THE UNIVERSITY
of EDINBURGH

Clinical Biomarkers of Response to Neoadjuvant Endocrine Therapy in Breast Cancer: Exploring the Potential of Gene Expression Data Integration

Arran K Turnbull

A thesis submitted in fulfilment of the requirements of the degree of Doctor of Philosophy to the University of Edinburgh
April 2013

Abstract

Introduction

Aromatase inhibitors (AIs) have an established role in the treatment of estrogen receptor alpha positive (ER^+) post-menopausal breast cancer. However, response rates are only 50-70% in the neoadjuvant setting and lower in advanced disease. There is a need to identify pre- or early on-treatment biomarkers to predict sensitivity which outperform those currently used, in a move towards stratified treatments and improved patient care. Given the heterogeneity known to exist in the breast cancer population, and the limited availability of matched pre- and on-treatment clinical material, this study also sought to develop novel data integration approaches allowing for the inclusion of similar previously published datasets, thus maximising the power of this study.

Experimental Design

Pre- and on-treatment (at 14 days and 3-months) biopsies were obtained from 34 post-menopausal women with ER+ breast cancer receiving 3 months of neoadjuvant letrozole. Illumina Beadarray gene expression data from these samples were combined with Affymetrix GeneChip data from a similar published study (n=55) and cross-platform integration approaches were evaluated. Dynamic clinical response was assessed for each patient from periodic 3D ultrasound measurements during treatment.

Results

Despite intrinsic differences between different microarray technologies, suitably similar studies can be directly integrated for robust and meaningful meta-analysis with improved statistical power. After mapping probe sequences to Ensembl genes it was demonstrated that, ComBat and cross platform normalisation (XPN), significantly outperform mean-centering and distance-weighted discrimination (DWD) in terms of minimising inter-platform variance. In particular it was observed that DWD, a popular method used in a number of previous studies, removed systematic bias at the expense of genuine biological variability, potentially reducing legitimate biological differences from integrated datasets. A pipeline for the successful integration of microarray datasets from different platforms was developed.

Using this approach a classifier of clinical response to endocrine therapy in the neoadjuvant setting based on the expression of 4 genes was developed which predicted response with 96% and 91% accuracy in training (n=73) and independent validation (n=44) datasets respectively. An early on-treatment biopsy was found to improve predictive power in addition to pre-treatment alone.

Conclusions

Using a novel data integration approach developed as part of this study, a model comprising 4 novel biomarkers for accurate and robust prediction of clinical response to AIs by two weeks of treatment has been generated and validated. On-going work will investigate the applicability to other anti-estrogens, and the adjuvant setting and will assess the potential for a new therapy response test.

Declaration

The work in this thesis is solely my own work unless otherwise indicated, where credit to the contributor is given. No part of the work has been submitted for any other degree.

Arran Turnbull, April 2013

Acknowledgements

I would like to thank my supervisors Andrew Sims PhD and Prof J. Michael Dixon for their excellent support and guidance and I would also like to give special thanks to Dr Alexey Larionov and Charlene Kay for their much appreciated support and contribution to my project.

In particular I would like to thank the following staff and students at the Breakthrough Research Unit (Edinburgh) for their contributions to my project:

Tissue Collection

J. Michael Dixon and Lorna Renshaw

Tissue Storage and Ethics

Experimental Cancer Medicine Centre (ECMC) supported by Cancer Research UK (CRUK) and the Chief Scientist Office (CSO)

Sander Henzing, Paul Mason, Alex MacLellan, Laura Johnston, Lynne Johnston, Laura Banks, Monica McGill and Lorna Renshaw

Histopathology

Jeremy Thomas

Tissue Sectioning

Helen Caldwell, Robert Morris, In hwa Um, Laura Johnston, Lynne Johnston and Susan Hardy

Contributions to Lab Work

Charlene Kay, Peter Mullen, Victoria Webber, Jyoti Nanda, Fiach O'Mahony, Alexey Larionov, Elaine McLay, Helen Caldwell, Carlos Martinez Perez and Carol Ward

Illumina Microarray Hybridisation (WTCRF)

Louise Evenden and Lee Murphy

Illumina Microarray Hybridisation (Ark Genomics)

Alison Dowling

Contributions to Data Analysis

Andy Sims, Robert Kitchen and J. Michael Dixon

Validation Data Analysis (Breakthrough London)

Anita Dunbier and Mitch Dowsett

Thank you to Prof J. Michael Dixon and Prof David Harrison for organising funding for this project. Funding was provided by Breakthrough Breast Cancer and from an NHS endowment.

I would especially like to thank all my friends at the Breakthrough lab for their support and for making the last three years so enjoyable. In particular: Colette, Romina, Helen, Danielle and Kyle who were all working on their projects at the same time as me and shared my office for most of the last three years, Danny and Victoria who I shared an office with for most of my thesis writing and who kept me laughing with our 2.00pm chats, and Jyoti and Fiach (and everyone in the Edinburgh Urological Cancer team) who I have enjoyed working and sharing a lab with during my project.

Finally, and most importantly, I would like to thank my parents, my two brothers: Michael and Stephen and my adopted Spanish brother Carlos for their total and ineffable support in everything I have chosen to do.

Publications

Chapter 2 was the basis for a published manuscript:

"*Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis*"

Turnbull AK, Kitchen RR, Larionov AA, Renshaw L, Dixon JM, Sims AH, (2012)

BMC Medical Genomics 5:35"

List of Contents

Abstract 2

Declaration 4

Acknowledgements 5

Publications 7

List of Contents 8

Abbreviations 16

1. Background 20

1.1. Breast Cancer 20

1.1.1. Incidence and Survival 20

1.1.2. Histological Subtypes and Staging 20

1.1.3. Molecular Subtypes 22

1.1.4. Breast Cancer Treatment 27

1.1.5. Estrogen Biosynthesis and ER^+ Breast Cancer 29

1.1.6. Signalling in ER^+ Breast Cancer 31

1.2. Endocrine Therapy and Resistance to Treatment 32

1.2.1.1. Endocrine Therapy 32

1.2.2. Resistance to Endocrine Therapy 36

1.2.3. Mechanisms of Resistance 37

1.2.3.1. ER and PR Expression in Tumours with Innate Resistance 37

1.2.3.2. PI3K-AKT-mTOR Pathway and Somatic Mutations 38

1.2.3.3. Genetic Traits Affecting Drug Action 39

1.2.3.4. ER-Independent Signalling 39

1.2.3.5. ER Post-translational Modifications 40

1.2.3.6. Differential ER Binding 41
1.2.3.7. Epidermal Growth Factor Receptors 41
1.2.3.8. Breast Cancer Stem Cells 43
1.2.3.9. Cell Cycle 43
1.2.3.10. Cell Survival and Apoptosis 44
1.2.3.11. NFκB Signalling and Inflammation 44
1.2.3.12. Hyper-sensitivity to Residual Estrogen 46
1.2.3.13. Epithelial-Mesenchymal Transition 47
1.2.3.14. Epigenetics 48
1.2.3.15. Autophagocytosis 48
1.2.3.16. Heterogeneity of Resistance 49
1.2.4. Combating Resistance 49
1.3. Studying Resistance and Predicting Response 51
1.3.1. Challenges with Studying Endocrine Therapy Resistance 51
1.3.1.1. Cell Line Studies 51
1.3.1.2. Adjuvant Setting 52
1.3.1.3. Neoadjuvant Setting 52
1.3.2. Predicting Response 53
1.3.2.1. Pathological and Clinical Response 53
1.3.2.2. Proliferative Response 54
1.3.2.3. Molecular Response Markers 54
1.3.2.4. Multigene Signatures 55
1.3.2.4.1. Gene Expression Profiling 55
1.3.2.4.2. Commercial Profiling: Prognostic Assays 56

1.3.2.4.3. Prognostic Signature Research 58
1.3.2.4.4. Lessons from Prognostic Signature Research 60
1.3.2.4.5. Predicting Response to Endocrine Therapy 60
1.3.3. Considerations when Developing Predictive Gene Signatures 64
1.3.3.1. Sample Size and Patient Heterogeneity 64
1.3.3.2. Microarray Bias 65
1.3.3.3. Independent Validation 65
1.3.3.4. Response Criteria 66
1.3.3.5. Heterogeneity in the Resistant Patients 67
1.3.3.6. Clinical Application of a Predictive Signature 67
1.4. Aims and Approaches 68
2. Direct Integration of Intensity-level Data from Affymetrix and Illumina Microarrays improves Statistical Power for Robust Reanalysis 71
2.1. Acknowledgements and Contributions 71
2.2. Introduction 71
2.2.1. Combining Datasets for Meta- or Re-analysis 71
2.2.2. Dataset Size and Statistic Power 72
2.2.3. Directly Integrating Affymetrix and Illumina Data 73
2.3. Materials and Methods 75
2.3.1. Data Generation 75
2.3.1.1. Affymetrix Dataset 75
2.3.1.2. Illumina Dataset (Affymetrix Subset) 76
2.3.1.3. New Illumina Dataset 76
2.3.2. Published MAQC and Breast Cancer Datasets 76

2.3.3. Data Processing and Analysis 77
2.4. Results 78
2.4.1. Direct Cross-platform Integration of MAQC Data 78
2.4.2. Increasing Statistical Power through Integration of Clinical Datasets 85
2.4.2.1. Matched Clinical Samples Common to both Platforms 85
2.4.2.2. Integrating Independent Clinical Datasets 89
2.4.3. Published Affymetrix and Illumina Datasets can be Successfully Integrated 97
2.5. Discussion 100
2.6. Conclusion 102
3. Understanding and Predicting Response to Endocrine Therapy Early in Breast Cancer Treatment 103
3.1. Acknowledgments and Contributions 103
3.2. Introduction 103
3.3. Materials and Methods 105
3.3.1. Patients 105
3.3.2. Response Assessment 106
3.3.3. Tumour Samples 107
3.3.4. RNA Processing and Microarray Hybridisation 107
3.3.5. Published Datasets 108
3.3.5.1. Affymetrix Dataset: Letrozole 108
3.3.5.2. Illumina Dataset: Anastrozole 108
3.3.6. Data Processing and Analysis 109
3.3.6.1. Random Forest (RF) 110
3.3.6.2. Classification and Regression Trees (CART) 111

3.3.7. Immunohistochemistry 113
3.3.7.1. Ki67 113
3.3.7.2. CD45 Staining 113
3.4. Results 113
3.4.1. Integration of Microarray Data 113
3.4.2. Clinical and Pathological Assessment of Response 118
3.4.2.1. Patterns of Clinical Response 118
3.4.2.2. Determination of Overall Clinical Response 120
3.4.2.3. Clinical Features of Response 122
3.4.2.4. Pathological Response 123
3.4.2.5. Breast Cancer Molecular Subtypes and Response 126
3.4.3. Molecular Profile of Quick Stable Response 127
3.4.3.1. Down-regulated Genes and Processes 129
3.4.3.2. Up-regulated Genes and Processes 130
3.4.3.3. Comparison of 'Quick Stable' and 'Slow' Response Gene Changes 131
3.4.4. Predicting Response to Therapy 132
3.4.4.1. Early and Late Gene Changes in Major Functional Groups 132
3.4.4.2. Proliferation and Response to Therapy 134
3.4.4.3. Inflammation, Stromal Remodelling and Response to Therapy 137
3.4.4.4. Differential Gene Expression between Responsive and Non-responsive Tumours 141
3.4.4.5. A Gene Expression Model for Early Prediction of Response 143
3.4.4.6. Acquired Response/Non-response 152
3.4.4.7. Independent Validation of Predictive Model 153

3.4.4.8. The Clinical Need for a 14 Day Biopsy 156
3.4.4.9. Improving on Previous Models 158
3.5. Discussion 162
3.5.1. Currently Used Measures and Biomarkers of Response 162
3.5.1.1. ER as a Biomarker 162
3.5.1.2. Early Clinical Response 162
3.5.1.3. Early Pathological Response 163
3.5.1.4. Molecular Subtyping and Response 163
3.5.1.5. HER2 as a Biomarker 164
3.5.2. Proposed Predictors of Response 165
3.5.2.1. Ki67 165
3.5.2.2. Proliferation Gene Expression 166
3.5.2.3. Immune Signature and Leukocyte Infiltration 166
3.5.2.4. Predictive Gene Expression Signatures 168
3.5.3. The 4 Gene CART Model for Predicting Response 169
3.5.4. Genes in the 4 Gene CART Model 171
3.5.4.1. Glycoprotein 130 (GP130) 171
3.5.4.2. Abnormal spindle-like microcephaly-associated protein (ASPM) 173
3.5.4.3. Minichromosome Maintenance Complex Component 4 (MCM4) 173
3.5.4.4. Nerve Growth Factor Receptor Associated Protein 1 (NGFRAP1/BEX3) 174
3.5.5. Improving on Previous Work 176
3.5.5.1. Increased Sample Number and Definition of Response 176
3.5.5.2. Improved Model Design 177

3.5.5.2.1. Data Integration and Reproducibility of Findings 177
3.5.5.2.2. CART Algorithm 178
3.5.6. Clinical Application 178
3.5.6.1. Clinical Need for an Early On-treatment Biopsy 178
3.5.6.2. Treating Non-responsive Patients 179
4. Conclusions and Future Work 180
4.1. Applicability to Other Settings and Therapies 180
4.1.1. Other Endocrine Therapies 180
4.1.2. The Adjuvant Setting 180
4.2. Translational Potential 181
4.2.1. Technology 181
4.2.1.1. Gene Expression Based Technology: qRT-PCR 181
4.2.1.2. Immunohistochemistry 182
4.2.2. Collecting, Processing and Storing Tissue 182
4.2.2.1. FFPE Tissue 182
4.2.2.2. Final Needle Aspiration 182
4.3. Understanding the Mechanisms underlying GP130 Expression 183
4.4. Investigating Mechanisms of Non-response 184
4.4.1. HER2 Signalling 184
4.4.2. Immune and Stromal Cell Infiltration 184
5. References 185
6. Supplementary 217
6.1. R-Scripts 217
6.2. Proliferation Set Gene List 222

6.3. 131 Genes Input to CART Gene List 224

6.4. LREC Approval 228

6.5. R&D Approval 230

Abbreviations

AI	Aromatase inhibitor
AR	Androgen receptor
AR^+	Androgen receptor positive
AR^-	Androgen receptor negative
AUC	Area under the curve
BCS	Breast conversing surgery
BL1	Basal-like 1 subtype
BL2	Basal-like 2 subtypes
BMI	Body mass index
CART	Classification and regressing tree
CB	ComBat
CC	Centroid classification
CC	Cell cycle subtype
CDF	Custom definition file
CAN	Copy number aberration
CIN	Chromosome instability
CPR	Complete pathological response
CSC	Cancer stem cell
DC	Dendritic cell subtype
DCIS	Ductal carcinoma in situ
DNA	Deoxyribonucleic acid
DWD	Distance weighted discrimination
E	Estrogen
ECM	Extracellular matrix
EGF	Epidermal growth factor

EMT	Epithelial to mesenchymal transition
EP	EndoPredcit
ER	Estrogen receptor alpha
ER^+	Estrogen receptor alpha positive
ER^-	Estrogen receptor alpha negative
ERβ	Estrogen receptor beta
ERE	Estrogen response element
F	Fulvestrant
FDA	Food and drug administration
FFPE	Formalin fixed paraffin embedded
GGI	Genomic grade index
GIDE	Global index of dependence on estrogen
GPCR	G protein coupled receptor
H&E	Haematoxylin and eosin
HER2	Human epidermal growth factor receptor 2
$HER2^+$	Human epidermal growth factor receptor 2 positive
$HER2^-$	Human epidermal growth factor receptor 2 negative
HSP	Heat shock protein
IDC	Infiltrating ducal carcinoma
IGF	Insulin-like growth factor
IHC	Immunohistochemistry
IM	Immunomodulatory subtype
IMPACT	Immediate preoperative Anastrozole or combined with Tamoxifen
IR	Immune response subtype
LAR	Luminal androgen receptor subtype
LR	Logistic regression

M	Mesenchymal subtype
MaSC	Mammary stem cell
MC	Mean centering
MDS	Multi-dimensional scaling
MSL	Mesenchymal stem-like subtype
NCCN	National comprehensive cancer network
NGF	Nerve Growth Factor
PFS	Progression free survival
PR	Progesterone receptor
PR^+	Progesterone receptor positive
PR^-	Progesterone receptor negative
QN	Quantile normalisation
qRT-PCR	Quantitative real-time polymerase chain reaction
RF	Random forest
RFS	Recurrence free survival
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
RP	Rank product
RS	Recurrence score
RTK	Receptor tyrosine kinase
SAM	Significance analysis of microarrays
SERD	Selective estrogen receptor down-regulator
SERM	Selective estrogen receptor modulator
SR	Steroid receptor subtype
T	Tamoxifen
TAM	Tumour-associated macrophage

TNM	Tumour, node and metastasis (breast cancer staging system)
UHRR	Universal human reference RNA
UICC	Union for International Cancer Control
WHO	World health organisation
XPN	Cross-platform normalisation

1. Background

1.1. Breast Cancer

1.1.1. Incidence and Survival

Breast cancer is a disease which according to Cancer Research UK official statistics affects approximately 48,000 woman in the UK each year and in 2010 resulted in over 11,000 deaths [1]. More woman than ever before in the UK are living with breast cancer because of a reduction in mortality over the last 20 years [2] (figure 1). This is due to better awareness, the introduction of screening in 1987 and improvements in treatment [2, 3].

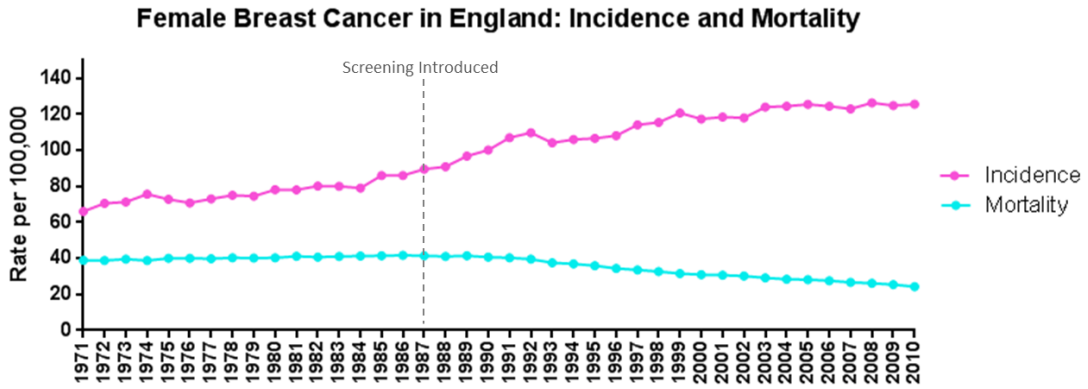


Figure 1. ***Female breast cancer in England: incidence and mortality between 1970 and 2010 (Office of National Statistics)****. Despite a reduction in mortality, since the introduction of screening in 1987 incidence of breast cancer has increased over the last 20 years.*

1.1.2. Histological Subtypes and Staging

Breast cancer is a heterogeneous disease both pathologically and genetically [4]. Several classification systems have been developed and adopted into clinical practice to subdivide breast cancers into groups with similar molecular and histological characteristics, with the intention of developing more targeted therapeutic strategies [5].

Using histology based on microscopic examination, breast cancer can be broadly categorised as either *in situ* or invasive carcinoma: characterised by invasion through the basement membrane. In the case of *in situ* carcinoma, the majority of such cases are classified as ductal (involving the terminal duct lobular unit and the ductal tree) with far fewer lobular (limited mainly to the terminal duct lobular unit) cases reported [6] (figure 2A). The terms ductal and lobular were selected because of the belief that ductal cancers arose in ducts and lobular cancers arose in the lobules. Subsequent work showed all cancer arises in the terminal duct lobular unit. The nomenclature remains in everyday use because ductal and lobular define two groups of cancers with different cellular morphology and different clinical behaviours [7]. *In situ* carcinoma is often regarded as 'pre-cancer' given its non-invasive phenotype. However, in some cases these tumours can develop into invasive cancer [8]. Improvements in screening have led to a rise in the detection and hence treatment of pre-cancer tumours before they have the opportunity to progress.

Invasive carcinomas are a heterogeneous group of tumours which have been classified into 7 histologically distinct subgroups: infiltrating ductal, invasive lobular, ductal/lobular, mucinous, tubular, medullary and papillary carcinomas [6]. Of these, infiltrating ductal carcinoma (IDC), sometimes designated as 'no special type', is the most common accounting for 70-80% of all invasive breast tumours [9]. It is characterised histologically by invasion of the primary tumour through the basement membrane of the duct. IDC is further classified as well-differentiated (grade 1), moderately differentiated (grade 2), and poorly differentiated (grade 3) using a system adapted from that originally developed by Bloom and Richardson in the 1950s, which takes into account nuclear pleomorphism, glandular/tubule formation and mitotic index (presence of hyperchromatic and mitotic nuclei) [10-13] (figures 2B and C). It was subsequently modified by Ellis and Ellston in 1991 (republished in 2002) [11] and all breast cancers are typed and graded according to their system. A staging system known as TNM based on tumour size (T), invasion of lymph nodes (N) and presence of metastasis (M) and this is commonly applied to define how advanced the disease is. An

alternative, numbered staging system comprises 4 stages of breast cancer from I (early stage) to IV (advanced stage) and these 4 stages are classified based on similar criteria applied to TNM staging [14].

