Tamoxifen containing media for WT and VA, as expected. But in the MCF-7 clones, C18 and C24, there was less apoptosis in the Tamoxifen media when compared with their normal media.

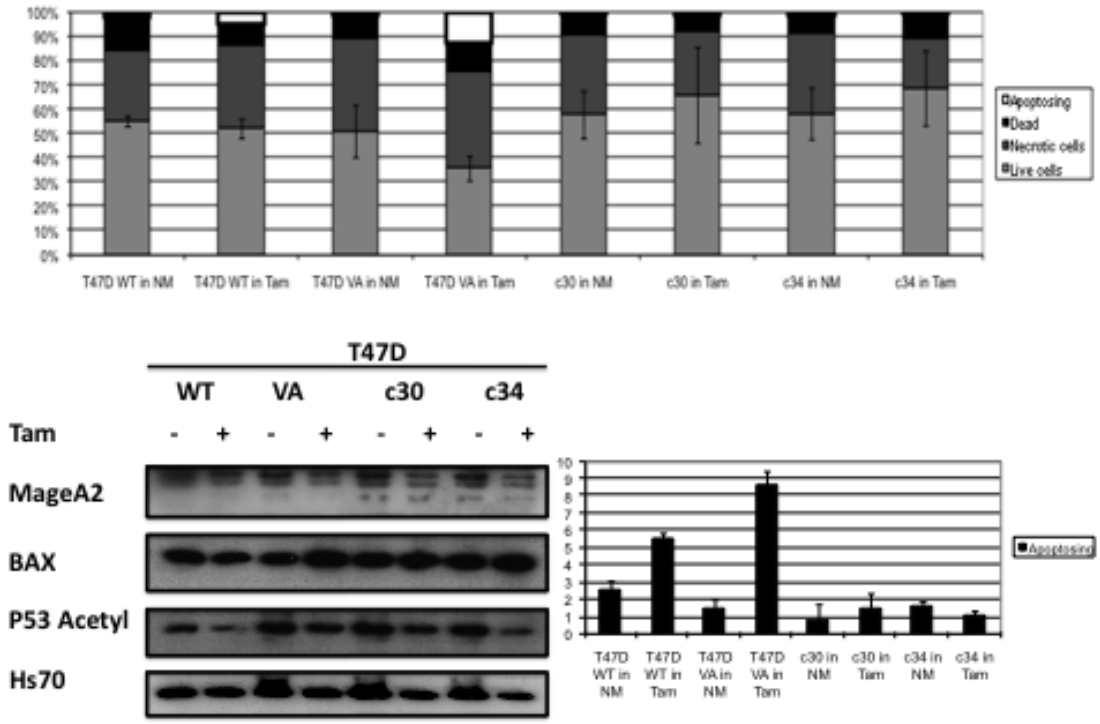


Figure 3.8 ***MAGEA2*-expressing clones show reduced apoptosis in Tamoxifen media compared to controls.** For each line (wt T47D, VA, c30, c34) 100,000 cells/well were plated in 6-well plates, and grown in normal media either without or with 10^{-7}M Tamoxifen, using triplicate wells for each line and condition. Cells were harvested at 24 hours and were processed using ApopNexin Annexin V FITC Apoptosis Kit (Millipore) according to the manufacture's protocol and analysed on a FACScalibur Flow Cytometer. (A, C) T47D positive clones have more total live cells (A) and less apoptosing cells (C) than WT and VA after exposure to Tamoxifen containing media, as shown by Annexin V and PI study. (B) Immunoblot showed that positive T47D clones have an increase expression of BAX in both their normal media as well as Tamoxifen containing media as compared with WT and VA. There was no convincing difference in the acetylated p53 between the positive clones and WT/VA.

3.7.4. Co-Immunoprecipitation (Co-IP) study showed MAGEA2 interacts with p53 to regulate its pathway

As mentioned in the introduction, previous studies have shown MAGEA2 is able to associate with p53 providing a cell survival advantage and chemotherapy resistance in melanoma cells and U2OS cell models (Monte *et al.*, 2006). I therefore examined, using co-immunoprecipitation (coIP) assays, if MageA2 is able to complex with p53 in MAGEA2-expressing clones and whether this reduces p53 signalling in Tamoxifen-containing media leading to reduced growth arrest and apoptosis.

T47D-derived lines were used for the coIP experiments as they express quite high levels of p53. Although they carry mutant p53, this mutation (L194F) still retains DNA binding activity and the ability to induce p53 target genes and wt T47D cells can still carry out p53-dependent apoptosis ((Chopin *et al.*, 2002; Kato *et al.*, 2003; Toillon *et al.*, 2002). Whole cell lysates were prepared from T47D/VA control cells and the MAGEA2-expressing clone 30. Cells were grown in their normal media, and immunoprecipitated (IP) for MageA2 or p53. To perform immunoprecipitate, I used IgG as control antibody, and p53-antibody (total p53) and MAGEA2 antibody on T47D and c30 lysates. Each IP was western blotted for both antigens which revealed that MageA2 and p53 formed a complex since p53 was detected in MageA2 immunprecipitates and p53 was found in MageA2 precipitates in lysates from clone 30 but not from control VA cells (Figure 3.9, top left panels). Control western blots of these lysates were also probed for acetylated p53 (p53Ac) which showed lower levels of p53Ac in clone 30 cells, although total p53 levels were similar in both lines. Levels of the p53 target gene $p21^{WAF}$ were also reduced in clone 30 compared to VA cell lysates (Figure 3.9, right panels). This Co-IP study therefore supports the hypothesis that MageA2 interacts with p53, leading to reduced levels of acetylated, transcriptionally active p53 and hence reduced expression of target genes such as p21 resulting in continued growth in tamoxifen-containing media.

Under normal circumstances, deacetylase inhibitor (Trichostatin A) increases acetylation of p53, as seen in the VA lane in Figure 3.9B, but with exogenously expressed MAGEA2, this acetylation of p53 is not seen, and subsequently p21 is down-regulated. Down-regulation of p21 is not readily seen in MCF-7 that stably overexpressed MAGEA2 in their normal media. Down-regulation of p21 in up-regulated MageA in other cell line study (such as in ovarian study) is also seen in a publication with melanoma (Liu *et al.*, 2008), and unpublished data (as communicated by Prof McNeish's team).

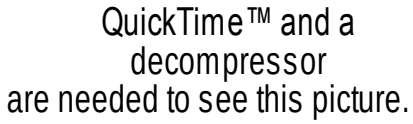


Figure 3.9 MAGEA2 interacts with p53 to regulate its pathway. Whole cell lysates were prepared from T47D VA and clone 30 cells and 30 μg of lysate was immunoprecipitated for either MageA2 or p53. (A) Co-immunoprecipitation indicated that MAGEA2 was immunoprecipitated with antibody to MAGEA2 or p53. IgG was used as control antibody alongside Also, overexpression of MAGEA2 down-regulated the acetylation level of p53 and its downstream effector, p21. **(B)** Deactylase inhibitor (Trichostatin A) treatment also showed that exogenous expression of MAGEA2 prevents p53 acetylation and subsequently p21 up-regulation.

3.7.5. MAGEA2 and p53 co-localise in cytoplasmic compartment

To date, there has not been any published data on the intracellular immunocytochemistry localisation of MAGEA2 in cells. We have shown in our study that the presence of MageA2 appeared as early as 12 hours after exposure to Tamoxifen containing media in both T47D and MCF-7 wild-type. The overexpressing clones of both cell lines co-localised with p53. The distribution of MageA2/p53 complexes was generally cytoplasmic apart from a few cells, which appeared to be either undergoing apoptosis, or had undergone apoptosis. In these cells, MAGEA2/p53-complexes are localised in the nucleus.

In MCF-7 VA control cells, there was no or little MAGEA2 in keeping with RT-PCR results indicating low or no expression of MAGEA2 in MCF-7 WT (Figure 3.5). As MCF-7 wild type does not carry mutant p53, cells grown in their normal media do not express p53 protein, as seen in our confocal immunoflourescent results (Figure 3.10). Upon exposure of MCF-7 VA control cells to Tamoxifen, expression of MageA2 was detected. Futhermore p53 localisation followed the same distribution as MageA2 (Figure 3.10B, 2nd row). In the MCF-7 VA, after exposure to Tamoxifen containing media, most of the cells were undergoing apoptosis. In keeping with the immunohistochemistry study, we observed that the MageA2 and p53 localisation was in the nucleus (Figure 3.10B, 3rd row). In healthy dividing cells, MCF-7 c18, MageA2 and p53 were localised in the cytoplasm. This finding is in keeping with our survival analysis of the immunohistochemistry results; cytoplasmic MageA2-staining was correlated with a worse prognosis, hence there appears to be greater cell survival potential.

In T47D vector alone control cells, there is moderately low expression of MageA2 as expected as T47D wild type cells have a detectable level of MAGEA2 by real-time PCR (see Figure 3.11B). As T47D carry mutant p53, the intensity of p53 protein expression was seen in T47D VA (control cell) even when grown in their normal medium. The intensity of MageA2 and p53 increase in overexpressing clones (C30) particularly in Tamoxifen containing media.

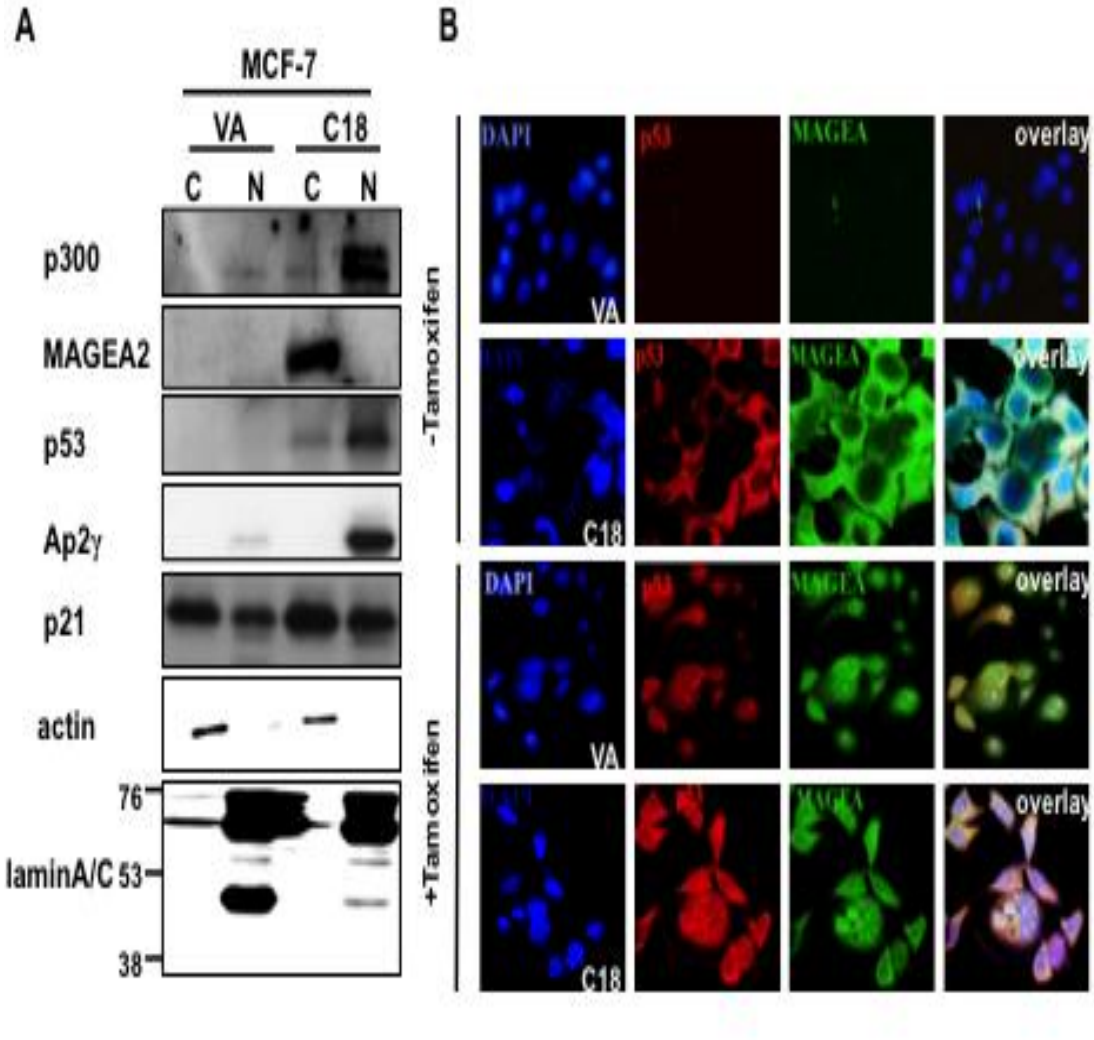


Figure 3.10 MageA2 co-localises with p53 in MCF7. (A) Western blot indicates that MAGEA2 is localized in the cytoplasmic compartment of MCF-7 cells. Interestingly, overexpression of MAGEA2 up-regulated the protein level of p53 in MCF-7 cells. Cells (MCF VA and MCF-7 MAGEA2 clone, C24) were stained with anti-MAGEA2 and p53 (DO-1). Cells were treated with either with or without 10^{-7}M taxmoifen for 24 hours. Cellular localisation was determined by confocal immunofluorescence microscopy, MAGEA2 is shown in green, and p53 is shown in red. Nucleus was stained with DAPI (blue). In healthy surviving cells, MageA2 is mainly localised in the cytoplasm of MAGEA2 overexpressed cells. In contrast, MageA2 and p53 became localised to the nuclear compartment of both cell lines in apoptosing-cells, or post apoptotic cells.

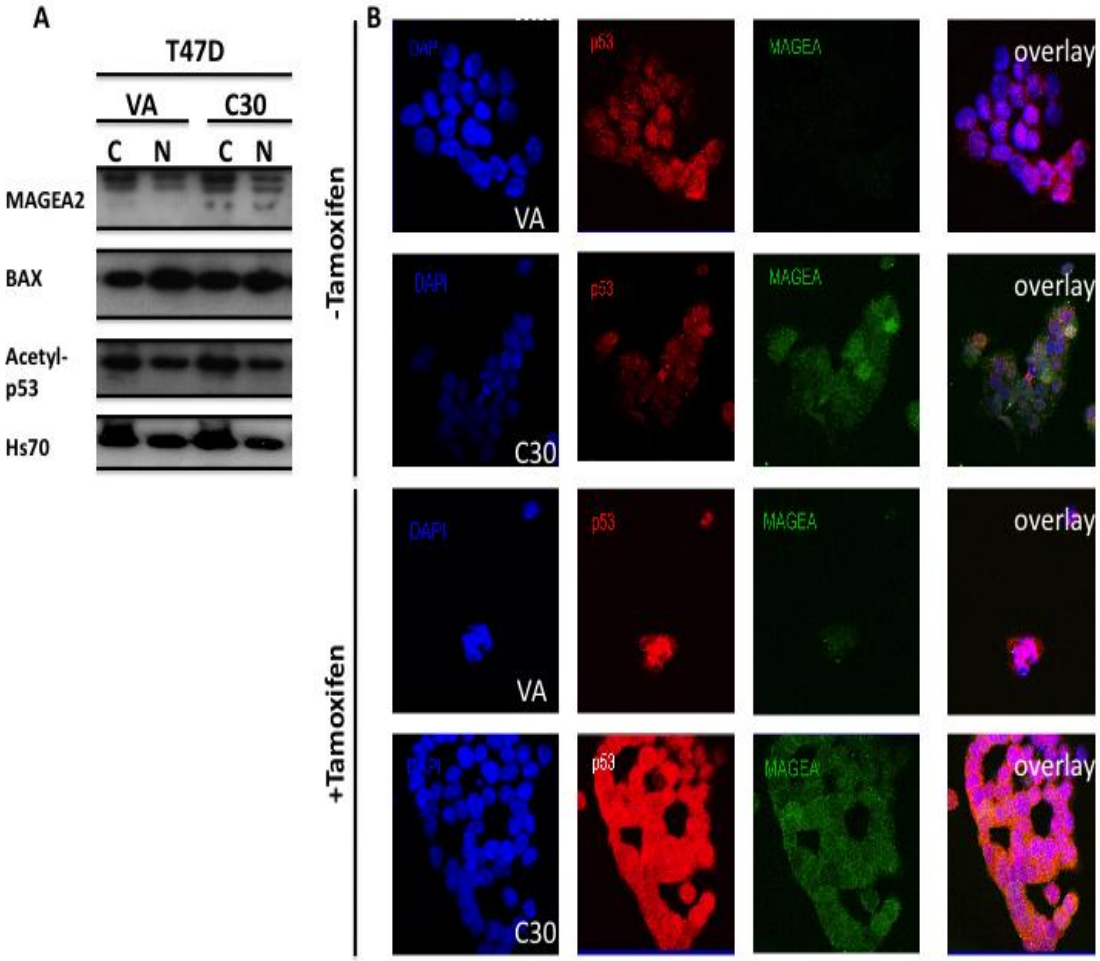


Figure 3.11 MageA2 co-localises with p53 in T47D. (A) Western blot indicates that MAGEA2 is localized in the cytoplasmic compartment of T47D cells (MageA2 band as seen as the bottom most band with this antibody. Overexpression of MAGEA2 up-regulated the protein level of p53 in T47D cells is not so easily appreciated as T47D (WT/VA) normally carry detectable mutant p53. Cells (T47D VA and T47D MAGEA2 clone, C30) were stained with anti-MAGEA2 and p53 (DO-1). Cells were treated with either with or without 10^{-7}M taxmoifen for 24 hours. Cellular localisation was determined by confocal immunofluorescence microscopy, MAGEA2 is shown in green, and p53 is shown in red. Nucleus was stained with DAPI (blue). In healthy surviving cells, MageA2 is mainly localised in the cytoplasm of MAGEA2 overexpressed cells (see C30 in normal and also Tamoxifen media). In contrast, MageA2 and p53 became localised to the nuclear compartment of both cell lines in apoptosing-cells, or post apoptotic cells (see VA in Tamoxifen).

Transwell Migration study

Transwells were used to used for the study of chemotactic response of MCF-7 (VA) and MCF-7 MAGEA2-expressing clone, C24. MAGEA2-expressing clones were significantly more chemotactic than control (VA) to media with serum (fetal bovine serum) at 18 hours. The migration towards serum was measured as total cells at the bottom of the well (cells which penetrated the filter), as a percentage of total cells (cells at the top well plus cells from the bottom well). The experiments were done in triplicates and at two separate times. The percentage of cells, which migrated to the bottom in the MAGEA2-expressing clone was visibly and objectively higher than MCF-7 vector alone (VA). The Student's t-test showed there was a statistical difference between the two groups with a p-value of 0.00023.

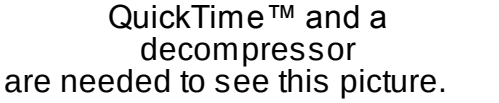


Figure 3.12 MAGEA2-overexpressing clone, C24 has a significantly higher migratory chemotactic ability when compared with VA in MCF-7 cell line. 200,000 cells were plated into 24-well chemotaxis chamber and polyvinyl-pyrolidone-free polycarbonate filters with 8 μm pore size (Costar®) and performed essentially as describe (see Chapter 2.11.4). After 18 hours, cells from the bottom well, and cells from the top well were trypsinised and counted separately.

3.7.6. MageA2 Immunohistochemistry Study

An immunohistochemistry study was performed on breast cancer tissue. A cohort of 129 patients, who had completed Tamoxifen treatment and had a complete clinical follow up. Seven cases were from relapsed patients' paraffin slides, which were provided by our collaborators at Charing Cross hospital. Of the 129 patients, a third were Tamoxifen-Sensitive (TS) while the majority 66% were Tamoxifen-Resistant (TR). A summary of the origin of the paraffin slides and the response rate in primary tissue and relapsed tissue is listed in Table 3.4.

Validation sets	Group A	Group B	Group C	Group D
Sample origin	25 cases GSTH	51 cases GSTH	7 paired primary & relapsed from CXH	46 cases primary & 1 relapse BTLH
Response	31% of TR primary tissue		62% Relapsed samples	
Sample type	Paraffin (1985-1997)	Paraffin (1985-1997)	Paraffin (1994-2004)	TMA (1985-2000)

Table 3.4 MageA2 was positive in 31% of Tamoxifen-Resistant primary breast tissue, and 62% of relapsed breast cancer tissue. Total of 129 cases of paraffin slides were stained by using automated Ventana (see material and methods); at 1:50 dilution. Most paraffin slides were made from standard paraffin blocks, which have been kept in archive library (some for a duration of >15 years).

As described above, the microarray study indicated MAGEA2 to be significantly up-regulated in Tamoxifen resistant in-vivo breast cancer cell lines. Hence MageA2 is likely to play as significant role in cell survival in relapsed tissue. For this reason, analysis for immunohistochemistry staining was separated into categories; in the primary breast tissue (n=121), and a rare but smaller collection of relapsed tissue (n=8). Relapsed tissues specimens are difficult to obtain, given that most patients with relapsed or metastatic breast cancer do not return to surgery as their subsequent line of treatment. Rather, they are restaged with radiology, and treated for metastatic breast cancer. In the relapsed samples were from ipsilateral breast cancer or locally relapsed breast cancer. The sensitivity for MageA2 was 62% in the relapsed tissue cohort, twice the sensitivity of primary tissue (31%). In the paired cases of primary and relapsed paraffin cases (7 cases from Charing Cross Hospital, and 1 case from Bart's and the London Hospital) in tamoxifen resistant cases (n=8), only 3 out of 8 primary tissues were MageA2 positive, while 5 out of 8 relapsed cases were MageA2 positive. The specificity for Tamoxifen-Resistant (TR) cells is considered high as *only* one sample of the Tamoxifen-Sensitive (TS) cells cohort was positively stained.

When the positive cases were analysed in detail, MageA2 was either distinctly positive or not there at all (Figure 3.13D). The intensity of the positive stained slides were clearly positive, and localised to breast tumour cells. (There are however two types of staining, either cytoplasmic or nuclear staining). We analysed the series for survival with MAGEA2-positive against negative patients, as well as distinguishing between cytoplasmic or nuclear staining. The results showed a statistical significant survival advantage in the positively stained versus the negatively stained tissue in the group A category (Guy's and St Thomas' Hospital) and group B (Bart's and the London, and Charing Cross Hospitals), with p=0.0455 and p=0.0286 respectively (see Figure 3.13A and B).

There was also a significant overall survival difference between the samples with cytoplasmic and nuclear staining, with nuclear favouring a better survival, p=0.0488 (Figure 3.14B). This observation led us to carry out immuno-flourescent analysis (IF) with the intention to localise MageA2 within a cell with respect to p53.

Analysis was also independently undertaken by the statistical department at the Wolfson Institute (within the Charterhouse Square) to provide the exact value of MageA2 as a biomarker for predicting tamoxifen resistant in primary breast tissue (see Table 3.5). MageA2 is a good biomarker with a high specificity (80%) and a good sensitivity (38.5%) to predict tamoxifen resistant (TR) primary breast cancer tissue. MageA2 has a positive predictive value of 89%, which suggests that when MageA2 is positively stained, it has 89% likelihood that the primary tissue is predictive of TR. MageA2 has diagnostic odd ratio of 2.59 (95% CI; 1.07 to 6.28). Youden index is a measure for assessing the quality of a diagnostic test and is calculated from the sensitivity and specificity as Youden index = sensitivity + specificity -1. Positive predictive value (PPV) is very important proportion of a diagnostic; it is the probability that a patient has the disease when restricted to those patients who test positive. It is calculated as PPV = TP/(TP + FP). Whereas the positive likelihood ratio (PLR) tells you how much the odds of the disease increase when a test is positive and is calculated as PLR = Sensitivity/(1-Specificity). The diagnostic accuracy refers to the ability of a test to identify a condition of interest. It is (TP + TN) / (TP + FN+FP + TN). Diagnostic odds ratio (DOR) summarises the performance diagnostic test. It is calculated as DOR=(TP*TN)/(FP*FN). The higher the odd ratio, the better the 'test'.

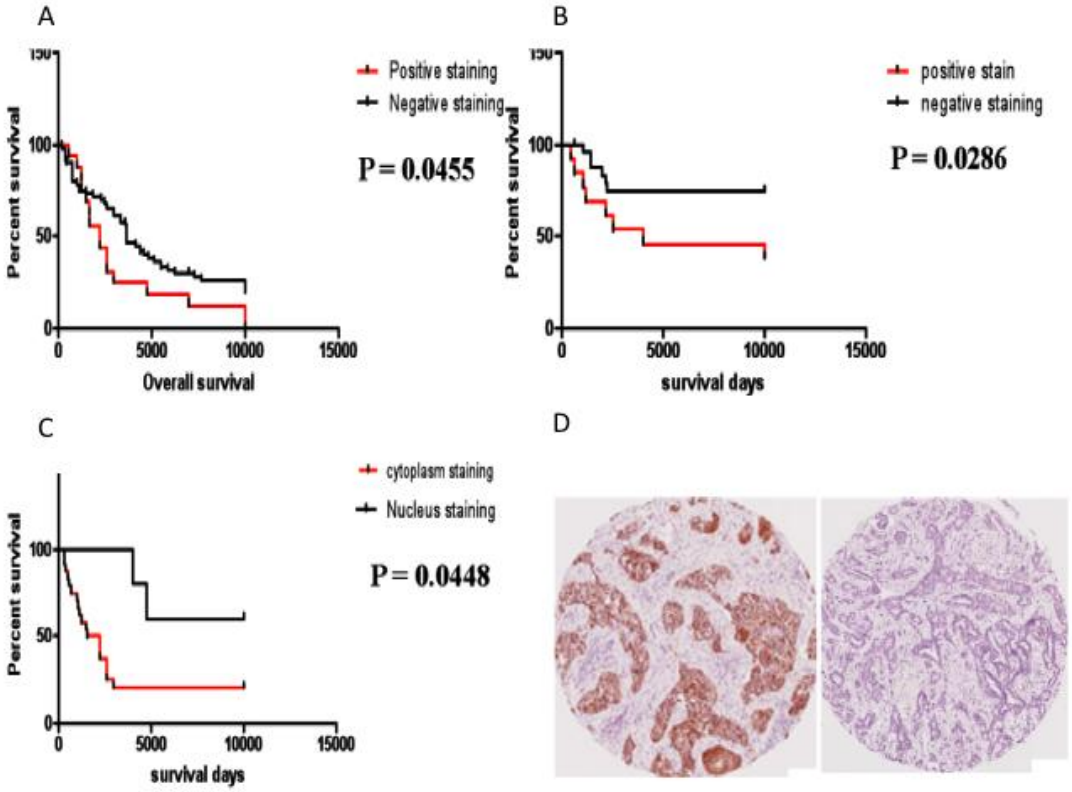


Figure 3.13 Kaplan-Meier survival curves showing the relationship between positive and negative staining of MAGEA2 on disease-free and overall survival. Graph (A) composed of patients from Guy's and St Thomas' hospital (GSTH) and (B) of patients from Bart's and the London hospital (BTLH). A significant OS difference (p=0.0455 and 0.0286 was found respectively) between the MageA2-positively stained and the MageA2-negative cohort. (C) Graph C is the combined analysis of GSTH and BTLH stained positive patients (only). A significant difference was detected in OS between the cytoplasmic and the nuclear MageA2-stained tissue; p-value=0.0448, with the cytoplasmic staining cohort conferring a worst prognosis. (D) Immunohistochemistry staining of MAGEA2; left image shows positive staining and right image, negative staining.

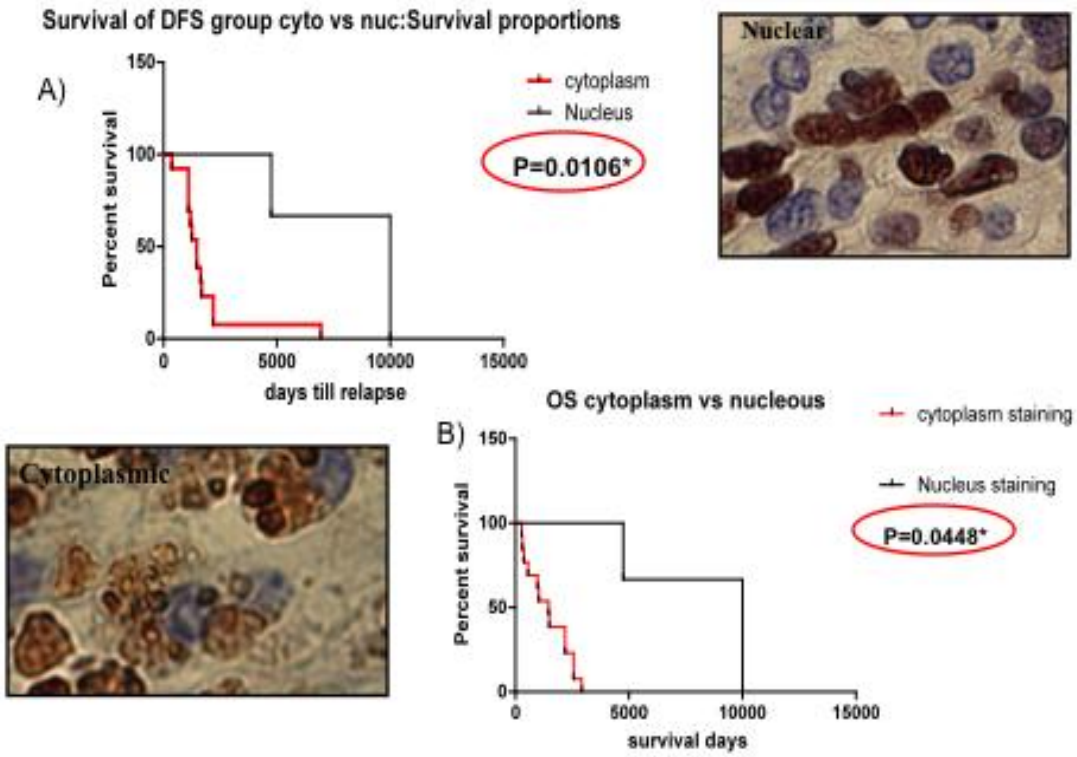


Figure 3.14 Cytoplasmic MageA2-staining predicted a worse prognosis than the nuclear MageA2-staining in disease free survival (DFS) and overall survival (OS). (A) Positive cytoplasmic MageA2-staining predicted a worse disease free survival prognosis compared with the positive nuclear MageA2-staining patients, with a significant p-value difference of p=0.0106*. (B) Positive cytoplasmic MageA2-staining predicted a worse overall survival prognosis compared with positive nuclear MageA2-staining, with a significant difference of p=0.0448*.

	Mage A2		
Calculation of	**Estimated Value**	**Lower CI**	**Upper CI**
Apparent prevalence	0.3502538	0.2838148	0.4212762
True prevalence	0.8172589	0.756098	0.8686062
Sensitivity	0.3850932	0.3095788	0.4649204
Specificity	0.8055556	0.639752	0.9180564
Diagnostic accuracy	0.4619289	0.3908159	0.5342049
Diagnostic odds ratio	2.594517	1.071532	6.282142
Youden's index	0.1906487	-0.05066922	0.3829768
Positive predictive value	0.8985507	0.8020815	0.9582291
Negative predictive value	0.2265625	0.1572867	0.3089065
Positive likelihood ratio	1.980479	0.9904406	3.960154
Negative likelihood ratio	0.7633326	0.6238766	0.9339614

Table 3.5 MageA2 is a good biomarker with a high specificity (80%) and a good sensitivity (38.5%) to predict Tamoxifen resistant (TR) primary breast cancer tissue. MageA2 has a positive predictive value of 89%, which suggests that when MageA2 is positively stained, it has a 89% likelihood that the primary tissue is predictive of TR. MageA2 has diagnostic odd ratio of 2.59 (95% CI; 1.07 to 6.28).

Functional study (in-vitro) of EGLN3 (PHD3)

EGLN3 is a member of the prolyl hydroxylase domain PHD family (also called EGLN family) that consists of three members, PHD1, PHD2 and PHD3 (Table 3.6). PHDs are Fe (II) and 2-oxoglutarate-dependent oxygenases that hydroxylate N- and C- terminal prolyl residues in HIF1*a* subunits. HIF (Hypoxia Inducible Factor) is a heterodimeric transcription factor composed of two subunits, HIF1α and HIF1β. (Percy *et al.*, 2003) Hydroxylation of HIF1*a*, under normal oxygen conditions, on specific prolyl residues in ODD (oxygen-dependent degradation) domains (Pro564 and Pro402) by PHDs generates a binding site for the pVHL (von Hippel-Lindau protein)-ubiquitin E3 ligase, tumour suppressor protein that promotes ubiquitination

Gene Name	Synonyms	Intracellular localisation
EGLN1	*PHD2, HPH-2*	cytoplasmic
EGLN2	*PHD1, HPH-3*	nuclear
EGLN3	*PHD3, HPH-1, SM-20* (Rat)	cytoplasmic and nuclear

Table 3.6 HIF prolyl hydroxylase nomenclature and intracellular localization.

and subsequent proteasomal degradation of HIF1α (del Peso *et al.*, 2003), (Hagg and Wennstrom, 2005) (Figure 3.15). PHDs function as intracellular oxygen sensors due to the fact that the prolyl hydroxylation reaction is oxygen dependent. Under low oxygen conditions, HIFα is not hydroxylated and is translocated to the nucleus, where it dimerises with HIF1β and activates the transcription of HIF target genes such as *GLUT1*, *p53*, *VEGF* and EGFR (Chung et al., 2009; Pietras et al., 2010; Rigopoulos et al., 2010; Sendoel et al., 2010). HIF1β is not oxygen dependent (Rankin and Giaccia, 2008). PHD activity depends not only on the availability of molecular oxygen, but also on the availability of amino acids (Serra-Perez *et al., 2010*). EGLN1 appears to be the primary HIF prolyl hydroxylase under normoxic conditions. EGLN2 and EGLN3 have only partial effects on HIF1α hydroxylation and therefore stability, and it is possible that they are mainly induced under specific conditions like hypoxia.

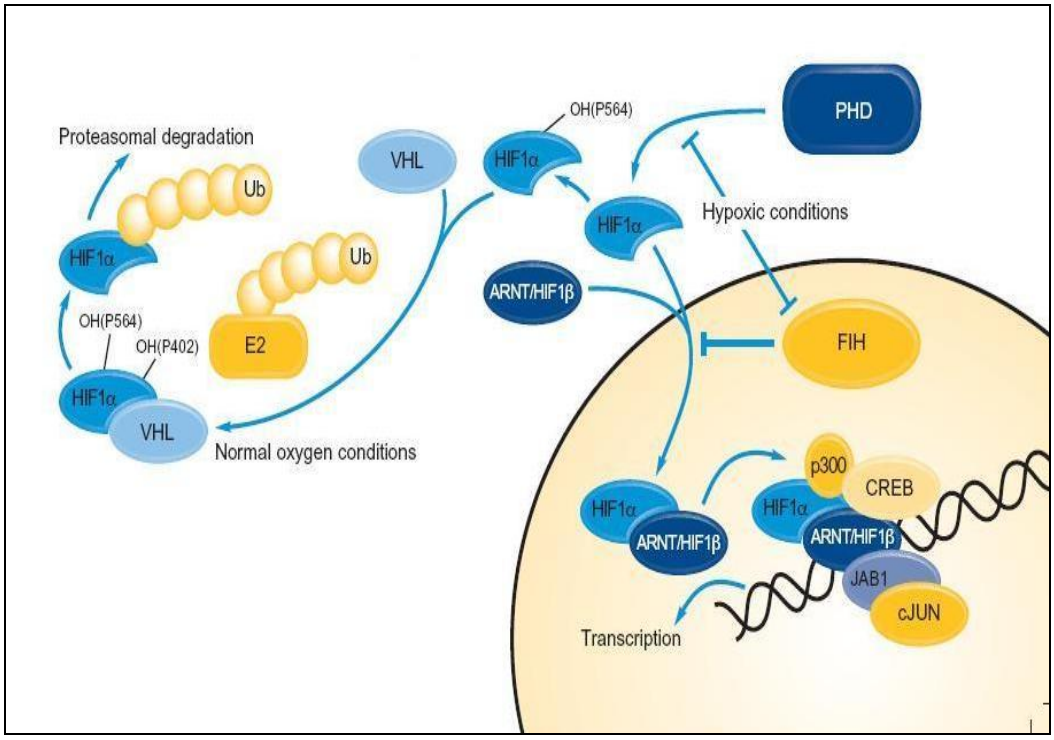


Figure 3.15 Regulation of HIF1α by PHDs and pVHL. Under normal oxygen conditions, HIF1a is hydroxylated on certain prolyl and asparagyl residues by PHDs and FIH (Factor Inhibiting HIF1) respectively, resulting in pVHL binding which leads to the polyubiquitination and proteasomal degradation of HIF1α. In hypoxic conditions, PHD and FIH are inactive and HIF1α is stabilized and translocates into the nucleus where it activates transcription [Figure taken from Abcam].

Amatschek *et al*, measured the levels of *EGLN3* by microarray analysis and RT-PCR in a number of normal tissues and cancers. They established that *EGLN3* is highly expressed in renal cell cancer and lung squamous cell cancers whereas in normal breast tissue and breast cancer it was not expressed, or expressed at very low levels (Amatschek *et al.*, 2004). Finding upregulation of *EGLN3* in our Tamoxifen resistant cell line study was unexpected, but may represent a hypoxia-independent cell survival pathway.

3.7.7. EGLN3 mRNA and protein overexpression in TR cell lines

The microarray study on TR cells (see Figure 3.5 and Table 3.3) had shown that EGLN3 was up-regulated by 3.6 fold compared with wild-type. In order to confirm the microarray data, mRNA levels of *EGLN3* in a number of different wt and TR cell lines were measured by RT-qPCR. *EGLN3* expression showed a range of fold up-regulation in almost all the TR cell lines compared with the TS breast cancer cell lines (Figure 3.16A). Of noteworthy, SKBR3, intrinsically Tamoxifen-resistant, and ER-negative breast cell line, also showed an increase in Egln3 in its TR counterpart. MDA-MD-361, a ER-positive breast cell line had no change in the level of Egln3 protein in its TR counterpart. To validate that *EGLN3* mRNA overexpression correlates with increased expression at the protein level, Western blotting was performed on breast cancer cell lines (Figure 3.16B). Egln3 upregulation was founding each case. In addition the oestrogen deprived cell lines, ODZR and ODT47D also had increase levels of Egln3 protein.

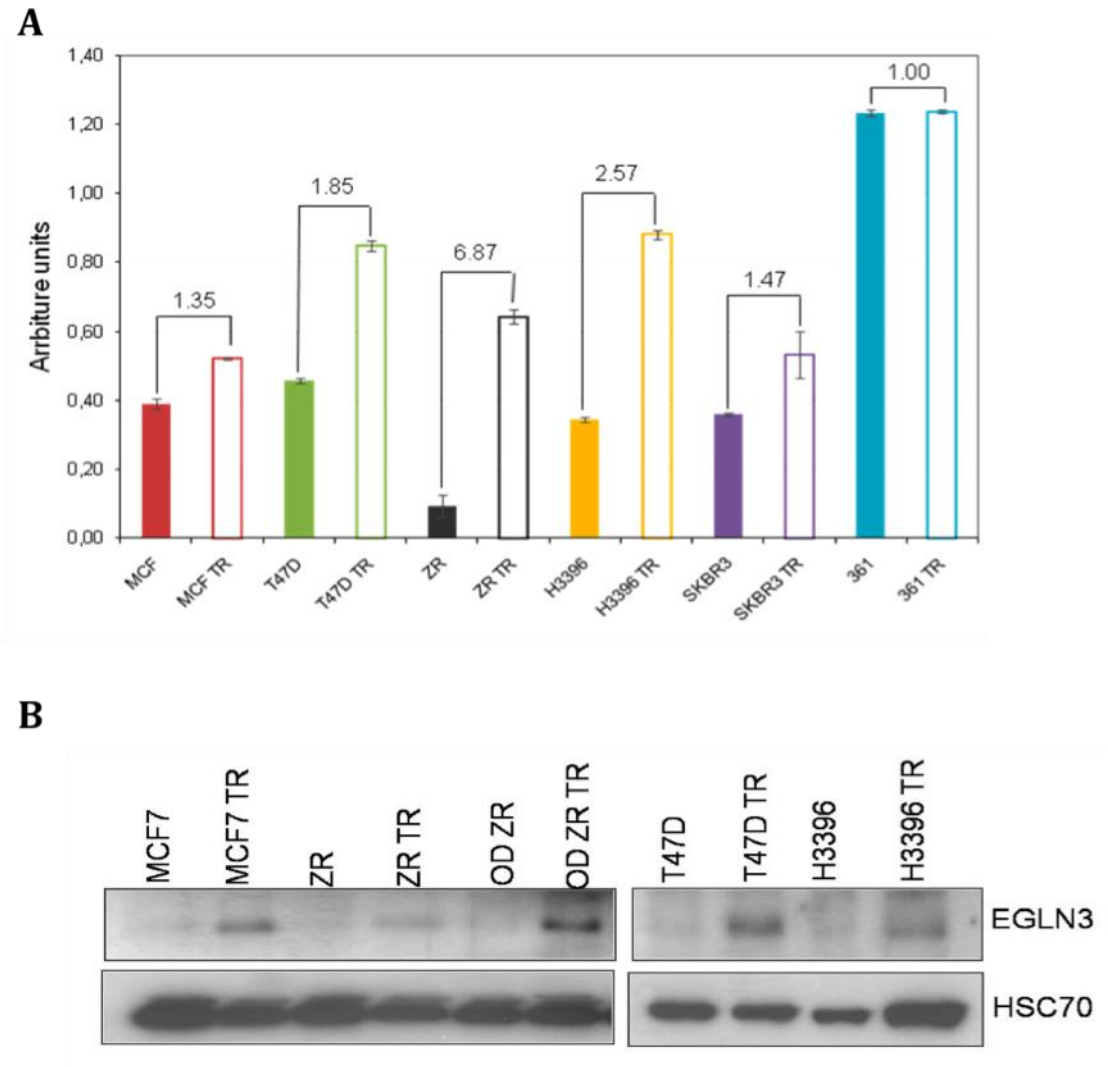


Figure 3.16 EGLN3 overexpression in TR breast cancer cell lines measured and assessed by qPCR and immunoblotting. (A) RNA was extracted from TS and TR derivatives of breast cancer cell lines as indicated. cDNA was prepared and the levels of *EGLN3* and *GAPDH* mRNA levels were quantified by qPCR. All cDNA products were diluted 1:3. Results were analysed using the standard curve method and were normalized against GAPDH levels. TR cells had previously been generated by growing wild type cells in media supplemented with tamoxifen for 6 months. Fold changes in expression between TS and TR lines is indicated for each pair of cell lines.

(B) Whole cell lysate (10 μg) from wt and TR breast cancer cell lines were separated by SDS-PAGE and analysed by Western blotting, probing with antibodies against EGLN3 (24 kDa, the lower band) and HSC70 (70 kDa; loading control) as indicated.

3.7.8. Generation of EGLN3 overexpressing lines

To examine functionally the association of EGLN3 with Tamoxifen resistance in breast cancer, stable T47D and MCF7 EGLN3 overexpressing clones were produced by transfecting wt cells with a pcDNA3.1/EGLN3 expression vector. For control lines I used the pcDNA3.1 transfected vector alone (VA) lines generated previously for the MAGEA2 study.

A number of individual, G418-resistant stable clones were established in each cell line. Before using these clones, they were verified for expression of both EGLN3 mRNA and protein.

RT-qPCR results (Figure 3.17A) show that all but one of the individual clones overexpressed *EGLN3* mRNA compared with the wt and VA controls. For T47D, almost all the clones express similar levels of *EGLN3* except for T47D C15, which was used as an additional negative control. For MCF7, all clones overexpressed *EGLN3* but C3 showed the highest levels compared with all the other clones. Protein levels of EGLN3 were also higher in the clones compared with the wt and VA controls as shown in Figure 3.17B. Furthermore, protein levels were broadly in line with mRNA levels. All the T47D clones showed similar expression and MCF7 C3 clearly expressed greater levels of EGLN3 than the other clones. Subsequently, EGLN3 clones MCF-7 c3 and c7, and T47D c8 and c12 were used for functional study.

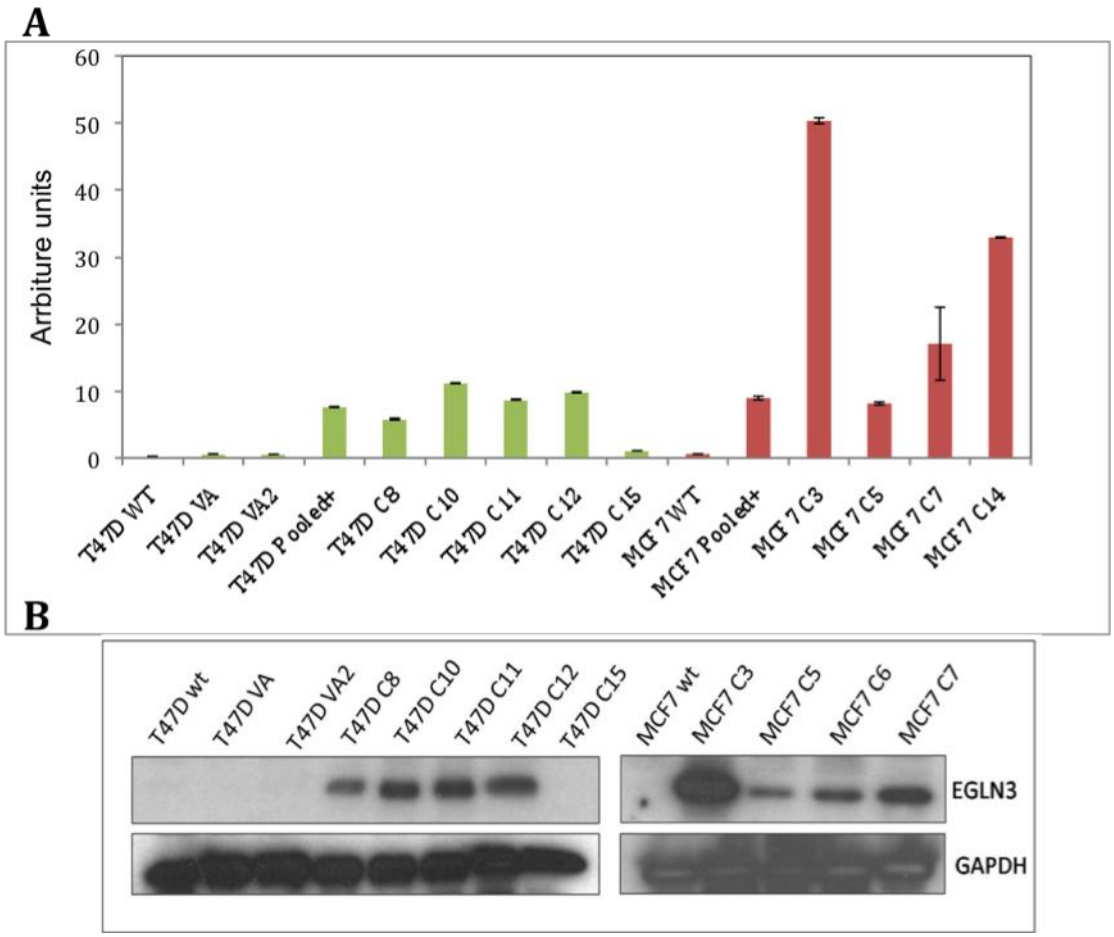


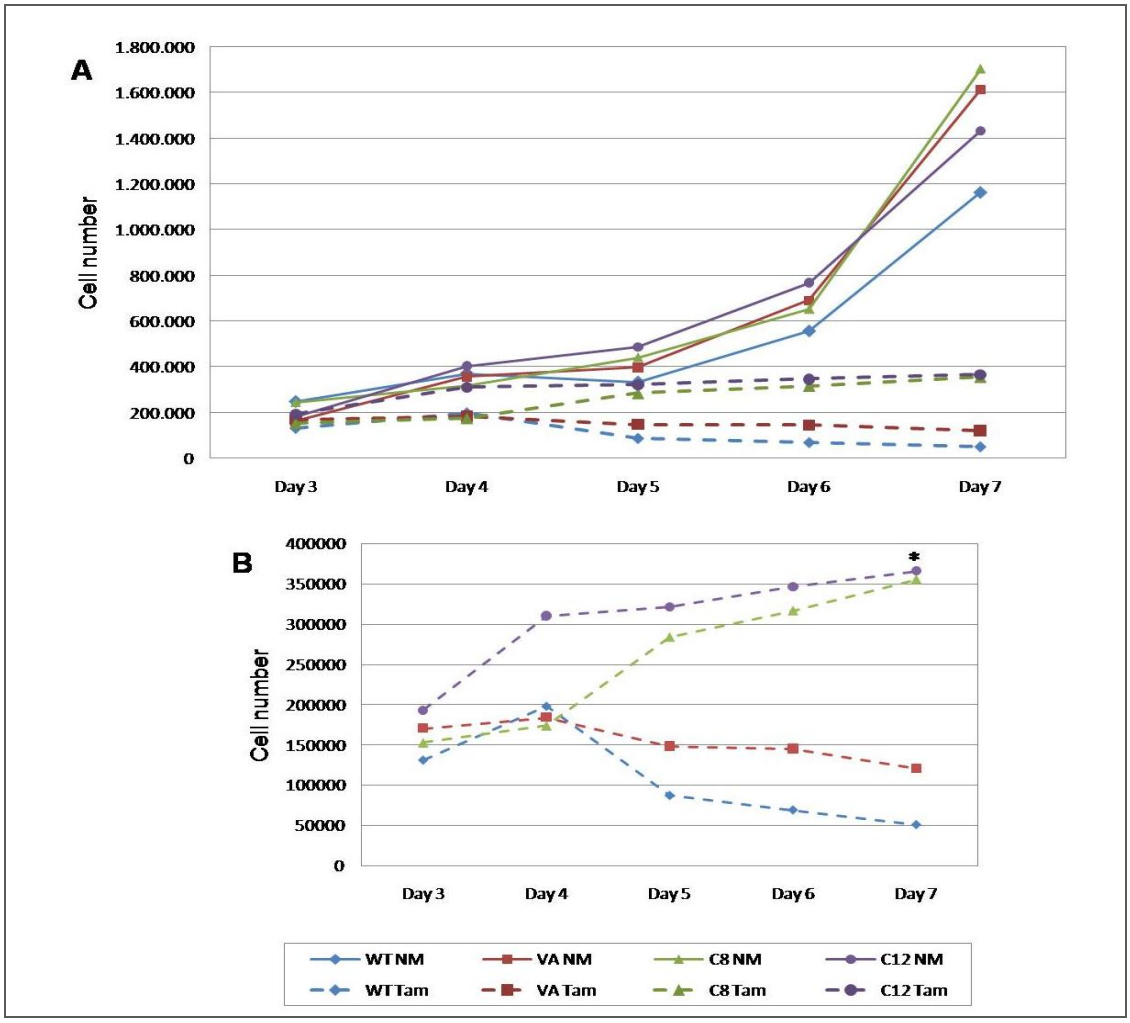
Figure 3.17 ***EGLN3*** **overexpression in T47D and MCF7 clones validated by qPCR and immuno-blotting**. T47D and MCF7 wt cells were transfected with pcDNA3.1/*EGLN3* vector and grown in selective media with G418. (A) cDNA was produced using RNA extracted from wt and VA cells and individual clones numbered as indicated. Levels of *EGLN3* and *GAPDH* mRNA were quantified by qPCR. Wt cells and vector alone (VA) cells were used as negative controls. *EGLN3* levels were calculated using the standard curve method and were normalized against *GAPDH* levels. Pooled + is a pool of positive clones. (B) WCL (20 μg) from T47D and MCF7 wt, VA and *EGLN3*-expressing cells were separated by SDS-PAGE. EGLN3 (24kDa) and GAPDH (32kDa) levels were detected by Western blot using the appropriate antibodies as described in Materials and Methods. Both wt cells and VA clones represent negative controls and GAPDH was used as a loading control.

3.7.9. T47D and MCF7 EGLN3 positive clones are less sensitive to tamoxifen

In order to characterize the clones that overexpress *EGLN3* and also to see how Tamoxifen affects their viability and proliferation, growth curves were produced. T47D and MCF7 wt cells and VA controls plus two *EGLN3* clones for each line were grown in normal and Tamoxifen-containing media and the number of cells was counted daily over 7 days.

For T47D cells, clones and controls proliferated at a similar rate in normal media (Figure 3.18A). However, in Tamoxifen-containing media, while the control cells declined in number over the 7 day time course, the *EGLN3* expressing clones continued to proliferate although at a much slower rate than in normal media (Figure 3.18B). This proliferation advantage reached significance in both *EGLN3* clones (p-value ≤ 0.05).

Slightly different results were observed for MCF7 cells where the *EGLN3*-expressing clones appeared to have a growth advantage in both normal and Tamoxifen containing media (Figure 3.19A). Although this is significant in both normal and Tamoxifen containing media, only in the presence of Tamoxifen is the p-value ≤ 0.01.



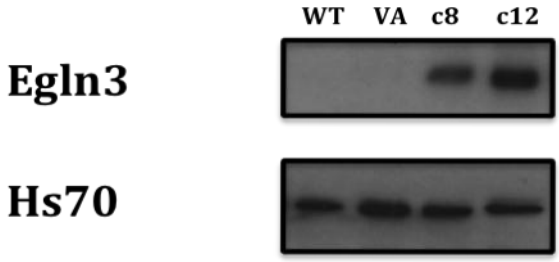
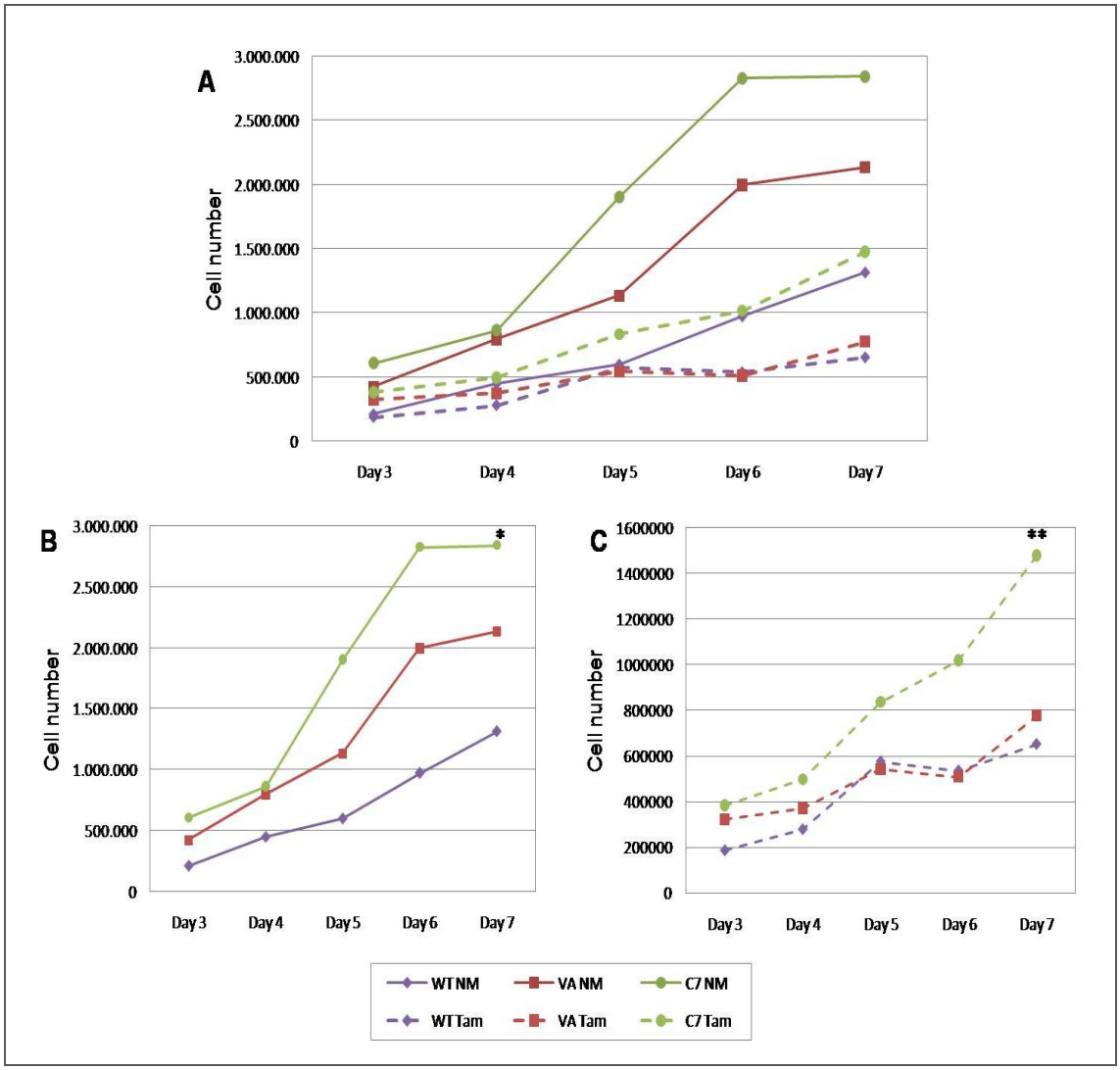


Figure 3.18 ***EGLN3*** **expressing T47D cells have a proliferation advantage in Tamoxifen-containing media.** 75,000 cells of wt, VA, *EGLN3* C8 and C12 were plated in 6 well plates in triplicate for each condition and grown in normal media, (NM) or treated with 10^{-7}M tamoxifen. Cells were harvested on days 3-7, trypsinised and counted (CASY counter). (**A**) Combined results of cell number for T47D cells in NM and Tam media from day 3 to day 7. (**B)** Expanded graph of cells grown in Tamoxifen from A to show differential growth of *EGLN3* clones compared to controls. *P≤0.05. (C) Western blot of 20μg/lane probe for Egln3 and Hs70 loading control as indicated.



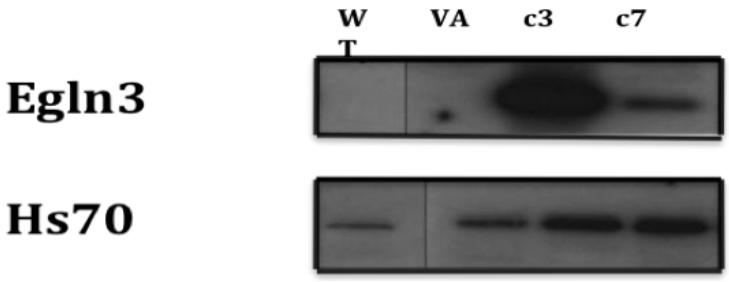


Figure 3.19 Proliferation advantage of MCF7 *EGLN3* clones. MCF7 wt, VA and *EGLN3* C7 cells were plated as described in Figure 3.1 and counted. (**A)** Combined results of cell number from day 3 to 7 in normal and tamoxifen containing media. (**B)** Expanded graph for proliferation in normal media * $p \leq 0.05$. (**C)** Expanded graph for proliferation in tamoxifen-containing media ** $p \leq 0.01$. (D) Western blot of 20µg/lane probed for Egln3 and Hs70 loading control as indicated.

In order to study whether overexpression of *EGLN3* affects cell viability and survival, the number of necrotic and apoptotic cells after incubation with Tamoxifen was determined using Annexin V and Propodium Iodide (PI) staining as used

previously to study MAGEA-expressing cells. Annexin V binds specifically to phosphatidylserine (PS), which is translocated from the inner membrane during early apoptosis. PI is used to distinguish apoptotic from necrotic cells. Necrotic cells are characterized by permeabilisation of the cell membrane (lysis) and therefore they can be detected by the DNA binding dye PI. Annexin V in conjunction with PI is used to differentiate live, apoptotic and necrotic cells.

For T47D wt and VA cells the proportion of live cells after incubation for 24 or 48h in tamoxifen was lower, and the proportion of apoptotic cells had increased compared to cells maintained in normal media. In contrast, for both *EGLN3-expressing* clones, C8 and C12 the number of live cells (light green columns) was higher in tamoxifen-containing media compared to normal media, further confirming that *EGLN3*-expressing cells have an advantage in tamoxifen-containing media (Figure 3.20A). The results for MCF7 derived lines were very similar to those for T47D (Figure 3.20B).

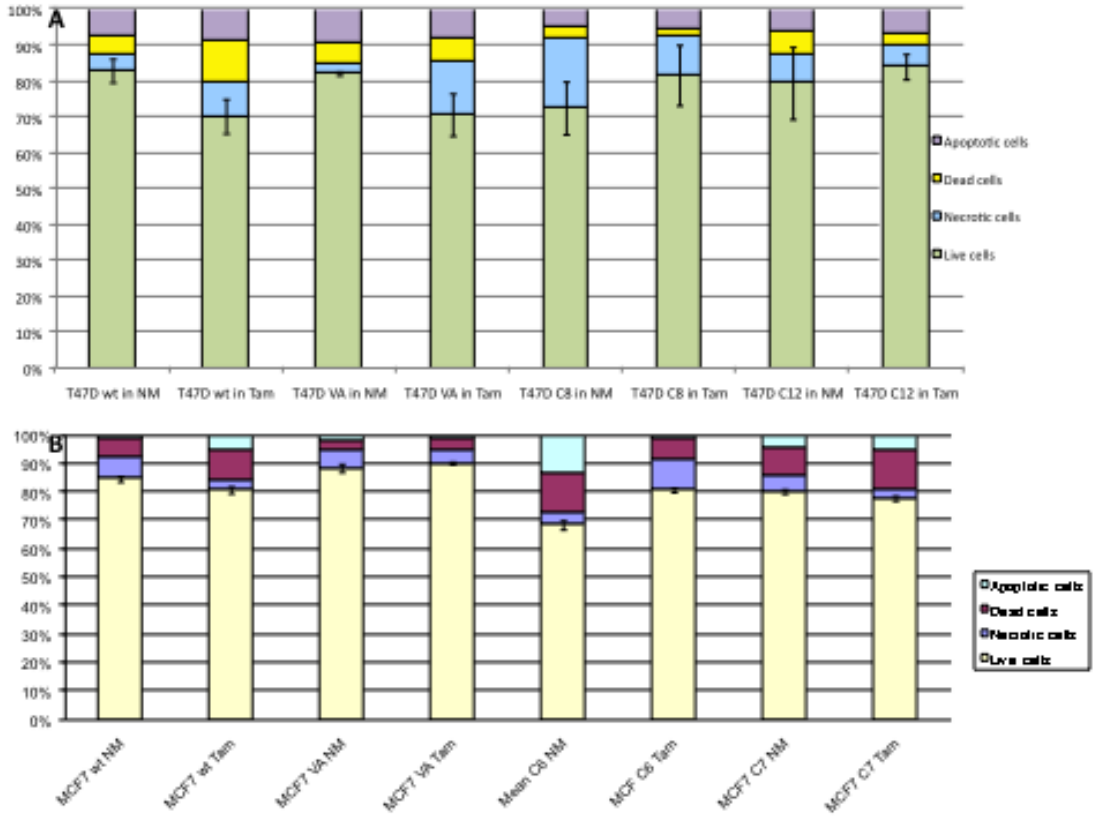


Figure 3.20 ***EGLN3*** **overexpression results in an increased proportion of live cells in Tamoxifen media.** T47D and MCF7 wt, VA and *EGLN3* clones were plated in 6-well plates at 200,000 cells/well and incubated for 24 and 48 hours in normal (NM) or Tamoxifen containing media (TAM). At harvest, cells were stained with Annexin V and PI and then were analysed by Flow Cytometry (see Chapter 2.11.3). In this assay live cells are non-fluorescent, apoptotic cells stain only with Annexin V and necrotic cells stain with both Annexin V and PI. All the incubations were done in triplicate. (**A)** T47D cells, 24h incubation. (**B)** MCF7 cells, 24h incubation.

3.7.10. Localisation of EGLN3 expression in the presence and absence of Tamoxifen

I wanted to see if the localization of Egln3 differs between VA from the EGLN3-overexpressing clones in both cell lines. Immunoflourescent study was undertaken after probing the cells with Egln3 antibody and other potentially associated proteins (as postulated by publications), such as HIF1α, phosphorylated-Rb and also MageA2. The later being a random screening as we were also working on MageA2 at the same time, and the ingenuity pathway analysis suggested that the two pathways (MageA2 and Egln3) might be linked.

As seen from Figure 3.21, EGLN3 is overexpressed in Tamoxifen resistant cells. Immunofluorescence microscopy was used for the localization of EGLN3. By using the same method of antibody staining, HIF1α, MAGEA2 and pRb were localized. Furthermore, investigated the effect of EGLN3 overexpression on these proteins. The same set of cells, MCF-7 VA and C7, and T47D VA and C12 (C7 and C12 are positive-EGLN3 clones respectively) were stained in their normal media and the Tamoxifen containing media.

T47D and MCF7 VA and *EGLN3* cloned cells upregulated EGLN3 and MAGEA2 when they were treated with Tamoxifen for 24h. This result can be compared with the protein levels found by Western blot in (Figure 3.16). EGLN3 is expressed in both the cytoplasm and the nucleus. HIF1*a* is not expressed or is expressed at very low concentrations in all the T47D cells. In fact, it is down-regulated in Egln3 clones (C12, in Normal Media, NM) as seen in Figure 3.22 and at higher resolution of Figure 3.20.