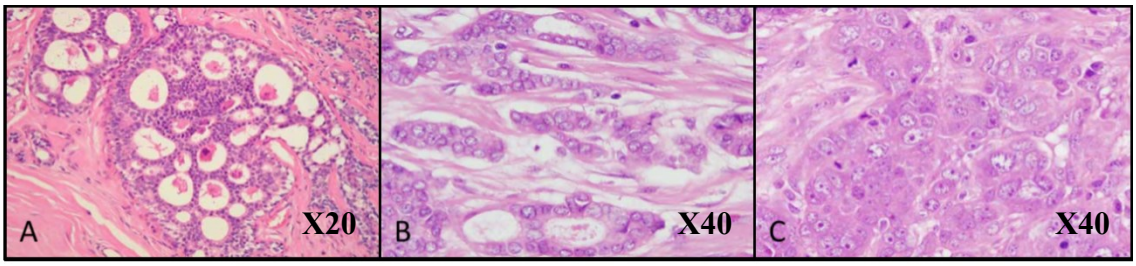


Figure 2. ***Light microscopic examination of hematoxylin and eosin (H&E) stained formalin fixed paraffin embedded (FFPE) sections of human breast tissue****. A: low power magnification of ductal carcinoma in situ (DCIS). B&C: medium power magnification of invasive infiltrating ductal carcinoma (IDC). B: grade 1 disease with noticeable tubule formation. C: grade 3 disease with pleomorphic nuclei and solid nests of tumour with no tubule formation. Images were provided courtesy of Jeremy Thomas (Consultant Breast Pathologist, Western General Hospital).*

1.1.3. Molecular Subtypes

The use of molecular markers to further classify carcinomas has been readily accepted as standard clinical practice in breast cancer diagnosis. Currently the immunohistochemical (IHC) determination of estrogen receptor alpha (ER), progesterone receptor (PR) and epidermal growth factor receptor 2 (HER2, ERBB-2) protein levels is used to guide treatment and determine which patients are likely to respond to targeted therapies (i.e. endocrine therapy for ER^+ patients and trastuzumab, pertuzumab or lapatinib for $HER2^+$ patients) [15-18]. Currently the Allred score, based on the proportion and intensity of positively stained cells, is applied to IHC staining to determine levels of ER and PR [19]. HER2 is assessed for over-expression using a scoring system ranging from 0 to 3+, where 0 and 1+ are considered negative for over-expression, and 3+ is considered positive [20, 21]. Samples that are 2+ have a further test to compare the copies of the HER2 gene with the centromere on chromosome 17 by fluorescence in situ hybridisation

(FISH). If the score exceeds 2.2 then HER2 is regarded as amplified and the cancer is classified as $HER2^{+}$. Many IHC 2+ samples have low amplification scores and so are classified as $HER2^{-}$ [20-22].

The advent of high-throughput technologies such as gene expression microarrays has led to the identification of distinct intrinsic molecular subtypes which have been confirmed and classified as: (i) basal-like (ER^{-}, $HER2^{-}$, PR^{-}; with expression of the basal cytokeratins: KRT5, KRT6A, KRT14, KRT16, KRT17, KRT23 and KRT81), (ii) $HER2^{+}$ (ER^{-}, PR^{-}, $HER2^{+}$), (iii) normal breast like (genetic profile similar to normal breast tissue; $ER^{+/-}$, $HER2^{-}$), (iv) luminal subtype A (ER^{+}, PR^{+}, $HER2^{-}$) and (v) luminal subtype B ($ER^{+/low}$, $PR^{+/-}$, $HER2^{-/+}$) [23-27]. Luminal breast cancers are also known to express cytokeratins: KRT7, KRT8, KRT18 and KRT19 [28]. More recently, additional intrinsic subtypes have been defined. A triple-negative (ER^{-}, $HER2^{-}$, PR^{-}) 'claudin-low' subtype, similar in molecular profile to the basal-like subtype, was reported which appeared to have an epithelial to mesenchymal transition (EMT) phenotype characterised by loss of claudins 3/4/7 and E-cadherin but an increase in zinc finger protein SNAI2 (SLUG), zinc finger protein SNAI1 (SNAIL), vimentin and zinc finger E-box-binding homeobox 1 (ZEB1) [29, 30]. Other studies have led to definition of the apocrine subtype (ER^{-}, PR^{-}, $HER2^{+/-}$), characterised by high expression of androgen receptors (AR^{+}), a molecular feature also reported in luminal breast cancers [31]. The apocrine subtype consists of those ER^{-} tumours out-with the basal-like group, which are reported to be AR^{+}. Amplification of HER2 is also a common feature within the apocrine subtype [32] (figure 3C).

A number of researchers have postulated that these molecularly distinct tumour subtypes may represent transformation of normal breast stem-like cells with arrest at specific stages of development, or alternatively direct transformation of mature cell types into less differentiated forms [33, 34]. The proposed model begins with an ER^{-} mammary stem cell (MaSC) which maintains itself through self-renewal and differentiates itself into committed progenitors. These progenitors essentially give rise to progeny which are mature ductal and alveolar cells (luminal epithelial cell lineage) and mature

myoepithelial cells, which surround the luminal epithelium and contact the basement membrane [34]. Molecular profiling based on each subpopulation cell type was carried out and compared with a human dataset containing representative populations of the 5 intrinsic subtypes (basal-like, claudin-low, $HER2^+$, luminal A and luminal B). The findings showed enrichment of the mature luminal cell signature in the luminal A, luminal B and $HER2^+$ subtypes and enrichment of the MaSC signature (mesenchymal signature) in the claudin-low subtype. The basal-like subtype was found to be enriched for the luminal progenitor signature, suggesting that basal-like tumours become arrested at a specific step in luminal development [33] (figures 3A&B).

Breast cancer is known to be a heterogeneous disease and much work has been done to further classify and define the ER^- subset of tumours, beyond intrinsic subtyping, using gene expression microarrays. One gene expression study of 587 triple-negative breast cancers revealed 6 molecularly distinct subgroups: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR). BL1 and 2 subtypes were both characterised by high expression of cell cycle and DNA damage response genes with BL2 also showing increased expression of growth factor receptor signalling pathways including the epidermal growth factor (EGF), nerve growth factor (NGF), c-Met (MET), Wnt/β-catenin and insulin-like growth factor 1 receptor (IGF1R) pathways. The IM subtype was enriched for immune cell signalling pathways (B-cell, natural killer cell (NK), T-cell and dendritic cell (DC) receptor signalling pathways), cytokine signalling (interleukin (IL) 12 and 7 pathways) and core immune signal transduction pathways (nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB), tumour necrosis factor (TNF), janus kinase/signal transducer and activator of transcription (JAK/STAT) pathways). Both M and MSL subtypes were enriched for EMT genes, with the MSL subtype also characterised by a decrease in proliferation associated genes and increase in stem cell associated genes. Like the apocrine subtype the LAR subtype had increased expression of androgen receptors. Comparison of these 6 subtypes with their classification based on the 5 intrinsic subtypes revealed that the majority of BL1

classified tumours were basal-like and the majority tumours in the other subtypes (BL2, M, MSL and IM) were a mix of either basal-like, normal-like or $HER2^+$. The LAR subtype tumours were mostly classified as luminal A, suggesting a luminal origin [35]. However, it should be noted that classification based on the existence of an apocrine subgroup (ER^-, AR^+) was not carried out in this study and may be a better classification for the LAR tumours (figure 3D).

Another study focused on ER^- breast cancer used an integrative analysis of three large microarray studies to define 5 molecular subtypes: steroid hormone response (SR^+), cell cycle (CC^+), immune response (IR^+), cell cycle with no immune response (CC^+/IR^+) and extra-cellular matrix (ECM^+). The SR^+ subtype was found to over-express HER2 and AR and has a molecular profile similar to the apocrine subtype discussed previously. In contrast the other 4 subtypes have low HER2 and AR expression. The CC^+ subtype was found to have increased expression of cell cycle and proliferation associated genes similar to the CC^+/IR^+ subtype which, in addition, was found to highly express immune response genes. The IR^+ subtype was characterised by low expression of cell cycle and proliferation associated genes and high expression of immune response genes. The ECM^+ subtype was found to over-express extra-cellular matrix associated genes. Comparison of these 5 subtypes with classifications using the intrinsic subtypes revealed that the majority of SR^+ classified tumours were $HER2^+$, while the majority of those classified as CC^+ and CC^+/IR^+ were basal. The IR^+ classified tumours were a mix of basal and $HER2^+$ and the ECM^+ classified tumours were largely basal and normal-like [36] (figure 3E). While these recently described subtypes have brought to light the concept of heterogeneity within the ER^- breast cancer population and have allowed for the identification of tumours with similar molecular profiles, their clinical relevance remains to be demonstrated.

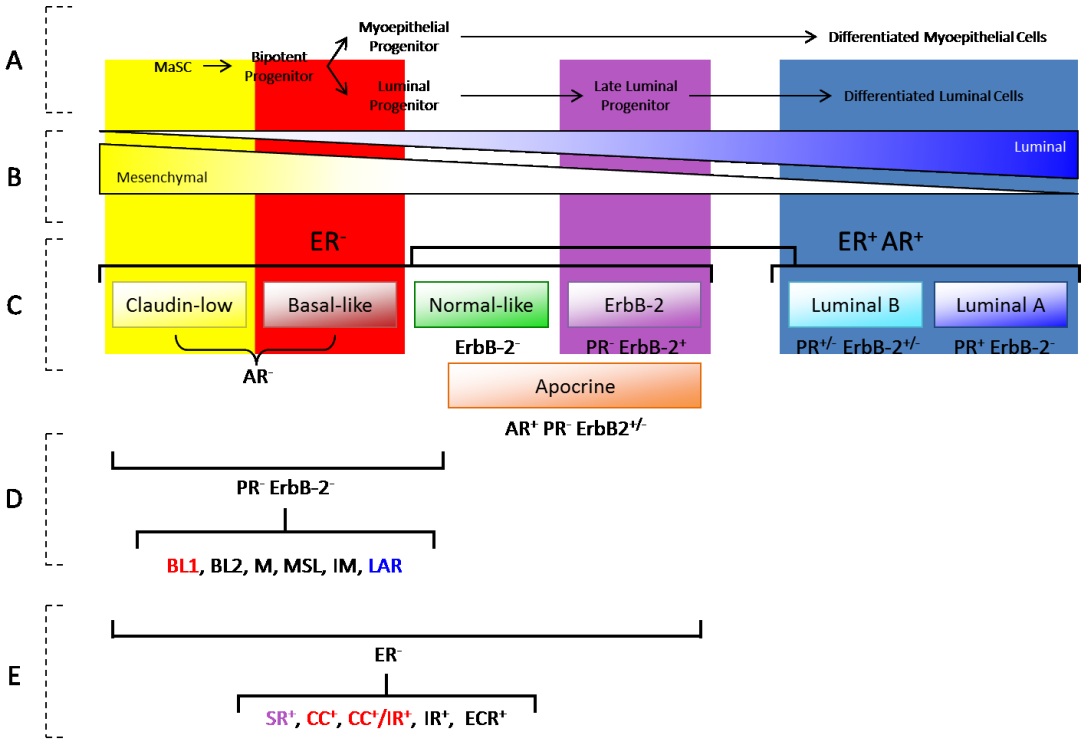


Figure 3. ***Molecular subtypes of breast cancer****. A: Model of the human mammary cellular differentiation hierarchy found in normal breast, linked to intrinsic subtypes. B: Gradient of enrichment of expression patterns defining mesenchymal and luminal signatures linked to intrinsic subtypes. C: Intrinsic breast cancer subtypes. D&E: classifications of ER- breast cancer, colour coded by association of classification to intrinsic subtypes (red=basal-like, purple=HER2, blue=luminal A, black=mix of intrinsic subtypes). D: 6 subtypes of triple-negative breast cancer; BL1=basal-like 1, BL2=basal-like 2, IM=immunomodulatory, M=mesenchymal, MSL=mesenchymal stem-like, LAR=luminal androgen receptor [35]. E: 5 classifications of ER- breast cancer; SR+=steroid hormone response, CC+=cell cycle, IR+=immune response, CC+/IR+=cell cycle with no immune response, ECR+=extra-cellular matrix [36].*

These various molecular subtypes have been identified and validated by several groups using microarray-based gene expression analysis and hierarchical clustering. They have allowed for the stratification of molecularly distinct diseases, some of which have been shown to display highly significant differences in overall survival prediction and disease-free survival, with the basal-like and claudin-low subgroups having the worst prognosis [24, 25]. Even within the ER^+ population, these approaches have allowed for a

more accurate molecular characterisation of the luminal A and B subtypes, improving on subtype assignment which was based on subjective IHC status of ER and HER2 alone, with a multi-gene molecular profile redefining luminal B as highly proliferative compared with luminal A and improving the association of luminal B with poor prognosis [24, 25]. Indeed within the luminal population work has been done to further characterise tumours so that therapeutic agents and new biomarkers to identify molecularly distinct subgroups with poor prognosis groups can be realised. Recently quiescin sulfydryl oxidase 1 (QSOX1), an enzyme involved in post-translational modifications, was found to be associated with poor prognosis in patients with luminal B breast cancer and may represent an important biomarker and target in that population [37, 38]. Indeed, another study reported that the transcriptional co-repressor ZNF703 was a driver, specific to luminal B disease [39]. A recent study proposed that luminal A and luminal B (HER2 negative) tumours should be further classified based on genomic status relating to diploid/aneuploid and chromosome instability (CIN). Two subgroups were identified: diploid/CIN- or aneuploid/CIN+ and the latter was linked to poor prognosis in the luminal population [40]. Furthermore, recent work linking copy number variation and gene expression to long-term outcome in primary breast tumours has given rise to an ER+ subgroup characterised by an 11q13/14 acquired somatic copy number aberration (CNA) which has a significantly favourable prognosis compared to a subgroup which was devoid of CNAs [41]. Continuing work to further understand the ER^+ population aims to characterise novel molecular subgroups of tumours with poor prognosis and identify new biomarkers and targets in a move towards improved, personalised treatment.

1.1.4. Breast Cancer Treatment

Five main treatment options are currently available for breast cancer: surgery, radiation therapy, chemotherapy, hormone therapy and targeted anti-HER2 therapy. In the majority of cases optimal clinical management involves surgery combined with one or more additional therapies [42]. The optimal treatment strategy is determined by considering tumour grade, disease stage (TNM), molecular characteristics of the cancer

and other patient factors such as age, menopausal status and overall health. To aid these considerations, standard clinical assessments at diagnosis often include calliper, mammographic and 3D ultrasound measurements, and core biopsy of tumour masses allowing for histopathological assessment.

Surgery to remove the tumour remains the most effective treatment for breast cancer and consists of either breast conserving surgery (BCS), followed by high-energy x-ray radiation therapy, or mastectomy [43-46]. In cases where tumour size exceeds that recommended for BCS, chemotherapy or hormone therapy can be given in the neoadjuvant setting (before surgery) to reduce tumour size sufficient to allow breast conserving surgery (BCS) [47].

Chemotherapy involves the cyclical administration of combinations of cytotoxic drugs (alkylating agents, anthracyclines, antimetabolites and taxanes) to kill cancer cells. It can be used in the adjuvant (after surgery), neoadjuvant or palliative setting [45, 48, 49]. Hormone therapy, also referred to as endocrine therapy is a treatment which removes or blocks the action of hormones such as estrogen and progesterone which are recognised as key molecular drivers in luminal breast cancers. Endocrine therapies are indicated for use in both the adjuvant and neoadjuvant settings and in the management of advanced or metastatic disease where surgery is not usually a viable option [45, 50]. Targeted anti-HER2 therapy, refers to giving monoclonal antibodies (such as trastuzumab) or tyrosine kinase inhibitors (lapatinib) that target key cellular receptors, such as HER2 which is overexpressed in around 15-20% of breast cancers and has an important role in tumorigenesis [51, 52]. Targeted anti-HER2 therapies are often combined with chemotherapy in $HER2^{+}$ patients (which are mostly ER^{-}) and have been shown to reduce risk of recurrence and risk of death after 3 years [53]. Other therapies designed to target specific pathways believed to be involved in progression and recurrence are currently being assessed for efficacy and benefit to survival in breast cancer. These include everolimus (RAD001) designed to target the mammalian target of rapamycin (mTOR) pathway by blocking the action of the key signalling protein mTOR complex 1 (MTORC1). Initial results showed a dramatic reduction in expression of proliferation-

associated genes particularly in highly-proliferative tumours, indicating that some patients with more aggressive tumours may benefit most from this therapy [54].

1.1.5. Estrogen Biosynthesis and ER$^+$ Breast Cancer

The hormone estrogen (17β-estradiol, E_2) has been identified as a key molecular stimulant in the development of ER^+ breast cancer, which constitutes around 70-80% of all breast cancers [55]. In both pre- and post-menopausal woman estrogen production occurs locally in the normal tissues of subcutaneous fat, the breast, muscle tissue and bone where it is produced by the enzymatic breakdown of androgens (androstenedione and testosterone) by aromatase [55-61]. Indeed, expression of aromatase has been shown to occur in fibroblasts within breast cancer tissue [62]. Residual levels of estrogen are also commonly found circulating in the blood plasma and are around 20-fold higher in post-menopausal women compared with pre-menopausal women despite the loss of ovarian estrogen production [57, 58, 63]. In post-menopausal women there is a noticeable correlation between risk of breast cancer and levels of circulating estrogen in the blood plasma [64]. Aromatase transcription and protein levels have been shown to differ between the quadrants of the breast in woman with breast cancer, being considerably increased in the quadrant containing the tumour [65, 66]. Evidence suggests that this is primarily due to tumour cell released factors such as prostaglandin E2 (PGE2) and inflammatory cell released cytokines such as interleukin 6 (IL6), interleukin 11 (IL11) and tumour necrosis factor alpha (TNFα), which enhance the activity of aromatase in fibroblasts through intracellular cyclic adenosine monophosphate (cAMP) signalling and regulation of the aromatase gene (CYP19) via an alternative non-standard promoter (promoter II) for the gene [55, 67, 68].

In hormone-dependent cancers, estrogen taken up from the blood plasma or from local production diffuses into the cancer cell and binds ER, in so doing causing the dissociation of heat shock proteins from the ER molecule. The ligand-bound molecule dimerises, associates with other co-activator or co-repressor proteins and subsequently binds to conserved estrogen response element (ERE) sequences within the promoter regions of genes over which it exerts transcriptional control [55, 69]. ER is a nuclear

receptor encoded by gene ESR1; it contains two distinct transactivation domains: activation function (AF)-1 in the amino-terminal region and AF2 in the carboxy-terminal region [70-72]. AF1 is regulated by growth factors acting through the mitogen activated protein (MAP) kinase pathway whereas AF2, located in the ligand binding region of ER is activated by estrogen [70, 73]. Full agonist activity requires both AF1 and AF2 to be active [74]. Studies have shown that ERE bound ER is ubiquitinated and targeted for proteosomal degradation suggesting that each ER molecule is destined for only one cycle of estrogen signalling [75]. EREs were first identified flanking the regions of estrogen-regulated vitelogennin genes in *Xenpous lavis* and have since been identified in the promoter regions of several hundred human genes, with a minimum consensus sequence consisting of a 13 base-pair palindromic inverted repeat: 5'-GGTCAnnnTGACC-3' (n: any nucleotide) [76-78]. Some EREs have been identified which have imperfect palindromic sequences, differing from the consensus sequence by one or more nucleotides and are often less responsive to ligand-bound ER than consensus sequence EREs [79]. The association of ligand-bound ER with EREs is thought to be achieved via two mechanisms: (i) 'direct binding' in which the molecule directly binds to EREs and associates directly with co-activator/co-repressor molecules and the RNA polymerase II transcription initiation complex, or (ii) 'tethering' whereby the ligand-bound ERα does not bind DNA but rather interacts with another DNA-bound transcription factor either stabilising the DNA-binding of that factor, or recruiting additional co-factors to the complex [80, 81]. The latter is thought to be the mechanism by which ligand-bound ER associates with transcription factor SP1 [80]. Via these transcriptional associations estrogen can induce proliferation of cancer cells which over-express ER (figure 4). It should also be noted that non-ligand-bound ER complexes can also bind to EREs without the ability to modulate gene expression [70, 82-84].

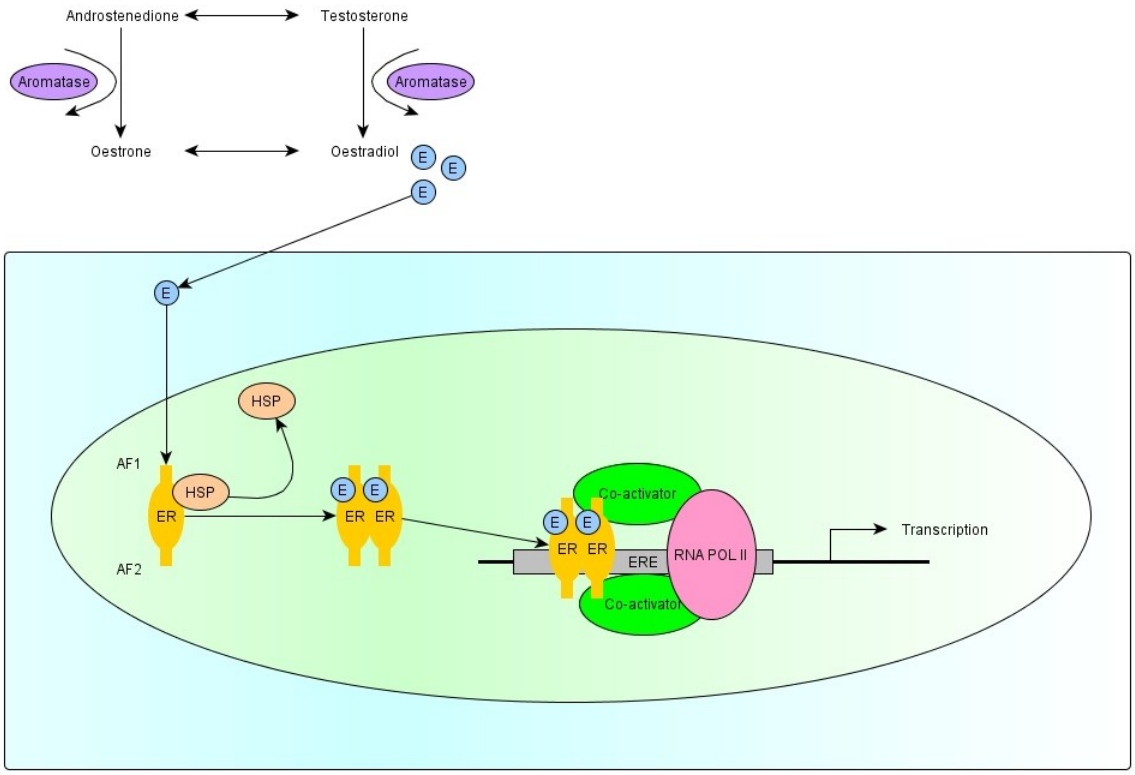


Figure 4. ***Mechanism of action of estrogen in ER$^+$ breast cancer****. Estrogen (E) is derived from the enzymatic breakdown of androgens and sequestered into tumour cells where it binds ER causing the molecule to displace heat-shock proteins and homodimerise. The ligand bound ER dimer subsequently binds to ERE in the promoter region of estrogen responsive genes and recruits co-activators which aid in promoting transcription.*

1.1.6. Signalling in ER$^+$ Breast Cancer

Although the exact mechanisms by which estrogen drives proliferation remain to be fully elucidated, a number of potential mechanisms have been proposed [85]. In normal human breast tissue, cells which do not express ER proliferate via paracrine signalling, whereas in tumours an autocrine action occurs in which ER$^+$ cells proliferate [86]. It has been reported that estrogen promotes transition through the G1/S phase of cell cycle via a number of pathways involving the induction of cyclin-D1 expression by ligand-bound ER, mediated by one or more transcription factors including: Jun oncogene (c-Jun), V-fos FBJ murine osteosarcoma viral oncogene homolog (c-Fos) and activating transcription factor 2 (ATF-2) at the activator protein 1 (AP1) promoter site, or via an

SP1 transcription factor dependent pathway [87-93]. Furthermore, ligand-bound ER has also been shown to bind cyclin-D1 and, as a complex, regulate the expression of cyclin-D1 and other downstream gene targets [91, 94]. Cyclin-D1 subsequently binds and activates cyclin-dependent kinase (CDK)-4 and CDK6, which regulate G1/S phase transition through the phosphorylation of retinoblastoma 1 (RB1). The latter can no longer inhibit E2F/transcription factor DP1 complexes, thus allowing them to activate the transcription of S-phase entry genes such as those encoding cyclin-E1 and cyclin-A1 [87, 88, 95, 96]. Other reports suggest that ligand-bound ER may promote G1/S phase transition by induction of V-myc myelocytomatosis viral oncogene homolog (c-MYC) which leads to activation of cell division cycle 25 homolog A (CDC25A) and CDK4 gene transcription [97-99]. Active CDC25A dephosphorylates CDK2 leading to the inhibition of RB1 and retinoblastoma-like 2 (p130) and transcriptional activation of E2F/DP1 complexes which in turn up-regulates S-phase entry genes [88, 100, 101]. Alternatively, it has been proposed that ER could activate G1/S phase transition via redistribution and down-regulation of cyclin-dependent kinase inhibitor 1A (p21) and cyclin-dependent kinase inhibitor 1B (P27KIP1) thus removing their inhibitory control over key cell cycle progression proteins such as CDK2 [101, 102]. It is thought this might be achieved by ubiquitin targeting for proteosomal degradation or by nuclear export via membrane-bound ER associated with mitogen-activated protein kinase 1 (ERK2) and exportin 1 (CRM1) [103, 104].