In the T47D and MCF-7 VA, the up-regulation of MAGEA2 is most distinct when exposed to Tamoxifen containing media. The positive T47D and MCF-7 EGLN3 clones have shown up-regulation of MAGEA2 but to a lesser extend.
Egln3 protein co-localised with phos-Rb protein as seen in both Figure 3.21 and 3.22.

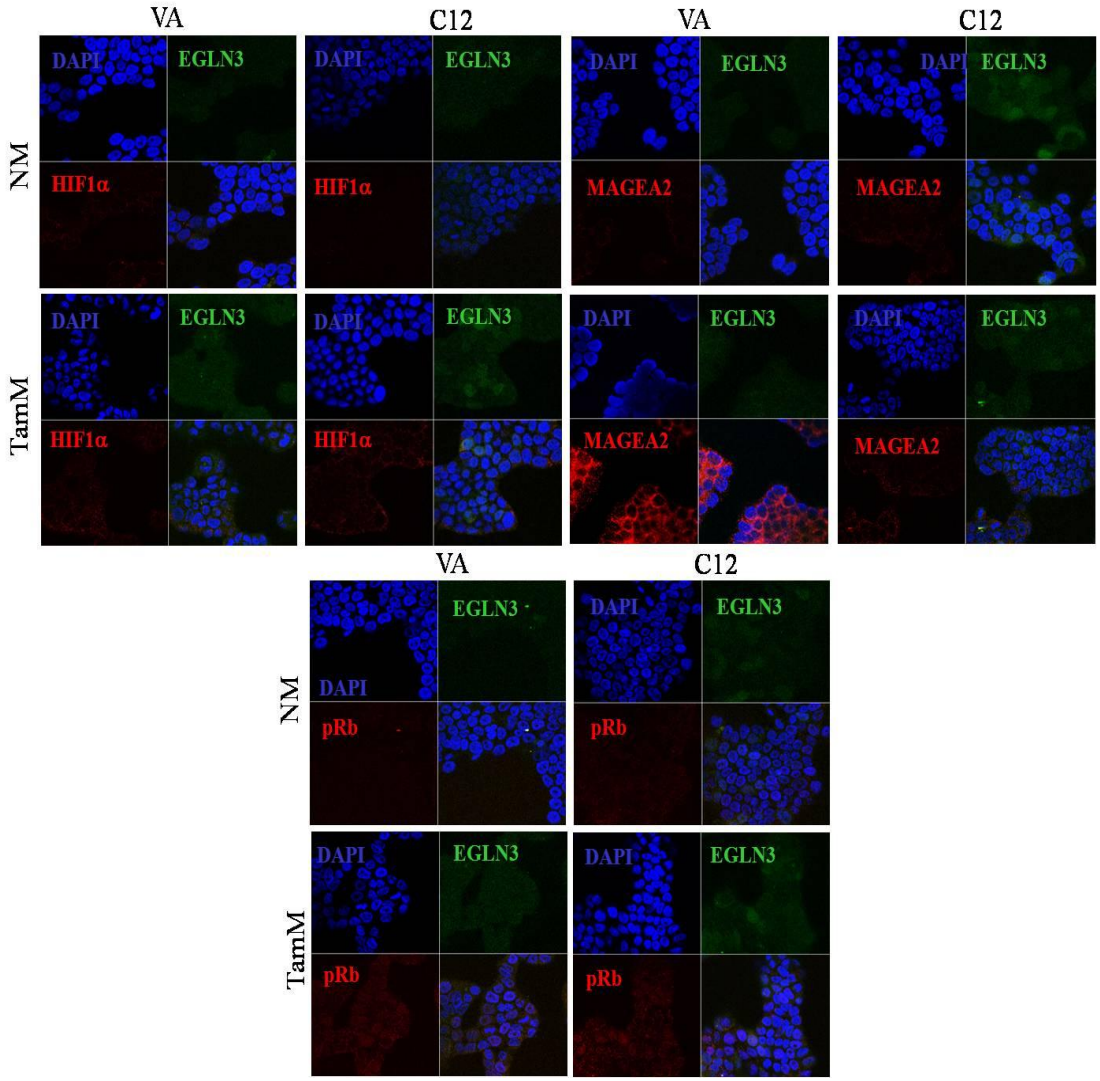


Figure 3.21 EGLN3 localisation in T47D cells and overexpression when cells are treated with Tam. Four million cells (T47D VA and T47D C12) were seeded on 9cm plate containing five sterile glass coverslips and incubated for 24 hours in normal or tamoxifen containing media, fixed with paraformaldehyde, permeabilised, blocked, probed with primary antibody overnight and then with secondary antibody. Coverslips were then stained with DAPI and examined using the immunofluorescence microscope (Confocal). DAPI is stained blue (nuclei), mouse Ab green (EGLN3) and rabbit Ab red (HIF1α, MAGEA2, pRb). NM=normal media, TamM=Tamoxifen containing media.

Differentially, in MCF7 cells, HIF1α is expressed at higher levels and it seems that is coexpressed with EGLN3, with a slight predominance in the cytoplasm. Finally, pRb is coexpressed and co-localised with EGLN3 especially in cells treated with tamoxifen (Figures 3.19 and 3.20).

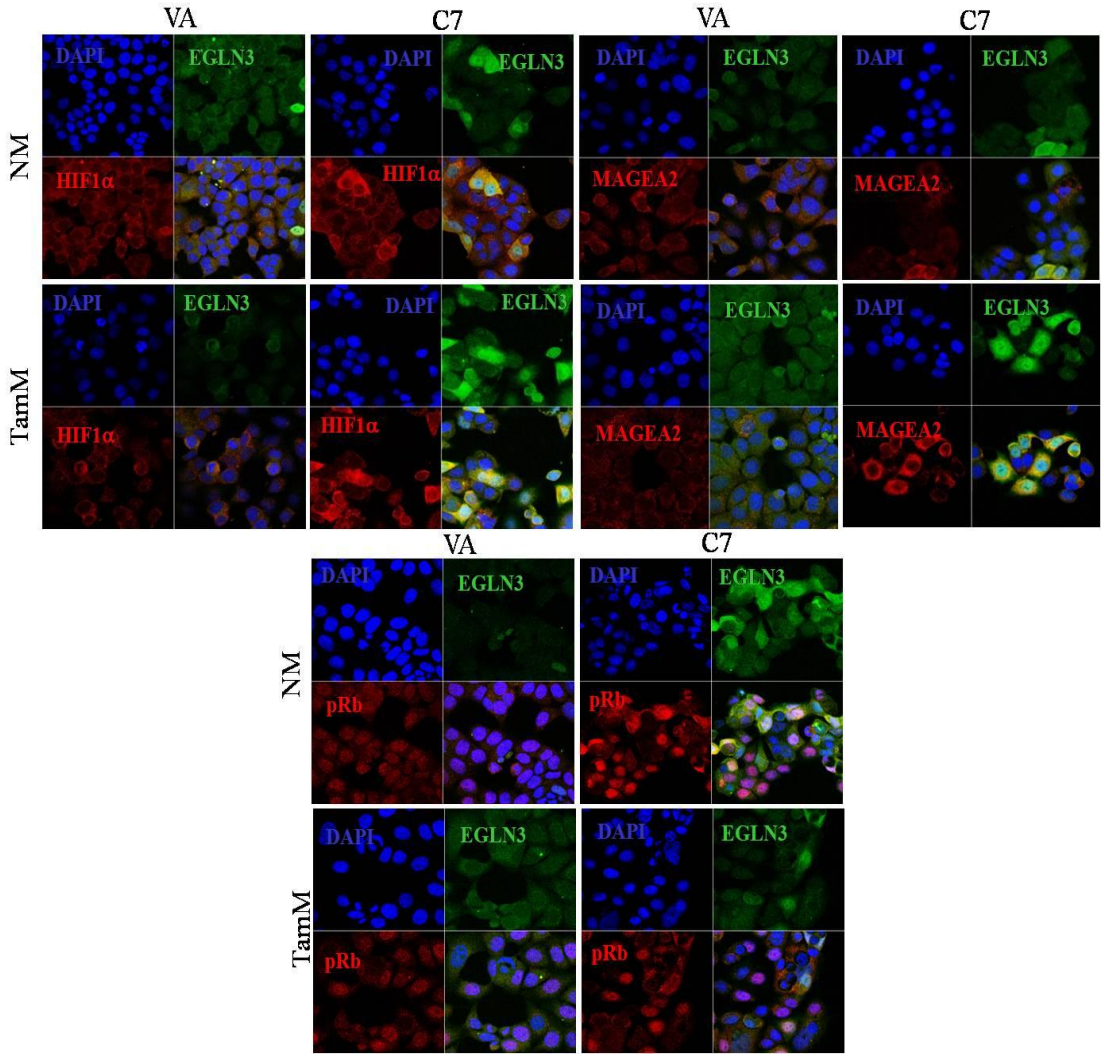


Figure 3.22 EGLN3 localisation and overexpression in MCF7 cells. MCF7 VA and C7 cells were prepared as described in Figure 3.8 for the T47D cells, probed with the same antibodies and examined in the immunofluorescence microscope (Confocal).

3.7.11. Does EGLN3 expression alter levels of hypoxia-associated proteins?

In order to look for a connection between EGLN3 expression and other molecules associated with the hypoxic response that may also be implicated in tamoxifen resistance we examined a number of proteins by Western blotting including HIF1α, HIF1β (ARNT), acetyl-p53, pAkt, BNIP3 and GLUT1. As described in the introduction, the EGLN3 family of proteins is mainly implicated in the hypoxia pathway since HIF1α is their most well defined substrate. Moreover, EGLN3 is normally a HIF1α target gene during the response to hypoxia. For this reason I first looked at the proteins that are main players in that pathway, HIF1α and HIF1β. In T47D clones that overexpressed EGLN3 there was a decrease in HIF1α levels compared to controls. HIF1β and GLUT1 (another HIF1α target gene) protein levels were also examined in both cell lines. GLUT1 levels were constant in all lines while HIF1β levels appeared to be increased in the EGLN3 clones compared to most of the controls (Figure 3.23).

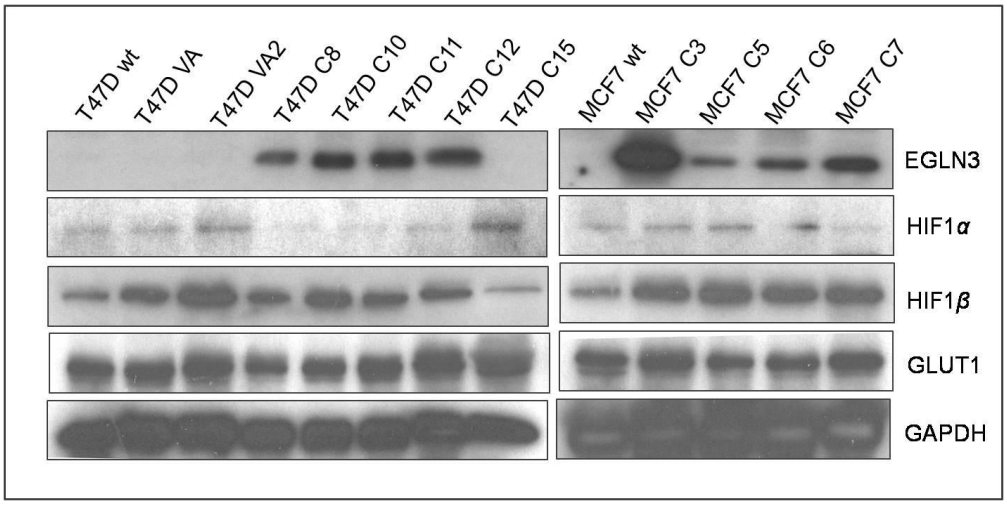


Figure 3.23 Expression of hypoxia pathway proteins in EGLN3 lines. WCE (20μg) from T47D and MCF7 wt, VA and EGLN3 overexpressing cells were separated by SDS-PAGE. EGLN3 (24kDa), HIF1α (120kDa), HIF1β (87kDa), GLUT1 (55kDa) and GAPDH (32kDa) levels were detected by Western blot using the respective antibodies as indicated. Both wt cells and VA clones were used as negative controls. T47D C15 did not express EGLN3 (see Figure 3.15) and was used as an extra negative control. GAPDH was used as a loading control.

Figure 3.24 Expression of acetyl-p53, pAkt and Raf in EGLN3 clones. WCE were examined by Western blot as described in Figure 3.25 for acetyl-p53 (53kDa), pAkt (60kDa), Raf (68kDa) and GAPDH (32kDa) proteins. GAPDH was used as loading control.

HIFα is a usually is diverted to ubiquitinate with VHL protein by EGLN1 and 2 (or also known as PHD 1 and PHD 2) and subsequently degraded. We expected EGLN3 to decrease HIFα as well. But with the HIFα availability reduced, the HIF transcription factor should also be deactivated, or not activated. Hence, what alternative pathway is actually stimulating cell proliferation, in the absence of HIF-transcription, p53, P13K/Akt and Ras/Raf pathways were studied in our VA and EGLN3-overexpressing clones for the reasons that we were looking for alternative pathways (Figure 3.24) which may be responsible for the cell survival advantage.

For MCF-7, HIF1α protein is not down-regulated at all in EGLN3-overexpression in the clones. As stated in Introduction Chapter, a number of cellular oncogenes have been reported to promote HIF stabilization, which in turn might facilitate solid tumour growth. Some oncogenes, such as activated Ras, block HIF prolyl hydroxylation and thereby promote HIF accumulation (Chan *et al.*, 2002). In contrast PI3K/AKT can promote HIF accumulation without an apparent change in HIF hydroxylation, possibly through activation of mTOR and HIF hydroxylation. There is a marginal increase in pAkt in the both the T47D and MCF-7 EGLN3 expressing clones compared with VA.

P53, PI3K/Akt and Ras/Raf pathway proteins are all implicated in cell proliferation, survival signaling and resistant to Tamoxifen, and therefore might play and important role in Tamoxifen resistant role caused by EGLN3 overexpression. In my experiments, I have not found substantial connection between these pathways with the over-expression of EGLN3.

3.7.12. Investigating if MAGEA2 is downstream from EGLN3 and if it is responsible for cell proliferation in Tamoxifen Resistance

In some experiments where samples from EGLN3-expressing clones had been examined together with MAGEA2 clones, I noted that MageA2 levels seemed to be higher in Egln3 clones than the controls. Interestingly while MAGEA2 was not expressed in wt or VA cells in either cell line as expected it was overexpressed in all the EGLN3 positive clones in both MCF-7 and T47D derived cell lines (Figure 3.25). Subsequently the MAGEA2 clones were examined for EGLN3 expression by qPCR but EGLN3 was not detected in any of the MAGEA2 expressing lines (data not shown).

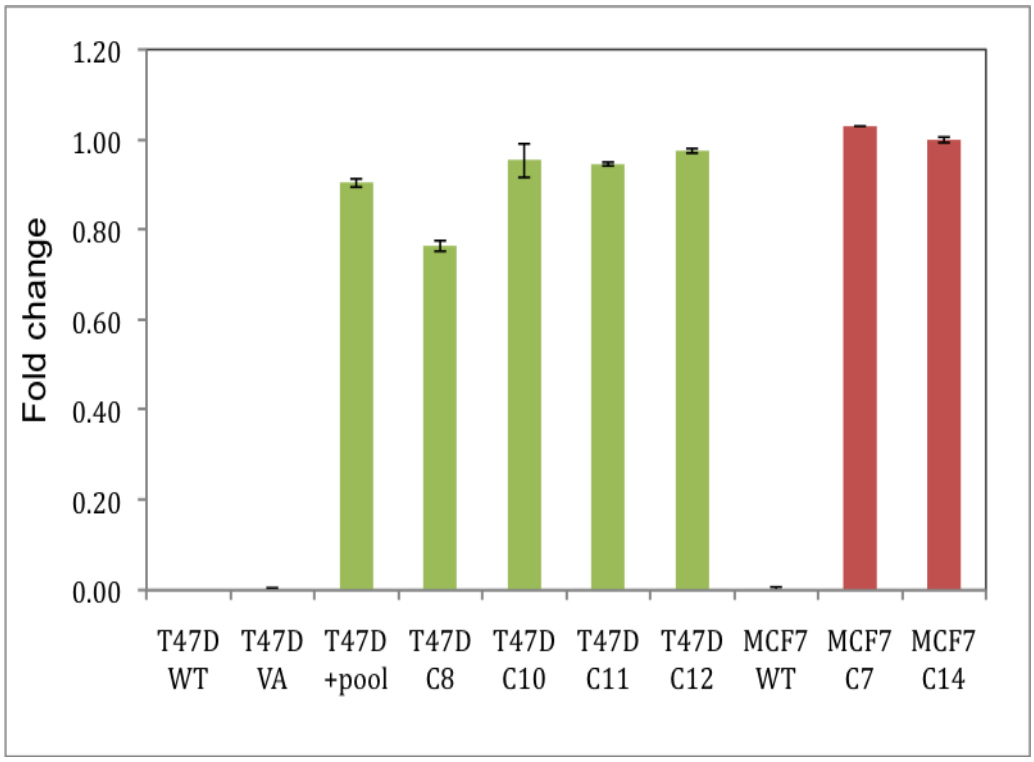


Figure 3.25 ***MAGEA2*** **overexpression in** ***EGLN3*** **positive clones.** cDNA was prepared from RNA extracted from T47D and MCF7 wt, VA and *EGLN3* clones as indicated and the level of *MAGEA2* and *GAPDH* mRNA levels were quantified by qPCR (see Chapter 2.9.2). All cDNA products were diluted 1:3. Results were analysed using the ΔCt method and were normalized against *GAPDH* levels. + pool is a pool of *EGLN3* positive T47D clones.

As I found there was an up-regulation of mRNA of MAGEA2 in all the EGLN3 clones (see Figure 3.25), but no up-regulation of EGLN3 in MAGEA2 positive clones, I set out to test if MageA2 is expressed downstream of Egln3 or if MageA2 is up-regulated concurrently with Egln3 (indirect). I used the Smart pool-Dharmacon siRNA (4 siRNA: Genome and on-Target; L-004274-00-0005) for Egln3 knockdown and achieved 90% protein knockdown (Figure 3.26). I immunoblotted for MageA2 with the same lysate, which showed no knockdown. A proliferation study was undertaken using the suphorhodamine assay (SRB) to evaluate the effect of the clones on cell proliferation in normal media and tamoxifen media. This method relies on the uptake of the negatively charged pink aminoxanthine dye, SRB by basic amino acids in the cells. The greater the number of cells, the greater the amount of dye is taken up. After fixing, when the cells are lysed, the released dye will give a more intense colour and greater absorbance when measured by spectrophotometer. The benefit of this assay is that viable cells can be quantitated within 12 hr to 72 hrs. As Egln3 has been shown to support cell survival, in the knockdown of Egln3 we expected the reverse finding.

The knockdown of Egln3 in positive Egln3 clones (rescue knockdown) cells were less in numbers and by SRB assay as shown in our graph in Figure 3.26. However, MageA2 is up-regulated in siRNA (Egln3) cells compared with non-silenced cells, which suggest that althought MageA2 is found to be up-regulated in mRNA levels in Egln3 positive clones, the knockdown of Egln3 did not reverse the MageA2 levels at 48 hr. We conclude that siRNA of Egln3 in positive Egln3 clones resulted in less cell viability and MageA2 is not downstream of Egln3 expression.

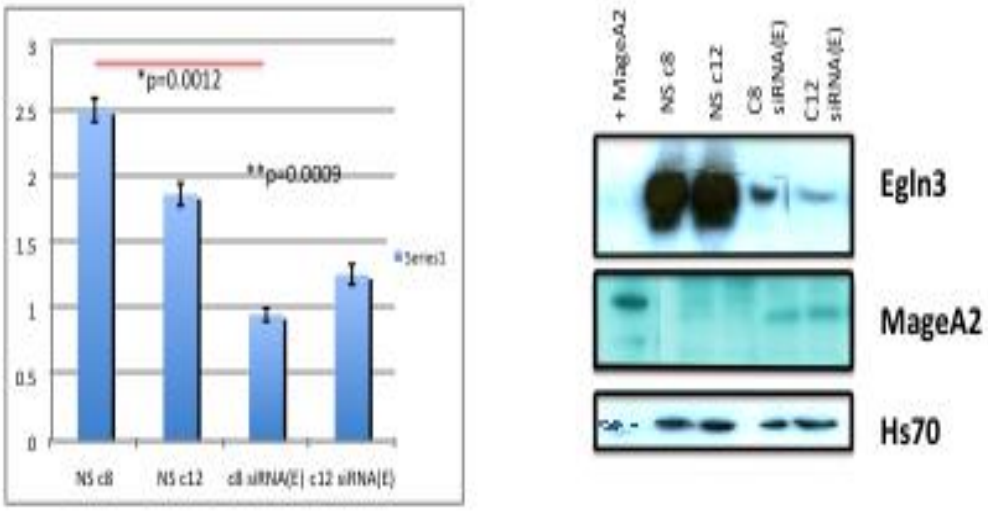


Figure 3.26 Rescue knockdown of Egln3 in Egln3-positive T47D cell lines resulted in decrease cell viability within 48 hrs. (A) Graph showed that T47D Egln3-overexpressing clones c8 and c12 which had successful Egln3-siRNA knockdown was significantly less viable than non-silencing (control) c8 and c12, with a p-value of 0.0012 and p 0.0009 for c8, and c12 respectively, by suphorhodamine (SRB) assay. The experiment design included two positive Egln3 clones, c8 and c12 in T47D. Cells were seeded at a density of 5 x 10^3 cells/well in 100 µl in 96-well plate in triplicates. Cells were then transfected with either non-silencing vector (All star) or by siRNA EGLN3 as detailed above. Cells were incubated for 48 hrs after transfection with AMAXA Nucleofection. Then read at 492 nM with a spectrometer.

In rescue knockdown of Egln3, MageA2 protein expression is retained or even increased, therefore expression of MageA2 is not a consequence of Egln3 expression. (B) Immunoblot of 30µg of WCL (whole cell lysate) of the non-silencing (NS) c8, NS c12, with siRNA (Egln3) c8, siRNA (Egln3) c12, the far left is a postive control for MageA2. Hs70 was used as a loading control.

3.7.13. Immunohistochemistry study of Egln3-staining

3.7.13.1. Survival analyses for positive EGLN3-staining in primary breast cancer tissue

Tissue samples of the same cohort of 129 patients as used for the MAGEA2 immunohistochemistry study, were stained for Egln3 using the same immunohistochemistry protocol. The slides were scored by myself and an independent trained pathologist, Dr Yaohe Wang. The antibody was optimised on a paraffin block made of MCF-7 EGLN3-expressing clone (clone 7), and tonsil as recommended by the protocol. The dilution were optimised by testing it against 1:25, 1:50, 1:100 and 1:200. I used 1:300 for primary antibody and 1:200 goat anti-rabbit secondary antibody.

The histopathology scores were calculated using an equation which combined the intensity of staining with that of percentage of tumour cells involved. A final score of either 0, 25, 50, 75 or 100% multiplied by 1, 2 or 3 (according to intensity) was calculated. The highest possible score was 300, while the lowest score was 0. Most slides were positively stained and had an intensity of 1, with areas of 25%. A cut-off of 150 was used as a positive EGLN3 score.

Using the positive EGLN3 score cases, I analysed the disease-free survival (DFS) and overall survival (OS) using Prism (Version 6.0). Analysis of outcome data was based on information received as of March 2009. Kaplan-Meier curves were used to quantify the values of disease free and overall survival. This software calculated the median days of survival, and also the p-value between the positively stained and the negatively stained cases using the Gehan-Breslow-Wilcoxon test.

There was no significant difference between the positively stained and the negatively stained tissue samples in either DFS or OS as shown in Figure 3.27. Moreover, EGLN3 appeared to be present in almost ever slide, in every type of cell. However it was much more intense in the Tamoxifen-resistant group, and breast carcinoma cells. This suggests that it may be a protein which cells require for sustainance but is expressed in abundance in TR cell type.

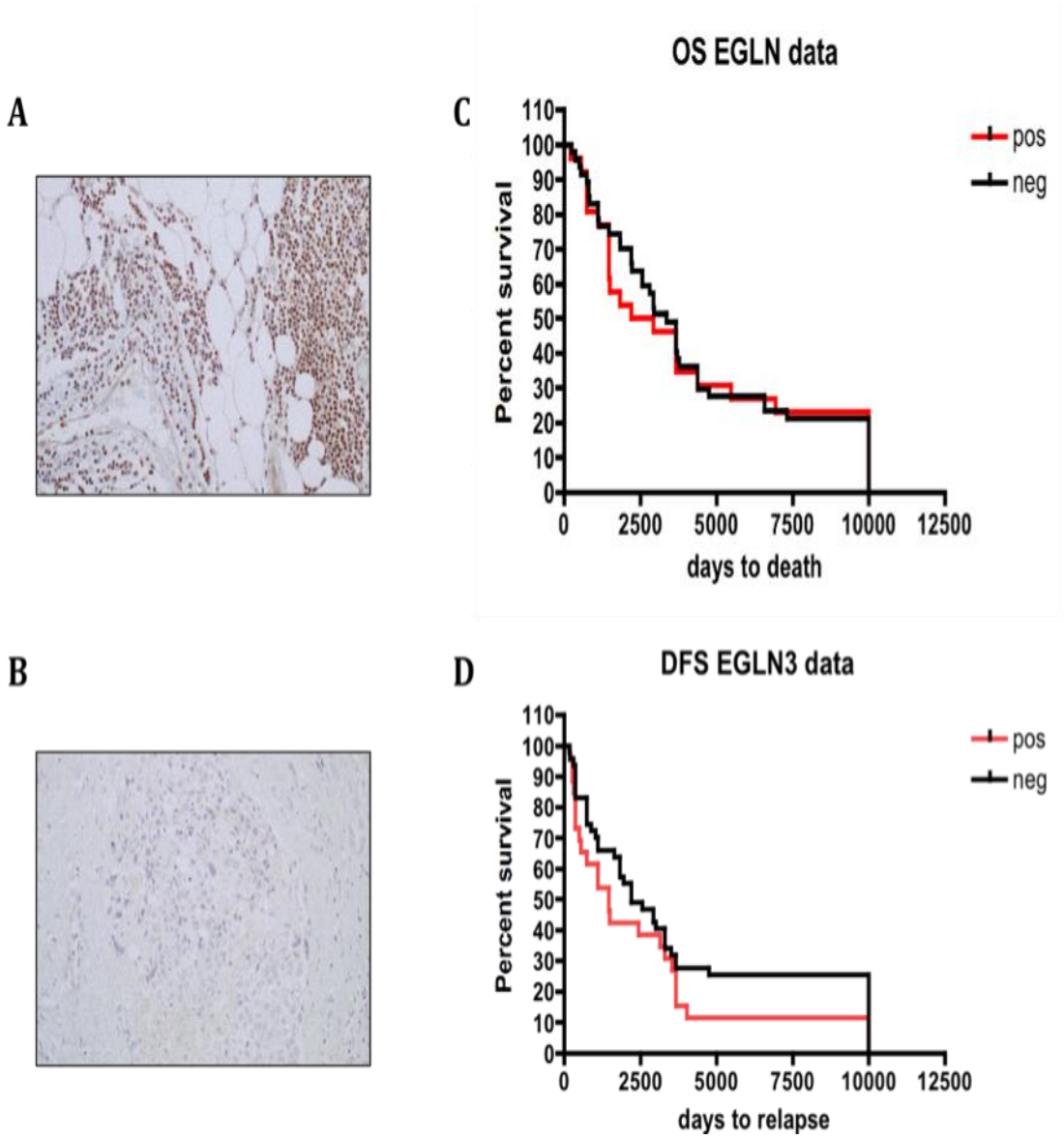


Figure 3.27 No significant difference in overall (OS) or disease free survival (DFS) between positive and negative EGLN3 samples. Immunohistochemistry staining of 1:300 dilution of Egln3 (Ab30782) using the automatic Ventana machine (see Chapter 2.14). (A) Positive EGLN3-staining on the breast cancer primary tissue. (B) Negative EGLN3-staining. (C) Using the Gehan-Breslow-Wilcoxon equation,the OS was not significant between the positive EGLN3-stain from the negative EGLN3-stain with a p value of 0.5748, with a HR=1.058 (95% CI; 0.6022-1.858). (D) There was no significant in DFS between the positive EGLN3-stan from the negative EGLN3-stain, with a p-value of 0.1816; HR=1.493.

When the data were analysed by an independent statistical faculty (Wolfson Institute, under J Cuzack), Egln3 as a biomarker was found have a sensitivity of 42.8% in Tamoxifen resistant primary breast cancer tissue, with a specificity of 66%, with a diagnostic odd ratio of 1.5. This means that Egln3 has greater negative predictive value (85%) than it does as a positive predictor, i.e. a negatively stained TR primary breast cancer tissue is highly unlikely to be TR. The table below gives the results of formal statistical analyses by an independent statistic institute.

	Egln3		
Calculation of	**Estimated Value**	**Lower CI**	**Upper CI**
Apparent prevalence	0.4111675	0.3417272	0.483323
True prevalence	0.8172589	0.756098	0.8686062
Sensitivity	0.4285714	0.3509848	0.5088368
Specificity	0.6666667	0.4902975	0.8144382
Diagnostic accuracy	0.4720812	0.4007205	0.5442947
Diagnostic odds ratio	1.5	0.7014947	3.207437
Youden's index	0.0952381	-0.1587177	0.323275
Positive predictive value	0.8518519	0.755511	0.921038
Negative predictive value	0.2068966	0.1372891	0.2920237
Positive likelihood ratio	1.285714	0.7835736	2.109644
Negative likelihood ratio	0.8571429	0.6563409	1.119379

Table 3.7 Statistical analysis showed Egln3 as a stand alone biomarker had a good sensitivity (42%) but a poor specificity (66%) for predicting Tamoxifen-resistant (TR) in primary breast cancer tissue. The negative likelihoofd ratio of 85% is high, which indicate that a negative Egln3 is likely to be a true negative, hence a high possibility of the patient being Tamoxifen sensitive (TS).

3.7.13.2. Survival analysis for combined MageA2 and Egln3-staining in primary breast cancer tissue

As MageA2 was incidentally found in EGLN3 overexpressing clones in MCF-7 and T47D breast cell lines, we decided we will analyse the survival of the cohort who had positive staining of both MageA2 and Egln3. Eighteen of the 129 patients were MageA2 and Egln3 positive. This made up 18 out of 72 total Tamoxifen Resistant. This supported a sensitivity rate of 25% of the TR tissue samples.

When survival analysis was carried out (Figure 3.28), the combined MageA2 and Egln3 positive stained tissue of the patients had a trend for poorer prognosis in DFS and OS but did not reached statistical significance. The p-value 0.239 between the double positive staining with the patients with double negative staining (HR=0.4041; 95%CI: -0.09to 0.89). The DFS graphs suggested that the most distinct difference in the double-positive staining is seen within the first 4000 days from day of diagnosis, but the four lines soon converge on longer follow-up.