1.2. Endocrine Therapy and Resistance to Treatment

1.2.1.1. Endocrine Therapy

Endocrine therapy constitutes a major treatment modality in ER^+ breast cancer and is seen as an attractive alternative to chemotherapy which has more associated toxicity [105-108]. Indeed, studies have shown it to provide more benefit in the adjuvant setting in post-menopausal women with ER^+ breast cancer than doxorubicin or taxane-containing chemotherapies [109, 110]. The endocrine system can be manipulated via the exogenous administration of hormone antagonists designed to inhibit the biosynthesis

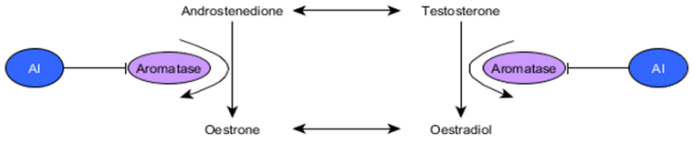
and/or activity of the key hormone estrogen. Endocrine therapies are considered to be cytostatic rather than cytotoxic, leading to reduced proliferation and reduction of growth rate [111]. At the simplest molecular level, they achieve this through the arrest of cell cycle in G1/S phase [112]. Several types of endocrine therapies exist and are used commonly in the treatment of ER^+ breast cancer in post-menopausal women.

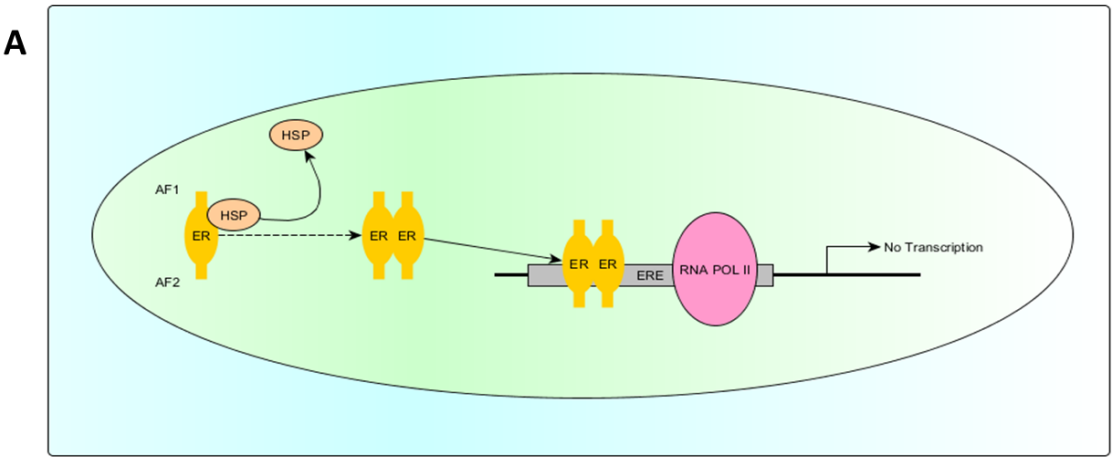
Tamoxifen (nolvadex) and raloxifene (evista) are examples of selective estrogen receptor modulators (SERMs). These function to disrupt the estrogen signalling pathway by competitive intra-nuclear binding to ER causing a conformational change to the subsequently formed ER dimer involving the shift of helix 12 into an adjacent co-activator site (AF2), thus blocking the binding of the co-activator, which significantly reduces the level of estrogen-regulated gene transcription [74, 113, 114]. However, this complex has been shown to exhibit partial estrogen-agonist properties due to the remaining activity of AF1 [55, 74]. A newer yet similar class of endocrine therapies exist which are known as selective estrogen receptor down-regulators (SERDs) and these include fulvestrant (faslodex). Fulvestrant functions to down-regulate ER by competitive binding to ER dimers and by causing immobilisation of ER in the nuclear matrix which is accompanied by degradation via the ubiquitin-proteasome pathway [115]. It has been shown to be more potent than tamoxifen *in vitro* and does not exhibit the partial estrogen-agonist properties, associated with tamoxifen, in murine models. This is due to the fact that both AF1 and AF2 activities are supressed, blocking co-factor recruitment at the ERE site of estrogen responsive genes [74, 116, 117].

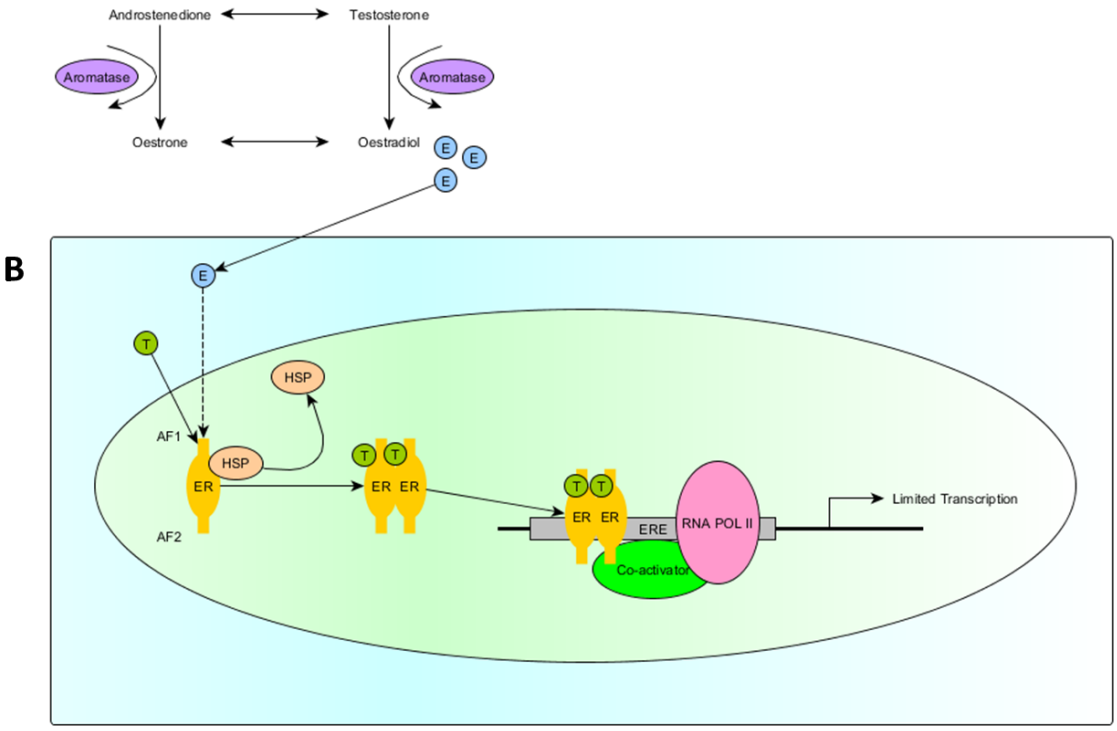
Another major group of endocrine therapies in routine clinical use are third-generation aromatase inhibitors (AIs) which comprise two drug types. Firstly, there are irreversible steroidal inhibitors (type 1) including exemestane (aromasin) which are androstenedione analogues. Secondly, there are non-steroidal inhibitors (type 2) which include letrozole (Femara) and anastrozole (Arimidex) [55, 118-123]. AIs seek to disrupt estrogen signalling by either: irreversible and inactivating binding (type 1), or reversible and competitive binding (type 2) to the aromatase enzyme; thus significantly reducing local estrogen biosynthesis and hence intra-tumoural levels of estrogen [55, 124-126]. Indeed,

in the adjuvant setting letrozole and anastrozole have been shown to be more effective than tamoxifen with a significant reduction in the rate of relapse [123, 127, 128]. Endocrine therapies can also be used in the treatment of ER^+ breast cancer in pre-menopausal women where their use is usually combined with drugs such as goserelin (Zoladex) designed to supress ovarian estrogen production [129].

In post-menopausal women, adjuvant treatment represents the major clinical setting for endocrine therapy use where long-term systemic treatment is targeted against micro-metastatic disease or progression of the primary disease [50]. Indeed several studies have reported overwhelming evidence of a high correlation between the adjuvant use of endocrine therapy and reduction in the risk of recurrence [130]. Endocrine therapy can also have an important role in the neoadjuvant setting where systemic treatment maybe indicated for 3 to 4 months prior to surgery in post-menopausal women with large and/or technically inoperable tumours. This treatment is intended to shrink the tumour such that surgery becomes possible [50, 106]. The Immediate Preoperative Anastrozole or Combined with Tamoxifen (IMPACT) study was a phase 3 clinical trial which showed that 46% of 124 ER^+ post-menopausal women initially recommended for mastectomy became eligible for BCS following 3 months of neoadjuvant anastrozole [131]. Indeed by measuring proliferation levels of malignant cells using the expression of nuclear antigen Ki67, studies have shown a reduction in proliferation in approximately 90% of ER^+ primary breast tumours treated with and responsive to AIs, confirming that these tumours derive significant proliferative stimulus from estrogen and that this can be potently supressed by endocrine therapy [132, 133]. Lastly, endocrine therapy can be used in the treatment of advanced or metastatic disease to prolong survival, either as a monotherapy or as part of a sequenced treatment regimen with chemotherapy, palliative surgery or radiotherapy [134].


Androstenedione
Testosterone
AI
Aromatase
Aromatase
AI
Oestrone
Oestradiol


A
HSP
AF1
HSP
ER
ER ER
ER ER
ERE
RNA POL II
No Transcription
AF2


Androstenedione
Testosterone
Aromatase
Aromatase
Oestrone
Oestradiol
E
E
E
B
T
HSP
AF1
HSP
ER
T T
ER ER
T T
ER ER
ERE
RNA POL II
Limited Transcription
Co-activator
AF2

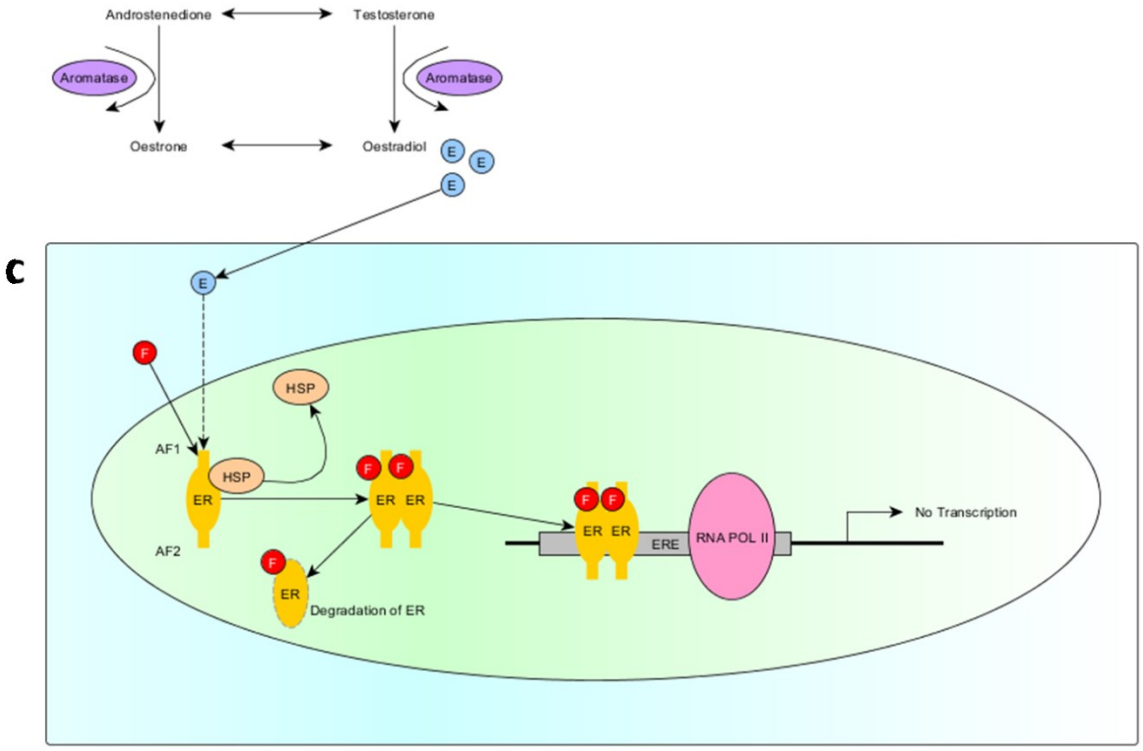


Figure 5. ***Endocrine Therapy Mode of Action****. A: Treatment with aromatase inhibitors (AI) which block the conversion of androgens to estrogen by inactivating binding to the aromatase enzyme. Non ligand bound ER homodimers can still bind EREs without the potential to activate gene expression. B: Treatment with tamoxifen (T) which competes with estrogen to bind the ER homodimers. Tamoxifen bound ER has much reduced potential to activate gene transcription, however can exhibit partial agonist activity due to the activity of AF1 which is not blocked by tamoxifen. C: Treatment with fulvestrant (F) which also competes with estrogen to bind ER homodimers. Fulvestrant bound ER has blocked AF1 and AF2 activity and as a result transcription of estrogen responsive genes is potently supressed. Fulvestrant also promotes immobilisation of ER in the nuclear matrix which is accompanied by rapid degradation via the ubiquitin-proteasome pathway.*

1.2.2. Resistance to Endocrine Therapy

Despite the benefits of endocrine therapy in the treatment of ER^+ breast cancer, resistance to treatment eventually occurs in a large number of patients and represents a severe encumbrance to optimal clinical management [50]. Clinically, resistance can manifest as a relapse or cancer recurrence, during or after completion of adjuvant

therapy; following surgery or in rare cases after complete pathological response (elimination of all cancer tissue) following a period of drug therapy. Alternatively, in the neoadjuvant setting, resistance can be observed as clinical progression of primary disease, usually constituting an increase in primary tumour size or disease spread to regional nodes or beyond to more distant metastatic sites. Pathological changes such as increased tumour grade or increased proliferation are indicators of potential resistance to therapy. In the neoadjuvant setting, resistance occurs as either a primary lack of response (no change or an increase in tumour size and no evidence of pathological response) early in treatment, implying innate resistance, or later following a period of response, suggesting acquired resistance [135]. Studies have shown that recurrence on adjuvant endocrine therapy occurs in approximately 10-15% of patients with early stage ER^+ breast cancer within 5 years [136]. In the neoadjuvant setting 5-10% of patients exhibit innate resistance with a further 25-30% developing acquired resistance to treatment within 5 years [137, 138]. The majority of patients with advanced or metastatic ER^+ disease acquire resistance within 2-3 years of starting endocrine therapy [139].

1.2.3. Mechanisms of Resistance

While little is understood about the exact mechanisms underlying clinical response to endocrine therapy, even less is known about the mechanisms leading to innate or acquired resistance. Several different mechanisms have been described which set the basis for continued research aimed at improving the outcome of endocrine treatment.

1.2.3.1. ER and PR Expression in Tumours with Innate Resistance

Studies have suggested that innate resistance may be linked to lower levels of ER, which might suggest that the drive to proliferation of these cancers is not as highly dependent on estrogen as those expressing higher levels of ER. The current use of the Allred score for assessing ER levels by IHC staining categorises all tumours with greater than 10% of positively stained cells as ER^+ despite the enormous variation within that group and it also does not cope well with intratumoural heterogeneity, with parts of the tumour less ER^+ than others [140, 141]. However, rather than a mechanism explaining resistance this

suggests that endocrine therapy alone may be an inappropriate treatment choice for these patients particularly if there is heterogeneous ER staining.

Clinically, the decision to treat with endocrine therapy is primarily based on ER status; however, IHC levels of PR are also determined at diagnosis. While the majority of ER^+ patients are also PR^+, a subset is PR^-, and some studies have linked this genotype with innate resistance. Indeed, tumours which are ER^+ PR^- display a poorer response to endocrine therapy and a more aggressive phenotype than ER^+ PR^+ tumours, and studies have shown that the ER^+ PR^+ population has a significantly better prognosis compared with the ER^+ PR^- group [142]. Studies looking at differential chromosomal loss and gain between ER^+ PR^- and ER^+ PR^+ tumours have shown loss of regions containing genes associated with tumour suppression and apoptosis in the PR^- genotype. Furthermore, gains have been identified in regions in PR^- tumours which encode genes including: mitogen activated protein kinase kinase kinase 3 (MAP3K3), ribosomal protein S6 kinase beta 1 (RPS6KB1), zinc finger protein 217 (ZNF217). Amplification of these genes could lead to activation of the phosphoinositide 3 kinase - protein kinase B-mammalian target of rapamycin (PI3K-AKT-mTOR), which has been implicated with endocrine therapy resistance [143].

1.2.3.2. PI3K-AKT-mTOR Pathway and Somatic Mutations

PI3K is activated by growth factor receptor tyrosine kinases (RTKs) and G-protein coupled receptors (GPCRs). PI3K phosphorylates phosphatidylinositol 4,5-bisphosphate (PIP2) to produce PIP3 which recruits several molecules such as PDK1 and AKT to the plasma membrane which, on activation, drive cell cycle progression and survival [144-146]. The pathway is negatively regulated by PTEN and INPP4B which dephosphorylate PIP3 and PIP2 respectively [147, 148]. AKT activates mTORC1 which regulates protein synthesis. The PI3K-AKT-mTOR pathway interacts with ER both directly and indirectly. AKT can phosphorylate ER which increases estrogen-induced, tamoxifen-induced and ligand independent ER transcriptional activity [149, 150]. In addition, PI3K promotes c-Jun phosphorylation, which complexes with c-Fos to form the AP-1 complex, which is known to be involved with ER transcription [151-153].

Studies have also shown that the PI3K pathway can be activated by HER2, the over-expression of which has been linked to a weaker response to endocrine therapy and poor prognosis following adjuvant therapy (see section 1.2.3.7). In this model, activation of the PI3K pathway confers resistance to tamoxifen, fulvestrant and deprivation of estrogen [149, 154, 155]. Somatic mutations, which represent the most common in ER^+ breast cancer, have been described in key members of the PI3M-AKT-mTOR pathway including PI3KCA, PIK3CB, AKT1, AKT2, PTEN and INPP4B which have been linked to aberrant activation and potential dependence on the pathway [156].

1.2.3.3. Genetic Traits Affecting Drug Action

Some studies looking at tamoxifen have alluded to resistance being linked to inherited genetic traits affecting drug metabolism. Some patients carry recessive alleles of a gene encoding a dysfunctional isoform of a protein known as cytochrome P450 2D6 (CYP2D6) which is involved in the conversion of tamoxifen to its active metabolite form endoxifen [157, 158]. Around 8% of Caucasian women carry these recessive alleles and are likely to respond less well to tamoxifen treatment than women carrying wild-type alleles of the gene [158]. Similar findings have been suggested in the study of aromatase inhibitors with innate resistance to letrozole being linked to polymorphisms in the gene encoding aromatase (CYP19) [159, 160]. Currently the importance of drug metabolising enzymes and their genetic variation is thought not to be an important factor in endocrine resistance and as such no research in this area has yet reached the clinic.

1.2.3.4. ER-Independent Signalling

Other reports suggest that innate resistance may be a feature of ER^+ tumours in which proliferation is regulated by an ER-independent signalling mechanism. Studies have shown that acquired resistance can occur in tumours with low levels of ER resulting from loss of expression or mutations in its encoding gene (ESR1), suggesting an ER-independent driving mechanism for proliferation [161, 162]. However, loss of ER is only seen in approximately 15-20% of resistant breast cancers and the incidence of ESR1 mutations is even lower, with less than 1% of resistant cases reported having this

genotype [163-166]. Alternative mechanisms involve expression of truncated isoforms of ER such as ERα36 or other estrogen-related receptors such as estrogen-related receptor gamma (ERRγ), both of which have been associated with reduced response to tamoxifen. Tamoxifen works by inactivating E2 binding to wild-type ER, resistance in the case of ERα36 over-expression could be a feature of lower binding affinity to the truncated isoform of the molecule. Studies have shown that expression of ERα36 can be induced by bone morphogenetic protein 2 (BMP2), a member of the bone morphogenetic protein family of proteins which are recognised as having roles in regulation of cell fate and cancer development, suggesting a potential role for this molecule in endocrine therapy resistance [167]. Resistance related to ERRγ over-expression might suggest an important role for this molecule in an alternative estrogen signalling pathway [168, 169]. Furthermore, it should be noted that estrogen receptor exists as two distinct isoforms: ER (ERα) and ERβ. The exact role of ERβ is not clear, however studies have shown that tamoxifen can bind ERβ and that tamoxifen-bound ERβ can activate Ap1 regulated genes, possibly by altering the balance of associated co-activators and co-repressors at the promoter site [170, 171]. Indeed, increased ERβ expression has been reported in tamoxifen resistant breast cancers and data from recent study suggested that the ratio of ERα to ERβ may be important in predicting response to tamoxifen and anstrozole in the neoadjuvant setting [172, 173].

1.2.3.5. ER Post-translational Modifications

A number of post-translational modifications of ER have been reported, including phosphorylation, methylation and sumoylation which influence its interaction with other members of the ER-signalling pathway. It has been suggested that aberrations in the post-translational modification of ER could be linked to endocrine resistance [174, 175]. ER can be phosphorylated at a number of different sites including serine-118, serine-167 and threonine-311 within the AF1 binding domain as well as in other domains. Phosphorylation and activation of ER at key positions can result from a number of pathways including: the MAPK/ERK pathway in response to growth factors such as epidermal growth factor (EGF), the PI3K-AKT pathway in response to insulin-like

growth factors and the p38-MAPK pathway in response to stress or various cytokines [176, 177]. As tamoxifen can still bind partially activated ER, over-expression and cross-talk between these pathways regulating ER activation might explain the partial agonist capabilities of the drug.

1.2.3.6. Differential ER Binding

A recent study looked at genome-wide ER binding events in primary breast tumours of patients sensitive and resistant to tamoxifen and revealed that in tamoxifen-resistant cancers ER is still recruited to the chromatin and binds regulatory regions in a pattern that is unique to resistant tumours [178]. The resistant phenotype may be due to selection and expansion of a resistant subpopulation of cells, or alternatively could involve the rapid reprogramming of ER binding by forkhead box protein A1 (FOXA1), which has a known role in ER-chromatin interactions in response to growth stimuli [178-181]. Forkhead motifs and EREs were found to be enriched within the DNA regions which showed increased ER binding in tamoxifen-resistant cell lines and in primary tumour specimens of patients with a poor clinical outcome, providing further evidence for the FOXA1-mediated reprogramming model of ER binding [178]. These findings suggest that ER may have an important role to play in tamoxifen resistance by binding to a distinct set of regulatory elements giving rise to a unique gene expression profile which drives tumour progression and confers resistance to therapy.

1.2.3.7. Epidermal Growth Factor Receptors

Resistance to endocrine therapy is common in ER^+ breast cancers that over-express HER2 [182]. Many studies have reported cross-talk between ER and receptor tyrosine kinases (RTKs) such as epidermal growth factor receptor 1 (ERBB-1; EGFR/HER1) and HER2, which are receptors for epidermal growth factor (EGF) and insulin-like growth factor 1 (IGF-1, somatomedin 1) [183, 184]. Over-expression of these receptors suggests that tyrosine kinase signalling is driving proliferation and evasion of apoptosis in these cancers, representing either a primary mechanism in the case of innate-resistant tumours or a switch in driving mechanisms to evade the action of endocrine therapy in tumours

with acquired resistance [154, 185]. Studies have also shown that EGF-signalling can lead to an EGF-induced ER with regulatory control, dependent on AP1, over a set of genes commonly over-expressed in $HER2^+$ breast cancer, which are distinct from those regulated by estrogen-induced ER [182]. These data indicate that proliferation may be driven by a distinct EGF-dependant mechanism which is independent of estrogen signalling. One group has suggested that EGF-induced ER might arise from EGF-signalling in response to soluble stromal factors including fibronectin and matrix metalloproteases, secreted into the tumour microenvironment by fibroblasts, which associate with tumour cell membrane-bound β1 integrin thus activating the PI3K/AKT pathway and MAPK/ERK pathway leading to ER phosphorylation and activation [186]. Furthermore, this indicates that patients with ER^+ and $HER2^+$ breast cancer could benefit from endocrine therapy combined with immune therapy targeted at EGF receptors. Indeed, this approach has been shown to work in cell line models [128], although studies in humans have been disappointing. This mechanism fits well with studies which have suggested that ligand bound-ER is associated with repression of ERBB-1 and HER2 [187]. Depletion of estrogen by endocrine therapy could lead to expression of these molecules via loss of activated ER repression. Studies have shown that expression of a transcription factor known as paired box gene 2 (PAX2) is associated with reduced incidence of endocrine therapy resistance due to its role in ER-mediated repression of HER2 [188]. One study reported that the response of cells to tamoxifen is regulated by competition between the ER co-activator amplified in breast cancer 1 (AIB-1) and PAX2 binding to the *cis*-regulatory elements in intron 4 of HER2. Indeed, they showed that a decrease in expression of PAX2 in tamoxifen resistant cells correlated with an increase in HER2 expression. Furthermore immunohistochemical staining of tamoxifen treated ER^+ breast cancer tumours showed that PAX2 expression in the absence of AIB-1 correlated with recurrence free survival (RFS) and a low rate of HER2 expression. Conversely patients which expressed higher levels of PAX2 and AIB-1 had a higher rate of RFS [188, 189].

It should also be noted that the mechanisms behind over-expression of HER2 are complex and have been shown to involve genetic and epigenetic modification, as well as alterations in upstream regulators such as under-expression of X-linked tumour suppressor forkhead box P3 (FOXP3) and transcription factor GATA4 [190, 191].

1.2.3.8. Breast Cancer Stem Cells

There is now a large body of evidence suggesting an important role for stem cells in the development of breast tissue and that there exist cancer stem cells (CSCs) in tumours [192]. In breast cancer these undifferentiated, clonogenic cells are associated with an increased invasive and metastatic phenotype, however, their frequency is dependent on tumour grade, disease stage and molecular subtype (discussed previously: section 1.1.3) [193-197]. Normal breast stem cells are thought to be basal-like and mainly ER^-. Consequently, it is thought that CSC development is not influenced greatly by hormones such as estrogen and that these cells may be resistant to endocrine therapy as a result of low ER expression, with any partial response attributed to paracrine signalling from nearby differentiated ER^+ tumour cells. Normal breast stem cells are regulated by EGF receptor and other growth factor receptor signalling. Some groups have suggested that the observed increase in EGF receptor expression in endocrine therapy resistant tumours may reflect a greater proportion of CSCs selected by endocrine therapy [192]. Furthermore, one study showed that letrozole treated tumours appeared to have an expression signature that was more like 'claudin-low' tumours and that post-treatment tumours were enriched for stem cells compared to pre-treatment samples [198].

1.2.3.9. Cell Cycle

Response to endocrine therapy essentially manifests at a molecular level as a G1/S phase arrest in cell cycle, a feature which is devoid in resistance cancer cells [112]. The resistant phenotype has been linked to amplification and/or over-expression of cyclin-D1 and cyclin-E1 [199]. Increased expression of cyclin-D1 is associated with activation of Cdk4 and Cdk6 and progression to the S-phase of cell cycle. Indeed, studies have linked over-expression of cyclin-D1 to tamoxifen resistance and over-expression of cyclin-E1

to letrozole resistance [200, 201]. Another study identified a link between over-expression of cyclin-E2 and resistance to tamoxifen [202]. Studies have also reported modifications in other key players associated with the estrogen-dependent regulation of cell cycle (discussed previously: section 1.1.6) including C-MYC, RB1, p21 and P27KIP1, which have all been linked to endocrine therapy resistance [101, 203, 204]. A more recent study has identified that mutations in the gene encoding tumour protein 53 (TP53) are also linked to resistance to aromatase inhibitors [205].

1.2.3.10. Cell Survival and Apoptosis

Another important feature of endocrine therapy resistance is evasion of apoptosis. While the mechanisms remain poorly understood, studies have reported increased expression of anti-apoptotic proteins such as B-cell lymphoma 2 (BCL2) and decreased levels of pro-apoptotic proteins such as BCL2 homologous antagonist killer (BAK), BCL2 interacting killer (BIK) and caspase 9 in tamoxifen resistant models [206, 207]. Conversely, some studies have shown that loss of BCL2 expression can occur in antiestrogen-resistant cell lines and this has been linked to increased sensitivity to chemotherapeutic agents including cisplatin [208]. In addition, the PI3K-AKT pathway which is associated with cell survival was found to be stimulated by upstream regulators in endocrine resistant models (see section 1.2.3.2) [206]. Additionally, NF-κB has been shown to have a role in blocking apoptosis through cross-talk with ER and regulation of the baculoviral IAP repeat-containing protein 3 (BIRC3) gene [209].

1.2.3.11. NFκB Signalling and Inflammation

NFκB plays an important role in processes such as cell survival and proliferation [210]. It can promote proliferation through regulation of key cell cycle genes including cyclins and CDKs and can mediate growth and survival signals via the PI3K-AKT pathway [211, 212]. NFκB has been reported as over-expressed in some endocrine therapy resistant breast tumours and many groups have alluded to its role in endocrine therapy resistance [213-218].

Some studies have reported cross-talk between ER and NFκB signalling in which cooperative binding to transcriptional response elements can lead to specific gene expression [219]. Conversely, ER has also been shown to inhibit NFκB via a mechanism involving displacement of NFκB co-regulators such as CREB-binding protein (CBP) at the site of the NFκB response element [220]. One study suggested the involvement of 'tumour growth factor beta (TGFβ) activated kinase 1 / MAPK7-interacting protein 2' (TAB2) in tamoxifen resistance in which a phosphorylated active form of TAB2 exports the co-repressor protein 'nuclear receptor co-repressor 1' (NCoR) from the nucleus, thus translocating it from EREs, resulting in loss of response to tamoxifen [221]. Tab2 has also been implicated in the activation of NFκB via the 'IκB kinase' (IKK) complex in response to 'interleukin 1' (IL1) stimulation [222].

Alterations in the NFκB cascade have also been identified in endocrine therapy resistant cells. In tamoxifen resistant cells, expression of the NFκB subunit p50 was increased however remained unchanged in fulvestrant resistant cells compared with the sensitive cells. Conversely, expression of the p65 subunit was found to be increased in fulvestrant resistant cells but remained unchanged in tamoxifen resistant compared with sensitive cells. The most abundant form of NFκB is the p50-p65 heterodimer and these findings suggest that resistant cells may utilise different strategies for up-regulating the activity of this molecule [210]. Furthermore, the phosphorylation levels of p65 at its serine-536 site were found to be increased in endocrine therapy resistant cells. Phosphorylation at this site is necessary for optimal activity of the molecule and was shown to enhance its transactivation potential [223].

Some risk factors identified for breast cancer, including increased age and menopause, are associated with increases in indicators of systemic inflammation such as increased levels of circulating pro-inflammatory cytokines [224, 225]. Other risk factors such as pregnancy and obesity have been linked to the promotion and maintenance of a local inflammatory microenvironment in the breast [226]. Tumour associated macrophages (TAMs) have been found to comprise up to 50% of a breast tumour mass in some patients. Indeed, increased macrophage infiltration in breast tumours has been positively

correlated with increased angiogenesis as well as reduced relapse-free and overall survival [227, 228]. One study in which TAMs were co-cultured with ER^+ breast cancer cells revealed an increase in cancer cell invasiveness compared with cultures with no TAMs present. This was reported to involve a mechanism in which an inflammatory cytokine known as 'tumour necrosis factors alpha' (TNFα), produced by macrophages, led to activation of NFκB and 'c-Jun N-terminal kinase' (JNK) pathways [229]. Indeed, high expression of the macrophage marker 'cluster of differentiation 68' (CD68) in breast cancer is associated with poor prognosis and lack of response to endocrine therapy [230]. Furthermore, increased circulating levels of tumour necrosis factor alpha (TNFα) have been associated with advanced tumour stage, lymph node metastasis and local invasion [231, 232]. TNFα has been shown to stimulate proliferation in some ER^+ cell lines through increased expression of cyclin-D1 by a mechanism dependent on NFκB [233, 234].

Upon TNFα stimulation the p65 NFκB subunit is phosphorylated at serine-536 via the IKK complex. Upon IL1 stimulation the PI3K-AKT pathway mediates phosphorylation of p65 at serine-536 via an unknown mechanism, although 'TANK-binding kinase' (TBK), IKK and p38 have all been implicated with IL1-induced p65 phosphorylation at serine-536 [235, 236]. TNFα-induced transcriptional activity of NFκB has been found to be significantly increased in endocrine therapy resistant cells compared to sensitive cells. This is thought to be due to the increased expression of NFκB subunits and increased levels of p65 phosphorylation. Interestingly, PR (discussed previously: section 1.2.3.1), has a known anti-inflammatory role in breast cancer cells and the loss of its expression in a subset of endocrine therapy resistant ER^+ cell lines which over-express NFκB has been reported [237, 238].

1.2.3.12. Hyper-sensitivity to Residual Estrogen

Hyper-sensitivity to estrogen has also been suggested as a potential mechanism of endocrine therapy resistance [55]. Aromatase inhibitors function to dramatically reduce estrogen biosynthesis however residual amounts of estrogen remain in tissues [239]. Cell line studies have shown that after prolonged deprivation of estrogen some cells develop

hyper-sensitivity to residual estrogen. Reports have suggested that this phenomenon is associated with an increase in expression of HER2 and subsequent over-activity of the MAPK pathway resulting in changes in the phosphorylation of ER conferring its hyper-sensitivity to residual estrogen [240, 241]. Other research has indicated that mutations in ESR1 may give rise to mutated ER which has increased interactions with the proto-oncogene tyrosine-protein kinase (SRC) family of co-activators and changes in promoter binding dynamics linked to hypersensitivity to estrogen [242].

1.2.3.13. Epithelial-Mesenchymal Transition

Epithelial-mesenchymal transition (EMT) is a morphological change which has been shown to occur in some epithelial tumours [243, 244]. EMT is associated with a loss of differentiation and loss of intracellular adhesion, a feature of an invasive phenotype characterising advanced metastatic disease [245-247]. Intracellular adhesion is an important feature of tissue architecture maintenance and can limit cell movement and proliferation. It is primarily mediated through the adherens junction (AJ) which are complexes of calcium-dependent transmembrane cadherin receptors, the cytoplasmic domains of which link to the actin cytoskeleton via α-catenin and β-catenin [248]. One study reported that development of tamoxifen resistance in cell lines is associated with an enhanced motile and invasive phenotype characterised by loss of intracellular adhesion and partial EMT. This phenomenon is thought to be brought about by the EGF-signalling mediated activation of β-catenin and the subsequent increase in expression of β-catenin regulated genes known to be involved with tumour progression. Inhibition of EGF-signalling in the same cells reduced β-catenin activity and promoted intracellular adhesion [249]. This suggests a possible role for EGF-signalling (involving β-catenin) in the manifestation of an aggressive phenotype of endocrine therapy resistant tumours. In another study, over-expression of the transcription factor zinc finger protein SNAI1 (Snail) resulted in EMT. In this case ER expression was also lost as a result of Snail binding to the promoter region of ESR1 and causing deacetylation of histone H3K9 [250]. This represents a potential EMT-associated mechanism by which endocrine resistance could develop and might suggest the involvement of epigenetic mechanisms.

1.2.3.14. Epigenetics

Epigenetic mechanisms including DNA methylation have been shown to be responsible for determining and maintaining cell fate and for the stable differentiation of cells [251]. An increasing body of evidence is building to suggest an important role for DNA methylation in cancer including the silencing of tumour suppressing genes, activation of oncogenes and promotion of metastasis [252-254]. More recently, changes in DNA methylation have been linked to endocrine therapy resistance [255, 256]. Studies comparing endocrine therapy (tamoxifen or fulvestrant) resistant and sensitive cell lines have identified a pattern of methylation characterised by promoter hypomethylation in the resistant cell line compared with the sensitive [255]. This mechanism of hypomethylation was further implicated in a report showing the development of tamoxifen resistance in sensitive cell lines treated with a DNA methylation inhibitor (5-azacytidine) [256]. One study also suggested the hypermethylation of the ERβ gene is associated with tamoxifen resistance [257]. In addition, it has been shown that ER itself participates in epigenetic control. When ER binds EREs within the genome it recruits cofactors involved with epigenetic control including: nuclear receptor corepressor 1 (NCOR1), NCOR2, steroid receptor coactivator 1 (SRC1) and amplified in breast cancer 1 (AIB-1) [258, 259]. Tamoxifen resistance has been linked with dysregulation of these co-factors suggesting a possible role for epigenetic mechanisms in endocrine resistance. Indeed, a recent study suggested a novel concept that prolonged tamoxifen exposure could induce epigenetic silencing of a set of estrogen responsive genes with functions associated with the negative control of proliferation [260].

1.2.3.15. Autophagocytosis

Autophagocytosis is a cellular catabolic degradation process involving the lysosomal machinery whereby aggregated proteins, unfolded or misfolded proteins or damaged subcellular organelles are degraded in response to stress or nutrient deprivation in an attempt to restore metabolic homeostasis. While the role of autophagocytosis in endocrine therapy resistance remains poorly understood, the process is known to be both pro-survival and pro-death with the final cell fate dependent on its extent and duration

[261, 262]. Some studies have shown that inhibition of autophagocytosis can re-sensitise resistance cells to tamoxifen suggesting that it might have a role in resistance in some cancers [261]. One study identified a protein known as 'heat shock protein B beta 8' (HSBP8) as having a role in regulating autophagocytosis in endocrine therapy resistant cells [263].

1.2.3.16. Heterogeneity of Resistance

A wide range of distinct mechanisms have been described that have been implicated with endocrine therapy resistance in breast cancer these include: lower levels and heterogeneity of ER expression; post-translational modifications of ER and differential binding of ER; ER-independent signalling including EGF-signalling and NFκB-signalling; genetic aberrations affecting drug action and control of cellular processed such as cell cycle and apoptosis; phenomena such as stem cells, EMT, epigenetics, estrogen hyper-sensitivity and autophagocytosis. Together, these findings suggest that resistance to therapy is likely to be complex, heterogeneous and may differ between primary and acquired resistance and even between endocrine therapy types. Most studies have shown that a proportion of patients who become resistant to tamoxifen can respond when treatment is switched to an aromatase inhibitor and vice versa, albeit to a lesser extent [55]. Groups have highlighted the heterogeneity in gene expression between resistant tumours from series of patients all of whom were treated with letrozole indicating that even within cancers specifically resistant to certain drugs heterogeneity is evident [264].

Much work remains to be done in order to understand fully and characterise the intricacies and interplay of the mechanisms which underlie endocrine resistance in breast cancer.

1.2.4. Combating Resistance

A number of treatment strategies have been used clinically and assessed in trials to counteract endocrine therapy resistance in breast cancer, including alternating or combining agents [265-267]. While combining tamoxifen and aromatase inhibitors

simultaneously does not appear of benefit, patients resistant to tamoxifen have been shown to respond when treatment was switched to an aromatase inhibitor [55].

Another approach has involved augmenting standard endocrine therapy with agents designed to re-sensitise resistant tumours to endocrine therapy by targeting pathways and molecules recognised as drivers of resistance. One such approach has been the combination of endocrine therapies with HER2 targeted therapies such as trastuzumab and lapatinib, which have shown some promise in endocrine therapy resistant cancers which over-express HER2 [268, 269]. A number of studies have also shown that the use PI3K-AKT-mTOR pathway targeted therapies such as the mTOR inhibitor everolimus and EGFR inhibitor gefitinib can reverse the PI3K-AKT-mTOR mediated resistance to endocrine therapy (discussed previously: sections 1.2.3.1/2/5/6/9/10) when used in combination with endocrine therapy [268, 270-274]. Indeed, everolimus is now in wide use in combination with exemestane (a steroidal aromatase inhibitor) after being shown to improve progression-free survival (PFS) in second or third line treatment of patients with ER^+ metastatic breast cancer [156, 275].

Lab based studies have also yielded some promising results. One cell line study revealed that cells resistant to tamoxifen can be re-sensitised to its growth-inhibitory effects by targeting and blocking the action of the NFκB pathway [210]. Another group suggested that by inhibiting O-6-methylguanine-DNA methyltransferase (MGMT), a DNA repair protein they identified as being over-expressed in some tamoxifen resistant cell lines, sensitivity to tamoxifen could be restored [276]. Another reported that a CDK2 inhibitor could reverse endocrine therapy resistance in tumours over-expressing cyclins-E1 and E2 [202].

A novel approach involved the withdrawal of endocrine therapy. Cell line studies have provided evidence for the growth inhibitory effects of withdrawal from endocrine therapy, however there is little clinical evidence to support this. One group reported a clinical trial which suggested that resistance to endocrine therapy could be minimised by intercalating therapy with periods of withdrawal [277].

A further treatment option was suggested in which low-dose estrogen is intercalated with aromatase inhibitors [278]. Indeed, before the development of drugs such as tamoxifen, high-dose estrogen represented a major therapy for the treatment of hormone-dependent breast cancer. It was thought to work by inducing apoptosis via extrinsic Fas/Fas ligand and intrinsic mitochondrial pathways [279]. Another study reported that the apoptosis inducing action of estrogen involves endoplasmic reticulum stress response and inflammatory response genes [280]. One group demonstrated in hormone-dependent xenograft models that loss of response to letrozole was accompanied by up-regulation of HER2 and MAPK pathways and down-regulation of ER and aromatase activity, which was reversed by replacing aromatase inhibitors with low dose estrogen treatment for a short period of time, thus re-sensitising the cells to estrogen and hence aromatase inhibition [278]. The major problem with these lab based cell line studies is that they take no account of the wide range of cancers seen in clinical practice. Their value at best is to provide leads for pathways or mechanisms of resistance that might be targeted.

1.3. Studying Resistance and Predicting Response

1.3.1. Challenges with Studying Endocrine Therapy Resistance

Resistance to endocrine therapy has been investigated using a number of different approaches which extend from fundamental cell line studies in culture or as xenografts in immunosuppressed animals, to clinical adjuvant and neoadjuvant treatment studies [50].

1.3.1.1. Cell Line Studies

A number of cell line based studies, both as *in vitro* cultures or as *in vivo* xenografts have been used to investigate endocrine therapy resistance and have involved endocrine therapy treated breast cancer cells transfected with the aromatase gene (CYP19) [128, 281, 282]. Such studies have elucidated some important findings and have advantages including the ability to assess dynamic changes with numerous time-points which is not practical in studies involving patients [283]. However, they have several limitations

including an inability to accurately model the heterogeneity known to be present between and within individual primary breast tumours due to their clonal nature. They are also unrepresentative of the local tumour microenvironment found in patients, in particular they lack stromal and immune components [50]. Cell line models do not represent a realistic model to evaluate endocrine resistance as seen in clinical practice, as many of the therapies such as aromatase inhibitors cannot be studied. Furthermore, many of the mechanisms identified in such cell lines have been found to have no or limited clinical utility.