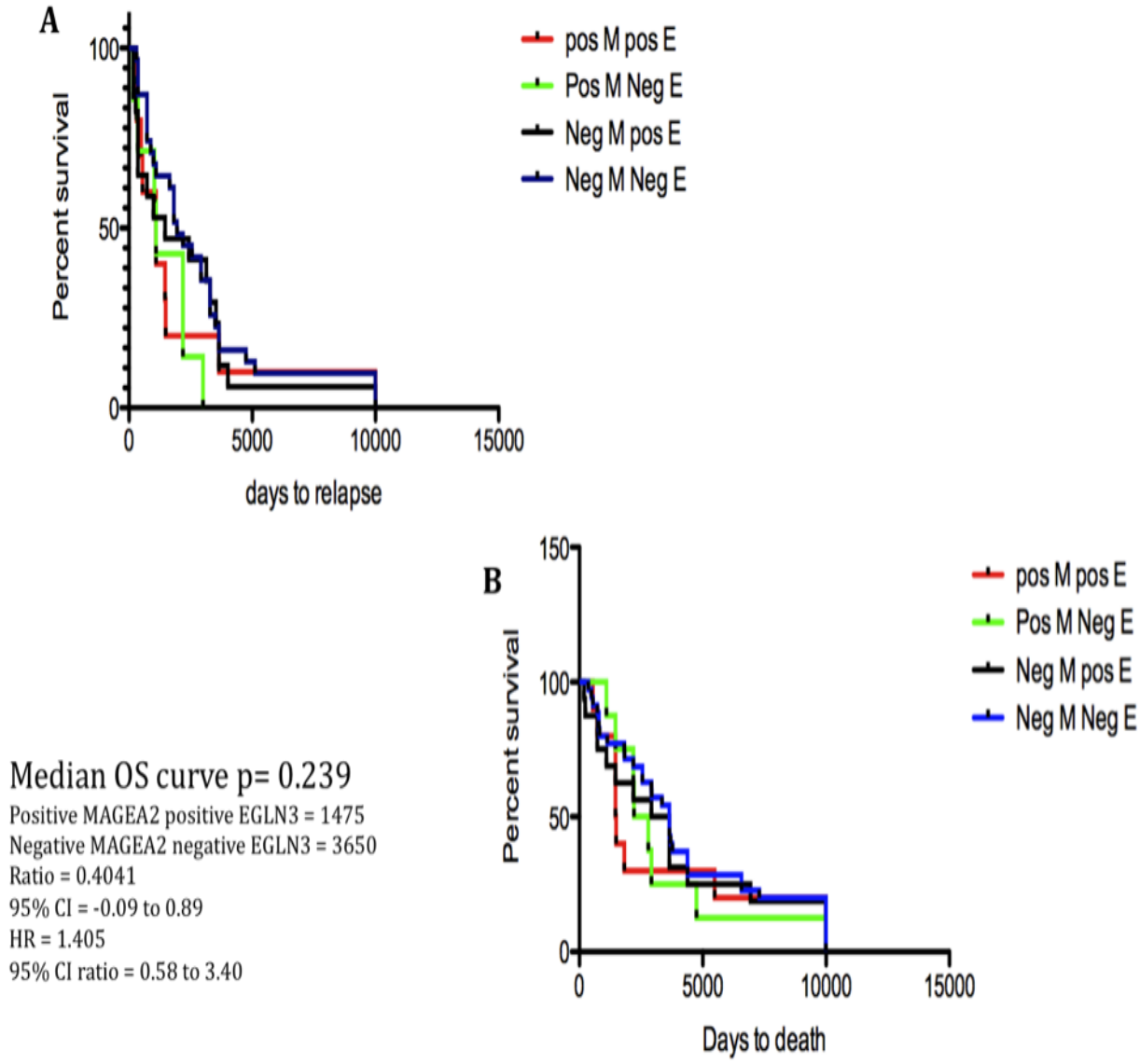


Figure 3.28 (A and B) Double positive-staining (MageA2 and Egln3) in human tissue (n=122) had a trend towards a poorer prognosis in DFS and OS, but this did not reach statistical significance. (A) Disease free survival graph depicting the survival of the four subgroups; Red=combined positive MageA2 and Egln3 staining, Green=Positive mageA2, and negative Egln3 staining, Black=Negative MageA2, and positive Egln3, Blue=Double negative MageA2 and Egln3. There is a trend for the double positive staining (red) to predict for poorer prognosis, but this did not reach statistical significance. (B) Overall survival graph showed there was a trend for double positive staining to predict worse OS, but this did not reach statistical significance.

CHAPTER 4: ANALYSIS OF BREAST CANCER HUMAN TISSUE USING EXON GENE-EXPRESSION AND GENOME-WIDE SNP6.0 PROFILING.

4. Study design

The main aim of this study was to use array technology to try to determine a set of genetic markers in breast cancer patients that is predictive of response to tamoxifen using a small but unique set of samples identified in the breast tumour bank from Guys and St Thomas / King's College London (GSTFT/KCL) Breast Tissue Bank. Patients were selected from those recruited between 1984-1991 into the European Oncology Research Trial Consortium (EORTC) 10850 & 10851 studies and access was granted under the LREC Ref 06/Q0603/25.

Cases from these trials were selected because:

1. These were elderly patients (all greater than 70 years old) - a population for whom the prolongation of tamoxifen effectiveness is highly clinically relevant.
2. There is long-term, complete clinical follow up data for these patients maintained in a curated, computerised database.
3. All the tumours were ER+ and patients were randomised to receive either radical mastectomy or wide local excision (WLE) followed by 20mg tamoxifen daily.
4. No further treatment was administered for their disease, therefore clinical outcome represents either tamoxifen sensitive (TS; defined as cases with no recurrence for 10 years) or tamoxifen resistant (TR; defined as relapsed within 5 years) breast cancer.
5. A subset of patients had a positive margin after their WLE primary surgery but chose not to have re-excision surgery, and were instead put on tamoxifen. Interestingly, this small group of patients were all found to be tamoxifen responsive with no recurrence for 10 or more years. Microarray data from these cases may therefore be particularly representative of a tamoxifen sensitive cohort of patients.
6. All primary tumours were formalin fixed but in a proportion of cases part of the tumour was cryopreserved in RNAse-free conditions in liquid nitrogen and therefore potentially suitable for recovery of RNA for array analysis.
7. Two of the patients had paired samples; where frozen tissues from both primary surgery and the relapsed stage were obtained. The relapsed fresh

tissues will give informative microarray data, which may be indicative of resistance mechanisms.

Cases for which frozen material may be available and therefore potentially suitable to use for an array-based study were identified by a previous clinical fellow (Charlotte Moss, PhD Thesis, 2008), using the selection criteria outlined above (see Table 4.1). A "training set" of 25 fresh frozen specimens (17 Tamoxifen resistant, 8 Tamoxifen sensitive) were selected for analysis on Exon Expression arrays, SNP 6.0 genome-wide chops and microRNA analysis. Although limited in number, the uniqueness of these samples (in particular the TS), make this study valuable. Our hypothesis was that by studying this unique set of tamoxifen treated patients in highest molecular detail using these Array Chips, the maximum amount of information would be generated which would, over time, allow the data to be mined in multiple ways as new generations of analysis software becomes available.

4.1. Sample Handling

For those cases where frozen tissue was available, $0.5cm^3$ samples were made available to us. In addition to the frozen tissue, 25 freshly cut paraffin sections per patient and a slide with H&E staining were obtained. For validation purposes, 25 paraffin sections were also received from each of a further 50 unselected ER+ cases from the same trials (see Materials & Methods, section 2.13).

From the H&E sections, two independent trained physicians quantified the percentage of tumour from the block of fresh frozen tissue. These are documented Table 4.1. The average percentage tumour content was 75% per sample. The average weight per sample received was 160mg. As the percentage of tumour content was high, I decided against using laser capture microdissection (LCM) with the benefit that the tissue would undergo less manipulation. In the one case where 20% of the block was tumour, I have also included the whole tissue for extraction of RNA and DNA.

Trial No.	Tam.	Surgical margin	Tissue type	Net wt	0.5cm Frozen.	% tumour 1st estimate	%tumour 2nd estimate
5001	R	pos	Relapsed	188mg	Yes	70%	70% (2 pieces
5002	R		Relapsed	204mg	Yes	80%	60%
5003	R	pos	Primary	0	No tissue		70%
5004	R		Primary	140mg	Yes	85%	80%
5005	R	pos	Primary	193mg	Yes	80%	60%
5006	S	pos	Primary	0	No tissue		50%
5007	S	pos	Primary	0	No tissue		0% (n fat)
5008	R		Primary	184mg	Yes	65%	80%
5009	R	pos	Relapsed	194mg	Yes	85%	80%
5010	R	pos	Primary	165mg	Yes	90%	90%
5011	R	neg	Primary	157mg	Yes	90%	70%
5012	S	pos	Primary	152mg	Yes	30%	20%
5013	R	pos	Primary	128mg	Yes	90%	90%
5014	S	pos	Primary	270mg	Yes	90%	70%
5015	R		Primary	265mg	Yes	70%	80%
5016	R	pos	Primary	265mg	Yes	100%	100%
5017	R		Primary	0	No tissue		75%
5018	R		Primary	237mg	Yes	90%(2 pieces	80%
5018_2			Relapsed		Yes		
5019	R	pos	Primary	244mg	Yes	90%(2 pieces	80%
5020	R		Relapsed	0	No tissue		40%
5021	R	neg	Primary	144mg	Yes	90%	50%
5022	R	pos	Primary	165mg	Yes	80%	70%
5023	S	neg	Primary	256mg	Yes	90%	80%
5024	S	pos	Primary	148mg	Yes (only 15%)	No tissue	not stained
5025	S		Primary	100mg	Yes	20%	
5026	S		Primary	90mg	Yes	80%	
5027	S		Primary	100mg	Yes	20%	

Table 4.1 Characteristics of the 'Training' set samples

Two independent clinicians estimated the percentage of tumour from the H&E slide to determine the tumour content of the fresh frozen samples (7th and 8th column). The Tamoxifen Response status of each patient was given as R=resistant and S=Sensitive (2nd column). Where surgical margin was documented, pos=positive margin and neg=negative. The precise weight from the tissue is shown in the 5th column. Tissues were either obtained at primary diagnosis (before endocrine therapy) or at relapse. This is documented in the 4th column. Specimens highlighted in yellow represent the paired-samples; 5008 & 5009 and 5018 &5018_2 (diagnosis & relapsed tissue respectively).

4.2. Optimisation of RNA and DNA Extraction Method

I optimised the extraction of RNA and DNA by comparing a number of available methods as charted in Table 4.2 using fresh frozen normal breast tissue (obtained from Prof Louise Jones) and some normal liver, kidney and spleen tissue from rats. The human 'practice' tissue had been stored for approximately as long as our training set frozen tissues (15 years). Other studies that have used tissue from this same breast tissue bank (GSTH) have reported good quality RNA and DNA extraction (Loi *et al.*, 2008). The preferred method for extracting RNA was found to be using the RNeasy Mini Kit from Qiagen. This gave the purest RNA without compromising the quantity retrieved. The method settled on for extracting DNA, after testing several reagents/kits was using the DNeasy Mini kit from Qiagen. Again, this method gave the most consistent quality of DNA with the quickest protocol. Various other methods were assessed using rat organ tissues and normal breast tissues.

Quality of the RNA was checked with Agilent 6000 Nanochips to determine the RNA Integrity Number (RIN). Samples with a RIN value of less the 6 are not considered suitable by our Microarray Facility (Tracy Chapman, *personal communication*). Quality of DNA was measured using agarose gels. Quantity was measured using nanodrop photospectrometer. The ratio of absorbance at 260 and 280 nm was used to assess the purity of DNA and RNA in addition to the Agilent nanochip. A ratio of ~1.8 is generally accepted as "pure" for DNA; a ratio of ~2.0 is generally accepted as "pure" for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or contaminants that absorb strongly at or near 280 nm.

260/230 ratio is a secondary measure of nucleic acid purity. The 260/230 values for "pure" nucleic acid are often higher than the respective 260/280 values. They are commonly in the range of 1.8-2.2. If the ratio is appreciably lower, this may indicate the presence of co-purified contaminants.

I extracted RNA and DNA from all the training set samples using the optimized

procedures and confirmed all had good quality RNA and DNA using the above criteria.

Methods	Tissue	Amount	RNA total	DNA total	Quality RNA	Quality DNA
TRIzol alone	NBrT 1	60mg	11 µg		RIN=7	
TRIzol alone	NBRtT2	72mg	86 µg		RIN=7	
TRIzol alone	NBrT2/2	186mg	230 µg		RIN=6	
TRIzol alone	NBrT3	128mg	48.7µg		RIN=6	
TRIzol/RNeasy kit	NBrT4	90mg	27µg		RIN=7	
DNA/RNA midi kit (Q)	S(M)	60mg	106 µg	43 µg	RIN=0.3	good
DNA/RNA midi kit (Q)	K(M)	62mg	35 µg	14 µg	RIN=6.2	good
DNA/RNA midi kit (Q)	H(M)	50mg	9.3 µg	14.5 µg		good
DNA/RNA midi kit (Q)	Lung(M)	50mg	47 µg	65 µg		good
TRIzol/RNAeast kit	L(M)	50mg	49 µg		RIN=3.5	good
DNA/RNA midi kit (Q)	L(M)	30mg	2 µg	39 µg	RIN=2.2	good
DNA/RNA midi kit (Q)	S(M)	30mg	2.4 µg	2.9 µg	RIN=1.5	good
DNAeasy mini kit	L(M)	10mg		10.1 µg	N/A	good
DNAeasy mini kit	L(M)	10mg		4.5 µg	N/A	good
RNeasy mini kit	K(M)	10mg	15.2 µg		RIN=7.4	N/A
RNeasy mini kit	K(M)	10mg	21.7 µg		RIN=7.5	N/A
TRIzol/RNeast kit	K(M)	12.5mg	45.7 µg		RIN=7.4	good
TRIzol/RNeast kit	K(M)	12.5mg	46.3 µg		RIN=7.5	good

Table 4.2 Optimisation of RNA and DNA extraction from tissue. TRIzol/RNeasy=Improvised combined TRIzol and RNeasy protocol, DNA/RNA midi kit (Q)=Commercially available simultaneous extraction of RNA and DNA from Qiagen. The 2nd column specifies which tissues were used for the optimisation: NBrT=Normal Breast Tissue, S(M)=Spleen(Mouse), K(M)=Kidney(Mouse), H(M)=Heart(Mouse), Lung(M)=Lung(Mouse), L(M)=Liver(Mouse). The 3rd column shows the amount of starting tissue used. The 4th and 5th columns show the final amount of RNA and DNA respectively. Quality of RNA was assessed using the Agilent Bioanalyser, RIN=RNA Integrity Number (value from 1-10, 10 being the highest quality and most intact RNA). N/A=not applicable.

4.3. Exon expression array processing and quality control

Latest generation human expression Affymetrix array chips, the Human Gene®Chip Exon 1.0 ST, were used for our human samples of breast cancer in preference to the Human U 133 plus 2.0 used in the cell line study. The key differences between the two types of expression chips are described in the Materials & Methods (Section 2.8) together with a more detailed description of the array procedure. Two configurations of the WT Sense Target labelling assay are available. The 100ng total RNA protocol is recommended for analysis of the gene level as this protocol allows for the omission of the ribosomal RNA (rRNA) reduction procedure, and hence benefits from a larger number of high-quality probes selected from the entire transcript. The 1μg total RNA with the ribosomal reduction procedure is recommended for Exon level analysis. A recent report from the Patterson lab (www.affymetrix.com/userforum), found Exon-level results were still good without the ribosomal reduction stage (riboMinus). They hence have supported the use of the 100ng protocol in cases where RNA quantities are limited. In some samples where we had sufficient tissue, an optimising experiment was carried out using the same amount of starting total RNA with and without the ribosomal reduction stage. The resulting mRNA samples were then compared. I used real-time PCR to quantify for housekeeping genes such as 18S and GAPDH. I found that although residual rRNA levels in the non-ribosomal reduced were higher, the mRNA were otherwise very similar. I therefore processed all the samples using the 100ng protocol.

Briefly, 100ng of total RNA was extracted from human breast tissue and prepared into probes as per Affymetrix GeneChip®Whole Transcript (WT) Sense Target Labeling Assay (i.e. Exon) protocol. Biotinalyted cRNA (A260/A280 ratio between 1.8 and 2.1) was prepared for each experimental sample and hybridised to Affymetrix GeneChip®Whole Transcript (WT) arrays. There are two quality control steps in the preparation of the probes. One is the amount of the cDNA after the IVT steps, and the other is qualification of the fragmented ssDNA using Agilent. All probes reached acceptable standards. Hybridisation, scanning and image analysis was performed by Tracy Chaplin (Institute of Cancer); after scanning, array images were

assessed by eye to confirm scanner alignment, the absence of significant bubbles or scratches on the chip surface, and the absence of slides with very high background.

Raw data were assessed for quality using the Affymetrix Power Tool statistical package (Affymetrix Console, *url*) as described in Materials and Methods 2.8.

4.4. Normalisation and Transformation of Raw Array Data

The aim of these analyses was to identify differentially expressed genes in tamoxifen resistant samples (TR) from those of tamoxifen sensitive samples (TS). In order to minimise variation between arrays caused by biological and experimental factors, it was important to perform appropriate data transformation and normalisation. First of all, data were transformed into CEL files. The Affymetrix GeneChip®Whole Transcript (WT) (Exon Array) has a universal background correction known as Detection of Average Background (DABG).

Subsequent analysis involved normalisation and transformation of raw array data using the Robust Multi-array Average (RMA) (Irizarry *et al.*, 2003) from the Affymetrix Power Tool (APT) software. The RMA method was used for gene-level summary and probe-set level summary. Box plots of all of the normalisation methods confirmed that transformation of the data is such that the distribution of probe intensities for every array in the set of the twenty-five arrays used in this experiment are the same and fit the same distribution (See Figure. 4.1A). However, it is clear from the histograms that this distribution is a normal distribution, which is applied to any downstream parametric statistical analysis (Figure 4.1B). This confirms the data appears to be transformed in such a way that the distributions of probe intensities for every array in the set of twenty-five arrays used in this experiment are the same and the data resembles a normal distribution.

Further criteria set on the data filter included:

- removal of control and un-annotated probe sets, detection above background (BG) p value >0.05; removal of the genes which were absent or un-transcribed,

- removal of probe sets that do not change significantly i.e. remove flat profiles and coefficient of variance (cv)=0.01.
- removal of probe sets which did not change significantly in at least one of the samples with an absolute value at the 10^{th} percentile.

Page for Fig 4.1 A and B

4.5. Hierarchical Clustering Analysis

RMA normalised data were imported into the Affymetrix Power Tool (Part of the available Affymetrix Console) in order to allow a hierarchical clustering analysis to monitor overall patterns of gene expression between the normalised arrays (Materials and Methods 2.8.3; (Eisen *et al.*, 1998). Briefly, a hierarchical clustering analysis produces a map of results where probe sets are grouped together based on similarities in their patterns of normalised expression across the twenty-five arrays. As the processed data follows a normal distribution it is appropriate to use the Pearson Correlation Coefficient, which calculates the similarity measure based on a linear model. Because we are interested in genes that change their expression between the reference and test groups, data were further filtered to include the 10,000 probe sets with the highest variance across the twenty-five arrays.

The same algorithm was also applied to cluster the experimental samples for similarities in their overall patterns of gene expression. Arrays hybridised with the reference samples formed their own cluster, indicating that there is a pattern of differential gene expression in the subset of TR patients compared to the TS ones. There was a general pattern of TR arrays clustering apart from TS arrays (BTB357, 5025, 5026, BTB378 and 5012 and 5027 in Figure 4.2). There were three outlier TS samples (5014, 5024 and 5023) but even these clustered in the opposite arm of each hierarchy branch.

4.5.1. Principle Component Analysis

Principle Component Analysis (PCA) is a statistical method of identifying patterns in data, and expressing it in such a way as to highlight similarities and differences. Since patterns can be hard to find in data of high dimension, where graphic representation is not available, PCA is a powerful tool for analysing microarray data. The other main advantage of PCA is that once you have found these patterns in the data, you can compress the data by reducing the number of dimensions, without much loss of information. This is a non-parametric analysis and the result is unique

and independent of any hypothesis about data probability distribution. However, the latter two properties are regarded as weaknesses as well as strengths.

The differences in clustering between TR versus TS were also found to be distinct when using PCA (Materials & Methods 2.8.3.1.). Using a commercially available statistical analysis software from Partek®, PCA diagrams of the normalised data were generated, illustrated in Figure 4.3A. This represents another way of looking at how the data cluster with respect to each other. Our PCA analysis confirmed the clustering of the TR group away from the TS group, and to a lesser extent, clustering of the relapsed tissue specimens from the diagnostic tissue specimens (Figure 4.3C). To investigate operational error, as only 5 probes were able to be prepared per week, PCA was carried out according to the date on which the probes were processed (Figure 4.3B). The PCA analysis did not show any clustering from specimens processed on the same date, which demonstrates that there was minimal batch and operational influence on our data.

4.6. Predictive Tamoxifen Resistant Differential Gene Expression Analysis

RMA normalised data were imported into LIMMA in R software for statistical analyses. As well as fitting a normal distribution, another assumption of statistical analyses, such as the widely used Student's t-test, is that the variability of a gene is constant across treatment types. However, in the absence of accurate diagnosis methods, it is safest to assume that variability may differ between our reference (all the TS) and test (all the TR) groups. As we have confirmed that our data fits a normal distribution, it is appropriate to conduct a Welch's t-test, which is a parametric analysis that corrects for difference in variability. Most of the subsequent data analysis (for both Exon and SNP arrays) was done by Dr Claude Chelala (Bioinformatics group, Institute of Cancer) using in-house software. In a few stated incidences I have also used Partek software when it offered a better illustration of the results than the in-house program.

As we are interested in identifying genes that change their expression between the reference and test groups, LIMMA was used for differential expression between TR and TS samples. Using a 2-fold cut-off and a p value <0.05, the list generated was one of 20 genes with significant change (see Table 4.3). However, it was felt that it would be more informative to look at the top 500 significant genes based on p value <0.05 rather than filtering on arbitrary fixed fold change, as with the latter there is a risk of ignoring genes that change significantly but are below the arbitrary fold change threshold. Using LIMMA, the top 500 most significant genes have been generated (See Appendix A, where the top 100 genes are listed).

Cluster ID	symbol	r	s	s.vs.rFOLDCHANGE	s.vs.r raw pvalue
2378068	G0S2	52.4928647	445.017579	8.477677522	5.84E-10
2480383	EPAS1	152.74672	364.392569	2.385599961	5.75E-05
2664640	RFTN1	73.3328057	128.531107	1.752709528	0.000559351
2686458	ABI3BP	44.5452549	128.057824	2.874780372	1.07E-05
2940202	F13A1	68.4511617	199.437077	2.913567455	0.000216504
2955638	CLIC5	21.5540899	44.9616438	2.085991285	0.000275841
2995589	AQP1	97.080669	285.522694	2.941087008	2.31E-05
3028934	PIP	72.4331669	699.712819	9.660116336	0.000552224
3088486	LPL	26.4223271	200.02419	7.570271508	2.43E-09
3107603	BC071775	79.9504389	32.1709144	-2.485177696	0.000532114
3203524	AQP7	57.4379467	113.293375	1.972448202	0.000177956
3331355	SERPING1	234.313378	559.625208	2.388362164	5.25E-05
3359121	INS-IGF2	254.234179	724.337034	2.849093844	6.88E-05
3391149	CRYAB	41.7804754	165.527524	3.961839181	0.000217546
3442475	C1R	228.224422	527.116075	2.309639213	0.00069587
3510450	LHFP	136.530966	289.218648	2.118337368	0.000288209
3544525	FOS	173.940688	504.408602	2.8998885	4.33E-05
3821893	JUNB	178.453163	357.722207	2.004571958	0.000671074
3848039	C3	312.050753	965.586724	3.094325886	7.66E-05

Table 4.3 Top 20 most significantly altered genes in TS cases when compared with TR breast cancer samples. Genes whose expression had a minimum 2-fold change and a p-value of <0.05. The transcript cluster identification is in the left column. The gene symbols are in the 2nd column, while r=resistant and s=sensitive for normalised probe intensity. The TS versus TR fold changes are seen in the 5th column, while the last column shows the TS versus TR p-value.

4.7. Mining biological pathways using Ingenuity Pathway Analysis (IPA)

To explore the relationships between genes that form the TR gene signature at the molecular level, the top 500 genes were analysed using IPA, to generate the most common canonical pathways that involve these 500 genes (Appendix B). Figure 4.4A illustrates the gene-level analysis and Figure 4.4B illustrates the splice-variant (Exon level) analysis, which was analysed using Partek. The latter analysis uses similar filtering and p-value to minimise false discovery. The crucial difference is that the latter analysed the individual exon-probe intensities as individual events, hence reporting the most significantly altered exon/probe in the analysis, while the gene-level analysis uses the probes sets of a particular gene as a unit. Here, I have set the filter for greater than 50% of the probes sets, i.e. >2 out of the 4 probes per exon of a gene, and > 50% of the probe sets (exons) of a particular gene must be significantly altered to be included in the analysis.

The results from the two IPA analyses are quite different. When analysed at the 'gene-level' (Figure 4.4A), the pathways identified are similar with published data regarding associations with breast cancer, such as the Integrin signalling, PI3K/AKT signalling and p53 signalling. The canonical pathways associated with Exon-level (using the same chip and sample but analysed as the change of the intensity of each probe of the four probes per exon) analysis (Figure 4.4B) are associated more with metabolism, such as aminosugars metabolism, tyrosine metabolism, glycolipid metabolism, glycolysis, xenobiotics by cytochrome p450, β-alanine and phenylalanine and purine metabolism. There are two cardiac associated pathways, the β-adrenergic signalling and coagulation system. One possible hypothesis is that detectable Exon changes are more reflective of acute daily events. All of our training set patients were elderly (>70 years), and probably had existing cardiac conditions, or were taking correcting cardiovascular drugs such as β-blockers, anticoagulants (warfarin), thyroid-correcting drugs (thyroxine or carbamazipine) and various polypharmacy, which may have augmented the p450 cytochrome pathway.

There are two common conical pathways between the IPA analyses; PPARα/RXRα activation and β-adrenergic signalling. Peroxisome proliferator-activated receptor-alpha (PPARα) has been shown to increase fatty acid oxidation, decrease cytokine levels and is implicated in insulin production. A doubling of breast cancer risk among women with a PPARα polymorphism versus common homozygous alleles has been reported and PPARα has been shown to be involved in the growth inhibitory effect of arachidonic acid on breast cancers (Bocca *et al.*, 2008; Golembesky *et al.*, 2008). Beta-adrenergic signalling has no published association with breast cancer.

4.8. Affymetrix GeneChip® SNP6.0 arrays

The genomic aberrations of all 25 samples were also analysed using the GeneChip® Human SNP6.0 array. The experiments were performed according to the standard protocol (see Chapter 2.8.5). The raw median-normalised log2 ratio at each SNP was extracted from the Affymetrix Chromosome Copy Number Analysis Tool (CNAT Ver4.0.1) software. The data normalisation was performed against SNP6.0 data from

10 normal breast samples (LREC Ref 05Q403/199), which were prepared in an identical manner to the 25 tumour specimens.

Copy number analysis was performed using nine different algorithms (Materials & Methods 2.8.5). The threshold for genetic "gains" or "losses" was determined as a median log2 ratio ≥+0.162 or ≤-0.162, respectively (±3 standard deviations (SD) of the interquantile range) {Cavatorta, 2004 #50}; {Tonon, 2005 #24}. "High-level amplifications (HAs)" and "homozygous deletions (HDs)" were defined as a median log2 ratio ≥+0.419 or ≤-0.401, respectively, which corresponds to 97% or 3% quantiles. In order to avoid false-positive changes due to random noise in signal intensity at each SNP, we set a minimum physical length of at least 20 consecutive SNPs for putative genetic alterations. All identified altered regions were then verified by assessing the raw normalised data.

4.9. Quality control of SNP6.0 arrays

Figure 4.6 illustrates the PCA as part of QC analysis of the SNP array. The separation of TR from TS samples is distinct, when normalised using data from the 10 normal breast tissues. Unsupervised clustering was performed, again after normalisation with the 10 normal breast samples. The dendrogram in Figure 4.6 again shows clear clustering of the TR away from the TS samples, except for two samples. The two-paired samples from two separate patients (indicated by the green box) also clustered away from the other samples, showing that the SNP patterns in samples from the same patient have clear similarities, as expected. In the case of one pair of samples (5008 and 5009), where relapse occurred in less than 6 months, there was very little difference at all in the SNP patterns, offering another level of reassurance. The expression array patterns from these paired samples were however distinct (see Figure 4.2).

Space for Fig 4.5

Space for fig 4.6

.

4.10. Genome-wide analysis of CNV and data integration with expression EXON data

The physical position of all SNPs (n=906,600) on the arrays was mapped according to the Human Genome Sequence (NCBI, Build 36). Colleagues at the Institute of Cancer have developed their own visualization software to merge all genetic aberrations with the gene annotation from the Ensembl Ver.37 (http://www.ensemble.org) public database. Taking structural variation in the human genome into account, this software integrated the Copy Number Variation, CNV (http://projects.tcag.ca/variation/) data into our analysis (Iafrate *et al.*, 2004), (Redon et al., 2006). We then merged our EXON array dataset (6.5 million features; up to 4 features to an exon) with that of our SNP6.0 data. The software is able to find the most significant regions of CNV, which have the most significant changes in the expression study simultaneously. This is the first such analysis in breast cancer for tamoxifen resistance. All the raw data are available in the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/).

The Copy Number Analyser for Affymetrix GeneChip (CNAG) analysis for all 25 tamoxifen-treated human breast cancer samples identified chromosome regions of CNV throughout the whole genome of these breast cancer patients (Figure 4.7). These data were normalised against the 10 normal breast tissues, and "smoothed" using the Hidden Markov model (HMM) algorithm (see Chapter 2.8.5). The most frequent genetic gain was detected at 8q24.22 where 15 of the 17 tamoxifen-resistant patients had CNV gain of the *ST3GAL1* gene. However, this was also gained in the tamoxifen-sensitive cohort, in 4 out of the 8 patients. Other gains were found at 16p13, 20q13, 12q13, 1q25 and 1q32 with high frequency in TR patients (13 out of 17), but they were also found in TS. These may be SNP regions associated with breast cancer alone. Figure 4.7 illustrates the chromosomal location of CNV of the combined significant results with red regions being most amplified and blue regions being most deleted. The parameters were set at above 1.7 ratio of amplification and less than 0.7 for regions of deletion. The software was asked to only show results where more than 10 arrays had the aberration in a region. There is a visual appreciation that chromosomes 11, 17, 8, 1, 20 and 6 have the most significant regions of CNV aberration. Although this software only uses one algorithm (HMM),

it is regarded as one of the most robust of the available algorithms. Reassuringly, the analysis using Partek was concordant with results from our in-house bioinformatics software (illustrated in Figure 4.8), which identified the actual number of significant CNVs which occurred in a segment of each chromosome. This also required normalisation against the 10 normal breast tissues, and data was smoothed using the 10 stringent algorithms, which are used in the R-conductor software (see Chapter 2.8.5). The most significant CNV losses were on 11q22-25, with 13 out of 17 TR patients showing loss of *OR8D4,* NP_001013765.1 and NP_055530.2 genes. These losses were not seen in any of the TS patients. Hence these genes may be very specific to tamoxifen response.

Analysis by visual observation (Figure 4.7) and by filtering of significant overlapping regions between significant SNPs in the CNV and EXON expression data revealed three regions of interest. The filtering was generally done by:
1) Excluding all the regions/probes, which were not altered at all.
2) Selecting the regions with >20 SNP length.
3) The specific criteria (i.e.>than 10 out of all 15 TR) of gains regions in one cohort (i.e. TR) while setting the reciprocal loss in the other cohort (i.e. TS).
4) Analysing in conjunction with the merged data from expression array.

The three regions of interest thus highlighted were:
1) **8q24**: This region (see Figure 4.9) was represented in 184 out of the 282 SNP regions which were significantly altered (***DNA copy number gain***) and was found in 15 out of the 17 TR patients, and conversely in less than 3 out of the 8 TS patients. The gain in this region overlapped with two genes, which were downregulated in the exon expression array study; *RUNX1T1* and *ENPP2* (by –4.57 and –3.92 fold respectively). These genes were later validated by qPCR in our cohort of combined 122-tissue microarray and paraffin slides from tamoxifen treated patients (see Section 4.13). MicroRNAs that were mapped to this region of gain were hsa-mir-661, hsa-mir-338, hsa-mir-30b and 30d. According to the miRBase database they potentially regulate genes which have been found to relate to tamoxifen resistant breast cancer, including *AP2A1* (AP-2 complex subunit Alpha-1), *TAF2, RAD54L, SOCS1, PARP16, BRAF, MAP3K5, S100A10, MAGEE1, SIAH1, TP73, PDZK1IP1, MAP4K2, THRAP5, ALDH3A1, PARP10* and *MAGEC3.*

2) **<u>11q23-25</u>**: This region (see Figure 4.10) was represented in 21 out of the 99 SNP regions which were significantly altered (***DNA copy number loss***) and was found in more than 13 out of 17 TR patients, but also in more than 2 out of 8 of the TS patients. Four genes from the expression data were found to overlap with this region, *SNF1LK2, ZBTB16* were downregulated by –4.02 and –4.99 fold respectively while *OR10G7, OPCML* were upregulated by +3.99 and +3.92 fold respectively.

3) **<u>17q21-25</u>**: This region of chromosome 17 represented 1655 of the total 31206 SNP regions which were significantly altered. 2322 DNA copy number gains were found in all chromosomes in more than 10 out of 17 TR patients, of which 377 copy number gains regions (see Figure 4.11) were at 17q21-25. There is published data available on the amplification of genes on 17q22; such as *HOXB13, COL1A1,* which are positioned in the second of three regions of the 17q21 HER2 amplicons {Jansen, 2005 #34;Sgroi, 2004 #39;Jansen, 2005 #176}. In our merged SNP and EXON datasets, 3 genes from the expression data overlapped with this region; *SP2* (Transcription factor Sp2), *CHRDL1* and *XR_000549.1* were found to have –4.26, -6.69 and +4.60 fold changes respectively. This region also contains microRNA 657 (17q25.3) which potentially regulates the expression of many known genes related to tamoxifen resistance, including *HOXB13, TGFβ1, MMP9, THRAP5, ESRRB* (ER-related receptor β), *LOXL4, ERBB2*, and *KIAA1324L* which is related to an AP-2 regulated gene studied in our lab (Ka Yi Chan, PhD Thesis, 2010). Genes related to breast cancer, such as *TP53, ADAM2* and *S100A1* may also be regulated by has-mir-657. We have hence chosen this miRNA to study further (see Chapter 4.15).