1.3.1.2. Adjuvant Setting

Investigations involving adjuvant treatment are limited because the primary tumour has been surgically removed and is thus not available for on-treatment molecular analysis. Furthermore, measuring the response to treatment relies on monitoring survival and disease recurrence. These require long-term follow up with carefully documented study outcomes which are difficult to monitor successfully as recurrence may be a result of inherent cancer aggressiveness rather than acquired resistance to therapy. Additionally, for studies to produce meaningful results the patient sample size must be sufficiently large to be statistically valid and data collection on relevant outcomes takes some considerable time to acquire [50].

1.3.1.3. Neoadjuvant Setting

The neoadjuvant setting presents a number of advantages for investigating the characteristics of response and resistance. The primary tumour remains in place during treatment and clinical response can be determined by considering changes in tumour volume as determined by 3D ultrasound or mammographic measurements. In addition, tumours can be biopsied, often at multiple and sequential times, allowing for assessment of dynamic changes in gene expression or protein levels as treatment continues. These data can be related to clinical response allowing for a dynamic comparison of clinical and molecular response in both responsive and resistant tumours [50, 124, 284, 285].

1.3.2. Predicting Response

There is a clinical need to identify early during treatment those patients who are unlikely to gain clinical benefit from endocrine therapy, thus sparing them several months of ineffective therapy and unnecessary side-effects. This is of particular relevance in the neoadjuvant setting, where the goal is to shrink tumour size such that a cancer becomes operable or less extensive surgery becomes possible.

1.3.2.1. Pathological and Clinical Response

Following 3 or 4 months of neoadjuvant endocrine therapy, 60-80% of tumours show signs of pathological response. This usually includes both a decrease in tumour volume and a decrease in cellularity with an increase in fibrosis or formation of fibrous connective tissue and in some cases a decrease in histological grade (discussed previously: section 1.1.2) [286, 287]. Complete pathological response (CPR) to endocrine therapy is rare, but the rate of CPR increases with the duration of treatment [288]. It should also be noted that quantitative measurements of partial pathological responses rely on subjective assessments without robust formal criteria, and concordance between pathologists has been reported to be only in the range of 50-86% [289, 290]. Indeed, while the clinical relevance of grade is well characterised, 40-50% of tumours are histologically classed as grade 2, and interobserver reproducibility of grade is suboptimal, although there has been great efforts to improve reproducibility [291]. A number of pathologically responding tumours decrease in size and this constitutes the clinically used definition ‘clinically responsive’ [50]. Indeed, clinical and pathological response is significantly associated in most tumours [288]. However, it is worth noting that around 20% of tumours are discordant in this respect, showing either a clinical response (decrease in volume) with no evidence of pathological response, or responding pathologically while not changing in size [290, 292].

These clinical and pathological determinants of response require repeated measurements and histological assessments respectively and, while they offer an effective end-point for categorising overall response, they often do not manifest rapidly enough to be suitable

for predicting, early-on-treatment, which tumours are likely to respond or not to endocrine therapy [50].

1.3.2.2. Proliferative Response

Significant decreases in proliferation after 3 months are seen in approximately 80% of ER^+ tumours in response to endocrine therapy and can be seen as early as 10-14 days into treatment. Levels of proliferation are routinely established by measuring expression changes of Ki67. Some tumours however display different patterns of Ki67 expression such as: an initial decrease followed by recovery to pre-treatment levels, a delayed change, and in some cases there is little change over the three months of treatment [292]. Ki67, discovered in 1991, is a nuclear non-histone protein which was reported to be absent from quiescent cells and universally expressed in proliferating tissues. Based on these characteristics, Ki67 became established as a marker of proliferation. However, work carried out since then showed that expression of Ki67 varies in intensity throughout cell cycle which could impact the identification of proliferating cells [293]. Indeed some studies have shown that the expression of Ki67 in the G1 phase of cell cycle can be minimal [294, 295]. That said, expression of Ki67 has been shown to correlate positively with other markers of proliferation such as proliferating nuclear antigen and minichromosome maintenance protein 2 (MCM2) [296, 297]. While the association between Ki67 and response to chemotherapy has been demonstrated [298, 299], no significant association has been reported with neoadjuvant endocrine therapy [300-302]. Indeed, while an early proliferative response is positively and significantly correlated with both clinical and pathological response (sensitivity) to endocrine therapy, discordancy occurs in a number of cases rendering it a poor predictor of which patients (specificity) are likely to be resistant to endocrine therapy [292, 303].

1.3.2.3. Molecular Response Markers

The most commonly used predictive molecular marker for endocrine therapy is ER. It has a strong negative predictive value with ER^- tumours hardly ever responding to endocrine therapy and around 50-70% of ER^+ tumours responding well. However, still

around 30-50% of ER^+ patients will exhibit innate or acquired resistance to endocrine therapy (discussed previously: see section 1.2.2) [135]. Another molecular marker with predictive value is PR. This molecule is regarded as a classical marker of estrogenic activity and is reduced in 70-80% of cases treated with endocrine therapy [288, 292]. However, loss of PR expression can occur independently of clinical and pathological response [292, 304]. Indeed, studies have shown increased response rates in PR^+ tumours compared with PR^- tumours; however response to endocrine therapy can occur in both [304-306]. Similar results were found with other classical markers of estrogenic activity including pS2 [304]. Additionally, it has been shown that ER^+ tumours over-expressing HER2 are less likely to respond to endocrine therapy; however the number of such cases is small and they do not account for all resistant cases [307, 308].

The use of single molecular markers to predict response is far from adequate given their poor correlation with clinical and pathological response [309, 310]. Even expression of ER, clinically the most widely used molecular marker, has only around 50-70% accuracy in endocrine therapy response prediction in ER^+ cases.

1.3.2.4. Multigene Signatures

1.3.2.4.1. Gene Expression Profiling

The advent of high-throughput gene expression profiling technologies, such as gene expression microarrays, applied to translational research has revolutionised the way breast cancer is perceived; highlighting the importance of heterogeneity and the fact that distinct subtypes of the disease exist which can affect the same anatomical site [311]. Indeed much work has focused on developing multigene signatures using high-throughput gene expression microarray technology, which have primarily been used for molecular classification; thereby confirming the importance of key disease drivers such as ER and HER2 signalling (discussed previously: section 1.1.3), and for prognosis [311-313]. A number of prognostic studies in breast cancer to date have focused on identifying subsets of patients with such a good prognosis, where the absolute benefit is small compared with associated toxicity, that they could forgo adjuvant systemic

chemotherapy [314]. Two of the most advanced prognostic signatures designed for this purpose and available for clinical use are Mammaprint® and Oncotype DX®, both developed to estimate recurrence risk in node-negative early breast cancer [230, 315, 316].

1.3.2.4.2. Commercial Profiling: Prognostic Assays

The 70-gene Mammaprint® test uses fresh tissue and microarray technology to evaluate expression of genes associated with proliferation, invasion, metastasis, tumour stroma and angiogenesis in both ER^+ and ER^- cancers [316]. The test was given Food and Drug Administration (FDA) approval in 2008 to be used as a prognostic test (indicating risk of relapse within 5 years) for breast cancer patients greater than 61 years of age with node negative stage 1 or 2 tumours greater than 5cm in size. Those patients with bad prognosis are recommended for chemotherapy and endocrine therapy combined adjuvant treatment, whereas those with good prognosis are recommended to receive only adjuvant endocrine therapy [311]. It should be noted that this test is of limited clinical use in ER^- breast cancer with only 0-4% of such patients predicted to have a good prognosis [317-321]. An independent validation of the Mammaprint® signature demonstrated that prognostic accuracy is highly time-dependent and may be more suitable as a predictor of early relapse [316, 320]. Furthermore, no clinical validation studies of the Mammaprint® signature have been performed in randomised trial populations [322-324].

The 21-gene Oncotype DX® test uses 'formalin fixed paraffin embedded' (FFPE) tissue with quantitative real-time PCR (qRT-PCR) to measure expression of 16 genes associated with proliferation, estrogen regulation, HER2 and invasion in hormone receptor positive cancers, as well as 5 reference genes [230, 315, 325]. The test outcome is presented as a Recurrence Score (RS) ranging from 0-100 to predict the risk of 10 year distant recurrence. In clinical use the RS is subdivided into three groups: low (<18), intermediate (18-31) and high (>31), and several publications have shown that ER^+ breast cancer patients with low RS have a low risk of recurrence and derive little benefit from chemotherapy, whereas the reverse is true for those with high RS [325-329].

According to the National Comprehensive Cancer Network (NCCN) guidelines for breast cancer treatment, patients with low RS score should be treated with endocrine therapy alone, while patients with high RS score should receive a combination of endocrine therapy and chemotherapy. However, one issue with the Oncotype DX® test is that the optimum clinical management strategy for patients with intermediate RS scores remains at present unclear [311].

More recently, a molecular diagnostic test known as Map*Quant* DX^{TM} was launched to accurately measure tumour grade, risk of metastasis, predict response to chemotherapy and indicate proliferation. This test utilises microarray technology and measures expression of the 'Genomic Grade Index' (GGI), a published 97 gene signature [330-333]. When applied to an independent validation cohort the GGI had strong association with histological grade, although 9% of grade 1 tumours were classified as having high GGI and 14% of grade 3 tumours were classified as having low GGI. However, GGI was found to be more strongly associated with relapse free survival than histological assessment of grade. Furthermore, GGI was able to stratify histological grade 2 cancers into prognostically significant high or low grade groups. In independent validation studies GGI has also shown to be independently prognostic of outcome (risk of recurrence) in tamoxifen treated patients. Tumours with high GGI profiles were also reported to respond well to neoadjuvant chemotherapy despite having a significantly worse outcome compared with low GGI tumours [330]. The Map*Quant* DX^{TM} test was recently converted to an 8 gene assay (PCR-GGI) based on quantitative real-time PCR (qRT-PCR) measurements [334]. While the evidence for genomic grading is compelling, independent validation of the Map*Quant* DX^{TM} and PCR-GGI systems has not yet been documented. Furthermore, their ability to discriminate between high and low grade in ER^{-} tumours is limited [335, 336].

Theros® is a qRT-PCR based assay designed for lymph node negative ER^{+} breast cancer treated with surgery alone reported to assess risk of recurrence and benefit from endocrine therapy. The test measures the ratio of expression of homeobox gene Hox-B13 (HOXB13) to interleukin 17B receptor (IL17BR) [311]. It has been shown to

identify ER^+ tamoxifen treated patients with a high risk of recurrence and predict outcome in ER^+ patients, either adjuvant systemic therapy naïve or tamoxifen treated. However, its ability to predict benefit from neoadjuvant hormone therapy or from chemotherapy remains to be seen [337, 338]. Furthermore, the test has been shown to have limited power in lymph node positive disease [339]. It should also be noted that neither the Map*Quant* DX™/PCR-GGI nor Theros® tests have been included in the NCCN guidelines for breast cancer treatment [311].

1.3.2.4.3. Prognostic Signature Research

Several other gene signatures, not yet adopted as commercially available tests, have also been reported in an attempt to refine tumour classification and improve prognostication.

The 'invasiveness gene signature' (IGS) is based on the hypothesis that tumours are arranged in cellular hierarchy, initiated and maintained by a small population of cancer stem cells with self-renewing potential which express antigen CD44 and are low or negative for signal transducer CD24 [340, 341]. The signature consists of 186 genes associated with heightened risk of death and metastasis in different types of cancers; however, in breast cancer it was only significant for ER^+ grade 2 disease [342]. It is also noteworthy that the study was underpowered, with the signature being derived from only 5 metastatic pleural effusions. Furthermore, it is questionable whether the gene expression profile of a rare subgroup of cancer stem cells (accounting for only 0.6-5% of tumour breast cells) can be detected in samples which are largely tumour bulk (composed mostly of non-tumorigenic cells) [311, 340].

One group compared the expression profiles of stroma from the tumour bed (connective framework of tissue surrounding tumorigenic cells) of 53 breast cancers and stroma from normal tissue, identifying a list of 200 most variable genes between the two groups. They derived a 26 gene 'stroma-derived prognostic predictor,' which was found to be independently predictive of outcome, highlighting the importance of the stromal compartment in prognostication. However, the biological implications and applicability

of the 'stroma-derived prognostic predictor' in the clinical setting are yet to be validated [311, 343].

The 'sensitivity to endocrine therapy' (SET) index comprised a signature of 165 genes which were co-expressed with ESR1 in 437 patients (unrelated to treatment or outcome). The association of the signature with distant relapse risk was evaluated in 5 independent validation cohorts: 2 cohorts (n=225 and 298) of patients who received adjuvant endocrine therapy, a cohort who received adjuvant chemotherapy followed by endocrine therapy (n=122) and two cohorts who received no adjuvant systemic therapy (n=208 and 133). The SET index was found to be significantly associated with distant relapse and death risk in endocrine therapy and chemotherapy plus endocrine therapy treated cohorts, but not in systemic therapy naïve cohorts. While this signature has been extensively validated in independent cohorts and been shown to predict survival benefit, there is as yet no evidence suggesting accurate prediction of response to therapy or inherent prognosis [344].

A recent test known as EndoPredict (EP), based on gene expression has been validated to predict the likelihood of distant recurrence in patients with ER^+ $HER2^-$ breast cancer treated with adjuvant endocrine therapy. The test uses RNA levels of a panel of 8 genes (plus 3 reference genes), as determined from RT-PCR of FFPE tissue, which are used to calculate an EP score. EP score and nodal status together with HER2 status are combined to give a comprehensive risk score: EPclin. Using the EPclin score, 58%-68% of women from two large phase III trials classified as having high/intermediate risk of recurrence (according to clinical guidelines) were predicted to be low risk. The rate of occurrence of distant metastases in this group was found to be only 5%. Further validation work has shown reproducible performance of the assay and negligible inter-laboratory variation. The authors suggest that use of the EPclin score may reduce the indications for chemotherapy in ER^+ post-menopausal women with a limited number of clinical risk factors [345-348]. While this test show promises in identifying subgroups of patients with a low risk of recurrence on adjuvant therapy and has been shown to out-

perform current clinical guidelines, there is once again no evidence, as yet, to suggest that the test can predict response to therapy or inherent prognosis.

1.3.2.4.4. Lessons from Prognostic Signature Research

Numerous prognostic signatures have been proposed beyond those already discussed including signatures assessing: amplification of cyclin-D1, inactivation of p53, activation of PI3K pathway, MAPK cascade and prediction of invasion [342, 349-353]. Most prognostic signatures have significant agreement in their prediction of outcome and identification of similar groups of patients despite the fact that the overlap of individual gene lists is negligible [335, 336, 354-360]. Meta-analysis of different prognostic signatures, looking at genes, pathways and networks revealed that classification of patients into good or bad prognosis groups relies heavily on expression of proliferation genes. Indeed some signatures were found to perform better when only proliferation associated genes were used to predict prognosis [336]. Furthermore, this analysis confirmed that the prognostic power of most signatures, particularly the four commercially available assays (Mammaprint®, Oncotype DX®, Map*Quant* DX™ and Theros®) was limited to the $ER^+/HER2^-$ subgroup of breast cancers, providing more evidence for proliferation as the key determinant of prognosis in this subgroup [335, 336]. The study and use of prognostic signatures has revealed important characteristics of breast cancer including the fact that ER^+ disease comprises a spectrum of cancers which range from low proliferation with good outcome to highly proliferative cancers with poor outcome. Whereas ER^- tumours largely comprise distinct entities, driven by distinct molecular aberrations, whose prognosis may not be determinable using proliferation gene based signatures [328, 361, 362]. Indeed, recent studies have suggested the importance of immune response genes in the prediction of outcome in ER^- breast cancer [36, 363, 364].

1.3.2.4.5. Predicting Response to Endocrine Therapy

Unlike prognostic signatures which relate gene expression to outcome (risk of relapse), predictive signatures are measurements associated with response or lack of response to a

particular therapy [311]. Such research is aimed at determining the optimal systemic therapeutic regimen and the magnitude of benefit associated with it for individual patients, in a move towards individualised therapy [311, 354].

A number of predictive multigene signatures have been reported which directly relate gene expression to the clinical and pathological response to endocrine therapy in ER^+ patients. The majority of which have not yet been validated for clinical use. One of the first developed for endocrine therapy was a study to predict clinical response to tamoxifen in patients with recurrent disease (local/regional relapse or distant metastasis). Gene expression profiles of primary tumour samples from 112 patients who later recurred and received tamoxifen were derived. Of the 112 patients, 52 responded with an objective decrease in size of recurrent mass and 60 had progressive disease. A 44-gene signature was derived, which included genes associated with estrogenic regulation, apoptosis, extracellular matrix remodelling and immune response. The signature was reported to predict endocrine therapy outcome and time to progression in ER^+ patients with recurrent disease. In independent validation the gene signature was shown to predict response to tamoxifen in 77% of patients, outperforming the commonly used ER marker which predicts response in 50-60% of metastatic patients [365]. However, it should be noted that this signature has not yet been shown to be of significant predictive value in patients with early stage disease.

One of the first gene expression studies focused on response to aromatase inhibitors was published in 2007 by Mackay *et al* [366] and reported response to letrozole and anastrozole during short pre-operative (14 days) treatment using sequential biopsies from 34 patients. The study revealed that short term estrogen deprivation by aromatase inhibitors led to profound changes in transcriptomic profile, including genes associated with proliferation and estrogen signalling. While many of the changed genes reported had been previously identified in cell line studies, many additional estrogen responsive genes were identified in this study. Importantly the study revealed complex changes in matrix remodelling and stromal interactions which cannot be easily studied in cell lines. The reported response gene changes were integrated into a global index of dependence

on estrogen (GIDE), a measure of the number of genes with at least a 2-fold change on treatment. The GIDE was found to be significantly associated with on-treatment changes in Ki67 and to pre-treatment levels of HER2 [366]. While time-course studies in cell lines have revealed much about the response to estrogen deprivation, cell lines are not representative of the tumour microenvironment and may not be an optimal model for investigating endocrine therapy resistance (discussed previously: section 1.3.1.1). This pilot study was one of the first to use multiple biopsies from the same patient, allowing for an assessment of changes in gene expression rather than static expression levels at a given time point. By comparing on-treatment with pre-treatment in the same patient, this valuable approach has allowed for identification of key genes which are consistently changed over a number of patients as a direct result of a given therapy. While the design of this study is novel and may have potential for determining key predictive genes associated with sensitivity or resistance to endocrine therapy, several limitations are apparent. Firstly, the study is underpowered with only 34 patients, and no independent validation was performed to confirm the findings. Furthermore, given that the study was based on only 14 days of treatment prior to surgery, clinical response could not be directly evaluated, as changes in tumour size take longer than 14 days to manifest (discussed previously: section 1.3.2.1). Instead, changes in immunohistochemical Ki67 were used as a surrogate for clinical response, which despite correlation with both clinical and pathological response, has been shown to be a suboptimal predictor of response due to discordancy in some patients (discussed previously: section 1.3.2.2). As a result, there is currently no conclusive evidence demonstrating the predictive capacity of this signature.

In 2007, Miller *et al* [284] published a similar study designed to investigate changes in gene expression associated with short-term neoadjuvant letrozole therapy also using pre-treatment and on-treatment (14 days) biopsies from the same patient. From gene expression profiling of 58 patients, 143 genes were identified which were consistently changed between pre-treatment and 14 days biopsies. Using the most significantly and consistently changed genes patients were stratified into 4 distinct molecular groups by

hierarchical clustering, reinforcing the concept of heterogeneity of response [284]. However, the clinical significance of the molecular subgroups remains to be seen. The molecular changes observed in both Mackay *et al* [366] and Miller *et al* [284] studies were largely consistent, with aromatase inhibitor treatment leading to supressed expression of genes associated with proliferation and estrogenic signalling, and increased expression of genes involved with stromal remodelling, cell adhesion and immune response [50]. Importantly, these changes were identifiable within two weeks of treatment; long before clinical changes and pathological changes in morphology could be seen (discussed previously: section 1.3.2.1), raising the potential for determining early-on-treatment which patients are likely to respond and adapting therapeutic protocols in resistant patients.

In 2008, Harvell *et al* [367, 368] published one of the first gene expression signatures reported to discriminate between tumours clinically responsive and non-responsive to endocrine therapy which used clinical response after extended therapy (4 months) as the end-point criteria. They published a 50 gene signature which was later refined to a 25 gene signature based on pre-treatment gene expression levels. However, the signature was based on expression profiles from only 6 patients: 3 responsive and 3 non-responsive rendering the study significantly underpowered. In addition, no independent validation of their findings was carried out.

In 2009, Miller *et al* [124] published the largest study of its type at the time of publication in which they presented a gene expression signature reported to discriminate between tumours clinically responsive and non-responsive to aromatase inhibitors. Findings were based on the same microarray gene expression dataset used in their 2007 publication [284]. Neoadjuvant clinical response to letrozole in the 58 patients was evaluated by changes in 3D ultrasound measurements taken over a 3 month treatment period (immediately prior to surgery), giving rise to 37 responding tumours and 15 showing lack of response. Gene expression analysis revealed 205 co-variables consistently differentially expressed between clinically responding and resistant tumours, which distinguished between the two response groups. Of the 205 genes, 69

were differentially expressed in pre-treatment samples, 45 were differentially expressed in 14-days samples and 91 were significantly different when considering expression changes between pre-treatment and 14-day samples. Hierarchical clustering based on the 205 genes separated clinically responding and resistant tumours into two distinct groups. The most predictive genes were found to be associated with protein biosynthesis, in particular of ribosomal proteins. Interestingly, changes in proliferation associated genes and estrogenic signalling genes were found to occur in both clinically responding and non-responding cases [124]. The major limitation of this study is the lack of validation of predictive capacity in an independent cohort of patients which as yet has not been reported. In addition, the assessment of clinical response (at least 50% reduction in tumour volume by 3 months) was an arbitrary threshold and does not allow for a satisfactory and clear differentiation between clinically responsive and resistant tumours, e.g. a tumour with a reduction of 51% would be classed as clinically responding whereas a tumour with a 50% reduction would be resistant.

1.3.3. Considerations when Developing Predictive Gene Signatures

1.3.3.1. Sample Size and Patient Heterogeneity

A number of multigene signatures have been reported to predict subgroups of ER^+ patients unlikely to respond to endocrine therapy. However, many of these studies are underpowered, with findings based on relatively small numbers of patients in the training set and as a result may not be representative of the population [369]. When considering gene signatures derived from a single time point, such as before treatment, this problem may be confounded by patient heterogeneity (variables such as age, BMI, tumour size, tumour grade, tumour histological subtype, lymph node involvement, metastasis, additional medical conditions or drug regimens and inherent genetic differences). By considering consistent changes in gene expression between sequential biopsies (before and after/during treatment) from the same patient in a pairwise fashion, the issues of patient heterogeneity can be somewhat minimised and significantly changed genes are more likely to be directly related to treatment response. Nevertheless,

for statistical validity it is desirable to recruit the largest number and most representative patients as possible.

1.3.3.2. Microarray Bias

The predictive capacity of many reported signatures is often not reproducible in external datasets [313]. This can be in part be due to the small unrepresentative sample size in the original training set; however, it can also be related to issues of microarray bias. Several studies have alluded to microarray bias as a major contributory factor affecting the reproducibility of microarray data derived results [313, 369, 370]. Bias can occur at all stages of microarray experiments from patient selection and sample processing to choice of microarray platform. Bias of this nature can lead to results becoming dataset-specific. Some studies have reported methods which can lead to a reduction in microarray bias which have been shown to significantly improve reproducibility [369, 370].

1.3.3.3. Independent Validation

A major limitation of published predictive gene signatures is a lack of independent validation, proving that findings are real, that they are strongly associated with clinical outcome and out-perform or improve upon current clinically used predictors of response. Successful validation is an essential first step if any new predictive signatures are to be endorsed in the clinic to improve patient care [312]. Furthermore, most predictive signatures are derived from microarray data, which should be considered a surrogate for gene expression, implying a further need for validation of results in external independent datasets and by alternative methods for measuring gene expression such as qRT-PCR. Independent validation is often carried out using publically available external datasets which may or may not be similar in design to the pilot study. Important factors for consideration when selecting an appropriate validation study for endocrine therapy response might include: the patient cohort (age range, BMI range, menopausal status), disease features of the cohort (tumour size, spread to nodes, metastasis, ER/PR/HER2 status), treatment (drug, length of treatment window) and assessment criteria used for response. With this in mind, it should be noted that while novel experimental designs,

such as gene expression profiling of multiple biopsies taken from the same patients over a period of treatment, are promising and may yield important information, a lack of similar external experimental datasets with which to validate findings is likely.

1.3.3.4. Response Criteria

Predictive multigene signatures for neoadjuvant response to therapy are designed by comparing differences in gene expression, or differences in changed gene expression, between tumours defined as responsive or non-responsive. The predictive capacity of such signatures is highly dependent on the criteria used for response assessment. Such criteria include clinical, pathological and molecular assessment of response (discussed previously: sections 1.3.2.1 to 1.3.2.3) and, while these show reasonable concordance, they do not agree for every case. Therefore, if the aim of a neoadjuvant predictive signature is based primarily on the clinical need to reduce tumour size sufficiently to allow surgery if the cancer was inoperable or breast conservation if operable only by mastectomy then it would seem logical to use clinical response assessment (as determined by changes in periodic 3D ultrasound over the treatment period) as the primary end point [50]. However, pathological and proliferative responses do represent useful secondary end points. Although it should be noted that using proliferation as an end point might lead only to identification of proliferation associated genes, which have been indicated as being poor predictors of clinical response to aromatase inhibitors [124]. If clinical response (changes in tumour size or volume) is to be used as the main end point then attention must be given to the cut-offs applied for each response group. Rather than applying a single arbitrary cut-off where all tumours above a certain value are classed as resistant and all below this value are responsive, it may be beneficial to design a predictive signature based on 2 well characterised groups of tumours that respond well or not at all. For example, tumours which reduce by a least 70% by 3 months would be classed as responsive, while tumours which increase or decrease by no more than 50% by 3 months would be classed as non-responsive, all tumours in-between (intermediate clinical response) could be excluded from the development of a predictive signature design. This approach may lead to the identification of a predictive signature

that has greater power to differentiate between responsive and non-responsive tumours. It may also be able to stratify subsequently the intermediate clinical response group into clinically relevant responsive or non-responsive subgroups, ultimately identifying which individual patients are likely to benefit from alternative or combination therapy.

1.3.3.5. Heterogeneity in the Resistant Patients

Previous studies have alluded to heterogeneity within the clinically resistant group despite a similar clinical response to treatment [264]. However, predictive multi-gene signatures for endocrine therapy in ER^+ patients to-date have failed to take this heterogeneity into account, instead considering all clinically resistant cases as belonging to the same molecular group. For this reason, predictive capability is likely to be dataset-specific, with reproducibility highly dependent on the frequency and distribution of distinct molecular subtypes within validation datasets compared with the training set. The exploration and characterisation of distinct molecular subgroups within the clinically resistant patients may be an important consideration in the development of signatures with greater predictive accuracy and reproducibility. Indeed, it should be noted that the accurate elucidation of distinct molecular subtypes will doubtless require a large sample number; once again reinforcing the importance of this factor in the experimental design.

1.3.3.6. Clinical Application of a Predictive Signature

The application of gene expression technology in the field of breast cancer research has yielded much important information and has vastly improved understanding of the disease at a molecular level [313]. However, there are a number of important considerations to address before it can be readily applied in a clinical setting to aid diagnostic and treatment decisions [312]. Foremost, a predictive signature must demonstrate clear clinical benefit for patients. Furthermore, it must be shown, ideally with prospective independent validation, to out-perform or improve upon currently used parameters for guiding clinical decisions. In addition, thought must be given to the specific technology chosen for the assay, and its implementation. Technologies range

from expensive high-throughput commercial or custom microarray platforms to low cost lower-throughput qRT-PCR assays. While most predictive signature are developed from microarray data, there may be potential to convert resultant signatures to lower cost technologies such as qRT-PCR as was the case with Oncotype DX® and Map*Quant* DX™ [230, 315, 325, 328]. Another benefit of qRT-PCR is that the assay can be effectively performed on formalin-fixed paraffin embedded (FFPE) tissue with greater ease than with microarray technology. Importantly, FFPE tissue is far easier to collect, store and work with than fresh frozen tissue. A number of factors can affect the choice of technology. Firstly, costs can be prohibitive and studies must be conducted to assess the cost-to-benefit ratio [312]. Secondly, the number of genes included in the assay is a point for consideration, as large numbers of genes may preclude the application of lower-throughput technologies such as qRT-PCR. There is also a known issue with reproducibility; microarray bias (discussed previously: section 1.3.3.2) can significantly impact the reproducibility of results and can manifest from differing technical variables (platforms, RNA extraction, processing and hybridisation techniques) used across different sites [369]. It may be possible to minimise bias by instigating a strict common protocol for implementation of a predictive assay at different sites; however, commercial enterprises including Oncotype DX® and Mammaprint® address the issue by only offering their assay at one controlled site using the same validated technology and delivered by the same technicians.

1.4. Aims and Approaches

There is a clinical need to identify, before or early-on-treatment, subgroups of patients who are or will acquire resistance to endocrine therapy, avoiding months of ineffective therapy or disease progression and ultimately improve patient care. The primary aim of this study is to identify novel predictive biomarkers of clinical response to endocrine therapy in ER rich primary breast tumours, improving upon currently used clinical and pathological determinants of response and out-performing previously reported predictive models. The study will take into account molecular heterogeneity which is known to

exist within the clinically non-responsive population in order to improve predictive power and reproducibility [264].

This study will involve the generation of a new whole genome expression microarray dataset, representing the largest of its type, profiling multiple tumour biopsies taken from each of 35 patients before, after 14 days and again after 3 months (before surgery) of continuous neoadjuvant letrozole therapy. The primary end point of the experiment will be clinical response as defined by dynamic changes in 3D ultrasound volume measurements taken from each patient periodically over the 3 months of therapy. Pathological response will also be assessed for each patient.

For the purpose of predictive marker investigation, cut-offs for tumour volume reduction/increase (clinical response) will be assigned so as to create two clearly separated groups of consistently responsive and non-responsive tumours, excluding those tumours with intermediate and transient response profiles. It is hypothesised that the resultant signature of predictive biomarkers, based on the well characterised response groups, may have heightened predictive power and reproducibility.

While selecting strict cut-offs for clinical response seems a logical step to improve discovery of robustly predictive markers, it will inevitably limit the sample size of the dataset lowering statistical significance, especially in the number of non-responsive cases which are usually less frequent than clinically responsive cases [138]. Indeed, this is confounded by the fact that the availability of matched pre- and on-treatment samples from patients with complete clinical follow-up is limited. To address this issue by increasing sample size, this study will also involve the development and validation of novel raw intensity full genome microarray data integration approaches, which have value out-with this project as a useful bioinformatic tool. However specific to this study, it will allow for the successful integration of the new microarray dataset (this study) with a previously published dataset with the same design [124], thereby generating the largest dataset of its type. This part of the study will involve an investigation into the nature of the bias imposed by biological and technical variability, which has been shown to exist

between different microarray experiments and which will need to be sufficiently minimised for successful integration of the data [369, 370]. In this case, the issue of microarray bias is confounded by the fact that both datasets were derived from different microarray technologies/platforms, which has previously been described as an insurmountable problem [371]. If successful, a pipeline for the integration of microarray data from different platforms will be developed and validated then applied to investigate endocrine therapy resistance in ER^+ breast cancer.

There is a clinical need to predict response as early as possible. If a predictive signature is to out-perform current clinical or pathological determinants of response, it must accurately predict response before the clinical or pathological changes manifest or become evaluable. For that reason, the developed signature of predictive biomarkers, derived from the integrated data, will be based on: pre-treatment, 14 days on-treatment and changes between these two time-points only. Investigations will be carried out to determine which time-point or combination of time-points will give rise to the signature with the greatest predictive power. Previous studies have alluded to 'change by 14 days' as being most predictive, however, this has never been conclusively confirmed and if found to be the case may inform the clinical need for a 14-day biopsy [50].

The developed signature of predictive markers will be validated in an independent external microarray dataset, profiling multiple tumour biopsies taken from each of 44 patients before and after 14 days of continuous neoadjuvant anastrozole. Clinical response in this dataset was assessed from repeated tumour measurements taken periodically from each patient over 16 weeks of treatment [372, 373].

2. Direct Integration of Intensity-level Data from Affymetrix and Illumina Microarrays improves Statistical Power for Robust Reanalysis

2.1. Acknowledgements and Contributions

I would like to acknowledge the contribution made to this chapter by Robert Kitchen PhD, who carried out the nested analysis of variance described in the materials and methods (section 2.3.3), the results of which are displayed in figure 11 (section 2.4.1), which he produced for this study.

2.2. Introduction

2.2.1. Combining Datasets for Meta- or Re-analysis

In the clinical sciences, systematic review of multiple attempts to answer a specific research question is a valuable tool to synthesise high-quality empirical evidence in order to determine a consensus view. Such reviews or meta-analysis have greater statistical power to identify true effects from insignificant artefacts and, as such, are capable of identifying subtle effects that might be missed or deemed insignificant in smaller datasets. In the context of gene-expression analyses, meta-analysis of results from microarray studies has great potential, but also presents significant challenges due to differences in the platforms and analysis approaches employed in each study [369, 370, 374-376]. Direct integration of probe-level expression data from multiple studies is potentially even more powerful, but is further complicated due to differences in the conditions under which each dataset was generated, such as the amplification or labelling method, the scanner used or even just the date on which the samples were processed. A recent comprehensive review found that the aims of different microarray meta-analysis studies were quite distinct, with the majority combining p-values, effect size or ranked analysis, with only 27% (51 studies) seeking to directly merge the data [374]. Previous work has demonstrated that non-trivial systematic bias or 'batch effects' can occur within Affymetrix GeneChip and Illumina Beadarray experiments, potentially

arising from variation at multiple stages in the experimental workflow; from patient selection and sample acquisition to laboratory processing and subsequent analysis of derived data (figure 6). However, such systematic bias can largely be removed from each with appropriate batch correction methods [370, 376, 377].

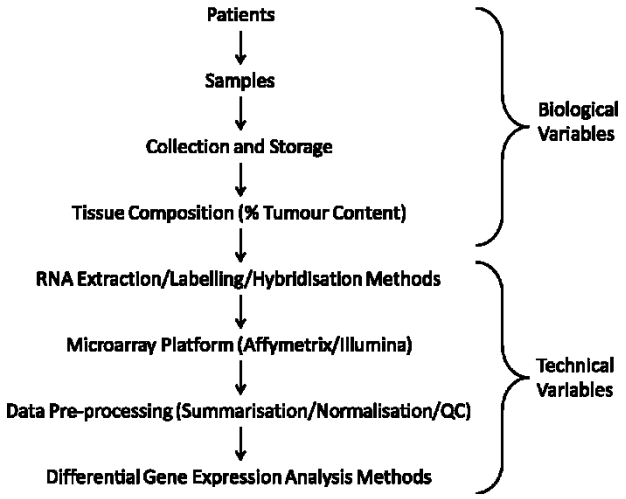


Figure 6. ***Overview of a microarray experiment workflow****. Biological and technical variation can occur at several stages, contributing to systematic bias. QC: quality control. Adapted from Sims et al (2009)[313].*

2.2.2. Dataset Size and Statistic Power

Gene expression profiling has been applied to many areas of translational cancer research, including identification of new drug-targets, monitoring response to treatment, revealing mechanisms of resistance, and predicting prognosis [313]. Although the majority of microarray datasets are now made publicly available, many studies are limited in size and therefore cannot accurately reflect the general population, as they lack statistical power [164, 378]. Collection of clinical material often remains the rate-limiting step, particularly with 'window-of-opportunity' studies that utilise matched *before*- and *after*-intervention samples from the same patient [54, 124, 377, 379, 380] (see section 1.3.3.1). Due to the reduced patient-patient variation, these studies can be highly effective for identifying consistent gene-expression changes, such as the effects of (neoadjuvant) cancer treatment.

The extensive patient- and tissue-diversity inherent in molecular studies of cancer, which often contribute to underpowered studies [378] and confounding [381], mean that it is currently not necessarily critical (or appropriate) to measure gene-expression at the greatest resolution or specificity now offered by exon-arrays and RNA-sequencing. Rather, it may be of greater utility to maximise the number of existing biologically independent observations by combining the growing numbers of datasets in the public repositories, than simply generating another small independent dataset with limited statistical power [313].

2.2.3. Directly Integrating Affymetrix and Illumina Data

Previous comparisons of expression measurements derived from Affymetrix and Illumina platforms have reported, 'generally consistent' [382], 'very high agreement' [383] or 'correspondence across platforms was high' [384]. However, these studies are often based on titrated or technical replicates rather than clinical samples and have not sought to integrate the data directly. Cross-platform analysis of microarray data has previously been shown to be possible and worthwhile, although this has normally been performed using transformed relative values [385], rather than by the direct integration of intensity-level data. Comparisons of relative values from two-colour microarray data has been shown to lead to fold change compression [384].

The Affymetrix and Illumina microarray platforms differ fundamentally in their design. Of particular note are the major differences in probe selection, design and physical attachment to the array or chip [383]. Affymetrix GeneChips are fabricated by *in situ* synthesis of 25-mer oligonucleotide probes, several of which span the entire length of a single gene and together make up a 'probeset' each of which provides an intensity reading for the gene (figure 7). Along with each probe, the Affymetrix GeneChip also includes one-base mismatch probes designed to act as controls for non-specific hybridisation [383, 386]. Illumina Beadarrays comprise single longer 50-mer oligonucleotide probes which are attached to micro-beads and spotted onto glass-slide arrays using a random self-assembly mechanism. Unlike the Affymetrix GeneChip, the majority of genes included on the Illumina Beadarray are represented by a single probe;

however, in contrast to the latter, each oligonucleotide is represented in the order of 30 times on each array, providing an internal technical replication, not found on Affymetrix GeneChips [383, 387]. Furthermore Affymetrix GeneChips exist as single microarrays on one chip (for a single sample) which are processed independently, whereas Illumina Beadarrays comprise several (6, 8 or 12) microarrays on one glass-slide and thus hybridisation and scanning can be performed simultaneously [383]. Considering these fundamental differences in the design, it is not clear whether data derived from Affymetrix and Illumina microarrays can or should be compared directly.

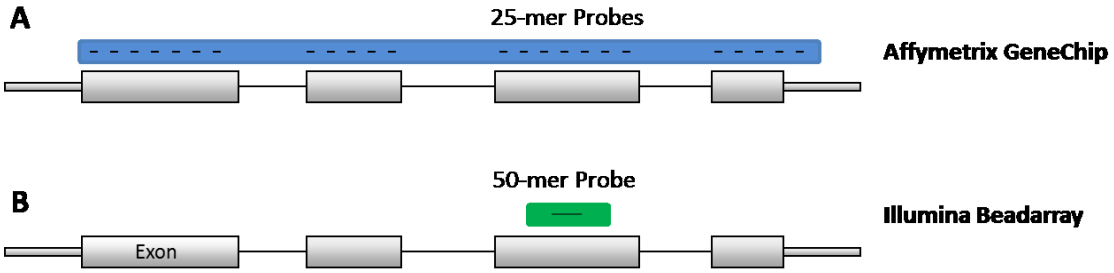


Figure 7. ***Schematic representation of the differences in probe design between Affymetrix GeneChips and Illumina Beadarrays****. A: Affymetrix GeneChips are comprised of multiple 25-mer probes which span the coding regions of each gene. B: Illumina Beadarrays utilise 50-mer probes and the majority of genes are represented by a single probe.*

The findings reported in this study demonstrate that it is possible to directly combine appropriate datasets at the intensity level to improve statistical power. Inter-platform bias can be sufficiently reduced to expose previously obscured biological variation and furthermore, data correction does not amplify meaningless noise in the results. Despite intrinsic differences between these technologies, suitably similar studies can be directly integrated for robust and meaningful meta-analysis, and to that end a validated workflow for the process has been formulated (figure 8).

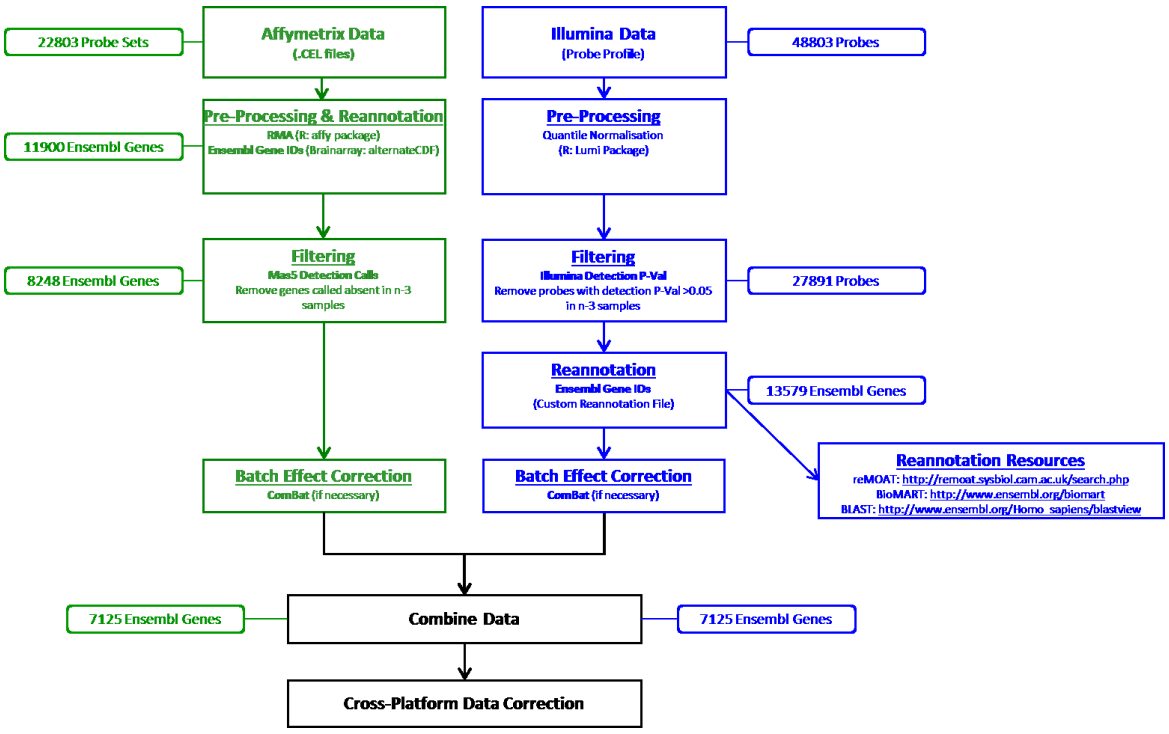


Figure 8. ***Workflow for direct integration of Illumina and Affymetrix intensity level gene expression data****. The same/similar processing steps were used wherever possible. Green: Affymetrix data processing and blue: Illumina data processing. The number of probes/genes remaining after each pre-processing step are shown in rounded rectangles.*

2.3. Materials and Methods

2.3.1. Data Generation

2.3.1.1. Affymetrix Dataset

Affymetrix gene expression data was generated from primary breast tumour core biopsies before, 10-14 days after and approximately 3 months following neoadjuvant Letrozole treatment as part of a previously described clinical study [124, 284]. The research was carried out in compliance with the Helsinki Declaration, with all patients giving informed consent to be included in the study which had been approved by the local ethics committee (LREC; 2001/8/80 and 2001/8/81 and 06/S1103/65 – see supplementary section). RNA was extracted, amplified and labelled as previously

described [284], before hybridisation to HGU-133A GeneChips (Affymetrix) according to the standard protocol.

2.3.1.2. Illumina Dataset (Affymetrix Subset)

RNA from a subset of 18 samples (pre-treatment, 10-14 days and 3 month samples from 6 patients defined as clinical responders to treatment) used in the aforementioned study [124, 284] was then amplified using the WT-Ovation FFPE System Version 2 (NuGEN), purified using the Qiaquick PCR Purification Kit (Qiagen), biotinylated using the IL Encore Biotin Module (NuGEN), purified using minElute Reaction Cleanup Kit (Qiagen) and quantified using a Bioanalyser 2100 with RNA 6000 Nano Kit (Agilent). cRNA was then hybridised to Human HT-12v3 expression Beadarrays (Illumina, Cambridge, United Kingdom) according to the standard protocol for NuGEN amplified samples.