Regions of SNP which encode microRNA were searched using the UCSC website. The link from this website to miRBase (Welcome Trust Sanger Institute) and TargetHumanScan (Whitehead Institute of Biomedical Research) were used to scroll through the related genes which each microRNA has been linked to. The available software packages score the strength of its association based on site-type contribution, 3' pairing contribution, local AU contribution and position contribution.

4.11. Ingenuity Pathway Analysis of the 8q24 region

The integrated analyses list showed 3 regions where significant copy number aberration; namely 8q24 (gain), 17q21-25 (gain) and 11q23-25 (loss). In the 282 SNP regions that were significantly changed in more than 10 of the 17 TR patients, but were seen in less than 3 out of the TS patients, 184 SNPs were located within the 8q24 region. This is a disproportionately high gain region, which led us to study this region further. When the genes associated with all 182 8q24 SNP regions, were computed into the ingenuity pathway biological analysis tool, the most significant canonical pathways were p53 and IGF-1 signalling, and androgen and oestrogen metabolism, (Figure 4.12). The p53 pathway was also found to be directly related to our *in vitro* TR mechanism (see Chapter 3.0 MAGEA2 mechanism of action via p53), while androgen and oestrogen metabolism, PTEN and IGF-1 signalling pathways are established published tamoxifen resistance related mechanisms (McCubrey *et al.*, 2006; Parisot *et al.*, 1999; Shoman *et al.*, 2005). Hence this analysis reinforced the fact that our array study is consistent with both our *in vitro* study and with published data at least using this biological mining website.

When I analysed the combined data from our TR cell line microarray study with the SNP array data using IPA, I found overlapping networks. The networks with the most significant molecules have been merged and is shown as Figure 4.13. The 3 networks comprise of 20 genes from combined datasets; microarray HU 133plus2.0, EXON 1.0 ST and SNP6.0, which includes *RUNX1*, *MGMT*, *MYBL1*, *PLP2*, *Camodulin*, *CALM2*, and centred around the *TP53* and Akt pathways. Interestingly, molecules which have supporting published data for tamoxifen resistant has also been found in the merged network, such as E2f and Akt.

2 page gap for figs 4.12 and 4.13

4.12. Validation of genes from integrated analysis of Exon and SNP arrays

4.13.1 Quantitative real-time PCR

Nine genes were further validated from the combined analysis. These were *EPHA7*, *PALM2*, *SNF1LK2*, *ZBTB16*, *OR10G7*, *OPCML*, *SP2*, *RUNX1T1* and *ENPP2*. In brief, these genes were found in SNP CNV regions, which were significantly altered in the majority of TR cases (TR>10, out of 17 cases) but not in TS cases (TS<3, out of the 8 cases). As described in section 4.11, these genes mapped within the altered regions and were also significantly altered in expression on the Exon arrays.

Quantitative real-time PCR was used for validation of cDNA from the training set normalised to GAPDH and compared with normal breast (10 samples). When more than one choice of probe sequence was given, I chose the 'inventorised' option with the shortest sequence (see Materials & Methods, section 2.9.2), unless if these probes not listed. For two genes, PALM2 and ZBTB16, the probe and primers were undetected despite trying again with higher concentration of starting cDNA. One possible explanation is that we may have not bought the optimised probe and primers. We intend to repeat these experiments again when we purchase new probes.

A summary of the results is shown in Table 4.5. All the genes were directionally positive in their correlation with the Exon array findings, for example where a gene was expected to be downregulated in TR cohorts, the qRT-PCR experiments showed downregulated mRNA of the particular gene when compared with their TS counterparts. This is illustrated for individual genes in Figures 4.13 and 4.14.

4.12.1. Immunohistochemistry

Commercially available antibodies against the encoded proteins of all 11 genes were obtained and each was optimised for immunohistochemistry on paraffin sections (IHC-P) using positive control specimens as suggested by the manufacturer. In two cases, the optimisation failed, however the other antibodies were used to stain my independent validation series of 129 cases of ER+ breast cancer (see Materials & Methods, section 2.13). Although the Ventana Discovery automated staining instrument was used in each case, only 5 sets of slides were suitable for scoring, with the antibodies giving a specific and relatively sensitive signal. The failure of the unsuccessful antibodies were due to either weak staining (despite optimisation) or indiscriminate staining (i.e. staining connective tissues and smooth muscles as well as breast tumour). This is summarised in Table 4.4.

CHR	CYTOBAND	SYMBOL	DESCRIPTION	FC (Exon)	q-PCR valid	IHC-P
6	q16.1	EPHA7	EPH receptor A7	3.94	P*=0.039	weak stain
9	q31.3	PALM2	PALM2-AKAP2	-4.12	Not done	Spec. & sens
11	q23.1	SNF1LK2	SNF1-like kinase 2	-4.02	p=0.0688	Spec. & sens
11	q23.2	ZBTB16	Zinc finger and BTB domain 16	-4.99	Unsuccessful	Spec. & sens
11	q24.1	OR10G7	Olfactory R, fam10, subfam.G, member7	3.99	p*=0.017	Spec. & sens
11	q25	OPCML	Opioid binding protein	3.92	p=0.169	Spec. & sens
17	q21.32	SP2	Sp2 transcription factor	-4.26	p=0.2593	Low specificity
8	q21.3	RUNX1T1	Runt-related TF 1; transloc to, 1	-4.57	p=0.339	Not done
17	q25.3	ENPP2	Ectonucleotide pyrophosph-tase/PDE 2	-3.92	p=0.710	Not done

Table 4.4 Summary of validation experiments

RT-PCR was performed on the 25 cases from the training set. The p-values were calculated with Whitney-Mann non-parametric statistical analysis (Prism) between the TR cases from the TS cases; two genes (marked *) reached significance. Immunohistochemistry (IHC) validation was performed for 7 proteins on an independent series of 122 cases (see Materials & Methods, section 2.13); 5 sets of slides where the staining was specific were scored. Antibody optimisation failed for 2 proteins (marked “not done”).

The sets of slides for *OCPML, OR10G7, SNF1LK2, PALM2* and *ZBTB16* were scored by myself and a qualified pathologist, Dr Yaohe Wang. Disease free survival (DFS and/or overall survival (OS) were then calculated for each gene by Kaplan-Meier curve, from the Prism software. As illustrated in Figures 4.14 to 4.18, genes were validated for their mRNA expression using Tagman probe and primers on the human primary breast cancer tissues cDNA. Using a non-parametric (one-tail) Mann-Whitney Equation, we analysed a scatter plot with a p-value calculated if there is a difference between the TR group from the TS group. This sets apart if the mRNA expression for that gene can differentiate the TR from the TS significantly. Following that, the independent 122 primary breast cancer tumours in their paraffin slides were validated for this gene of interest's protein expression by immunohistochemistry. The staining results were analysed for their effective prognosis by using Kaplan-Meier curve, using Prism software. Disease-free survival analysed if the positive staining of the slides had any prognostic value in predicting the period between diagnosis and time when disease recurs. Overall survival measures the period between diagnosis dates with time of death.

4.13. miRNAs in Tamoxifen Resistant

4.13.1. hsa-mir-657 from our integrated array analysis

As discussed in the Introduction, recent studies have highlighted the key regulatory roles of miRNAs in all fundamental cellular processes including cancer. From the integrated analyses of the SNP6.0 and Exon array data, hsa-mir-657 was highlighted as an interesting miRNA, with a potential association with predicting tamoxifen resistance. The regions of 8q24, 11q22.2-25, and 17q21-25; (the three most significant from our combined analysis) were mapped with the miRNA library by Dr Claude Chelala's software. There are currently 4,000 mature miRNA listed. Table 4.5 lists the most significant miRNAs, which mapped within these 3 regions of SNP aberration (see Chapter 4.11).

Chr	Band	From	To	miRNA	TR	Proteins negatively control
1	q32.1	203684053	203684149	hsa-mir-135b	10	S100P, ESRRB∗, BNIP1, ESRβ∗, ENPP7, MDM4
8	q24.22	135881945	135882032	hsa-mir-30b	10	SOCS1, PARP16, TAF2, RAD54L
8	q24.22	135886301	135886370	hsa-mir-30d	10	SOCS1, THRAP6, MAGEE1, PDZK1∗
8	q24.3	145091347	145091435	hsa-mir-661	11	TP73, THRAP5∗, MAGEC3
17	q25.3	76713671	76713768	**hsa-mir-657**	10	**HOXB13∗, TGF-β1, MMP9∗, PGF, THRAP5∗, ESRRB∗, ERRB2∗, TP53**
17	q25.3	76714278	76714344	hsa-mir-338	10	HOXA3, FGFR2, V-FOS MAP3K3∗, ADAM 17

Table 4.5 The five significant miRNAs found in more than 10 out of 17 TR and less than 3 out of 8 TS cases. The genes, which are negatively controlled by the miRNA, which are related to breast cancer and (*) specifically with tamoxifen resistance are listed in the last column as found from the miRBase database.

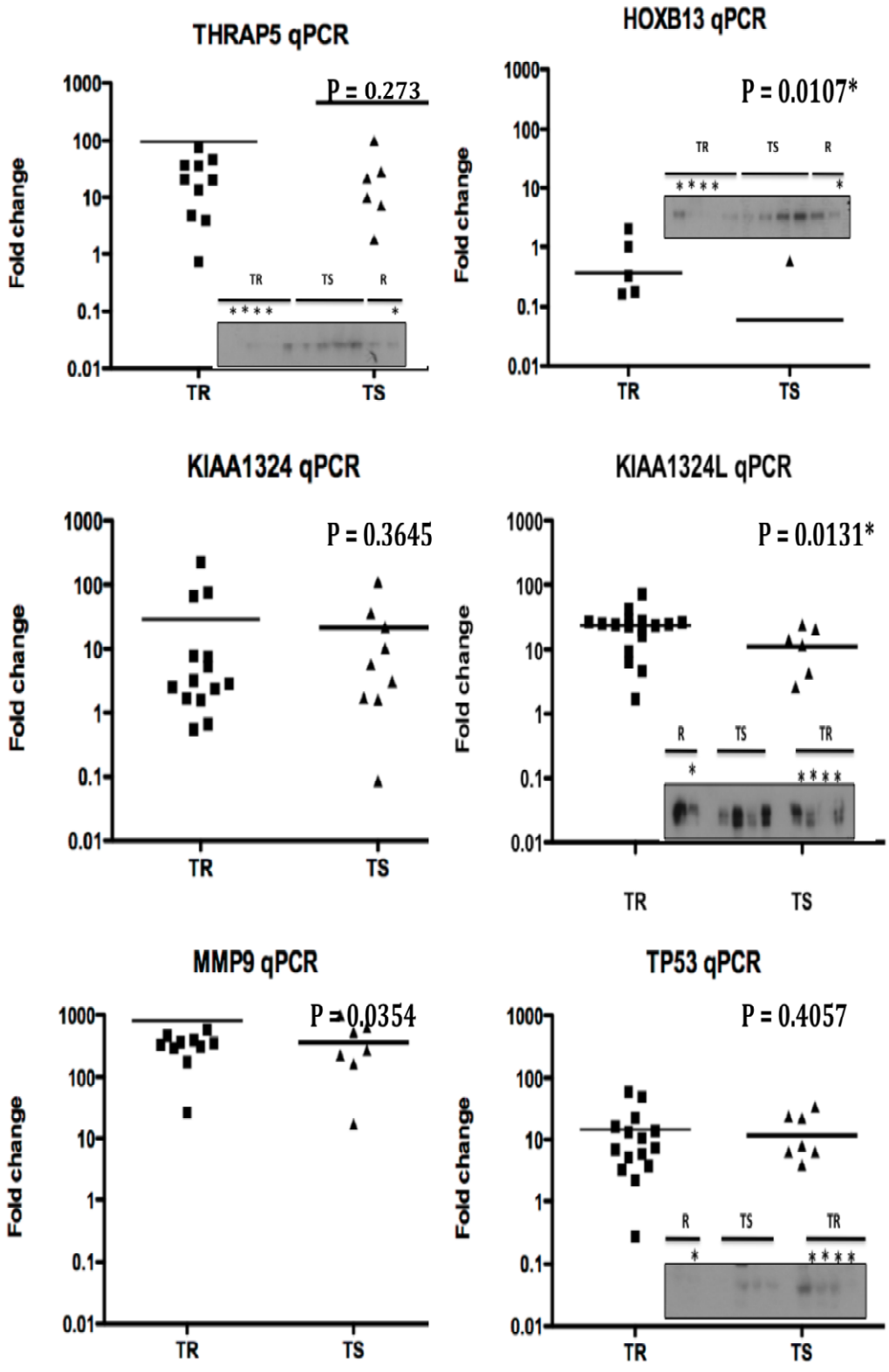
Hsa–mir-657 is located at 17q25.3, As listed in Table 4.5, this miRNA is associated with many breast cancer genes and seven genes in particular have a published association with tamoxifen resistance; *THRAP5, HOXB13, TP53, MAP2K2* and *AKR1B1*. For this reason I decided to study its relevance in our human primary breast study and in the TR breast cancer cell line series. Using the mirVANA PARIS kit, I simultaneously extracted protein lysates and miRNA from primary fresh frozen tissues from the training set (n=25). Real-time qPCR for hsa–mir-657 showed that the difference in mRNA expression between the TR and the TS groups was not significant (p-value of 0.3705). Interestingly however, the same 11 cases from the training set from the combined SNP and EXON analysis, were found to have mRNA expression of hsa-mir-657 by RT-qPCR. Only 2 of the 8 TS expressed this miRNA (25%), as compared with 9 out of 17 of TR (53%). The p-value does not reach significance different between the two groups due to the small size of our training set. Conversely, the sensitivity in the TR group was high.

I have shown:

1) A trend for increased qPCR detection of hsa-mir-657 in our human primary tissue of the TR cohort compared with the TS cohort.

2) Majority of the genes potentially regulated by hsa-mir-657; i.e. THRAP5, MMP9 and TP53 have no significant mRNA expression difference between TR and TS. HOXB13 and KIAA1324L mRNA expression in TR was significantly different from TS. For HOXB13, the mRNA difference may be due to the fact that sample size in the TS group was small, due to poor detection of the probe despite higher concentration of cDNA. As for KIAA1324L, this may well be a true significance.

3) The protein expressions of HOXB13, THRAP5 and KIAA1324 are reduced in cases with hsa-mir-657 as shown in immuno-blotting (Figure 4.16). P53 however was less consistent in its protein expression reduction in has-mir-657 carrying lysates. There are no good existing KIAA1324L and MMP9 antibodies for immuno-blotting.


THRAP5 qPCR
P = 0.273
Fold change
1000
100
10
1
0.1
0.01
TR
TS
R
****
*
TR
TS
HOXB13 qPCR
P = 0.0107*
Fold change
TR
TS
R
****
*
KIAA1324 qPCR
P = 0.3645
Fold change
TR
TS
KIAA1324L qPCR
P = 0.0131*
Fold change
R
TS
TR
*
****
MMP9 qPCR
P = 0.0354
Fold change
TP53 qPCR
P = 0.4057
Fold change

4.13.2. hsa-mir-221

MicroRNA, hsa-mir-221/222 was shown to negatively regulate ERα expression and associate with tamoxifen resistance *in vitro* (Zhao *et al.*, 2008). Transient knockdown of this miRNA caused breast cancer cells to re-express ER and recover sensitivity to tamoxifen. In contrast, ectopic expression of miR-221/222 rendered parental MCF-7 cells resistant to tamoxifen by reducing the protein levels of the cell cycle inhibitor $p27^{KIP1}$, and increasing ErbB2 expression (Miller *et al.*, 2008). I performed qPCR on our human primary tissues (training set) cDNA for hsa-mir-221, and for *ERBB2* as a cross study comparison to the references above. Here the p values for hsa-mir-221 and ERBB2 are of significance, 0.0484* and 0.0310* respectively for the TR versus the TS cohorts. This suggests that our study sample of patients is consistent with the expected findings from the published data stated above.

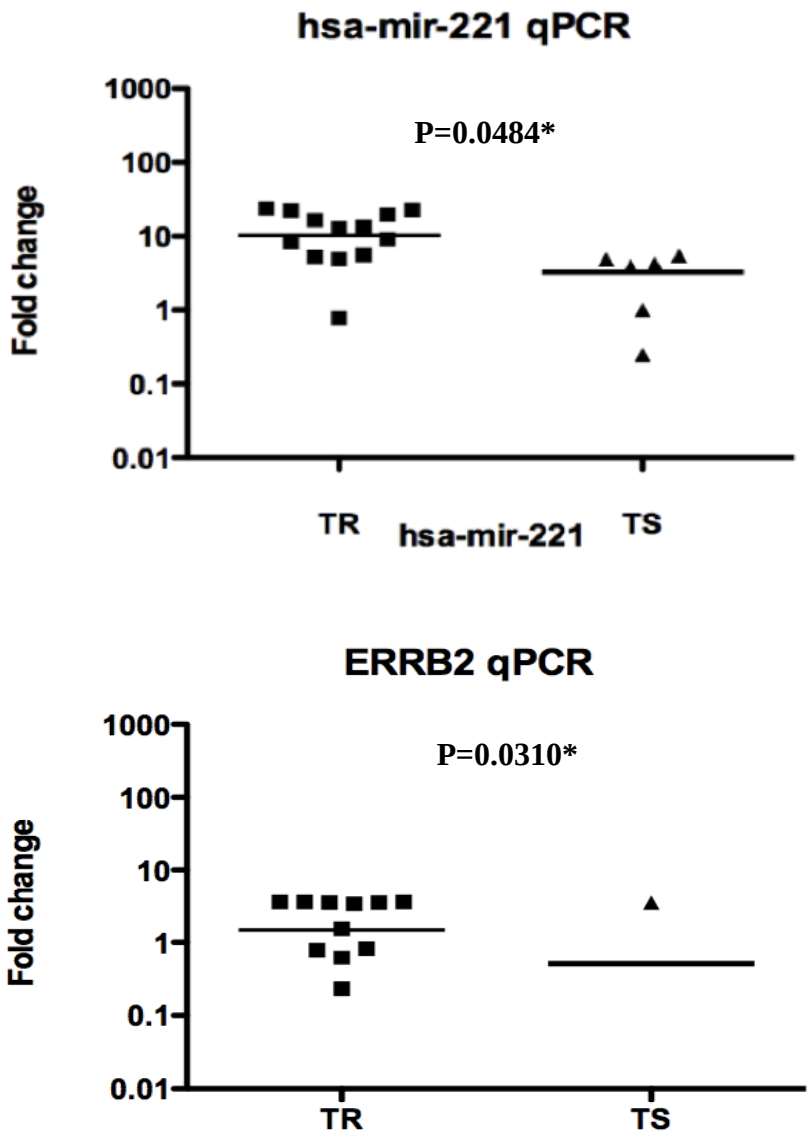


Figure 4.17 Expression of has-mir-221 and ERBB2 are significantly associated with TR. Scatter-plot showing the fold change of mRNA of hsa-mir-221 and ERBB2 real-time PCR of the tamoxifen resistant primary breast tissue (TR) as compared with the Tamoxifen sensitive primary breast tissue (TS), in n=25 of our training set. Stock solution of cDNA (5μl) of each patient were plated in triplicates and added with mastermix and Tagman®probe for hsa-mir-221 and ERBB2. Standard 40 cycles of real time PCR program using the ABI7500. P-value analysed using Mann-Whitney non-parametric equation, using Prism software.

4.13.3. miRNA in Tamoxifen Resistant Breast Cancer Cell lines

Here expression of the two miRNAs (hsa-mir-657 and hsa-mir-221) was also investigated in parental and TR breast tumour cell lines (see Chapter 3.0). As listed in table 4.16, we extracted miRNA and protein whole cell extract using miRVana PARIS kit from these parental and their TR counterpart breast cancer cell lines. I made cDNA from the miRNA extracted using the microRNA high-capacity cDNA kit (see Section 2.13) following their RT-PCR protocol. Real-time PCR were then performed on the samples with hsa-mir-221 and hsa-mir-657 primers. There is a consistent fold change increase of hsa-mir-221 in 6 out of 9 TR paired breast cancer cell lines. Some fold-change was more significant than others. These results were done in duplicates with very tight error bars (Figure 4.18).

A

	ER positive breat cancer cell lines							ER negative	
	MCF-7	T47D	MDA-MB-361	HC1500	HC38	H3396	ZR75-1	453	BT20
WT	35	1573	185	0	2	303	5619	23899	1
TR	70	1840	29	0	4	774	18887	4980	1.7
FC	2	1.2	0.15	0	2	2.5	3.3	0.2	1.7

B

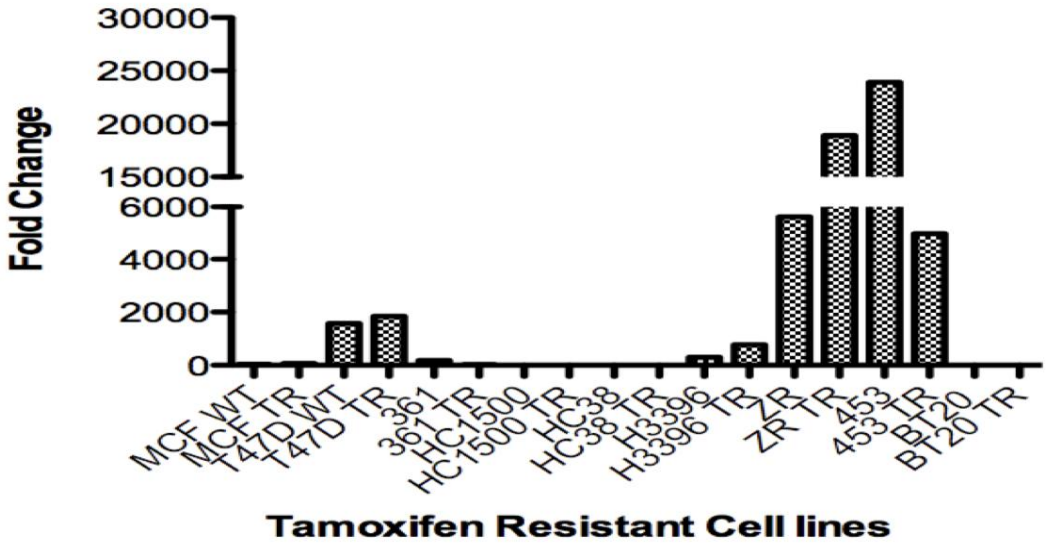


Figure 4.18 (A) Expression of hsa-mir-221 in TR and parental breast cell lines. In the table the cell lines are presented as ER-positive breast cancer cell lines; MCF-7, T47D, MDA-MB-361, HC1500, HC38, H3396 and ZR75-1, or as ER-negative; MDA-MB-453 and BT20. WT (wild-type) and TR (Tamoxifen resistant) are shown as their absolute amount after normalising with endogenous control, SNU24. Five µl of the stock cDNA is used for real-time PCR. (B) The fold change (FC) was measured as normalised (with miRNA endogenous control SNU24) expression of hsa-mir-221, as seen in the graph below the table.

One possible explanation for the undetectable has-mir-221 in HC1500 could be that it may not carry any miRNA hsa-mir-221 in its genome. An explanation of MDA-MB-361 and 453 had a down-regulation of hsa-mir-221 in their TR counterpart could be that as they are the only 2 breast cell lines (of my series shown in Figure 4.18A) which carries over-expression of ErbB2 in their wild-type form, this may have spared them from the route of hsa-mir-221, which we know from publication that hsa-mir-221 up-regulate ERBB2 mRNA expression but the reverse is not true.

In the experiment to detect hsa-mir-657 from this same TR breast cancer cell line series, we have been able to detect low levels. Although low, there is an up-regulated trend in the TR counterpart in cell lines apart from MDA-MD-361, H3396, HC38, ZR75-1 and MDA-MB-453 (Figure 4.19). HC38 and MDA-MB-453 both had undetected levels of hsa-mir-657 in their WT, which may suggest that these cells do not carry this particular miRNA.

A

	ER positive breast cancer cells							ER negative	
	MCF-7	T47D	MDA-MB-361	HC1500	HC38	H3396	ZR75-1	453	BT20
WT	undetected	0.001	0.01	0.01	undetected	0.346	0.02	undetected	0.00001
TR	0.007	0.06	undetected	0.08	undetected	0.05	0.01	undetected	0.0014
FC	0.007	60	undetected	8	undetected	0.144	0.5	undetected	140

B

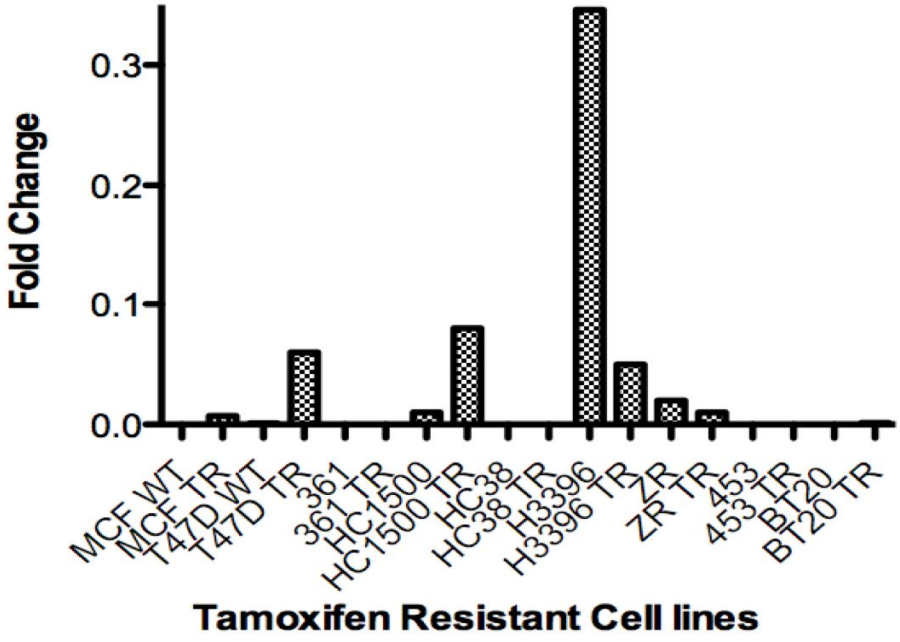


Figure 4.19 (A) Expression of has-mir-657 in TR and parental breast cancer cell lines. In the table the cell lines are presented as ER-positive breast cancer cell lines; MCF-7, T47D, MDA-MB-361, HC1500, HC38, H3396 and ZR75-1, or as ER-negative; MDA-MB-453 and BT20. WT (wild-type) and TR (Tamoxifen resistant) are shown as their absolute amount after normalising with endogenous control, SNU24. Five μl of the stock cDNA is used for real-time PCR. (B) The fold change (FC) was measured as normalised (with miRNA endogenous control SNU24) expression of hsa-mir-657, as seen in the graph below the table.

CHAPTER 5: DISCUSSION

5. Breast Cancer Derived Cell Lines And Human Primary Tissue Molecular Profiling Study

Tamoxifen is the oldest and most commonly used drug against oestrogen receptor α (ER+) breast cancer. Tamoxifen treatment in the adjuvant setting reduces the recurrence rate and improves overall survival; when used for treatment of metastatic breast cancer, it provides remission for over half the patients. De novo and acquired resistance to tamoxifen are important clinical problems since almost all metastatic patients, and up to 40% of adjuvant patients, will relapse and die from breast disease. Despite many studies using derived cell lines with selected tamoxifen resistance (TR), and many studies on resistant breast cancers, mechanisms of resistance are still not fully understood.

What we know so far is that progesterone receptor-negative (PR-) status in ER+ cases has been shown to be an independent predictive factor for benefit of adjuvant tamoxifen (Dowsett et al., 2005; Howell et al., 2005). Patients with highly expressed HER2/ErbB2-positive cancers also fail to benefit from tamoxifen treatment (Mc Ilroy et al., 2006; Piccart et al., 2001; Shou et al., 2004; Smith et al., 2005). There is an inverse relationship between Progesterone and HER2 receptors (Lal et al., 2005). It was suggested that the overexpression of HER2 may activate mitogen-activated protein kinase, which in turn activates ERα by phosphorylation at Ser^{118} {Yamashita, 2008 #33; Thomas, 2008 #36; Yamashita, 2005 #51; Thomas, 2008 #187; Yamashita, 2005 #188} and AIB1 (an ER co-activator) may be activated by signalling downstream of Her2, and in the presence of the two, the agonist activity of tamoxifen may be enhanced {Azorsa, 2001 #95;Reiter, 2004 #87;Tikkanen, 2000 #189;Azorsa, 2001 #190}. Stabilisation of the interaction between ERα and SRC1 by cyclin D1 were reported to be related to resistance in vitro (Nakuci et al., 2006).

These advancements are important as individual predictive biomarkers but the challenge for a full understanding of the mechanism of tamoxifen resistance is far from complete. Continual selection of pathways which constitute tamoxifen

resistance in a wider range of breast cancer derived TR cell lines (such as in this thesis), in parallel with confirmatory analysis of these markers in situ in large breast cancer patient cohorts will define the pathways leading to tamoxifen resistance. We hope by understanding these pathways reliable target(s) for therapy can be identified.

Overall, there was a high degree of similarity with the results of our TR (breast cancer cell lines) profiling with those of published data defining oestrogen-responsive genes in ER-positive breast cancer lines by microarray. ESR1, PGR, GATA3, GATA4, MYBL, GREB1, GREB7, PDZK1, SCUBE2, IGF1R, ErbB2, CYBSR1 and AR genes have all been identified in most ER receptor gene expression profiling studies (Ma et al., 2009). Loi et al (Loi et al., 2005) showed a chosen '13-gene cluster signature' to validate on independent cohorts with 32%-45% correlation with the Dutch group (Knauer et al., 2010) and 21-gene signature (Mamounas and Paik et al., 2010). These are explained in greater detail in section 5.4. As we had good Quality Control (QC) control for our probe preparation and our data (see result Chapters 3 & 4), as seen as a clear distinction of the test TR from reference TS groups, it is somewhat not surprising that our data had good correlation with the published group. Our bioinformatics analyses showed similar correlation, namely MYBL, GREB1, PDZK1, EGFR, ErbB2, ESR1 in our TR breast cancer cell line study, and c-MYC, PGR, ESR1, GATA3, IGF1, IGFBP6, AKR1C3 from the human primary breast tissue (predictive biomarker) when we analysed against the available breast cancer gene-expression database of the cohorts mentioned above from their publications.

For the breast cell line work, from which our microarray selected oestrogen dependent and oestrogen-independent (as we also array oestrogen-deprived breast cancer cell lines) significant molecules associated with TR, I have selected less characterised genes for further studies. This included genes that have no published association with tamoxifen resistance (TR) or breast cancer but some association with drug resistance in the literature. Our aim was to characterise the function of the genes and the pathways, which may be supporting TR. Full details of the TR cell

line microarray are seen in Appendices A. I have chosen the genes (i.e. MAGEA2 and EGLN3) with proliferation pathways to study further.

5.1. MageA2

MAGEA2 was chosen as the first gene, as it was 4-fold upregulated in TR, and had been suggested to have a role in chemotherapy (Etoposide) resistance in melanoma cell lines (Monte et al. 2006). *MAGEA2* was not detected in our breast cancer cell line panel (MCF-7, ZR75-1, MDA-MB 453, H3396, HCC1500, HCC38) apart from T47D (mutant p53), MDA-MB-361 (WT p53), SKBR3 (mutant p53) and MDA-MB-436 (latter two both ER-/PgR- cell lines). It was also not present in normal tissue apart from placenta and germ cells. It is however up-regulated in TR cell lines, which makes it a potential predictor gene for tamoxifen resistance (see Figure 3.3) as well as a target gene/protein for reversing tamoxifen-resistance. The presence of MageA2 in an ER-receptor negative cell line suggests that it is regulated independently of the ER and this is supported by evidence from the literature examining MageA expression in tumours (Cho *et al.*, 2006) Grigoriadis *et al.*, 2009). As seen in our microarray cell line study, MAGEA2 was up-regulated in all our tamoxifen-resistant cell lines (T47D TR, ZR TR, as well as the oestrogen-deprived TR cell lines, OD T47D TR and OD ZR TR).

Our work focused on the mechanistic relationships among MageA2, p53 and p21 and deacetylation of p53 in breast cancer cell lines. First, I showed that the stable expression of MAGEA2 in MCF-7 and T47D lines had a positive effect on proliferation and decreased apoptotic events in normal media and tamoxifen containing media (see Figure 3.7 and Figure 3.8). Compared to the difference in growth between the control (VA), and the MAGEA2 expressing clones, the significant difference was more marked (p-value=<0.001) in tamoxifen-containing media. Second, we showed a direct interaction between p53 and MageA2 in our overexpressing lines, with a consequence of reduced p21 levels (see Figure 3.9).

Reduced p21induction in these lines on exposure to tamoxifen suggested that the complex with MageA2 may repress the activation potential of p53. Monte M et al (2006) have proposed that MageA2 represses p53 by recruiting HDAC3 in a co-repressor complex with MageA2 acting as a scaffold protein bringing p53 and HDAC3 together. In my study (See Chapter 3.7.4), in co-immunoprecipitation experiments suggested that MageA2 interacts directly with p53, leading to reduced levels of acetylated, transcriptionally active p53 and hence reduced expression of target genes such as p21 resulting in continued growth in tamoxifen-containing media. The total available acetylated-p53 is reduced in the stable overexpressed MAGEA2 clones. .

Yang B et al (2007) also suggested Mage can suppress p53-dependent apoptosis but that a range of MageA-, B- and C- proteins had similar activity. We aim in future to examine if other Class I Mage antigens can also confer tamoxifen resistance by testing MageA3, MageB and MageC overexpressing cells in the same way I have done for MageA2-expressing lines.

In the MAGEA2 stable-expression clones I showed co-localisation with p53 (wild-type) using confocal immunoflourescence. There was a trend for cell survival when the staining of MageA2 was found in the cytoplasm (as seen in thriving TR cells). In contrast, in dying or apoptosing cells, such as control, VA in tamoxifen-containing media, MageA2 were concentrated in nucleus. This supports the hypothesis that MageA2 is able to sequester the available wild-type p53 in the cytoplasm essentially reducing available p53 (and acetylated p53) for interaction with growth inhibition/apoptosis-related target genes, thereby conferring a cell survival/proliferation advantage, as seen particularly in media containing tamoxifen.

In the immunohistochemistry validation of an independent cohort (n=125), patients staining positively for MageA (pan-MageA, Zymed antibody) were statistically significantly worse off in their overall survival compared with those who were

negatively stained for MageA (see Figure 3.13). In addition, when all of those who were positively stained were statistically analysed with the Kaplan-Meier curve for survival, those who stained for MageA in the cytoplasm did significantly worse than those who stained positive in the nucleus (p=0.0448). This study also showed that MageA had a 62% sensitivity rate in relapse TR tissue samples, and 38% in primary human tissue of patients we know are TR. Several cancer/testis antigens (including MageA) have already proved to be useful biomarkers for several types of cancer: breast ER negative receptor (Grigoriadis A et al, 2009), lung (Van den Eynde and van der Bruggen, 1997), melanoma (Carrasco *et al.*, 2008), pharyngeal (Pastorcic-Grgic *et al.*, 2009) and colorectal cancer (Toh *et al.*, 2009). The work of Grigoriadis et al suggested that MageA expression is most commonly found in ER-negative breast cancer. Interestingly, the proposition of MageA staining tumours in ER-positive cases doubled in their cohort of metastatic tumours compared to the primary cohort. This supports our suggestion that MageA may be a good biomarker for picking up TR cases. Even though all our cases were ER-positive, my data imply that MAGEA may be used to predict the patients who will be more likely to become TR, despite their favourable ER-positive receptor status. More importantly, MageA poses as a real candidate biomarker for tamoxifen resistance prediction (sensitivity of 38% (95%CI; 0.309-0.464), specificity of 80% (95%CI: 0.639-0.918, Positive predictive value=89%: 95%CI: 0.802-0.958). This is an exciting finding because we are now in an era, where we know Mage-A proteins have pivotal roles in many cancers, and in resistance to treatment. Moreover, there is a vaccine available against MAGEA3, which has a safe profile to date for Phase 2 treatment in non-small-cell lung cancer, colorectal cancer (Toh HC, 2009) and melanoma.

In this regard it is interesting that MAGEA3 was also upregulated in our collaborators microarray of tamoxifen and fulvestrant resistant breast cell lines (Dr Julia Gee, personal communication), and it is also found at the protein level in our TR lines. If we can also establish that overexpression of MAGEA3 also confers resistance to tamoxifen in cell lines this would indicate that the commercially available vaccine against MAGEA3 may be useful in the treatment of TR/metastatic

breast disease where MAGEA upregulation is detected. Since MageA antigens are highly homologous, it is likely that the vaccine may induce an immune response against a number of MageA proteins, so the precise expression profile may not be important.

I am currently undertaking an animal study designed to compare the proliferation of an MAGEA2-expressing MCF-7 clone with the VA control when both lines are grown as xenografts in overectomised mice. Mice were implanted with pellets containing either oestrodiole or tamoxifen or both. A different growth response has been seen in the tamoxifen pellet alone mice, with only the MAGEA2 clone forming tumours. As we have not reached the endpoint of our animal study yet, I have not described this aspect in the results chapters, but the aim is to test the MAGEA2-expressing cells in a more in vitro setting. Ultimately we may be able to use immunocompetent mouse model, to see if the vaccine will reverse the tamoxifen-resistance of the MAGEA2- (or MAGEA3-) expressing cells.

5.2. EglN3

ELGN3 is synonymous with egl nine homolog 3 (C. elegans) and is a member of the egg-laying-defective 9 (EglN) prolyl-hydroxylases. This was one of the most significantly upregulated genes in our TR study but with the least published data on the gene's relationship with breast cancer. EGLN3 is one of the three known prolyl hydroxylases which catalyse the hydroxylation of Hypoxia-induced factor (HIF). The heterodimeric HIF (a transcriptional regulator) is regulated by proteolysis of its alpha-subunits, following oxygen dependent hydroxylation of specific prolyl residues. Examination of the literature has shown that the prolyl hydroxylase, EGLN3 can be alternatively spliced to produce both active and inactive forms (Cervera et al., 2006).

It is highly interesting that EGLN3 was found to be 3.8-fold upregulated in our tamoxifen-resistant breast cancer cell lines. As little is known about how, and in what conditions EGLN3 is expressed, I carried out an in-vitro study on EGLN3-expressing breast cancer lines in MCF-7 and T47D. I aimed to study its role in proliferation, and apoptosis, in normal and tamoxifen-containing media. In addition I aimed to localise its distribution in cells to establish its site of action, when conferring a cell survival advantage.

EGLN3-expressing clones were found to have a proliferation advantage in both normal and in tamoxifen-containing media, and reduced apoptosis in tamoxifen containing media as demonstrated by two different proliferation assays, the cell count study and the Annexin V and PI study. The growth advantage of the EGLN3-overexpressing clones in tamoxifen-containing media may be in part attributed to the simultaneous increased expression of HIF1α, MageA2 (in particularly for MCF-7 C7) and most importantly pRb (in particular for T47D, see Figure 3.21). We have shown the knockdown EGLN3 in our clones (rescue knockdown) reversed the cells into sensitivity to tamoxifen again (see Figure 3.26).

It is intriguing that MCF-7 EGLN3-overexpressing clones possessed a significant proliferation advantage even in their normal media, with no notable change in the protein level of HIF1α (see Figure 3.25). In contrast, T47D EGLN3-overexpressing clones had a trend for increased proliferation (although this did not reach statistical significance) in their normal media, but the HIF1α protein level was notably down-regulated when compared with WT and VA (see Figure 3.25). Excess of EGLN3 could theoretically generate more of the unique binding site for ubiquitin ligase complex containing the von Hippel-Lindau (VHL) tumour suppressor protein, which results in HIF1α destruction. This should divert away from the pathway of the master transcription regulator of hypoxia inducible (HIF) genes. I hypothesise that there are alternate pathways, which is contributing to the proliferation advantage of the EGLN3-expressing clones.

A number of cellular oncogenes have been reported to promote HIF stabilisation, which facilitates tumour growth. Oncogenes such as v-Src and activated Ras, block HIF1α prolyl hydroxylation and therefore lead to HIF accumulation (Chan *et al.*, 2002). In contrast, the PI3K/AKT pathway promotes HIF accumulation without any change in HIF hydroxylation, possibly through the activation of mTOR and HIF translation (Zhong *et al.*, 2000), (Arsham *et al.*, 2003), (Brugarolas *et al.*, 2003), (Zundel *et al.*, 2000).

In T47D EGLN3-expressing clones, the HIF pathway is adequately switched off by the hydroxylation of HIF by EGLN3 even in normoxia condition. I showed that there is a compensatory increase in pAKT and Raf protein level (Figure 3.26) in T47D EGLN3-overexpressing clones, which might account for the slight cell proliferation advantage. Conversely, MCF-7 EGLN3-overexpressing clones may be using the Ras pathway as shown in Figure 3.26, where the protein level of Raf is up-regulated in the clones. This is consistent with the finding of accumulation of HIF1α due to inhibition of HIF prolyl hydroxylation despite the presence of EGLN3. It is still unclear how EGLN3 switches on the Ras/Raf pathway in MCF-7.

The next step would be to establish if the hydroxylation activity of EGLN3 is required to confer the proliferation advantage in overexpressing lines. We intend to study this by transfecting MCF-7 and T47D wild-type cells with an expression construct for "catalytic-dead" (H196A) EGLN3 (given as a gift by Dr W Kaelin) and perform a further in-vitro study on the proliferation of these clones in their normal and tamoxifen-containing media. In addition, Dr Kaelin's group in Boston is undertaking a screen of our EGLN3-expressing clones to see if they can identify any novel hydroxylation targets linked to tamoxifen resistance since they have preliminary data suggesting the existence of non-HIF targets in some breast cell lines. This group has recently published (Zhang et al, 2009) that the related EGLN2 regulates cyclin D1 activity and tumourigenesis in ER+ breast tumour lines, although the pathway between EGLN2 and CCND1 was not defined.

As EGLN3 was found to be upregulated as a consequent of the breast cancer cells developing Tamoxifen resistant, we also validated them across a series of 196 of independent primary breast cancer tissue of patients treated with Tamoxifen, as an adjuvant endocrine treatment, to assess if EGLN3 is a good predictive biomarker. EGLN3 has a positive predictive value of 85% (95% CI: 0.75-0.92), with a sensitivity of 42%, and specificity of 66% of all TR primary breast tumour.

The up-regulation of MAGEA2 mRNA and protein (as seen in Figure 3.21, 3.22 and 3.24) in EGLN3-overexpressing clones was somewhat a surprise finding in our study. We know that in the MAGEA2-overexpressimg clones, there was no increase in EGLN3 mRNA or protein (data not shown). I have established that MageA2 is not directly downstream of Egln3 in our rescue knockdown study of the EGLN3-overexpressing clones (see Figure 3.26). As discussed above, MageA2 has been shown to sequester p53, and subsequently reduce available acetylated p53. In Figure 3.26, I have shown that T47D EGLN3-overexpressing clones have reduced acetylated p53. This further supports our hypothesis that MageA2 is up-regulated concurrently in EGLN3-overexpressing clones. In our microarray study of TR cell lines, both of these genes were the most widely up-regulated. As to how precisely EGLN3 influences MAGEA2 expression remains unknown, and needs further study.

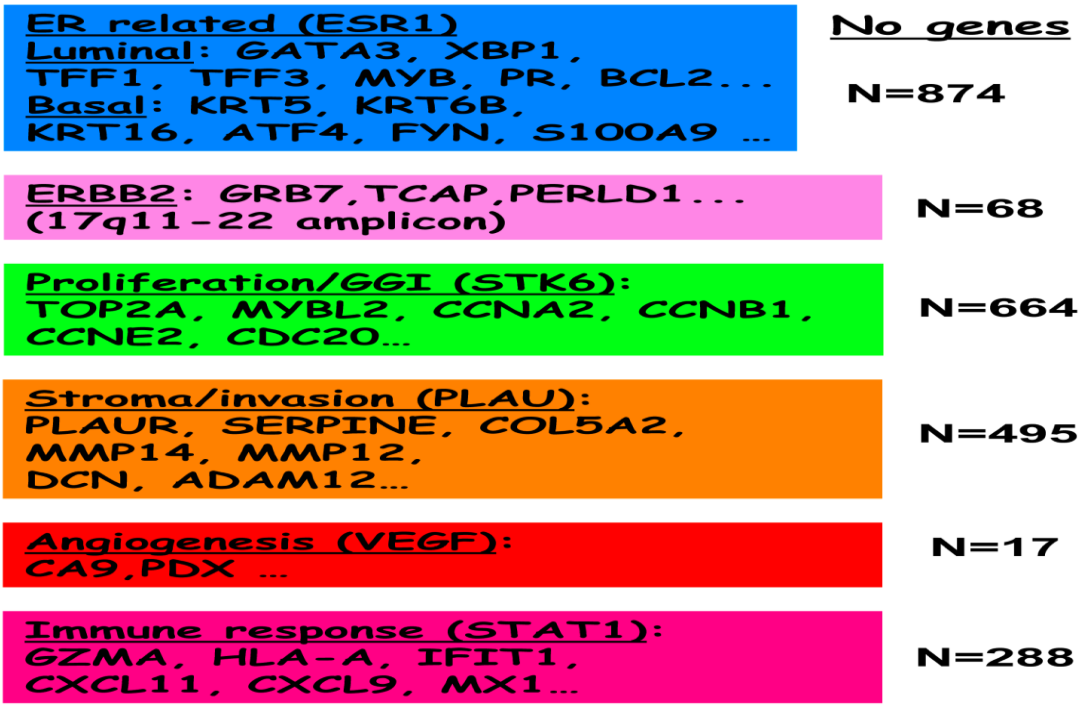
Should time and funding permit, we would like to repeat the rescue knockdown of EGLN3-expressing clones experiments, to be certain that MageA2 id definitely not directly downstream of Egln3. Then we will carry out knockdown with siRNA pool-EGLN3 experiments on the T47D TR and ZR TR cell lines. We will then test to see if this has any effect on MageA2, which we know is overexpressed in TR cell lines. We could also knockdown a wild-type breast cell line which carries a high level of EGLN3, such as MDA-MB-361 (see Figure 3.16), and screen the cell line for MageA2 when efficient knock down is achieved. Intriguingly, MDA-MB-361 was one of the four WT breast cancer lines, which expressed detectable levels of MageA2 (see Figure 3.5A).

5.3. HUMAN BREAST CANCER STUDY

The main aim of our study from the human TR breast cancer specimens was to try to identify a highly selective and specific predictive molecular signature, which would predict response to tamoxifen in primary breast cancer tissues of patients. Currently, there exist three groups who are playing major roles in molecular profiling in oestrogen receptor-positive breast cancer. The initial group under Christos Sotiriou (Loi S et al, 2008) developed a gene classifier that predicts clinical outcome in tamoxifen-treated ER+ breast cancer. The gene classifier consisted of 181 genes belonging to 13 biological clusters. Six of the 13 gene clusters presented pathways involving cell cycle, migration, angiogenesis, ER-related, ERBB2-related and proliferation (see Figure 5.1). In an independent set of adjuvant treated primary samples (n=362, ER+ breast cancer with adjuvant tamoxifen treatment), the classifier from the 'proliferation' gene set was able to define a distinctly poor prognostic group (p=0.00000001), which was termed gene expression grade index (GGI). The six clustered pathways have a high degree of similarity with the second group (Paik, 2006), whose work has led to the success of a predictive gene chip, which is now widely used in the United States, known as Oncotype DX (see Introduction Chapter 1.8.1). They use a 21-gene signature (see Figure 5.2) to calculate a recurrence score (RS) which predicts if ER-positive patients, who are lymph node negative may fall into a high risk category that will benefit adjuvant chemotherapy, which conventionally if classified by histological (i.e. lymph node-negative) parameters, would not be offered adjuvant chemotherapy. The 21-genes broadly cover the same sub-clustering of biological pathways as the GGI, namely oestrogen-related, ERRB2-related, proliferation, migration, and inflammation. Although each group used different genes in each cluster, the bulk of the gene signatures prior to pruning down to the crucial few had similarities.

A third group at The Netherlands Cancer Institute (van de Vijver et al., 2002) has generated a 70-gene signature predictive for lymph-node negative patients for ten-year metastasis-free survival. This has now been made available as a commercial

gene chip (Mammaprint®) approved by the FDA in the United States. Both the 21-gene (OncotypeDx®) signature and the 70-gene signature chip have overlapping genes which both the signatures share.



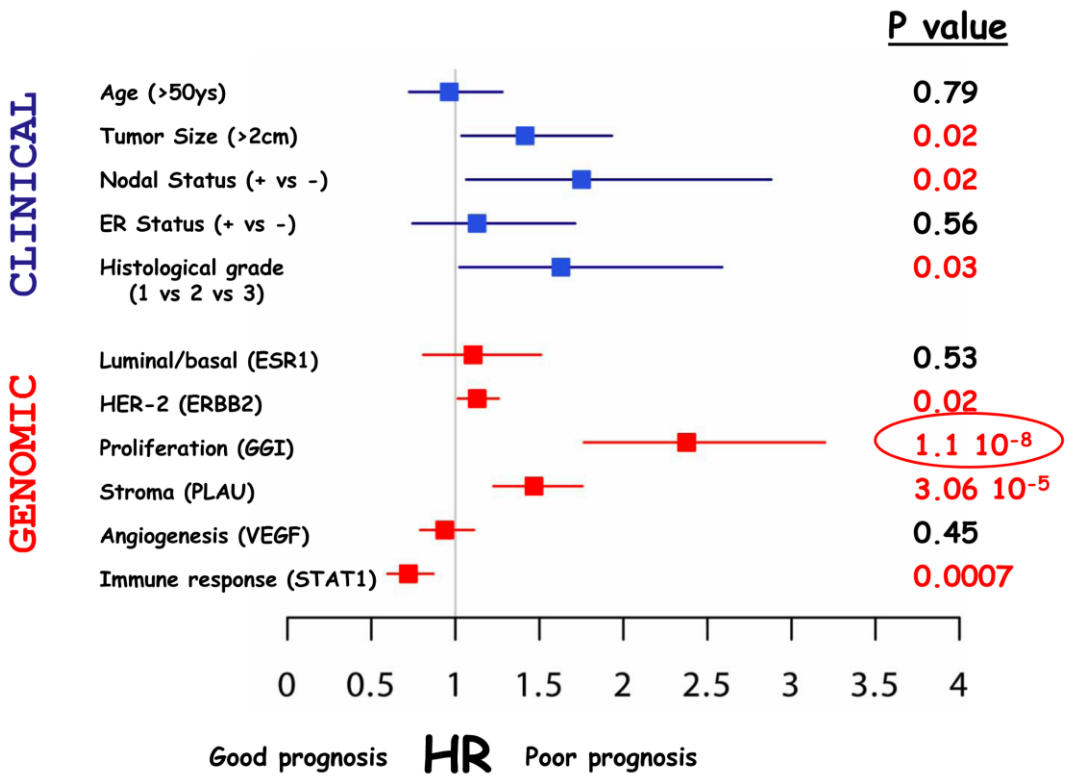


Figure 5.1 (A) The six clusters of genes, which have been found to predict prognosis of ER-positive breast cancer patients who had anti-endocrine treatment by the Jules Bordet Institute. The 6 biological clusters from (Gene Grade Index) were broadly labelled ER-related, ERBB2 related, proliferation/GGI, stromal invasion, angiogenesis and immune response. N=no of patients which carried the genes in each cluster. The analysis was found from n=2000 array study, or primary breats cancer treatment. (B) Forest plots of hazard ratios obtained from the six clusters when compared with conventional pathological classification with prognosis. Here, we see that the proliferation (GGI) and the stromal invasion cluster have the highest p-values in the prediction of prognosis.

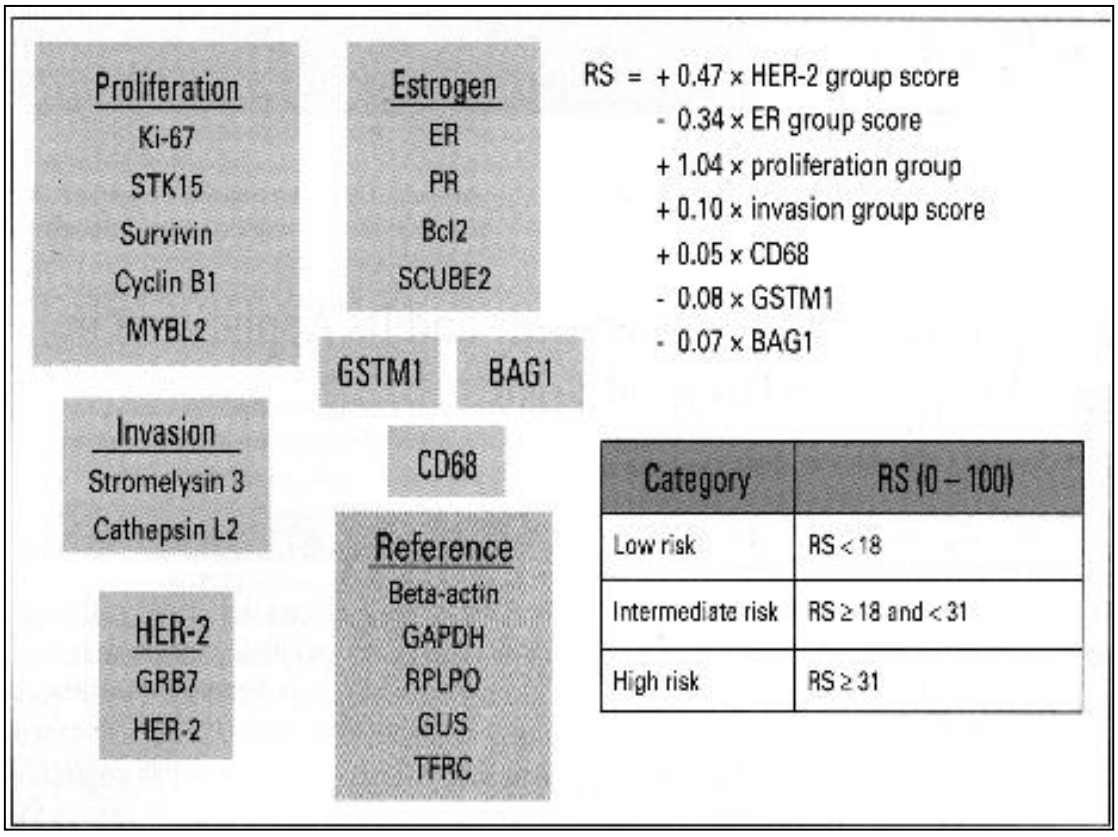


Figure 5.2 Oncotype Dx (Genomic Health) recurrence score (RS) genes and algorithm. HER, human epidermal growth factor; ER, oestrogen receptor; PR, progesterone receptor.

Predictive gene profiling and genomic signature studies have many limitations. As in any predictive study, the signatures which can directly cause an event are usually more diverse and variance between one person to another, influenced by their genetics i.e. racial background, cell or organ susceptibility, or influence by environment i.e. the epigenetic, nutritional or prescription poly-pharmacy such as cardiovascular drugs, the contraceptive pill etc. For this reason, predictive genetic studies ideally need to be large and prospectively planned to generate adequate statistical power. Here, we acknowledge the limitation of our small study with only 25 fresh frozen samples, but we have tried to maximise the outcome by integrating data from genome-wide SNP6.0 chips (1.2 Million SNPs, previous generation 500K had 500,000 SNP features) with Human Gene®Chip Exon 1.0 ST Array (6.5 million features; the previous generation of gene-level arrays had approximately 600,000). In addition, we have combined the in-vitro breast cell line TR gene expression data with that of our human predictive data to mine for related pathways with associated potential mechanisms of TR (from the breast cancer cell lines) with that of the predictive signature.

The three chromosomes, which had the most significant regions of copy number aberration, were chromosomes 11, 17 and 8 (see Chapter 4.11). The region 8q24 (146 out of 152 significant SNPs localised to chromosome 8) not only represents a large proportion of Copy Number Variation (CNV) regions of TR human predictive signatures, it also encompasses the canonical pathways which to date have been associated with tamoxifen resistance mechanisms of action namely p53 signalling, Integrin, IGF-1 signalling and hormone steroid pathways (see Chapter 4.12).

Genetic variation in this region of 8q24, a few hundred kilobases telomeric to the Myc amplicon (8q24.21; 128817498 to 128822853) has also been associated with increased susceptibility to prostate and colorectal cancer and has been described as a "gene desert". Recent work has identified a series of enhancer elements that act to regulate transcription of the Myc gene (Sotelo et al, 2010). Multiple genetic variants on chromosome 8q24, in particular rs13281615 have been found to be associated with an increase in breast cancer risk (Easton *et al.*, 2007). This specific region was amplified in my SNP data in ten out of 17 of our TR cases.

The group of Dennis Sgroi has studied the 17q amplification region (Goetz *et al.*, 2006) in specific subsets of ER-positive patients who either developed recurrence or who did not while on tamoxifen monotherapy. Our study design consists of the same patient criteria, apart from the fact that they had a larger cohort, and used different platforms. A novel approach would be to look into regions of significant SNP (copy number changes) aberration, which are sensitive across a high percentage of TR specimens and look into the 'flanking regions' for translocation footprints. There is already published data on the amplification of genes on 17q22; such as *HOXB13*, *COL1A1*, which are positioned in the second of three regions of the 17q12 HER2 amplicon {Jansen, 2005 #34; Sgroi, 2004 #39}. We have correlating significant data: 58.8%; 10 out of the 17 TR patients had amplification, whereas less than 2 out of 8 TS patients had amplification in SNP regions at the 17q21-25 arm. In fact this region is the second most aberrant site in our genomic analysis (see Chapter 4.11). The expression patterns of genes and variation in SNP regions from our study are

centred around chromosome 17q21 to 17q22 and possibly correlated with *HER2/ERBB2* amplification, whose over-expression is associated with reduction of response to first line endocrine therapy (Musgrove & Sutherland, 2009). We hypothesise that the control of TR may originate from this region, and set about to test the genes, Sp2 (see Chapter 4.12.1 and Table 4.7), their splice variants (future work should time permit), if any, in this region when we complete our full combined statistical analysis.

I have also identified at least 5 very specific biomarkers, PALM2 (9q31.3), SNF1LK2 (11q23.1), ZBTB16 (11q23.2), OR10G7 (11q24.1) and OPCML (11q25), which can be used for prognostic (predicts risk of recurrence) and predictive (i.e. predictive of tamoxifen response) of TR. Interestingly, 4 out of these specific biomarkers for TR are on chromosome 11q23-25. The statistical analysis of the immunohistochemistry validation on 125-independent cohort of primary breast cancer tumour from patients who had been treated with Tamoxifen, showed that these genes individually is an acceptably specific biomarker (OR10G7, 48% for TR, SNF1LK2, 66%, PALM2, 44% and ZBTB-16, 56%) and relatively high sensitivity (OR10G7, 54%, SNF1LK2, 45%, PALM2, 37% and ZBTB-16, 58%) but more importantly, they carry high negative predictive values (OR10G7, 68%, SNF1LK2, 71%, PALM2, 59%, and ZBTB-16, 73%). This means that as a biomarker, if the genes/proteins are stained negatively, it has a high likelihood to be true tamoxifen sensitive. (We intend to study this region further in the future, should time and funding permit). These have low false positive, (but extremely high specificity) but also a low sensitivity for all TR. In part this is in keeping with our claim which we made regarding the limitations of this study as predictive signatures are much more diverse.

5.4. The role of microRNA-657 in predicting tamoxifen resistant in primary breast cancer tissue

This part of the study utilised micrarray approach, in particular the new SNP6.0 where the probes include specific regions of miRNA within the genome, and the combined integrated analysis of mRNA expression array, Exon ST1.0 (with the largest number of probes in an expression chip), to further investigate novel signalling pathways regulated by microRNAs (miRNAs), which may be predictive of tamoxifen response.

I previously carried out a microarray study on breast cancer cell lines, which I made tamoxifen resistant. The significantly altered genes from this study are likely to represent genes involved in cellular mechanisms of escape from the control of tamoxifen. Whereas from the SNP6.0 and Exon array studies of the human frozen specimens, genomic susceptibility and predictive expression patterns inform us of the signatures which could make an individual more likely to develop resistance. As miRNA play a critical role in gene silencing, I set out to find a specific miRNA which has potential control a large set of ER-positive breast cancer related genes. I also took an interest in miRNAs, which can act as onco-miRs or tumour suppressors (see Table 4.6). Hsa-miR-657 was chosen as the most interesting miRNA for further study from our combined analyses. This miRNA has been found to potentially control the post-translational expression of many of our TR cell line genes (where the microarray study included oestrogen-deprived breast cancer cell lines made into TR: hence the pathway of TR is likely to be non-ER receptor dependent), including THRAP5, HOXB13 (also found by the Sgroi group as an accurate gene predictor of response to tamoxifen), MMP9, AKR1C3, ESRRB, ERBB2, TGF-β, and also TP53. A high percentage of our TR human primary breast cancer tissue specimens carried this miRNA, 10 to 11 out of 17 cases of TR. Conversely, less than 3 out of the TS cohort carried this miRNA.

There has been recent increased interest in miRNA and their potential role in endocrine therapy resistance. Four published reports have shown miR-221/222 to play a role in TR, though the two different groups showed different targets, $p27^{KIP1}$ (Miller *et al.*, 2008) and ER (Zhao *et al.*, 2008), that are negatively regulated at the protein level. The latter group found that miR-221 and miR-222 inhibit ERα translation by direct interaction with the 3'-UTR of the ERα and hence suggest a molecular mechanism of ERα regulation at the post-transcriptional level in breast cancer. They also showed that by knocking down miR-221/222, they restored the expression of ERα and sensitivity to tamoxifen induced cell growth arrest and apoptosis.

In my study, I have shown that our TR human primary tissues have significantly higher expression of hsa-mir-221 by qPCR than the TS group. In keeping with the publication of Miller (Miller *et al.*, 2008), I have shown that our elevated has-mir-221 primary human breast cancer TR group also has significantly higher ERBB2 mRNA expression (see Chapter 4.15.2). This part of my study suggests that our study sample of patients is consistent with published data.

The second part of our miRNA study, examined if hsa-mir-657 is detected in our TR patients. Due to a low level of mRNA expression, some of our TR samples, which we know from the SNP array study to carry this miRNA, did not detect any expression of hsa-mir-657 by qPCR. Although the qPCR validation was not significant between TR and the TS primary breast tissue samples, the trend for a higher level of hsa-mir-657 was found in the TR cohort. Interestingly, the protein expression for three proteins potentially regulated by hsa-mir-657, namely THRAP5, HOXB13 and KIAA1324L was decreased in the human primary tissue whole cell extract lysates (see Figure 4.16). Protein expression of p53 in the lysate of hsa-mir-657 carriers was not consistently down-regulated as expected, but this could be due to the fact that a crucial master transcriptional controller like p53 is likely to be controlled by more that one miRNA as well as multiple other regulatory

mechanisms. MicroRNA-657 is also elevated in half of our TR cell lines (in a set of ten breast cell lines and their paired TR counterpart) by qPCR (see Figure 4.19). These results suggest that further in-vitro studies, such as knockdown study using antagomiR-657 in tamoxifen-resistant cells to see any the effect on the reversibility of sensitivity to tamoxifen, is warranted.

Conclusions:

MAGEA2 and EGLN3 study:

1) From our microarray study of the tamoxifen-resistant ER-positive breast cancer cell lines, MCF-7, T47D and ZR75-1, MAGEA2 and EGLN3 is found to be 4.0 and 3.8 up-regulated. In my in-vitro study, I have made into stable MAGEA2 and EGLN3-expressing clones in MCF-7 and T47D cell lines and carried out functional studies. MAGEA2 and EGLN3-expressing clones (respectively) have proliferation advantage, and decrease apoptosis compared to their control, vector-alone. I have shown that MageA2 has a direct interaction with p53, sequestering p53, and decreasing acetylated-p53 in the cytoplasm and the nucleus.

2) In the immunohistochemistry validation study (n=196), MageA and Egln3 are good sensitive and specific predictive biomarker for tamoxifen-resistant in primary breast cancer tumours, with MageA having a sensitivity of 38% in primary breast tumours, and 61% (even though this is only a set of 9 patients as relapsed specimen are rare), and Egln3 having a sensitivity of 42% and sensitivity of 66% for picking up tamoxifen-resistant cases.

3) There is statistical significant p-value for the positively stained MageA cases in overall survival (OS); p=0.0455*. In addition, when the positive-MageA stained cases were analysed according to cytoplasmic versus nucleus staining, there was a statistical significance for the cytoplasmic MageA staining cohort to do worse than the nucleus-staining, p=0.0448*.

4) Egln3 had no statistical difference between the positively-stained from the negatively stained cases in disease-free and overall survival. However it is interesting that the EGLN3-expressing clones had elevated levels of MAGEA2 mRNA expression. And survival analysis with the Kaplan-Meier curve showed a trend for the combined positive, i.e. the positive-MageA and positive-Egln3 to do worse in overall survival when compared with negatively stained MageA and negatively stained Egln3, but this did not reach a statistical significance (see Figure 3.28).