2.3.1.3. New Illumina Dataset

A new Illumina gene expression dataset was also generated from primary breast tumour core biopsies before, 10-14 days after and approximately 3 months following neoadjuvant Letrozole treatment. RNA was extracted using the miRNeasy Mini Kit with RNAse Free DNAse treatment (Qiagen). RNA was then amplified, labelled, purified, quantified and hybridised as described above for the Illumina (Affymetrix Subset).

2.3.2. Published MAQC and Breast Cancer Datasets

Methods for the MAQC Illumina Human-6 Expression BeadChip (v1) and Affymetrix U133 Plus 2.0 array hybridisations are provided in the original study [384]. The data was retrieved from NCBI GEO accession GSE5350. Publicly available primary breast cancer datasets [388, 389] were downloaded datasets from NCBI GEO and ArrayExpress. Breast cancer subtypes were assigned using three signatures from Sørlie *et al.* (2003) [25], Parker *et al.* [27] and Hu *et al*. [390] as described previously [391].

2.3.3. Data Processing and Analysis

All data were processed using the R/Bioconductor software and packages [392]. A custom Chip Definition File (CDF) file [393] was used to map the Affymetrix data to Ensembl gene annotations and RMA implemented by the *affy* package used for normalisation. Illumina probe profiles were quantile normalised using the *lumi* package and mapped to Ensembl gene sequences using a composite list comprising mappings from reMOAT [394], Ensembl BioMart [395] and a custom BLAST sequence search of the online Ensembl gene database where there was agreement between at least two of the resources (details of resources used are provided in figure 8). Where multiple Illumina probes represented an Ensembl gene the mean expression level was calculated. Prior to reannotation the data were filtered using Illumina or Affymetrix probe detection P-values, removing probes that were undetected ($p>0.05$ in the total minus 3 samples). The number of probes/genes remaining after each pre-processing step are shown in figure 8. R-scripts for data pre-processing are given in supplementary section 6.1)

A number of batch-correction and cross-platform normalisation methods were evaluated, including mean-centering [369], ComBat [396], Distance Weighted Discrimination [397] and Cross-platform Normalisation (XPN) [398] (XPN and other batch correction methods can be implemented through ArrayMining Online [399] or via the *inSilicoMerging* R package) in order to determine the most effective method for reducing the bias imposed by the different platforms. Principal component analysis and hierarchical clustering analysis was performed using Cluster [400]. Significance analysis of Microarrays (SAM) [401] pairwise differential gene expression analysis was performed using the *siggenes* package (R/Bioconductor).

A linear additive model was applied to log-scale expression data to estimate the variances in the MAQC dataset. The variation introduced at a given level propagates additively throughout subsequent levels, allowing these variance contributions to be modelled. The total variance for a given gene was assumed to be the aggregate of individual contributions from the inter-sample, -platform, -laboratory, and -replicate variability. These contributions are assumed to be independent and randomly drawn

from log-normal distributions and, as all factors meet in unique combinations, a nested variance model is individually applied to each gene such that the model of the measured expression, X_{ijkl}, of each probe is defined as $X_{ijkl} = \mu + A_i + B_{ij} + C_{ijk} + D_{ijkl} + \epsilon_{ijkl}$. (i=1,…,s; j=1,…,t; k=1,…,u; l=1,…,v) where μ is the geometric-mean expression of the gene from the given sample-type, A_i is the effect attributed to the ith sample, B_{ij} is the random effect of the jth platform, C_{ijk} is the random effect of the kth lab, D_{ijkl} is the random effect of the lth replicate hybridisation, and ϵ_{ij} is the residual measurement error. Finally, s is the total number of samples, t is the number of platforms on which the samples were assessed, u is the number of labs processing the arrays, and v is the number of replicate samples in the corresponding platform processed in each lab. The variance of any given observation, X_{ijkl}, is $\sigma^2_A + \sigma^2_B + \sigma^2_C + \sigma^2_D + \sigma^2$; these components represent the inter-sample, inter-platform, inter-laboratory, and inter-replicate variance respectively. The estimation of σ^2_A, σ^2_B, σ^2_C, σ^2_D, and σ^2 is performed independently for each gene as stated in Snedecor and Cochran (1989) [402]. Models of this kind are formally defined in Neter, Wasserman and Kutner (1985) [403] and Oberg and Mahoney (2007) [404] and have previously been used to optimise gene-expression experimental design [405, 406]. All variance estimates were performed using a REML procedure implemented in the *nlme* package in R [407, 408].

2.4. Results

2.4.1. Direct Cross-platform Integration of MAQC Data

The Microarray Quality Control (MAQC) consortium [384] investigated the reproducibility of microarray-derived gene expression measurements by assessing performance across platforms, chips, and processing sites using a titration of Universal Human Reference RNA (UHRR) and Human Brain Reference RNA (UBRR). The complete MAQC Affymetrix and Illumina datasets were combined by re-annotating probes on each platform in terms of their Ensembl gene targets (discussed previously: see section 2.2.3 and figure 8). As expected, sample A (100% UHRR) replicates from the same platform were found to be more highly correlated with sample C (75% UHRR,

25% HBRR) replicates than the other samples. This was also the case for sample B (100% HBRR) and D (25% UHRR, 75% HBRR) replicates, reflecting their relative biological similarity (figure 9). Without adjustment, correlations between the same samples (A, B, C, or D) processed on different platforms were much lower (R=0.70-0.77) than the same samples processed only on the Illumina Beadarrays (R=0.98-1.00) or Affymetrix GeneChips (R=0.99-1.00) (figures 9 and 10).

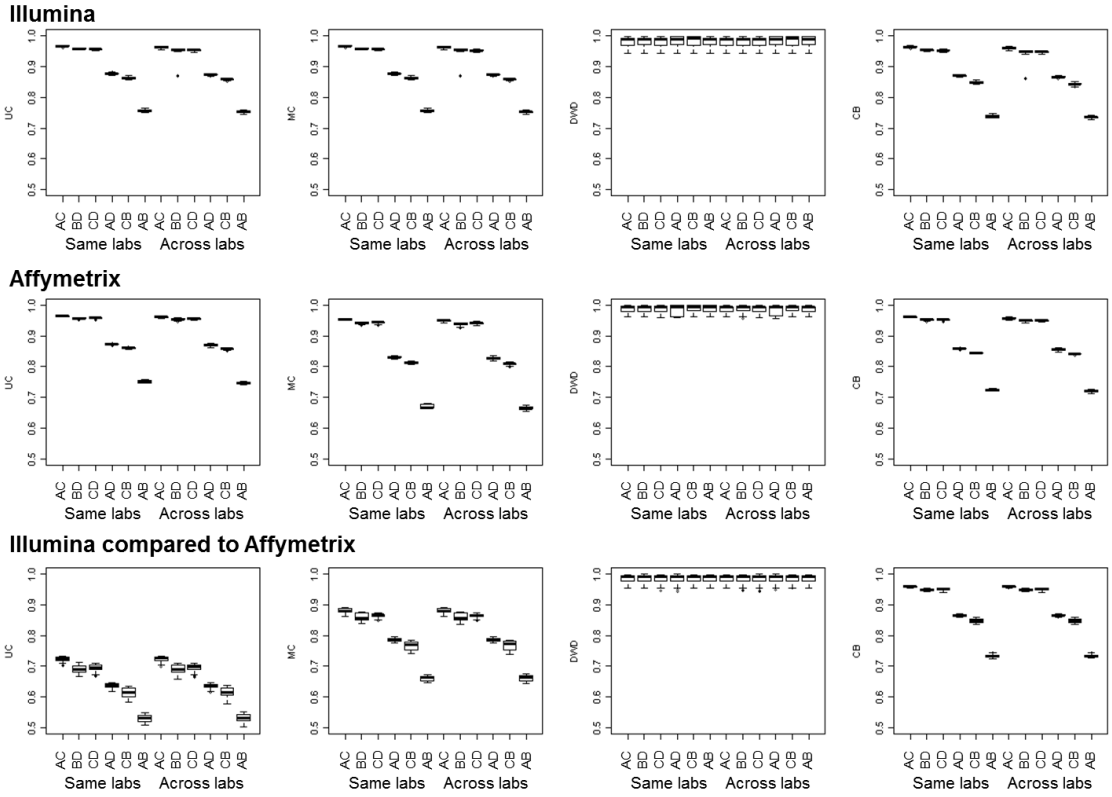


Figure 9. ***Boxplots showing the Pearson correlation coefficients within and between labs before and after correction by three different methods****. X-axis: comparison of samples: A = 100% UHRR, B = 100% HBRR, C = 75% UHRR + 25% HBRR, D = 25% UHRR + 75% HBRR. Y-axis: UC = uncorrected, MC = mean-centering, DWD = Distance Weighted Discrimination, CB = Empirical Bayes method (ComBat).*

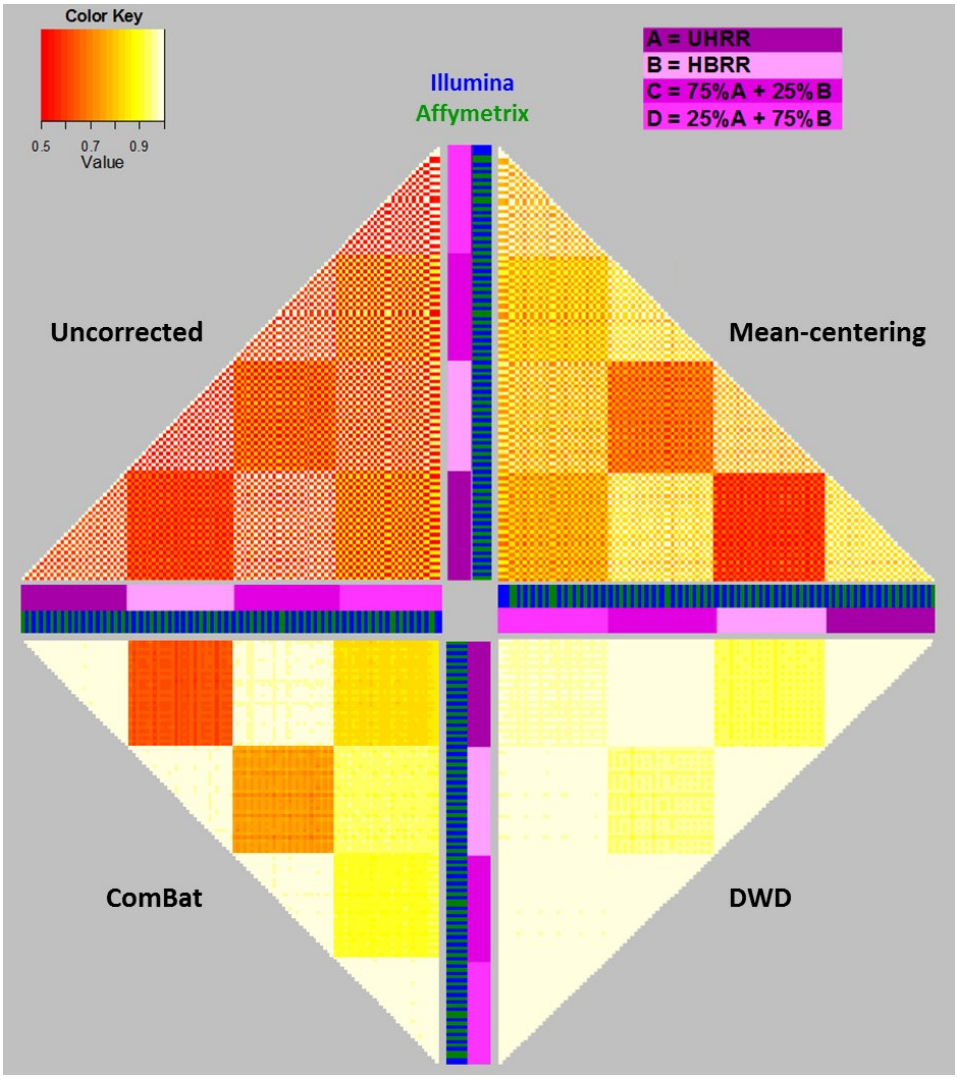


Figure 10. ***Affymetrix and Illumina data from the MAQC project can be directly integrated****. Pairwise Pearson correlation heatmaps demonstrate cross platform bias and the effects of three correction methods: mean-centering, Distance Weighted Discrimination (DWD) and an Empirical Bayes method (ComBat). R values range from low correlation (red) to high correlation (white) through shades of orange and yellow reflecting the overall similarity of expression profiles based on biological and platform-specific variation. The shades of purple to pink indicate the samples (A = 100% UHRR, B = 100% HBRR, C = 75% UHRR + 25% HBRR, D = 25% UHRR + 75% HBRR). Samples are ordered by replicate and lab name rather than by platform. Green bars for Affymetrix samples and blue bars for Illumina samples.*

Adjusting for the platform differences using the mean-centring method [369] provided only a marginal improvement compared to uncorrected data, whilst the Distance Weighted Discrimination (DWD) method [397] suppressed not only the platform-specific bias, but also legitimate biological variability between samples (figures 9 and 10). The greatest improvement was observed following correction by ComBat, a method that exploits variance moderation during data adjustment [396]. Similar correlations were found both across and within platforms, suggesting that whilst removing the platform bias, the ComBat method retains legitimate biological variation between the biologically distinct samples (figures 9 and 10). Another promising method, Cross-platform Normalisation (XPN) [398], could not be evaluated with these data due to the small number of independent biological replicates.

In addition to correlating expression values, variance estimates for each of the 15,781 Ensembl genes probed by the two platforms at the inter-sample, inter-platform, inter-laboratory, and inter-chip levels were calculated using a nested analysis of variance described in section 2.3.3 (figure 11). As expected, and in agreement with the correlation analysis, the difference between the platforms was responsible for the majority of the overall variance in uncorrected (58%), quantile-normalised (47%), and mean-centered (44%) expression data. Inter-platform variance was significantly reduced by both DWD and ComBat, to 15% and 7% of the total, respectively. Consistent with the correlation analysis, the DWD method also substantially reduced inter-sample variance, which is likely to obscure genuine biological differences between the samples (figure 11). Conversely, the ComBat method slightly increased inter-sample variance, potentially uncovering meaningful biological differences between the UHRR/UBRR titrations.

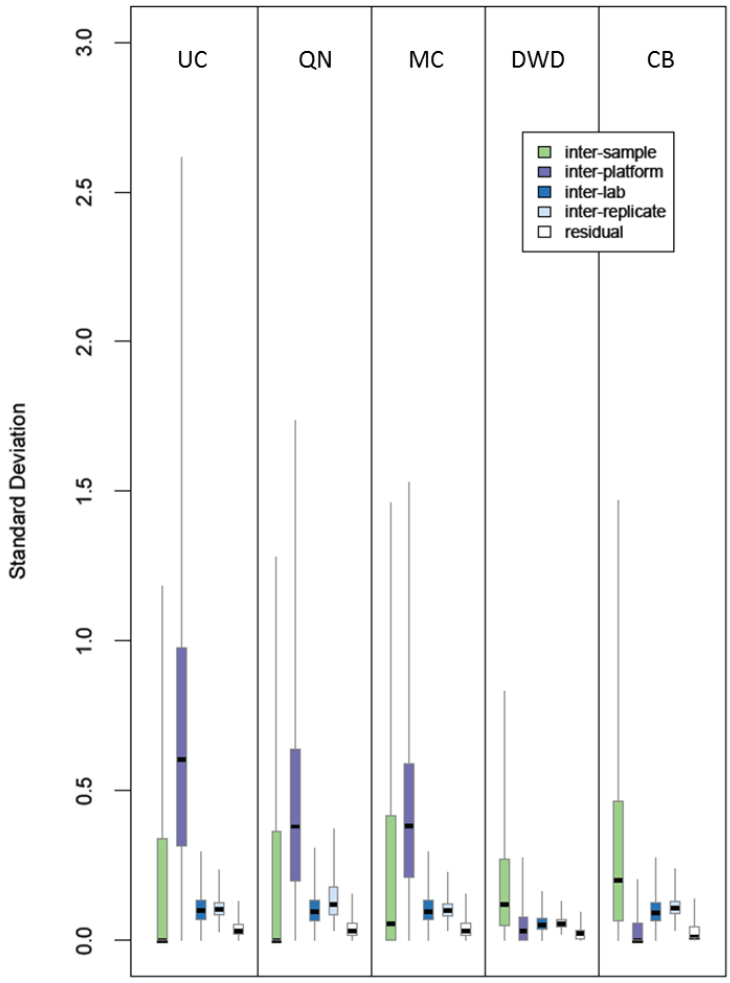


Figure 11. ***Cross-platform correction minimises technical variation whilst maintaining biological variation and differential expression****. UC = uncorrected, MC = mean-centering, DWD = Distance Weighted Discrimination, CB = Empirical Bayes method (ComBat). This figure was produced by Robert Kitchen PhD.*

To examine the effects of cross-platform integration on the identification of genes differentially expressed between UHRR and HBRR, Affymetrix and Illumina data were both analysed separately and as a combined dataset. Differential expression was assessed using the Significance Analysis of Microarrays (SAM) method [401]. Analysis of the 60 combined Affymetrix plus Illumina HBRR and UHRR samples together, resulted in lower false discovery rates and a greater number of statistically significant differentially expressed genes (figure 12) than when the Affymetrix or Illumina (15 'A'

and 15 ‘B’) samples were analysed separately. There were also many more overlapping genes in the combined analysis and either of the platforms following cross-platform correction, when identifying the top 1000 differentially expressed genes (as described previously [369]), again with ComBat performing best (figure 13). The overlap of differentially expressed genes identified by samples processed on either of the two platforms independently (15 ‘A’ and 15 ‘B’ samples) was also much more consistent following ComBat, than DWD or mean-centering correction (figure 14). Taken together, these results indicate that combining data across the two platforms increases specificity and reduces the number of predicted false positives, suggesting improved statistical power.

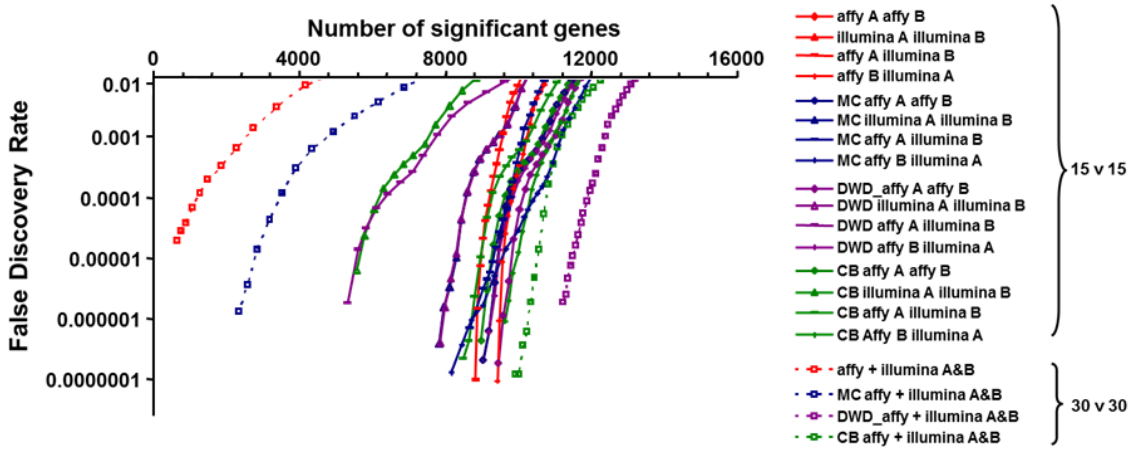


Figure 12. ***Plot showing the relationship between the false discovery rate and the number of genes identified****. Comparing UHRR (A) with HBRR (B) using either 15 Affymetrix or 15 Illumina replicates or both together. MC = mean-centering, DWD = Distance Weighted Discrimination, CB = Empirical Bayes method (ComBat).*

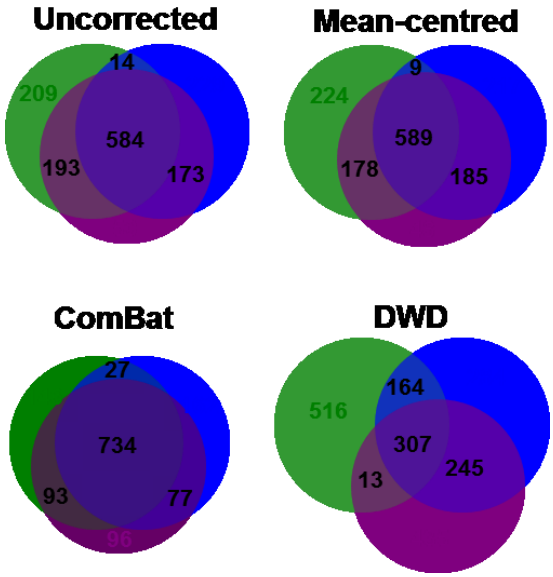


Figure 13. ***Venn diagrams demonstrate the overlap between the 1000 most differentially expressed genes between the MAQC UHRR and HBRR (A and B samples)****. The SAM method was used to identify the most differentially expressed genes with either Affymetrix (green) or Illumina (blue) alone, or Affymetrix and Illumina together (purple).*

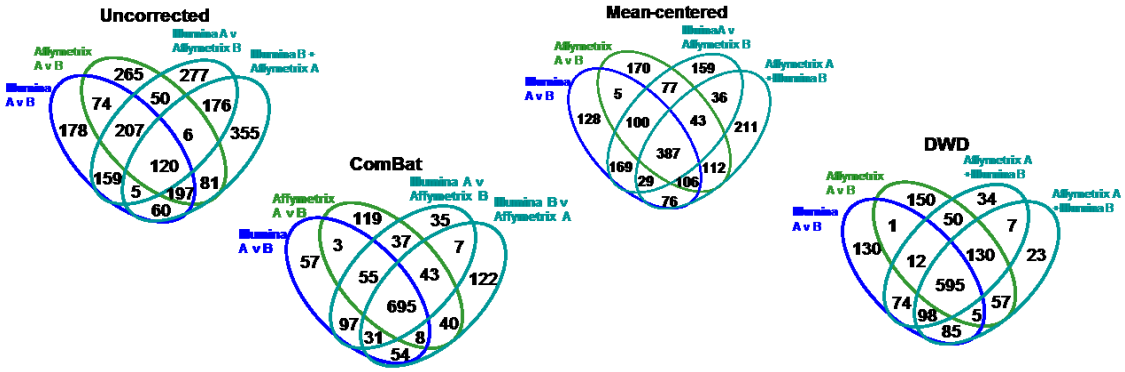


Figure 14. ***Venn diagrams showing the overlaps between the 1000 most significant differentially expressed genes****. The SAM method was used to identify the most differentially expressed genes (each comparison is 15 'A' samples versus 15 'B' samples). Green = Affymetrix comparisons, blue = Illumina comparisons, teal = Affymetrix versus Illumina (inter-platform) comparisons.*

2.4.2. Increasing Statistical Power through Integration of Clinical Datasets

2.4.2.1. Matched Clinical Samples Common to both Platforms

In order to evaluate the feasibility of directly comparing intensity level gene expression of clinical samples processed separately on the two platforms, a new dataset of Illumina Beadarray data was generated from RNA derived from breast tumour samples that were assessed as part of a larger published study using Affymetrix GeneChips [124, 284, 409] (figure 15). These samples comprised matched pre-treatment, 10-14 day, and 3 month primary breast tumours from 6 patients with a clinical response to neoadjuvant Letrozole.