Predictive signature from human breast cancer study:

1) From the human breast primary cancer tissue study, combined integrated analysis of the Exon expression array and SNP6.0 genomic array was undertaken with the aim of finding a predictive signature which includes the 3D-structure of breast cancer, which tissue array provides.

2) Five biomarkers have been found and validated, OPCML, OR10G7, SNF1LK2, PALM2 and ZBTB-16. Individually they have moderately low specificity but high sensitivity. However 4 out of 5 have very high negative predictive value, which means if stained negative, they are highly likely to predict tamoxifen sensitivity.

3) We are currently working in collaboration with the Wolfson Institute's statistical department to analyse if in combination, the 5 genes could give a reproducible predictive signature, which can predict accurately the probability of TR. This is very exciting as these genes are novel and have never been associated with the ER-related genes.

4) Four out of the 5 genes/protein, OPCML (11q25), OR10G7 (11q24.1), SNF1LK2 (11q23.1), PALM2 (9q31.3) and ZBTB-16 (11q23.2) is within the most significant region of copy number changes in the SNP6.0 studies. I am currently analysing with the detailed exonal mRNA changes in this region,

11q23-25, and also the other region of interest 17q21-22, to ascertain if there is a trigger variant, or a direct association of these genes in conferring TR.

miRNA study:

1) last but not least, the miRNA has shown promise as there is correlation with our human specimens and TR breast cancer cell lines with hsa-mir-221, and ERBB2 connection. We hope to expand this into a full in vitro study with our miRNA, hsa-mir-657 in the study of its role in tamoxifen tolerance.

The exact mechanisms that predict an individual to be tamoxifen-resistant is unknown, but in the last decade, we have come closer to assemble the pathways which may be associated with the downstream targets rendering a cell 'resistant', and also closer to the upstream genes activation which predispose an individual to TR. I believe in the near future, patients' care will include their inert inherited genes as part of their staging work-up, which will be an integrated factor in deciding the best possible treatment for that individual patient.

In an ideal world, a predictive chip would consist of few clusters of genes which covers all of the essential network pathways, that predicts response to a treatment or, predict prognosis. These two chips should be designed separately with specific targets. The training sets in array studies should be clearly defined.

If I were to design a chip predicting response to tamoxifen, I would like to include genes which are highly specific for TR, low false positive for tamoxifen sensitive, a reproducible high positive predictive value, and inclusive of all the known and the new novel genes (as comprehensive as any high-throughput genenic and genomic study vehicles can provide). The bearing of each gene will be calculated from a statistical formula, which should be validated on a large number of validation cohorts. Finally, all predictive genetic chips should be validated on a large independent cohort in phase 3 trials, with interim results available to the public.

6. Appendix A

NM_001386	8	8p22-p21	DPYSL2	dihydropyrimidinase-like 2	2.2124516	2.398451	3.318149576	2.618520848
NM_000992	3	3p21.3-p21.2	RPL29	ribosomal protein L29	2.377048	2.259531	2.489672898	2.087159692
NM_000980	19	19p13	RPL18A	ribosomal protein L18a	2.0281586	2.478309	2.098420623	2.212658777
NM_005567	17	17q25	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	2.1847384	-2.20147	0.567164335	-1.05516891
NM_004343	19	19p13.3-p13.	CALR	calreticulin	-2.012232	-2.57185	-2.444142065	-2.96287641
BG261322	2	2q11.2	EIF5B	eukaryotic translation initiation factor 5B	-2.152289	-2.04859	-0.9615939	-2.03465703
NM_004417	5	5q34	DUSP1	dual specificity phosphatase 1	3.272713	2.501547	2.511125335	2.828128158
NM_015383	1	1q12-q21.2	NBPF14	neuroblastoma breakpoint family, member 14	-2.482495	-2.34731	-2.074335603	-2.17446213
NM_002510	7	7p15	GPNMB	glycoprotein (transmembrane) nmb	3.0185007	-2.07107	3.555246213	-3.57630937
BF512200	3	3q25	MBNL1	muscleblind-like (Drosophila)	-2.436044	-2.15552	-1.188536273	-1.8407539
AF061337	X	Xp11.3-p11.2	DDX3X	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	-2.518827	-2.02372	-1.896422915	-2.06080062
AF010314	5	5q12-q13.3	ENC1	ectodermal-neural cortex (with BTB-like domain)	4.7599222	2.401336	4.047718386	1.959760113
NM_002084	5	5q23	GPX3	glutathione peroxidase 3 (plasma)	2.082958	-2.40925	2.440592669	-2.61692141
NM_006332	19	19p13.1	IFI30	interferon, gamma-inducible protein 30	2.2473576	2.066437	3.195750633	1.980403775
NM_001552	17	17q12-q21.1	IGFBP4	insulin-like growth factor binding protein 4	-2.85646	-5.02425	-4.214951091	-5.99938011
NM_006349	7	7q22.1	ZNHIT1	zinc finger, HIT type 1	2.9977887	2.151354	2.274312649	2.007800833
NM_002166	2	2p25	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix	2.1762404	2.113653	1.120548135	2.411279787
NM_005496	3	3q26.1	SMC4	structural maintenance of chromosomes 4	-3.559156	-2.27087	-2.573655864	-3.13263603
D42063	2	2q12.3	RANBP2	RAN binding protein 2	-2.164966	-2.20986	-1.868952354	-2.30889536
BF110993	1	1q25	TPR	translocated promoter region (to activated MET oncogene)	-2.926721	-2.22373	-2.695506724	-2.06730809
AW968555	X	Xp22.3	TBL1X	transducin (beta)-like 1X-linked	2.4288526	2.031994	2.841981672	2.241040589
NM_005881	16	16p11.2	BCKDK	branched chain ketoacid dehydrogenase kinase	2.8184181	2.510523	1.750661183	1.800709714
NM_004491	19	19q13.3	GRLF1	glucocorticoid receptor DNA binding factor 1	-3.373506	-2.30065	-2.984257439	-1.94865462
NM_000714	22	22q13.31	TSPO	translocator protein (18kDa)	2.0784775	2.295903	1.835014586	2.070033886
NM_003689	1	1p35.1-p36.2	AKR7A2	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde	2.0300397	2.535261	1.722096903	1.968308627
NM_001085	14	14q32.1	SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, an	2.0250667	-2.30128	0.916375438	-2.6911076
AA772747	11	11p15	POLR2L	polymerase (RNA) II (DNA directed) polypeptide L, 7.6kDa	2.8012013	2.064199	2.015185292	2.098437374
AI824012	21	21q11.2	NRIP1	nuclear receptor interacting protein 1	-2.870129	-2.53772	-1.285647678	-1.991115
NM_002801	16	16q22.1	PSMB10	proteasome (prosome, macropain) subunit, beta type, 10	3.8670283	2.073013	2.862266152	1.983772061
BF001670	13	13q33	EFNB2	ephrin-B2	3.9293173	2.063607	3.185365012	2.104887941
U16797	13	13q33	EFNB2	ephrin-B2	3.7130476	3.456397	2.280064968	3.533904592
NM_000666	3	3p21.1	ACY1	aminoacylase 1	2.596471	2.041848	1.883182838	1.796986451
NM_003721	19	19p12	RFXANK	regulatory factor X-associated ankyrin-containing protein	2.5186793	2.42672	1.941378469	2.055913575
NM_015180	14	14q23.2	SYNE2	spectrin repeat containing, nuclear envelope 2	-2.813489	-2.70576	-2.467162934	-2.81502078
NM_006012	19	19p13.3	CLPP	ClpP caseinolytic peptidase, ATP-dependent, proteolytic subur	2.3698628	2.571803	2.210028501	2.306018374
NM_004146	19	19p13.12-p13	NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18	3.2053032	2.201836	2.387459619	1.741195931
NM_006396	11	11q13.1	SSSCA1	Sjogren's syndrome/scleroderma autoantigen 1	2.9835159	2.13138	2.854401472	1.741611098
NM_005886	16	16q13	KATNB1	katanin p80 (WD repeat containing) subunit B 1	3.4653349	2.491474	2.387183469	2.084516397
NM_005886	16	16q13	KATNB1	katanin p80 (WD repeat containing) subunit B 1	3.8217877	2.741338	3.147921212	2.154248418
NM_002961	1	1q21	S100A4	S100 calcium binding protein A4	3.2194058	2.630335	3.595912352	2.022755692
NM_001380	10	10q26.13-q26	DOCK1	dedicator of cytokinesis 1	-3.303092	-2.00255	-3.273716495	-1.7667934
BC002791	9	9q34	FLJ35348	FLJ35348	2.4338135	2.052736	1.910632957	1.794422417
NM_005573	5	5q23.3-q31.1	LMNB1	lamin B1	-4.153992	-3.00838	-1.502928114	-2.94135516
NM_021136	14	14q23.1	RTN1	reticulon 1	3.505964	2.976013	0.424970867	3.077681524
NM_006829	10	10q23.2	C10orf116	chromosome 10 open reading frame 116	2.6548881	2.449518	0.647595912	1.849084507
NM_001661	17	17q12-q21	ARL4D	ADP-ribosylation factor-like 4D	2.1151045	3.581386	1.207171667	2.721324576
NM_012079	8	8q24.3	DGAT1	diacylglycerol O-acyltransferase homolog 1 (mouse)	2.4903318	2.026911	1.226889172	1.453179305
NM_002434	16	16p13.3	MPG	N-methylpurine-DNA glycosylase	2.6423451	2.294738	1.800367747	1.896815744
NM_001983	19	19q13.2-q13.	ERCC1	excision repair cross-complementing rodent repair deficiency,	3.4563898	2.045748	2.402443745	1.701622356
NM_001983	19	19q13.2-q13.	ERCC1	excision repair cross-complementing rodent repair deficiency,	3.2551325	2.052034	2.252858174	1.695528401

AGI	Chr	MAP	symbol	NAME	T47.TRvsWT	ZR.TRvsWT	OD_T47.TRvsWT	OD_ZR.TRvsWT
NM_152251	8	8p23.1	DEFB106A	defensin, beta 106A	-2.303863	-2.07515	-2.952479607	-1.82840775
NM_148914	7	7q11.23	ABHD11	abhydrolase domain containing 11	3.4092279	2.261543	3.533456333	1.945562347
NM_018685	7	7p15-p14	ANLN	anillin, actin binding protein	-3.200678	-2.42706	-1.749637602	-2.88422721
NM_021163	7	7p22.1	RBAK	RB-associated KRAB zinc finger	-2.521513	-3.10283	-1.925837792	-2.99687278
NM_005361	X	Xq28	MAGEA2	melanoma antigen family A, 2	2.976205	5.085235	4.43491211	5.660534339
BC010931	19	19p12	MEF2B	MADS box transcription enhancer factor 2, polypeptide B (my	2.9100767	2.300935	1.980847104	2.01743305
BC029395	16	16q23.3	MPHOSPH6	M-phase phosphoprotein 6	2.0566703	2.808625	0.817554552	3.788724073
BC034024	NA	NA	NA	NA	-2.490703	-3.03997	-2.430945958	-2.96997786
AF394782	5	5q31.1	RAPGEF6	Rap guanine nucleotide exchange factor (GEF) 6	-3.41714	-3.09488	-2.795112604	-4.12868558
BC008034	NA	NA	NA	NA	-3.01115	-2.09403	-3.210803735	-1.44186946
CA430188	3	3p25.3	LOC440944	hypothetical gene supported by AK128398	-2.895388	-2.52515	-3.470269404	-2.13589778
BC019880	9	9q21.11	RP11-262H14.	hypothetical locus MGC21881	2.2817604	2.171967	2.460978493	1.706224505
AK094613	10	10q22-q23	RPS24	ribosomal protein S24	-4.143952	-2.18267	-3.65667624	-2.53458711
BU683892	7	7p15.2	CBX3	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	-3.301972	-3.0286	-2.596197945	-3.0388309
AU116818	9	9q22.31	FAM120A	family with sequence similarity 120A	-2.048728	-2.11641	-1.402960122	-0.98316694
AI332397	3	3q28	EIF4A2	eukaryotic translation initiation factor 4A, isoform 2	-3.431634	-2.89815	-4.403718313	-2.4808583
BQ025347	5	5q32	CSNK1A1	casein kinase 1, alpha 1	-2.522647	-2.34677	-3.187134078	-2.65900571
AI377389	5	5q32	CSNK1A1	casein kinase 1, alpha 1	-2.652119	-3.32542	-3.052916633	-2.44631822
AI201248	17	17q11.2	RHOT1	ras homolog gene family, member T1	-3.654612	-3.12368	-4.084760969	-2.40762425
AI692624	NA	NA	NA	NA	-2.707543	-3.25297	-2.563067316	-1.68865136
AI732587	3	3p12-p11.1	CGGBP1	CGG triplet repeat binding protein 1	-2.13367	-2.52157	-2.195429671	-1.39690224
H48516	13	13q14.3	DLEU2	deleted in lymphocytic leukemia, 2	-2.882624	-2.98087	-3.64866142	-3.2251828
AA580691	14	14q24.3	RBM25	RNA binding motif protein 25	-2.402223	-2.83152	-3.373068393	-2.13739696
BU789637	21	21q22.3	RUNX1	runt-related transcription factor 1 (acute myeloid leukemia 1;	-2.167181	-2.15456	-2.045603006	-2.04474277
AK092750	11	11q13.1	EHBP1L1	EH domain binding protein 1-like 1	-2.462456	-2.06703	-2.34671307	-2.22764862
BC042832	16	16p12.2	LOC641298	PI-3-kinase-related kinase SMG-1 - like locus	-2.534769	-3.35793	-2.537447502	-2.45135254
BI857154	NA	NA	NA	NA	-2.328308	-2.01416	-1.824453529	-0.96931492
BI832220	1	1q31.3	C1orf53	chromosome 1 open reading frame 53	2.2922443	2.079746	2.760227673	2.025525712
BI832220	1	1q31.3	C1orf53	chromosome 1 open reading frame 53	3.8996102	3.05838	4.326575319	3.185717832
BE708432	11	11q13.1	MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-	-2.772665	-2.35864	-3.132996614	-1.55456563
BG109249	15	15q13.2	FAM7A2	family with sequence similarity 7, member A2	-3.617214	-2.31288	-2.043990331	-2.57264147
AL711520	5	5q13.2	LOC643367	region containing SMA4; hypothetical protein LOC153561	-2.485055	-3.90696	-2.237664133	-3.20586887
AK095307	6	6p12.3-p11.2	FBXO9	F-box protein 9	-2.442157	-2.30716	-2.684253353	-1.85941182
BC009786	1	1p36.33	NOC2L	nucleolar complex associated 2 homolog (S. cerevisiae)	-2.361922	-3.32417	-2.033907592	-3.11984453
T56980	8	8q24.12	TRPS1	trichorhinophalangeal syndrome I	-3.532769	-4.06233	-4.63854429	-1.83590741
AF147429	17	17q23.3	DDX42	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42	-2.625148	-2.70842	-2.12063906	-1.81716267
BC036606	6	6pter-p22.3	ZFAND3	zinc finger, AN1-type domain 3	-3.245913	-2.30931	-2.699112915	-2.31133857
AF085868	8	8q24.12	TRPS1	trichorhinophalangeal syndrome I	-2.143403	-2.19189	-2.915399104	-1.64938012
W76548	15	15q15.3	HISPPD2A	histidine acid phosphatase domain containing 2A	-2.058659	-2.20782	-2.286178802	-1.93059686
AL832146	1	1q23	PBX1	pre-B-cell leukemia homeobox 1	-2.456437	-4.13521	-1.981102969	-2.69994519
AK021715	NA	NA	NA	NA	-2.552511	-2.11614	-2.334855404	-1.806809
AK092078	5	5q21.2	LOC285708	hypothetical protein LOC285708	-2.409628	-2.03141	-2.820227772	-1.78410186
AF277897	7	7p12	EGFR	epidermal growth factor receptor (erythroblastic leukemia vira	-4.868413	-3.55526	-4.840278451	-2.92275422
AF277897	7	7p12	EGFR	epidermal growth factor receptor (erythroblastic leukemia vira	-3.652517	-2.23589	-3.589048771	-2.09341588

AGI	Chr	MAP	symbol	NAME	T47.TRvsWT	ZR.TRvsWT	OD_T47.TRvsWT	OD_ZR.TRvsWT
NM_152251	8	8p23.1	DEFB106A	defensin, beta 106A	-2.303863	-2.07515	-2.952479607	-1.82840775
NM_148914	7	7q11.23	ABHD11	abhydrolase domain containing 11	3.4092279	2.261543	3.533456333	1.945562347
NM_018685	7	7p15-p14	ANLN	anillin, actin binding protein	-3.200678	-2.42706	-1.749637602	-2.88422721
NM_021163	7	7p22.1	RBAK	RB-associated KRAB zinc finger	-2.521513	-3.10283	-1.925837792	-2.99687278
NM_005361	X	Xq28	MAGEA2	melanoma antigen family A, 2	2.976205	5.085235	4.43491211	5.660534339
BC010931	19	19p12	MEF2B	MADS box transcription enhancer factor 2, polypeptide B (my	2.9100767	2.300935	1.980847104	2.01743305
BC029395	16	16q23.3	MPHOSPH6	M-phase phosphoprotein 6	2.0566703	2.808625	0.817554552	3.788724073
BC034024	NA	NA	NA	NA	-2.490703	-3.03997	-2.430945958	-2.96997786
AF394782	5	5q31.1	RAPGEF6	Rap guanine nucleotide exchange factor (GEF) 6	-3.41714	-3.09488	-2.795112604	-4.12868558
BC008034	NA	NA	NA	NA	-3.01115	-2.09403	-3.210803735	-1.44186946
CA430188	3	3p25.3	LOC440944	hypothetical gene supported by AK128398	-2.895388	-2.52515	-3.470269404	-2.13589778
BC019880	9	9q21.11	RP11-262H14.	hypothetical locus MGC21881	2.2817604	2.171967	2.460978493	1.706224505
AK094613	10	10q22-q23	RPS24	ribosomal protein S24	-4.143952	-2.18267	-3.65667624	-2.53458711
BU683892	7	7p15.2	CBX3	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	-3.301972	-3.0286	-2.596197945	-3.0388309
AU116818	9	9q22.31	FAM120A	family with sequence similarity 120A	-2.048728	-2.11641	-1.402960122	-0.98316694
AI332397	3	3q28	EIF4A2	eukaryotic translation initiation factor 4A, isoform 2	-3.431634	-2.89815	-4.403718313	-2.4808583
BQ025347	5	5q32	CSNK1A1	casein kinase 1, alpha 1	-2.522647	-2.34677	-3.187134078	-2.65900571
AI377389	5	5q32	CSNK1A1	casein kinase 1, alpha 1	-2.652119	-3.32542	-3.052916633	-2.44631822
AI201248	17	17q11.2	RHOT1	ras homolog gene family, member T1	-3.654612	-3.12368	-4.084760969	-2.40762425
AI692624	NA	NA	NA	NA	-2.707543	-3.25297	-2.563067316	-1.68865136
AI732587	3	3p12-p11.1	CGGBP1	CGG triplet repeat binding protein 1	-2.13367	-2.52157	-2.195429671	-1.39690224
H48516	13	13q14.3	DLEU2	deleted in lymphocytic leukemia, 2	-2.882624	-2.98087	-3.64866142	-3.2251828
AA580691	14	14q24.3	RBM25	RNA binding motif protein 25	-2.402223	-2.83152	-3.373068393	-2.13739696
BU789637	21	21q22.3	RUNX1	runt-related transcription factor 1 (acute myeloid leukemia 1;	-2.167181	-2.15456	-2.045603006	-2.04474277
AK092750	11	11q13.1	EHBP1L1	EH domain binding protein 1-like 1	-2.462456	-2.06703	-2.34671307	-2.22764862
BC042832	16	16p12.2	LOC641298	PI-3-kinase-related kinase SMG-1 - like locus	-2.534769	-3.35793	-2.537447502	-2.45135254
BI857154	NA	NA	NA	NA	-2.328308	-2.01416	-1.824453529	-0.96931492
BI832220	1	1q31.3	C1orf53	chromosome 1 open reading frame 53	2.2922443	2.079746	2.760227673	2.025525712
BI832220	1	1q31.3	C1orf53	chromosome 1 open reading frame 53	3.8996102	3.05838	4.326575319	3.185717832
BE708432	11	11q13.1	MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-	-2.772665	-2.35864	-3.132996614	-1.55456563
BG109249	15	15q13.2	FAM7A2	family with sequence similarity 7, member A2	-3.617214	-2.31288	-2.043990331	-2.57264147
AL711520	5	5q13.2	LOC643367	region containing SMA4; hypothetical protein LOC153561	-2.485055	-3.90696	-2.237664133	-3.20586887
AK095307	6	6p12.3-p11.2	FBXO9	F-box protein 9	-2.442157	-2.30716	-2.684253353	-1.85941182
BC009786	1	1p36.33	NOC2L	nucleolar complex associated 2 homolog (S. cerevisiae)	-2.361922	-3.32417	-2.033907592	-3.11984453
T56980	8	8q24.12	TRPS1	trichorhinophalangeal syndrome I	-3.532769	-4.06233	-4.63854429	-1.83590741
AF147429	17	17q23.3	DDX42	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42	-2.625148	-2.70842	-2.12063906	-1.81716267
BC036606	6	6pter-p22.3	ZFAND3	zinc finger, AN1-type domain 3	-3.245913	-2.30931	-2.699112915	-2.31133857
AF085868	8	8q24.12	TRPS1	trichorhinophalangeal syndrome I	-2.143403	-2.19189	-2.915399104	-1.64938012
W76548	15	15q15.3	HISPPD2A	histidine acid phosphatase domain containing 2A	-2.058659	-2.20782	-2.286178802	-1.93059686
AL832146	1	1q23	PBX1	pre-B-cell leukemia homeobox 1	-2.456437	-4.13521	-1.981102969	-2.69994519
AK021715	NA	NA	NA	NA	-2.552511	-2.11614	-2.334855404	-1.806809
AK092078	5	5q21.2	LOC285708	hypothetical protein LOC285708	-2.409628	-2.03141	-2.820227772	-1.78410186
AF277897	7	7p12	EGFR	epidermal growth factor receptor (erythroblastic leukemia vira	-4.868413	-3.55526	-4.840278451	-2.92275422
AF277897	7	7p12	EGFR	epidermal growth factor receptor (erythroblastic leukemia vira	-3.652517	-2.23589	-3.589048771	-2.09341588
BE930017	16	16p11.2	FUS	fusion (involved in t(12;16) in malignant liposarcoma)	-2.025682	-2.48964	-0.994339302	-2.74308497
U59479	14	14q21.1	PNN	pinin, desmosome associated protein	-2.933948	-2.04457	-2.246431135	-2.80543625
BC017896	17	17q21.32	LRRC37A2	leucine rich repeat containing 37, member A2	-2.057553	-2.15841	-2.933243153	-1.47085121
U66879	11	11q13.1	BAD	BCL2-antagonist of cell death	3.3134247	2.443962	1.924401006	1.981275341
NM_003754	11	11p15.4	EIF3S5	eukaryotic translation initiation factor 3, subunit 5 epsilon, 47	2.7702275	2.136514	2.505418188	2.078159702

7. Appendix B

NM_001005236	chr9	9q33.2	OR1L1	olfactory receptor, family 1, subfamily L, member 1	-2.253592588	0.00275
NM_001002907	chr11	11q11	OR8K1	olfactory receptor, family 8, subfamily K, member 1	-2.252712598	0.000315
NM_032997	chr10	10q21-q22	ZWINT	ZW10 interactor	-2.230629579	0.001974
NM_014411	chr7	7q31	C7orf54	chromosome 7 open reading frame 54	-2.221413156	0.001599
---	chr9				-2.205694357	0.000383
---	chr13				-2.187134807	0.000594
NM_024803	chr10	10p15.1	TUBAL3	tubulin, alpha-like 3	-2.144422341	9.97E-05
NR_003672	chr9	9q34.3	SNORA17	small nucleolar RNA host gene (non-protein coding) 7	-2.143826837	0.000684
AF395440	chr1	1p22	HEJ1	similar to DNAJ	-2.095295007	0.001188
NM_001001961	chr9	9q31.1	OR13C3	olfactory receptor, family 13, subfamily C, member 3	-2.087602135	0.002995
NM_005617	chr1	5q31-q33	RPS14	ribosomal protein S14	-2.064784881	0.001097
NM_000475	chrX	Xp21.3-p2	NR0B1	nuclear receptor subfamily 0, group B, member 1	-2.062081207	0.000663
BC096745	chrX	Xp22.13	CXorf20	chromosome X open reading frame 20	-2.058707448	9.14E-05
NM_033273	chr7	7p11.2	ZNF479	zinc finger protein 479	-2.054271396	0.000235
NM_020665	chrX	Xp22	TMEM27	transmembrane protein 27	-2.052087945	1.66E-05
NM_153442	chr10	10q26.2	GPR26	G protein-coupled receptor 26	-2.044038434	0.000646
AK125177	chr1	1q44	LOC14913	hypothetical LOC149134	-2.040681912	7.45E-05
NR_002775	chr11	11cen-q12	RPLP0P2	ribosomal protein, large, P0 pseudogene 2	-2.032820845	0.001143
NM_001001916	chr11	11p15.4	OR52J3	olfactory receptor, family 52, subfamily J, member 3	-2.030557016	0.000891
NM_016512	chr8	8p23-p22	SPAG11B	sperm associated antigen 11B	-2.020355688	6.68E-05
NM_002701	chr1	6p21.31	POU5F1	POU class 5 homeobox 1	-2.016754638	4.61E-05
---	chr21				-2.001816083	0.000529
NM_005212	chr4	4q21.1	CSN3	casein kappa	-1.987632475	0.00124
BC014249	chr7	7q11.21	VKORC1L1	vitamin K epoxide reductase complex, subunit 1-like 1	-1.977504002	0.000348
NM_000969	chr1	1p22.1	RPL5	ribosomal protein L5	-1.969957291	0.00042
NM_182911	chr2	2q11.2	TSGA10	testis specific, 10	-1.969202874	0.000276
NM_003725	chr12	12q13	HSD17B6	hydroxysteroid (17-beta) dehydrogenase 6 homolog (mouse)	-1.965285166	0.000713
NM_022129	chr10	10pter-q25	PBLD	phenazine biosynthesis-like protein domain containing	-1.953103974	0.00011
NM_003326	chr1	1q25	TNFSF4	tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally	-1.952823771	0.001861
NM_001001960	chr11	11q11	OR5W2	olfactory receptor, family 5, subfamily W, member 2	-1.946169121	0.002583
L20860	chr22	22q11.21-q	Sep-05	glycoprotein Ib (platelet), beta polypeptide	-1.943434795	0.000107
NM_133326	chr1	1q31.3	ATP6V1G3	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G3	-1.936282749	0.000288
NM_001031748	chr12	12q12	C12orf40	chromosome 12 open reading frame 40	-1.935421655	0.00141
---	chr5				-1.932944427	0.003087
NM_144651	chr8	8q11.22	PXDNL	peroxidasin homolog (Drosophila)-like	-1.929069919	0.000287
NM_178859	chr15	15q22.31	OSTbeta	organic solute transporter beta	-1.918614504	0.002618
NM_014269	chr4	4q34	ADAM29	ADAM metallopeptidase domain 29	-1.917924604	0.002386
AK124409	chr2	2p25.2	FLJ42418	FLJ42418 protein	-1.912696939	0.000696
NM_032523	chr2	2q31-q32.1	OSBPL6	oxysterol binding protein-like 6	-1.907920832	0.00014
NM_031277	chr13	13q12.12	RNF17	ring finger protein 17	-1.901709802	0.002864
NM_198529	chr17	17q11.2	EFCAB5	EF-hand calcium binding domain 5	-1.895357486	0.000653
NM_145238	chr1	1p34.3	ZSCAN20	zinc finger and SCAN domain containing 20	-1.887844635	0.001019
NM_175878	chr22	22q11.1	XKR3	XK, Kell blood group complex subunit-related family, member 3	-1.885588591	0.002527
NM_016370	chrX	Xq22.1-q2	RAB9B	RAB9B, member RAS oncogene family	-1.884519097	0.00045
NM_152764	chr16	16p13.3		chromosome 16 open reading frame 73	-1.882529005	0.00041
NM_001211	chr15	15q15	BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	-1.874914108	0.002353
NM_182575	chr19	19q13.33	IZUMO1	izumo sperm-egg fusion 1	-1.873626172	0.001164
NM_000590	chr5	5q31.1	IL9	interleukin 9	-1.872857165	0.000331
NM_018324	chr10	10p13	OLAH	oleoyl-ACP hydrolase	-1.871156367	0.00099
NM_199420	chr3	3q13.33	POLQ	polymerase (DNA directed), theta	-1.870292306	0.001568

8. REFERENCES

(1986). Hormones and cancer: 90 years after Beatson. *Cancer Surv* **5:** 435-687.

Afonso N, Bouwman D (2008). Lobular carcinoma in situ. *Eur J Cancer Prev* **17:** 312-6.

Ali S, Coombes RC (2002). Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer* **2:** 101-12.

Amatschek S, Koenig U, Auer H, Steinlein P, Pacher M, Gruenfelder A *et al* (2004). Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumor-specific genes. *Cancer Res* **64:** 844-56.

Atanackovic D, Altorki NK, Cao Y, Ritter E, Ferrara CA, Ritter G, Hoffman EW, Bokemeyer C, Old LJ, Gnjatic S. (2008) Booster vaccination of cancer patients with MAGE-A3 protein reveals long-term immunological memory or tolerance depending on priming. Proc Natl Acad Sci U S A;105(5):1650-5.

Anzick SL, Kononen J, Walker RL, Azorsa DO, Tanner MM, Guan XY *et al* (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277:** 965-8.

Arsham AM, Howell JJ, Simon MC (2003). A novel hypoxia-inducible factor-independent hypoxic response regulating mammalian target of rapamycin and its targets. *J Biol Chem* **278:** 29655-60.

Avisar E, Khan MA, Axelrod D, Oza K (1998). Pure mucinous carcinoma of the breast: a clinicopathologic correlation study. *Ann Surg Oncol* **5:** 447-51.

Bachman KE, Argani P, Sameuls Y et al. The PIK3CA gene is mutated with high frequency in human breast cancer. Cancer Biol Ther 2004;3:772-5.

Badve S, Nakshatri H (2009). Oestrogen-receptor-positive breast cancer: towards bridging histopathological and molecular classifications. *J Clin Pathol* **62:** 6-12.

Barnett GC, Shah M, Redman K, Easton DF, Ponder BA, Pharoah PD (2008). Risk factors for the incidence of breast cancer: do they affect survival from the disease? *J Clin Oncol* **26:** 3310-6.

Bartel DP (2009). MicroRNAs: target recognition and regulatory functions. *Cell* **136:** 215-33.

Baum M (2002). Has tamoxifen had its day? *Breast Cancer Res* **4:** 213-7.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125:** 279-84.

Berstad P, Ma H, Bernstein L, Ursin G (2008). Alcohol intake and breast cancer risk among young women. *Breast Cancer Res Treat* **108:** 113-20.

Bertucci F, Tarpin C, Charafe-Jauffret E, Bardou VJ, Braud AC, Tallet A *et al* (2004). Multivariate analysis of survival in inflammatory breast cancer: impact of intensity of chemotherapy in multimodality treatment. *Bone Marrow Transplant* **33:** 913-20.

Bhandare D, Nayar R, Bryk M, Hou N, Cohn R, Golewale N *et al* (2005). Endocrine biomarkers in ductal lavage samples from women at high risk for breast cancer. *Cancer Epidemiol Biomarkers Prev* **14:** 2620-7.

Bocca C, Bozzo F, Martinasso G, Canuto RA, Miglietta A (2008). Involvement of PPARalpha in the growth inhibitory effect of arachidonic acid on breast cancer cells. *Br J Nutr* **100:** 739-50.

Brugarolas JB, Vazquez F, Reddy A, Sellers WR, Kaelin WG, Jr. (2003). TSC2 regulates VEGF through mTOR-dependent and -independent pathways. *Cancer Cell* **4:** 147-58.

Buzdar AU (2005). Aromatase inhibitors: changing the face of endocrine therapy for breast cancer. *Breast Dis* **24:** 107-17.

Caballero OL, Chen YT. (2009) Cancer/testis (CT) antigens: potential targets for immunotherapy. Cancer Sci.;100(11):2014-21.

Campbell RA, Bhat-Nakshatri P, Patel NM, Constantinidou D, Ali S, Nakshatri H (2001). Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance. *J Biol Chem* **276:** 9817-24.

Carrasco J, Van Pel A, Neyns B, Lethe B, Brasseur F, Renkvist N *et al* (2008). Vaccination of a melanoma patient with mature dendritic cells pulsed with MAGE-3 peptides triggers the activity of nonvaccine anti-tumor cells. *J Immunol* **180:** 3585-93.

Cervera AM, Apostolova N, Luna-Crespo F, Sanjuan-Pla A, Garcia-Bou R, McCreath KJ (2006). An alternatively spliced transcript of the PHD3 gene retains prolyl hydroxylase activity. *Cancer Lett* **233:** 131-8.

Chan DA, Sutphin PD, Denko NC, Giaccia AJ (2002). Role of prolyl hydroxylation in oncogenically stabilized hypoxia-inducible factor-1alpha. *J Biol Chem* **277:** 40112-7.

Chapman EJ, Knowles MA. (2009) Necdin: a multi functional protein with potential tumor suppressor role? Mol Carcinog.;48(11):975-81.

Chandel NS Mitochondrial regulation of oxygen sensing. *Adv Exp Med Biol* **661:** 339-54.

Chen ME, Lin SH, Chung LW, Sikes RA (1998). Isolation and characterization of PAGE-1 and GAGE-7. New genes expressed in the LNCaP prostate cancer progression model that share homology with melanoma-associated antigens. *J Biol Chem* **273:** 17618-25.

Cheng TC, Chen ST, Huang CS, Fu YP, Yu JC, Cheng CW *et al* (2005). Breast cancer risk associated with genotype polymorphism of the catechol estrogen-metabolizing genes: a multigenic study on cancer susceptibility. *Int J Cancer* **113:** 345-53.

Cho HJ, Caballero OL, Gnjatic S, Andrade VC, Colleoni GW, Vettore AL *et al* (2006). Physical interaction of two cancer-testis antigens, MAGE-C1 (CT7) and NY-ESO-1 (CT6). *Cancer Immun* **6:** 12.

Chomez P, De Backer O, Bertrand M, De Plaen E, Boon T, Lucas S. (2001) An overview of the MAGE gene family with the identification of all human members of the family. Cancer Res.;61(14):5544-51.

Chopin V, Toillon RA, Jouy N, Le Bourhis X (2002). Sodium butyrate induces P53-independent, Fas-mediated apoptosis in MCF-7 human breast cancer cells. *Br J Pharmacol* **135:** 79-86.

Chung FY, Huang MY, Yeh CS, Chang HJ, Cheng TL, Yen LC *et al* (2009). GLUT1 gene is a potential hypoxic marker in colorectal cancer patients. *BMC Cancer* **9:** 241.

Clark AS, West K, Streicher S, Dennis PA (2002). Constitutive and inducible Akt activity promotes resistance to chemotherapy, trastuzumab, or tamoxifen in breast cancer cells. *Mol Cancer Ther* **1:** 707-17.

Coutts AS, Murphy LC (1998). Elevated mitogen-activated protein kinase activity in estrogen-nonresponsive human breast cancer cells. *Cancer Res* **58:** 4071-4.

Cuzick J, Forbes J, Edwards R, Baum M, Cawthorn S, Coates A *et al* (2002). First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. *Lancet* **360:** 817-24.

De Backer O, Arden KC, Boretti M, Vantomme V, De Smet C, Czekay S *et al* (1999). Characterization of the GAGE genes that are expressed in various human cancers and in normal testis. *Cancer Res* **59:** 3157-65.

del Peso L, Castellanos MC, Temes E, Martin-Puig S, Cuevas Y, Olmos G *et al* (2003). The von Hippel Lindau/hypoxia-inducible factor (HIF) pathway regulates the transcription of the HIF-proline hydroxylase genes in response to low oxygen. *J Biol Chem* **278:** 48690-5.

Desmedt C, Sotiriou C (2006). Proliferation: the most prominent predictor of clinical outcome in breast cancer. *Cell Cycle* **5:** 2198-202.

Domchek SM, Friebel TM, Garber JE, Isaacs C, Matloff E, Eeles R *et al* Occult ovarian cancers identified at risk-reducing salpingo-oophorectomy in a prospective cohort of BRCA1/2 mutation carriers. *Breast Cancer Res Treat*.

Dorssers LC, Van der Flier S, Brinkman A, van Agthoven T, Veldscholte J, Berns EM *et al* (2001). Tamoxifen resistance in breast cancer: elucidating mechanisms. *Drugs* **61:** 1721-33.

Dowsett M, Cuzick J, Wale C, Howell T, Houghton J, 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:** 7512-7.

Dowsett M, Ellis MJ (2003). Role of biologic markers in patient selection and application to disease prevention. *Am J Clin Oncol* **26:** S34-9.

Draper L (2006). Breast cancer: trends, risks, treatments, and effects. *AAOHN J* **54:** 445-51; quiz 452-3.

Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG *et al* (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447:** 1087-93.

Ehrlich M (2009). DNA hypomethylation in cancer cells. *Epigenomics* **1:** 239-259.

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

Esquela-Kerscher A, Slack FJ (2006). Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6:** 259-69.

Feng Y, Hao X, Mao H (2001). [Clinical significance of MDR gene expression in malignant pleural effusion and ascites and solid tumors]. *Zhonghua Yi Xue Za Zhi* **81:** 1484-7.

Fisher B, Costantino JP, Wickerham DL, Cecchini RS, Cronin WM, Robidoux A *et al* (2005). Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst* **97:** 1652-62.

Font de Mora J, Brown M. AIB1 is a conduit for kinase-mediated growth factor signaling to estrogen receptor. Mol Cell Biol 2000;5041-7.

Forbes JF, Cuzick J, Buzdar A, Howell A, Tobias JS, Baum M *et al* (2008). Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 100-month analysis of the ATAC trial
Tamoxifen and contralateral breast cancer. *Lancet Oncol* **9:** 45-53.

Ohman Forslund K, Nordqvist K. (2001) The melanoma antigen genes--any clues to their functions in normal tissues? Exp Cell Res.; 265(2):185-94.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM *et al* (2006). Copy number variation: new insights in genome diversity. *Genome Res* **16:** 949-61.

Gama-Sosa MA, Midgett RM, Slagel VA, Githens S, Kuo KC, Gehrke CW *et al* (1983). Tissue-specific differences in DNA methylation in various mammals. *Biochim Biophys Acta* **740:** 212-9.

Gee JM, Robertson JF, Ellis IO, Nicholson RI (2001). Phosphorylation of ERK1/2 mitogen-activated protein kinase is associated with poor response to anti-hormonal therapy and decreased patient survival in clinical breast cancer. *Int J Cancer* **95:** 247-54.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al* (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80.

Goetz MP, Rae JM, Suman VJ, Safgren SL, Ames MM, Visscher DW *et al* (2005). Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J Clin Oncol* **23:** 9312-8.

Goetz MP, Suman VJ, Ingle JN, Nibbe AM, Visscher DW, Reynolds CA *et al* (2006). A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clin Cancer Res* **12:** 2080-7.

Golembesky AK, Gammon MD, North KE, Bensen JT, Schroeder JC, Teitelbaum SL *et al* (2008). Peroxisome proliferator-activated receptor-alpha (PPARA) genetic polymorphisms and breast cancer risk: a Long Island ancillary study. *Carcinogenesis* **29:** 1944-9.

Grigoriadis A, Caballero OL, Hoek KS, da Silva L, Chen YT, Shin SJ *et al* (2009). CT-X antigen expression in human breast cancer. *Proc Natl Acad Sci U S A* **106:** 13493-8.

Gutierrez MC, Detre S, Johnston S, Mohsin SK, Shou J, Allred DC *et al* (2005). Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. *J Clin Oncol* **23:** 2469-76.

Hagg M, Wennstrom S (2005). Activation of hypoxia-induced transcription in normoxia. *Exp Cell Res* **306:** 180-91.

Harvey JA, Santen RJ, Petroni GR, Bovbjerg VE, Smolkin ME, Sheriff FS *et al* (2008). Histologic changes in the breast with menopausal hormone therapy use: correlation with breast density, estrogen receptor, progesterone receptor, and proliferation indices. *Menopause* **15:** 67-73.

Hemminki K, Shu X, Li X, Ji J, Sundquist J, Sundquist K (2008). Familial risks for hospitalization with endocrine diseases. *J Clin Endocrinol Metab* **93:** 4755-8.

Henderson BJ, Tyndel S, Brain K, Clements A, Bankhead C, Austoker J *et al* (2008). Factors associated with breast cancer-specific distress in younger women participating in a family history mammography screening programme. *Psychooncology* **17:** 74-82.

Hirsila M, Koivunen P, Gunzler V, Kivirikko KI, Myllyharju J (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J Biol Chem* **278:** 30772-80.

Hoskin JM, Carey LA, McLeod HL. CYP2D6 and tamoxifen, DNA matters in breast cancer. Nature Rev. Cancer 9, 576-586 (2009).

Howell A (2001). Preliminary experience with pure antiestrogens. *Clin Cancer Res* **7:** 4369s-4375s; discussion 4411s-4412s.

Howell A, Cuzick J, Baum M, Buzdar A, Dowsett M, Forbes JF *et al* (2005). Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet* **365:** 60-2.

Hu YF, Lau KM, Ho SM, Russo J (1998). Increased expression of estrogen receptor beta in chemically transformed human breast epithelial cells. *Int J Oncol* **12:** 1225-8.

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF *et al* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7:** 96.

Hurtado A, Holmes KA, Geistlinger TR, Hutcheson I, Nicholson R, Brown M, Jiang J, Howat W, Ali S, Carroll JS. Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. Nature 2008;456:663-666.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y *et al* (2004). Detection of large-scale variation in the human genome. *Nat Genet* **36:** 949-51.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al* (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-64.

Ivan M (2008). The ongoing microRNA revolution and its impact in biology and medicine. *J Cell Mol Med* **12:** 1425.

Izuo M (1992). [New drugs in endocrine treatment of breast cancer]. *Nippon Naibunpi Gakkai Zasshi* **68:** 538-49.

Jakesz R, Jonat W, Gnant M, Mittlboeck M, Greil R, Tausch C *et al* (2005). Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial. *Lancet* **366:** 455-62.

Johansson H, Gandini S, Guerrieri-Gonzaga A, Iodice S, Ruscica M, Bonanni B *et al* (2008). Effect of fenretinide and low-dose tamoxifen on insulin sensitivity in premenopausal women at high risk for breast cancer. *Cancer Res* **68:** 9512-8.

Johnson MD, Zuo H, Lee KH, Trebley JP, Rae JM, Weatherman RV *et al* (2004). Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res Treat* **85:** 151-9.

Johnston SR, Haynes BP, Smith IE, Jarman M, Sacks NP, Ebbs SR *et al* (1993). Acquired tamoxifen resistance in human breast cancer and reduced intra-tumoral drug concentration. *Lancet* **342:** 1521-2.

Johnston SR, Saccani-Jotti G, Smith IE, Salter J, Newby J, Coppen M *et al* (1995). Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer. *Cancer Res* **55:** 3331-8.

Kang S, Bader AG, Vogt PK. Phosphotidylinositol 3-kinase mutations identified in human cancer are oncogenic. Proc Natl Acad Sci U S A 2005;102:802-7.

Karnik PS, Kulkarni S, Liu XP, Budd GT, Bukowski RM (1994). Estrogen receptor mutations in tamoxifen-resistant breast cancer. *Cancer Res* **54:** 349-53.

Kato S (2001). Estrogen receptor-mediated cross-talk with growth factor signaling pathways. *Breast Cancer* **8:** 3-9.

Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R *et al* (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A* **100:** 8424-9.

Katzenellenbogen BS, Choi I, Delage-Mourroux R, Ediger TR, Martini PG, Montano M *et al* (2000a). Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology. *J Steroid Biochem Mol Biol* **74:** 279-85.

Katzenellenbogen BS, Katzenellenbogen JA (2000). Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Res* **2:** 335-44.

Katzenellenbogen BS, Montano MM, Ediger TR, Sun J, Ekena K, Lazennec G *et al* (2000b). Estrogen receptors: selective ligands, partners, and distinctive pharmacology. *Recent Prog Horm Res* **55:** 163-93; discussion 194-5.

Kim C, Paik S Gene-expression-based prognostic assays for breast cancer. *Nat Rev Clin Oncol* **7:** 340-7.

Knauer M, Mook S, Rutgers EJ, Bender RA, Hauptmann M, van de Vijver MJ *et al* The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat* **120:** 655-61.

Köditz J, Nesper J, Wottawa M, Stiehl DP, Camenisch G, Franke C, Myllyharju J, Wenger RH, Katschinski DM. (2007) Oxygen-dependent ATF-4 stability is mediated by the PHD3 oxygen sensor. Blood.;110(10):3610-7.

Kun Y et al. Classifying the oestrogen receptor status of breast cancers by expression profiles reveals a poor prognosis subpopulation exhibiting high expression of the ERBB2 receptor. hum mol. Genet. 12, 3245-3258 (2003).

Kyakumoto S, Kito N, Sato N (2003). Expression of cAMP response element binding protein (CREB)-binding protein (CBP) and the implication in retinoic acid-inducible transcription activation in human salivary gland adenocarcinoma cell line HSG. *Endocr Res* **29:** 277-89.

Lal P, Tan LK, Chen B (2005). Correlation of HER-2 status with estrogen and progesterone receptors and histologic features in 3,655 invasive breast carcinomas. *Am J Clin Pathol* **123:** 541-6.

Lavigne JA, Helzlsouer KJ, Huang HY, Strickland PT, Bell DA, Selmin O *et al* (1997). An association between the allele coding for a low activity variant of catechol-O-methyltransferase and the risk for breast cancer. *Cancer Res* **57:** 5493-7.

Lavinsky RM, Jepsen K, Heinzel T, Torchia J, Mullen TM, Schiff R *et al* (1998). Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc Natl Acad Sci U S A* **95:** 2920-5.

Lee RC, Feinbaum RL, Ambros V (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843-54.

Lee ZH, Kwack K, Kim KK, Lee SH, Kim HH (2000). Activation of c-Jun N-terminal kinase and activator protein 1 by receptor activator of nuclear factor kappaB. *Mol Pharmacol* **58:** 1536-45.

Leo JC, Wang SM, Guo CH, Aw SE, Zhao Y, Li JM *et al* (2005). Gene regulation profile reveals consistent anticancer properties of progesterone in hormone-independent breast cancer cells transfected with progesterone receptor. *Int J Cancer* **117:** 561-8.

Lin BC, Scanlan TS (2005). Few things in life are "free": cellular uptake of steroid hormones by an active transport mechanism. *Mol Interv* **5:** 338-40.

Liu W, Cheng S, Asa SL, Ezzat S (2008). The melanoma-associated antigen A3 mediates fibronectin-controlled cancer progression and metastasis. *Cancer Res* **68:** 8104-12.

Llave C, Xie Z, Kasschau KD, Carrington JC (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297:** 2053-6.

Loboda A, Jozkowicz A, Dulak J. (2010) HIF-1 and HIF-2 transcription factors-similar but not identical. Mol Cells.;29(5):435-42.

Loi S, Desmedt C, Cardoso F, Piccart M, Sotiriou C (2005). Breast cancer gene expression profiling: clinical trial and practice implications. *Pharmacogenomics* **6:** 49-58.

Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM *et al* (2008). Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9:** 239.

Loukinov D, Ghochikyan A, Mkrtichyan M, Ichim TE, Lobanenkov VV, Cribbs DH *et al* (2006). Antitumor efficacy of DNA vaccination to the epigenetically acting tumor promoting transcription factor BORIS and CD80 molecular adjuvant. *J Cell Biochem* **98:** 1037-43.

Love RR, Newcomb PA, Wiebe DA, Surawicz TS, Jordan VC, Carbone PP *et al* (1990). Effects of tamoxifen therapy on lipid and lipoprotein levels in

postmenopausal patients with node-negative breast cancer. *J Natl Cancer Inst* **82:** 1327-32.

Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* **11:** R7.

Maillot G, Lacroix-Triki M, Pierredon S, Gratadou L, Schmidt S, Benes V *et al* (2009). Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. *Cancer Res* **69:** 8332-40.

Markman M, Petersen J, Montgomery R (2008). An examination of the influence of patient race and ethnicity on expressed interest in learning about cancer clinical trials. *J Cancer Res Clin Oncol* **134:** 115-8.

Martin LA, Farmer I, Johnston SRD, Ali S, Dowsett M (2005). Elavated ERK1/ERK2/oestrogen receptor cross-talk enhances oestrogen-mediated signaling during long-term oestrogen deprivation. Endocrine-related Cancer 12; S75-S84.

Marxsen JH, Stengel P, Doege K, Heikkinen P, Jokilehto T, Wagner T *et al* (2004). Hypoxia-inducible factor-1 (HIF-1) promotes its degradation by induction of HIF-alpha-prolyl-4-hydroxylases. *Biochem J* **381:** 761-7.

Mauri D, Pavlidis N, Polyzos NP, Ioannidis JP (2006). Survival with aromatase inhibitors and inactivators versus standard hormonal therapy in advanced breast cancer: meta-analysis. *J Natl Cancer Inst* **98:** 1285-91.

Mc Ilroy M, Fleming FJ, Buggy Y, Hill AD, Young LS (2006). Tamoxifen-induced ER-alpha-SRC-3 interaction in HER2 positive human breast cancer; a possible mechanism for ER isoform specific recurrence. *Endocr Relat Cancer* **13:** 1135-45.

McCubrey JA, Steelman LS, Abrams SL, Lee JT, Chang F, Bertrand FE *et al* (2006). Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. *Adv Enzyme Regul* **46:** 249-79.

McMurray RW, Ndebele K, Hardy KJ, Jenkins JK (2001). 17-beta-estradiol suppresses IL-2 and IL-2 receptor. *Cytokine* **14:** 324-33.

Meng S, Tripathy D, Shete S, Ashfaq R, Haley B, Perkins S *et al* (2004). HER-2 gene amplification can be acquired as breast cancer progresses. *Proc Natl Acad Sci U S A* **101:** 9393-8.

Meteoglu I, Dikicioglu E, Erkus M, Culhaci N, Kacar F, Ozkara E *et al* (2005). Breast carcinogenesis. Transition from hyperplasia to invasive lesions. *Saudi Med J* **26:** 1889-96.

Metzen E (2007). Enzyme substrate recognition in oxygen sensing: how the HIF trap snaps. *Biochem J* **408:** e5-6.

Metzen E, Zhou J, Jelkmann W, Fandrey J, Brune B (2003). Nitric oxide impairs normoxic degradation of HIF-1alpha by inhibition of prolyl hydroxylases. *Mol Biol Cell* **14:** 3470-81.

Mikuls TR, Saag KG, George V, Mudano AS, Banerjee S (2005). Racial disparities in the receipt of osteoporosis related healthcare among community-dwelling older women with arthritis and previous fracture. *J Rheumatol* **32:** 870-5.

Miller TW, Perez-Torres M, Narasanna A, Guix M, Stal O, Perez-Tenorio G, Gonzalez-Angulo Ana M, Hennessy B, Mills GB, Kennedy JP, Lindsley GW, Artega CL. Loss of Phosphatase and tensin homologue deleted on chromosome 10 engages ErbB3 and IGF-IR signaling to promote antiestrogen resistance in breast cancer. Cancer Res. 2009 May 15;69(10):4192-4201.

Miller TE, Ghoshal K, Ramaswamy B, Roy S, Datta J, Shapiro CL *et al* (2008). MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *J Biol Chem* **283:** 29897-903.

Monte M, Simonatto M, Peche LY, Bublik DR, Gobessi S, Pierotti MA *et al* (2006). MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents. *Proc Natl Acad Sci U S A* **103:** 11160-5.

Moulder S, Hortobagyi GN (2008). Advances in the treatment of breast cancer. *Clin Pharmacol Ther* **83:** 26-36.

Musgrove E and Sutherland R (2009). Biological determinants of endocrine resistance in breast cancer. Nature Reviews, Cancer 9; 631-641.

Nakuci E, Mahner S, Direnzo J, ElShamy WM (2006). BRCA1-IRIS regulates cyclin D1 expression in breast cancer cells. *Exp Cell Res* **312:** 3120-31.

Orso F, Cottone E, Hasleton MD, Ibbitt JC, Sismondi P, Hurst HC *et al* (2004). Activator protein-2gamma (AP-2gamma) expression is specifically induced by oestrogens through binding of the oestrogen receptor to a canonical element within the 5'-untranslated region. *Biochem J* **377:** 429-38.

Osborne CK, Bardou V, Hopp TA, Chamness GC, Hilsenbeck SG, Fuqua SA *et al* (2003). Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* **95:** 353-61.

Osborne CK, Coronado E, Allred DC, Wiebe V, DeGregorio M (1991). Acquired tamoxifen resistance: correlation with reduced breast tumor levels of tamoxifen and isomerization of trans-4-hydroxytamoxifen. *J Natl Cancer Inst* **83:** 1477-82.

Osborne CK, Schiff R, Fuqua SA, Shou J (2001). Estrogen receptor: current understanding of its activation and modulation. *Clin Cancer Res* **7:** 4338s-4342s; discussion 4411s-4412s.

Paik S (2006). Molecular profiling of breast cancer. *Curr Opin Obstet Gynecol* **18:** 59-63.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M *et al* (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351:** 2817-26.

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC *et al* (2003). Control of leaf morphogenesis by microRNAs. *Nature* **425:** 257-63.

Parisot JP, Hu XF, DeLuise M, Zalcberg JR (1999). Altered expression of the IGF-1 receptor in a tamoxifen-resistant human breast cancer cell line. *Br J Cancer* **79:** 693-700.

Pastorcic-Grgic M, Sarcevic B, Dosen D, Juretic A, Spagnoli GC, Grgic M (2009). Prognostic value of MAGE-A and NY-ESO-1 expression in pharyngeal cancer. *Head Neck.*

Percy MJ, Mooney SM, McMullin MF, Flores A, Lappin TR, Lee FS (2003). A common polymorphism in the oxygen-dependent degradation (ODD) domain of hypoxia inducible factor-1alpha (HIF-1alpha) does not impair Pro-564 hydroxylation. *Mol Cancer* **2:** 31.

Piccart M, Lohrisch C, Di Leo A, Larsimont D (2001). The predictive value of HER2 in breast cancer. *Oncology* **61 Suppl 2:** 73-82.

Piccart-Gebhart, MJ et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. NEJM 353, 1659-1672 (2005).
Pietras A, Johnsson AS, Pahlman S The HIF-2alpha-Driven Pseudo-Hypoxic Phenotype in Tumor Aggressiveness, Differentiation, and Vascularization. *Curr Top Microbiol Immunol*.

Pogribny IP, Tryndyak VP, Boyko A, Rodriguez-Juarez R, Beland FA, Kovalchuk O (2007). Induction of microRNAome deregulation in rat liver by long-term tamoxifen exposure. *Mutat Res* **619:** 30-7.

Ponzone R, Montemurro F, Maggiorotto F, Robba C, Gregori D, Jacomuzzi ME *et al* (2006). Clinical outcome of adjuvant endocrine treatment according to PR and HER-2 status in early breast cancer. *Ann Oncol* **17:** 1631-6.

Powles TJ, Ashley S, Tidy A, Smith IE, Dowsett M (2007). Twenty-year follow-up of the Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial. *J Natl Cancer Inst* **99:** 283-90.

Rankin EB, Giaccia AJ (2008). The role of hypoxia-inducible factors in tumorigenesis. *Cell Death Differ* **15:** 678-85.

Rantanen K, Pursiheimo J, Högel H, Himanen V, Metzen E, Jaakkola PM. (2008) Prolyl hydroxylase PHD3 activates oxygen-dependent protein aggregation. Mol Biol Cell.;19(5):2231-40.

Rapp K, Klenk J, Ulmer H, Concin H, Diem G, Oberaigner W *et al* (2008). Weight change and cancer risk in a cohort of more than 65,000 adults in Austria. *Ann Oncol* **19:** 641-8.

Ravdin PM (1998). How can prognostic and predictive factors in breast cancer be used in a practical way today? *Recent Results Cancer Res* **152:** 86-93.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al* (2006). Global variation in copy number in the human genome. *Nature* **444:** 444-54.

Rigopoulos DN, Tsiambas E, Lazaris AC, Kavantzas N, Papazachariou I, Kravvaritis C *et al* Deregulation of EGFR/VEGF/HIF-1a signaling pathway in colon adenocarcinoma based on tissue microarrays analysis. *J BUON* **15:** 107-15.

Rodriguez-Gonzalez FG, Sieuwerts AM, Smid M, Look MP, Meijer-van Gelder ME, de Weerd V *et al* MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. *Breast Cancer Res Treat*.

Rody A, Karn T, Gatje R, Kourtis K, Minckwitz G, Loibl S *et al* (2006). Gene expression profiles of breast cancer obtained from core cut biopsies before neoadjuvant docetaxel, adriamycin, and cyclophoshamide chemotherapy correlate with routine prognostic markers and could be used to identify predictive signatures. *Zentralbl Gynakol* **128:** 76-81.

Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD *et al* (1995). Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *J Natl Cancer Inst* **87:** 446-51.

Saji S, Jensen EV, Nilsson S, Rylander T, Warner M, Gustafsson JA (2000). Estrogen receptors alpha and beta in the rodent mammary gland. *Proc Natl Acad Sci U S A* **97:** 337-42.

Sandau KB, Fandrey J, Brune B (2001a). Accumulation of HIF-1alpha under the influence of nitric oxide. *Blood* **97:** 1009-15.

Sandau KB, Zhou J, Kietzmann T, Brune B (2001b). Regulation of the hypoxia-inducible factor 1alpha by the inflammatory mediators nitric oxide and tumor necrosis factor-alpha in contrast to desferroxamine and phenylarsine oxide. *J Biol Chem* **276:** 39805-11.

Scalliet PG, Kirkove C (2007). Breast cancer in elderly women: can radiotherapy be omitted? *Eur J Cancer* **43:** 2264-9.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P *et al* (2004). Large-scale copy number polymorphism in the human genome. *Science* **305:** 525-8.

Semenza GL. (2010) Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. Oncogene;29(5):625-34.

Sendoel A, Kohler I, Fellmann C, Lowe SW, Hengartner MO HIF-1 antagonizes p53-mediated apoptosis through a secreted neuronal tyrosinase. *Nature* **465:** 577-83.

Serra-Perez A, Planas AM, Nunez-O'Mara A, Berra E, Garcia-Villoria J, Ribes A *et al* Extended ischemia prevents HIF1alpha degradation at reoxygenation by impairing prolyl-hydroxylation: role of Krebs cycle metabolites. *J Biol Chem* **285:** 18217-24.

Seth P, Krop I, Porter D, Polyak K. (2002) Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene.;21(5):836-43.

Shoman N, Klassen S, McFadden A, Bickis MG, Torlakovic E, Chibbar R (2005). Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Mod Pathol* **18:** 250-9.

Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H *et al* (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *J Natl Cancer Inst* **96:** 926-35.

Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ (2005). Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* **5:** 615-25.

Smith CL, Nawaz Z, O'Malley BW (1997). Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen. *Mol Endocrinol* **11:** 657-66.

Smith IE, Dowsett M, Ebbs SR, Dixon JM, Skene A, Blohmer JU *et al* (2005). Neoadjuvant treatment of postmenopausal breast cancer with anastrozole, tamoxifen, or both in combination: the Immediate Preoperative Anastrozole, Tamoxifen, or Combined with Tamoxifen (IMPACT) multicenter double-blind randomized trial. *J Clin Oncol* **23:** 5108-16.

Sotiriou C, Desmedt C (2006). Gene expression profiling in breast cancer. *Ann Oncol* **17 Suppl 10:** x259-62.

Speirs V, Malone C, Walton DS, Kerin MJ, Atkin SL (1999). Increased expression of estrogen receptor beta mRNA in tamoxifen-resistant breast cancer patients. *Cancer Res* **59:** 5421-4.

Suzuki R, Orsini N, Mignone L, Saji S, Wolk A (2008). Alcohol intake and risk of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis of epidemiological studies. *Int J Cancer* **122:** 1832-41.

Taylor MS (2001). Characterization and comparative analysis of the EGLN gene family. *Gene* **275:** 125-32.

Tikhomirova OS, Grudinina NA, Golubkov VI, Mandel'shtam M, Vasil'ev BV (2007). [Novel BRCA1 gene mutations in breast cancer patients from St. Petersburg]. *Genetika* **43:** 1263-8.

Titus-Ernstoff L, Troisi R, Hatch EE, Hyer M, Wise LA, Palmer JR *et al* (2008). Offspring of women exposed in utero to diethylstilbestrol (DES): a preliminary report of benign and malignant pathology in the third generation. *Epidemiology* **19:** 251-7.

Toh HC, Wang WW, Chia WK, Kvistborg P, Sun L, Teo K *et al* (2009). Clinical Benefit of Allogeneic Melanoma Cell Lysate-Pulsed Autologous Dendritic Cell Vaccine in MAGE-Positive Colorectal Cancer Patients. *Clin Cancer Res* **15:** 7726-7736.

Toillon RA, Chopin V, Jouy N, Fauquette W, Boilly B, Le Bourhis X (2002). Normal breast epithelial cells induce p53-dependent apoptosis and p53-independent cell cycle arrest of breast cancer cells. *Breast Cancer Res Treat* **71:** 269-80.

Tonetti DA, Jordan VC (1995). Possible mechanisms in the emergence of tamoxifen-resistant breast cancer. *Anticancer Drugs* **6:** 498-507.

Troudi W, Uhrhammer N, Romdhane KB, Sibille C, Amor MB, Khodjet El Khil H *et al* (2008). Complete mutation screening and haplotype characterization of BRCA1 gene in Tunisian patients with familial breast cancer. *Cancer Biomark* **4:** 11-8.

Tzukerman MT, Esty A, Santiso-Mere D, Danielian P, Parker MG, Stein RB *et al* (1994). Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. *Mol Endocrinol* **8:** 21-30.

Van den Eynde B, Peeters O, De Backer O, Gaugler B, Lucas S, Boon T. (1995) A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma.. J Exp Med.;182(3):689-98.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Van den Eynde BJ, van der Bruggen P (1997). T cell defined tumor antigens. *Curr Opin Immunol* **9:** 684-93.

Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI *et al* (2005). Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. *Cancer Res* **65:** 7751-62.

Ward HW (1973). Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *Br Med J* **1:** 13-4.

Ward HW, Arthur K, Banks AJ, Bond WH, Brown I, Freeman WE *et al* (1978). Anti-oestrogen therapy for breast cancer--a report on 300 patients treated with tamoxifen. *Clin Oncol* **4:** 11-7.

Widschwendter M, Siegmund KD, Muller HM, Fiegl H, Marth C, Muller-Holzner E *et al* (2004). Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res* **64:** 3807-13.

Yang B, O'Herrin S, Wu J, Reagan-Shaw S, Ma Y, Nihal M *et al* (2007a). Select cancer testes antigens of the MAGE-A, -B, and -C families are expressed in mast cell lines and promote cell viability in vitro and in vivo. *J Invest Dermatol* **127:** 267-75.

Yang B, O'Herrin SM, Wu J, Reagan-Shaw S, Ma Y, Bhat KM *et al* (2007b). MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines. *Cancer Res* **67:** 9954-62.

Yang B, Wu J, Maddodi N, Ma Y, Setaluri V, Longley BJ (2007c). Epigenetic control of MAGE gene expression by the KIT tyrosine kinase. *J Invest Dermatol* **127:** 2123-8.

Zendman AJ, Van Kraats AA, Weidle UH, Ruiter DJ, Van Muijen GN (2002). The XAGE family of cancer/testis-associated genes: alignment and expression profile in normal tissues, melanoma lesions and Ewing's sarcoma. *Int J Cancer* **99:** 361-9.

Zhang Q, Gu J, Li L, Liu J, Luo B, Cheung HW, Boehm JS, Ni M, Geisen C, Root DE, Polyak K, Brown M, Richardson AL, Hahn WC, Kaelin WG Jr, Bommi-Reddy A. (2009) Control of cyclin D1 and breast tumorigenesis by the EglN2 prolyl hydroxylase. Cancer Cell;16(5):413-24.

Zhao JJ, Lin J, Yang H, Kong W, He L, Ma X *et al* (2008). MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J Biol Chem* **283:** 31079-86.

Zhong H, Chiles K, Feldser D, Laughner E, Hanrahan C, Georgescu MM *et al* (2000). Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics. *Cancer Res* **60:** 1541-5.

Zhu X, Asa SL, Ezzat S (2008). Fibroblast growth factor 2 and estrogen control the balance of histone 3 modifications targeting MAGE-A3 in pituitary neoplasia. *Clin Cancer Res* **14:** 1984-96.

Zhu X, Asa SL, Ezzat S (2009). Histone-acetylated control of fibroblast growth factor receptor 2 intron 2 polymorphisms and isoform splicing in breast cancer. *Mol Endocrinol* **23:** 1397-405.

Zundel W, Schindler C, Haas-Kogan D, Koong A, Kaper F, Chen E *et al* (2000). Loss of PTEN facilitates HIF-1-mediated gene expression. *Genes Dev* **14:** 391-6.