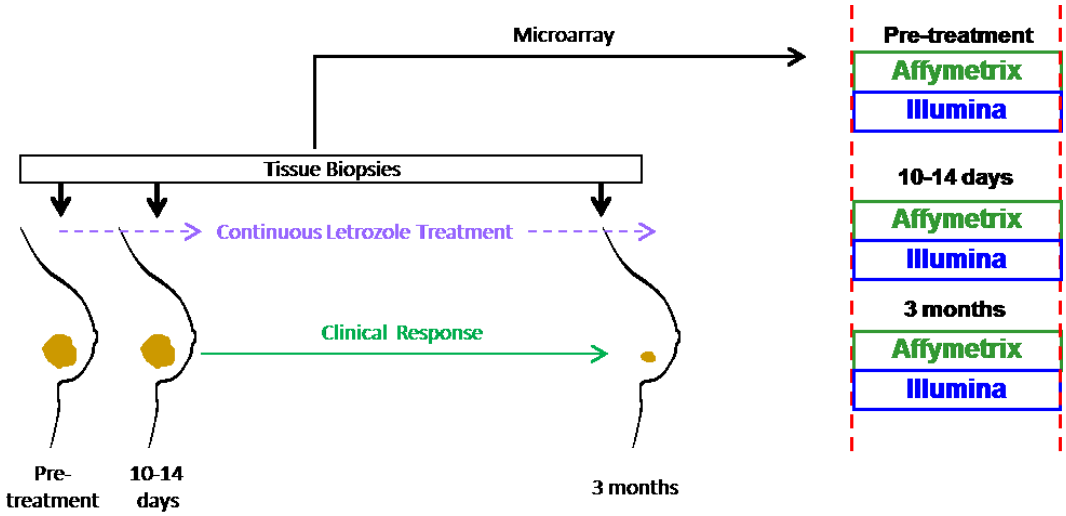


Figure 15. ***Matched clinical samples processed on Affymetrix and Illumina platforms.*** *Core biopsies of primary breast cancer tumours were taken at pre-treatment, 10-14 days on treatment and again after 3 months from each of 6 patients (18 samples) undergoing neoadjuvant letrozole treatment. All samples were processed and hybridised to both Affymetrix GeneChips and Illumina Beadarrays for comparison.*

As with the MAQC data, pairwise Pearson correlations of samples processed on the two platforms were significantly increased following correction with the ComBat method, which again outperformed mean-centering and DWD by maintaining variation between

biologically independent samples (figures 16 & 17). A fourth method, cross platform normalisation (XPN) [398] generated similar results to ComBat, although Pearson correlations for the majority of matched samples across both platforms were marginally higher (figures 17 and 18). In addition, a greater number of pairs of Affymetrix and Illumina samples clustered together with the XPN method than with ComBat (figure 18).

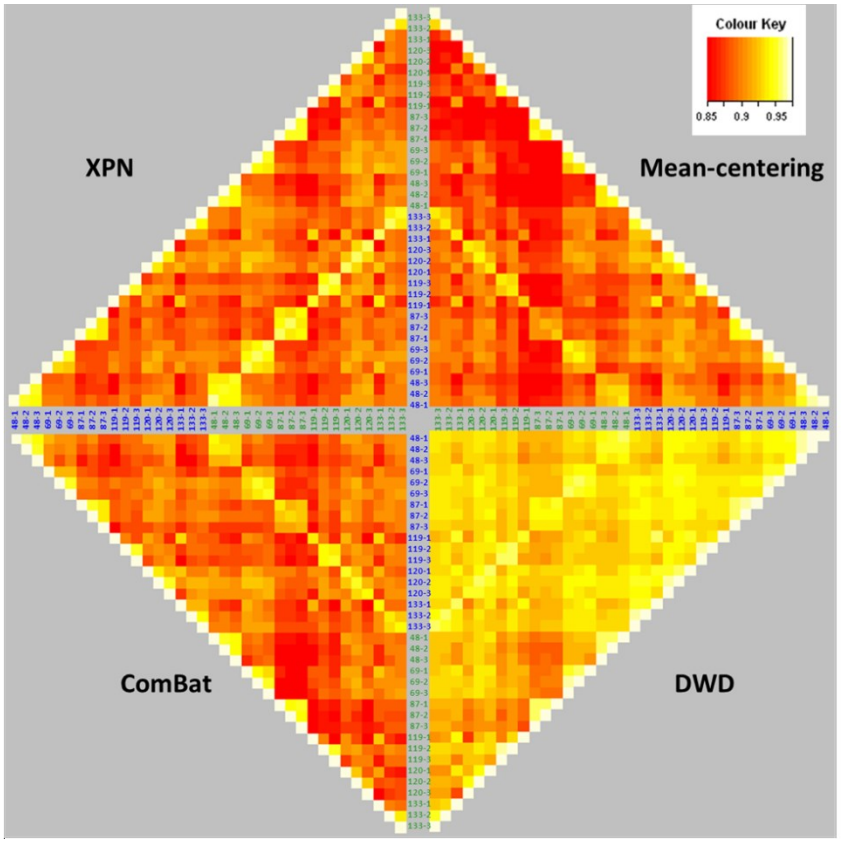


Figure 16. ***Affymetrix and Illumina data from clinical datasets can be directly integrated.*** *Pairwise Pearson correlation heatmaps from 18 matched clinical breast cancer samples from 6 patients demonstrate cross platform bias and the effects of four correction methods: mean-centering, Distance Weighted Discrimination (DWD), an Empirical Bayes method (ComBat) and cross platform normalisation (XPN). R values range from low correlation (red) to high correlation (white) through shades of orange and yellow reflecting the overall similarity of expression profiles based upon biological and platform-specific variation. The inner diamond represents the matched samples from the two platforms. Each patient is numbered with an ID (48, 69, 87, 119, 120 and 133) and each sample is also numbered pre-treatment (-1), 10-14 days (-2) and 3 months (-3) post treatment. Uncorrected data is NOT shown (to show it on the same colour scale as the other plots would not clearly demonstrate the differences between the correction methods).*

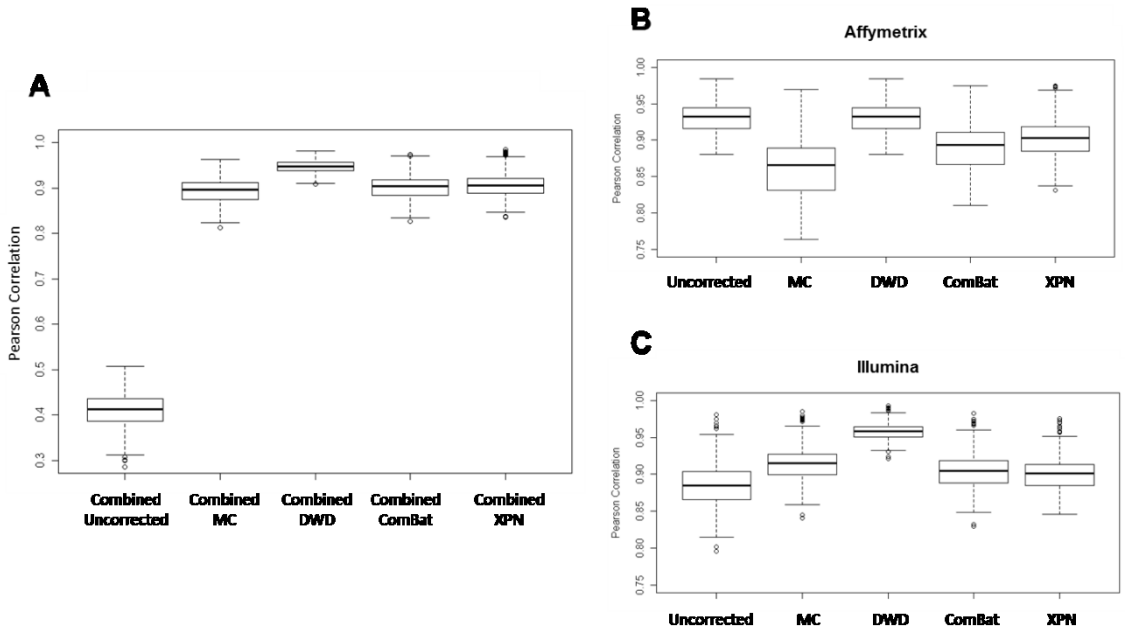


Figure 17. ***Comparison of Pearson correlation coefficients across different correction methods****. A: Boxplots showing the range of Pearson correlation coefficients between 18 matched samples (including pre-treatment, 10-14 days and 3 months from 6 patients) for different correction methods. B: Affymetrix dataset and C: Illumina dataset boxplots showing the range of Pearson correlation coefficients between all possible sample combinations for different correction methods. MC = mean-centering, DWD = Distance Weighted Discrimination, CB = Empirical Bayes method (ComBat), XPN = cross platform normalisation.*

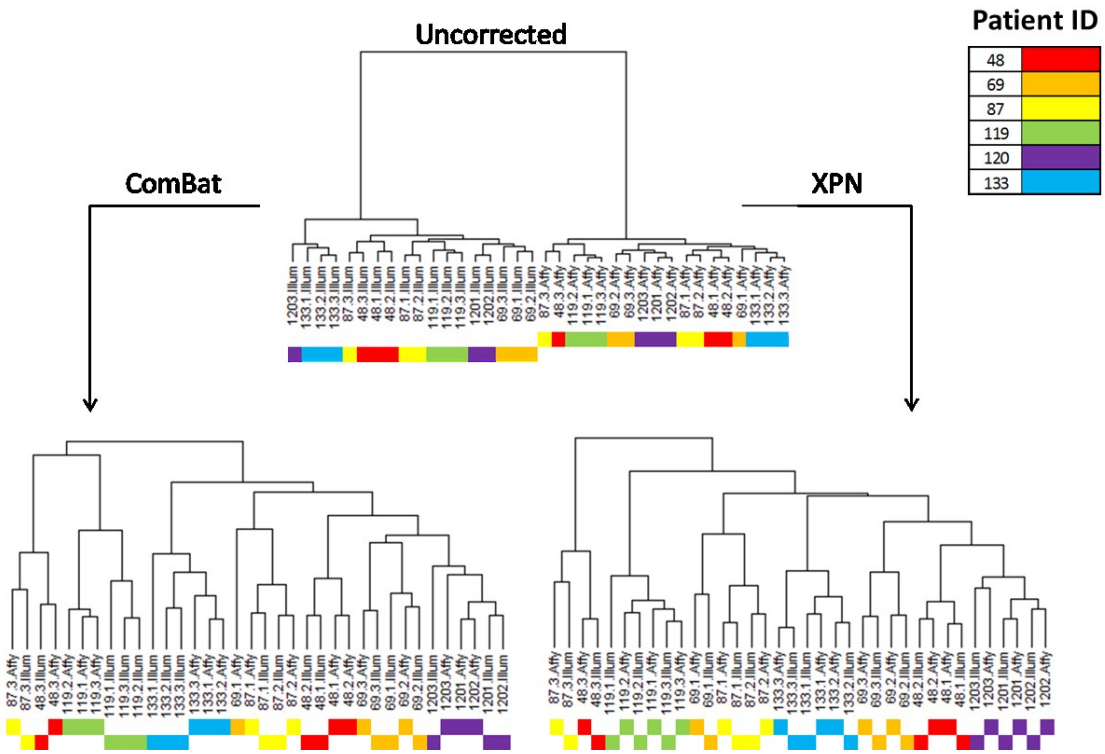


Figure 18. ***Hierarchical clustering of samples based on Pearson correlation after either ComBat or XPN correction****. Colour denotes samples from the same patient, the suffixes on patient ID's (48, 69, 87, 119, 120 and 133) denote as follows: '.1' = pre-treatment, '.2' = 10-14-days, '.3' = 3 months, '.Illum' = Illumina Beadarray data, '.Affy' = Affymetrix GeneChip data.*

2.4.2.2. Integrating Independent Clinical Datasets

The cross-platform dataset was expanded with 48 new Illumina pre-treatment and matched 3 month samples from 24 independent patients to give a total of 60 Illumina samples to compare with 60 pre-treatment and 3 months Affymetrix samples from the original dataset. All patients and tumours had similar characteristics and were shown to clinically respond to 3 months of neoadjuvant Letrozole treatment, with tumour ultrasound measurements showing a stable volume reduction of 70% over the 3 month period. The twelve pre-treatment and 3 month samples common to both microarrays were retained (figure 19).

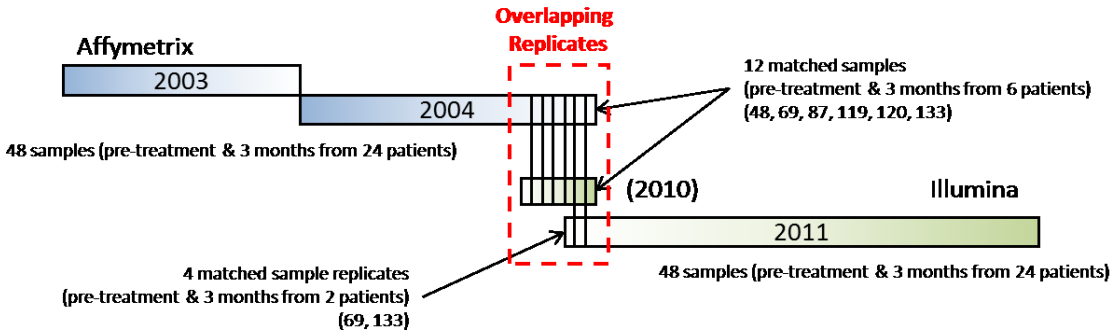


Figure 19. ***Integration of partially overlapping Affymetrix and Illumina datasets****. The Affymetrix dataset (original study) was generated in two separate batches, the first in 2003 and the second in 2004. The 2010 Illumina dataset comprises 12 (discounting 10-14 day samples) of the 18 matched samples (patient IDs: 48, 69, 87, 119, 120 and 133) common to both datasets (discussed previously: see section 2.3.2.1). The new Illumina dataset comprises an additional 48 samples from 24 independent patients and 4 replicates from the overlapping 12 matched samples (pre-treatment and 3 months from patients 69 and 133) common to both platforms (discussed previously: see section 2.3.2.1).*

Independent assessment of the Affymetrix and Illumina datasets in this study identified the need to correct for batch effects within the platforms (prior to integration) due to variation in date of sample processing, consistent with previous studies [369, 370, 376]. Before correction samples (based on expression of all genes after filtering) were found to cluster together by date of processing rather than by biological similarity and differences (based on expression of all genes after filtering). Batch correction was performed on each dataset independently using ComBat as described previously [369, 370, 376, 377, 396]. Following correction samples clustered more heterogeneously with no identifiable batch effects in multi-dimensional scaling (MDS) analysis. In addition the Illumina datasets included replicate matched samples, common to both datasets (2010 and 2011). Before correction identical samples from different processing batches had no identifiable association with each other, the result of masking by the batch effect. After correction, matched samples were found to cluster strongly with each other indicating that the batch effect bias had been reduced while maintaining true biological variation (figure 20).

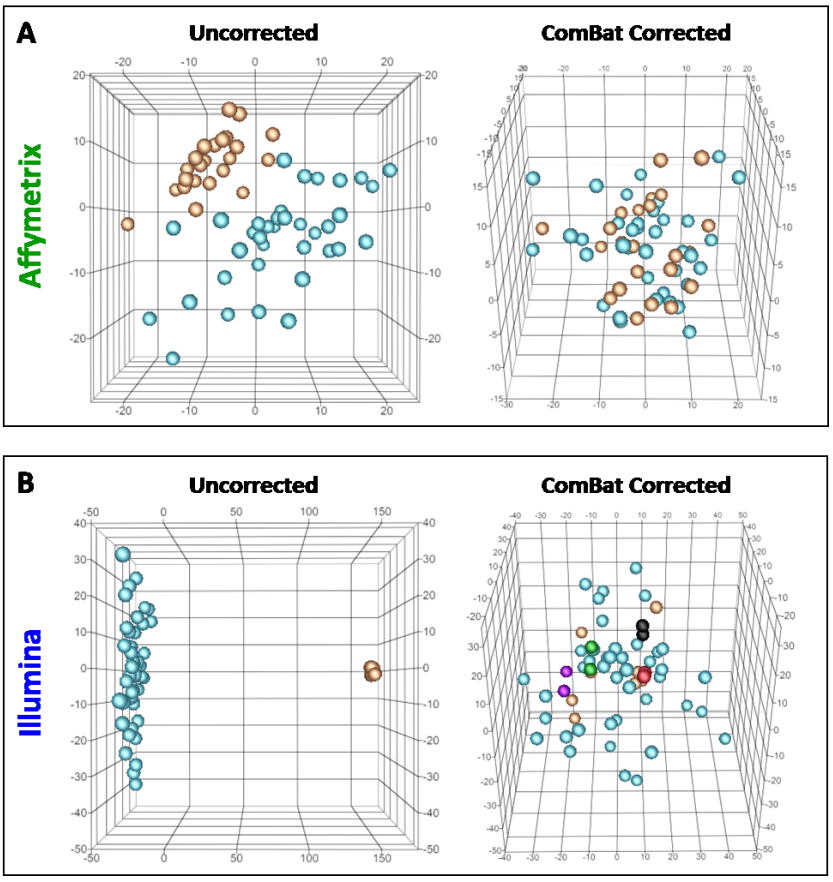


Figure 20. ***Multidimensional scaling to identify batch effects within datasets****. A: Affymetrix and B: Illumina 3D multi-dimensional scaling (MDS) plots of all samples based on all genes (after filtering; discussed previously: see section 2.2.3) before (left) and after (right) ComBat batch correction. All plots are scaled on 3 dimensions, plots after correction are tilted to better display the association of paired samples. Before correction samples on both platforms separate into clusters based on the date they were processed. Following correction artificial bias has been reduced and the clustering of samples is more heterogeneous. After correction of the Illumina datasets, pairs of matched replicate samples common to both datasets cluster with each other. Affymetrix (A): blue circles = 2003 samples, orange circles = 2004 samples. Illumina (B): blue circles = 2010 samples, orange circles = 2011 samples. Red/black/purple/green circles = represent pairs of identical matched samples common to both 2010 and 2011 Illumina datasets.*

Following intra-platform batch correction, Affymetrix and Illumina datasets were integrated, retaining only Ensembl gene ID's present in both datasets (discussed previously: see section 2.2.3). Without cross-platform correction, plotting the fold

changes between pre-treatment and 3 month samples across the two platforms results in reasonable concordance (R=0.68). However, following XPN correction there is a dramatic improvement in the correlation of fold changes (R=0.99) demonstrating that XPN has greatly reduced the variation between both platforms while maintaining a sufficient range of highly-concordant fold changes to account for biological variability. Importantly, the analysis of fold changes showed that XPN correction did not affect the magnitude of fold changes (figure 21).

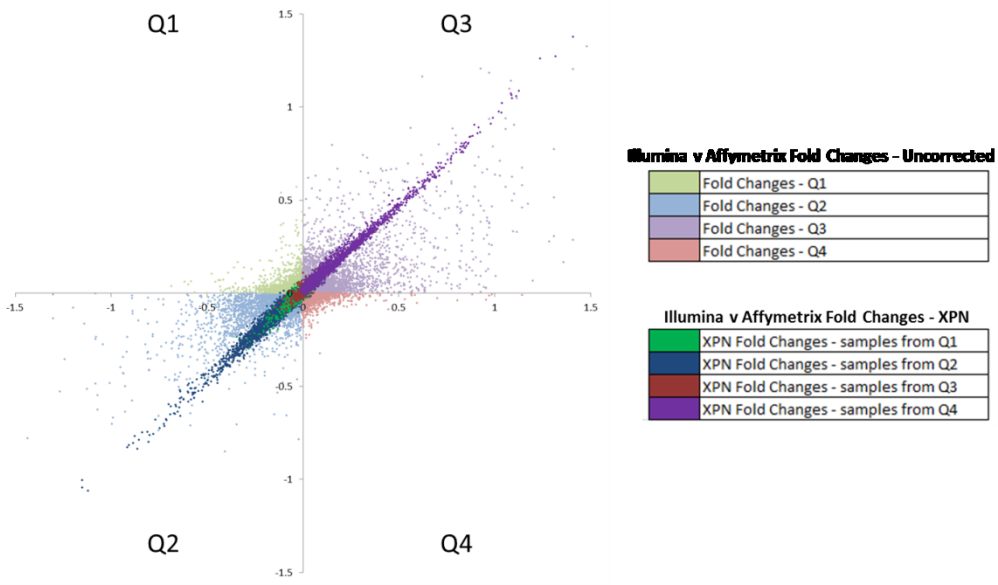


Figure 21. ***Comparison of Affymetrix and Illumina fold changes between pre-treatment and 3 month samples before and after XPN correction****. Scatterplot demonstrating the pre-treatment to 3 month fold changes between the Affymetrix and Illumina datasets before (light colours) and after XPN correction (dark colours). Light colours were applied to data points reflecting their position within the quadrants of the plot (Q1-Q4). Dark colours were applied to XPN corrected data points based on the same data points original quadrant location prior to correction: ie after correction (dark purple) the data points in quadrant 3 (Q3) were located in the same quadrant as before correction (light purple).*

Multidimensional scaling (MDS) demonstrated that the samples common to the Affymetrix and Illumina datasets cluster together and that intra- and inter-platform batch effects have been minimised (figure 22). Prior to XPN correction, samples from the Affymetrix and Illumina datasets form independent clusters; however, after correction

pre-treatment samples from the same patient cluster closely together as do the 3 month samples from the same patient. XPN correction significantly reduces the bias between samples from different platforms, but the pre-treatment and 3 month samples from the same patients still cluster independently, indicating that the true biological differences (due to treatment) are maintained.

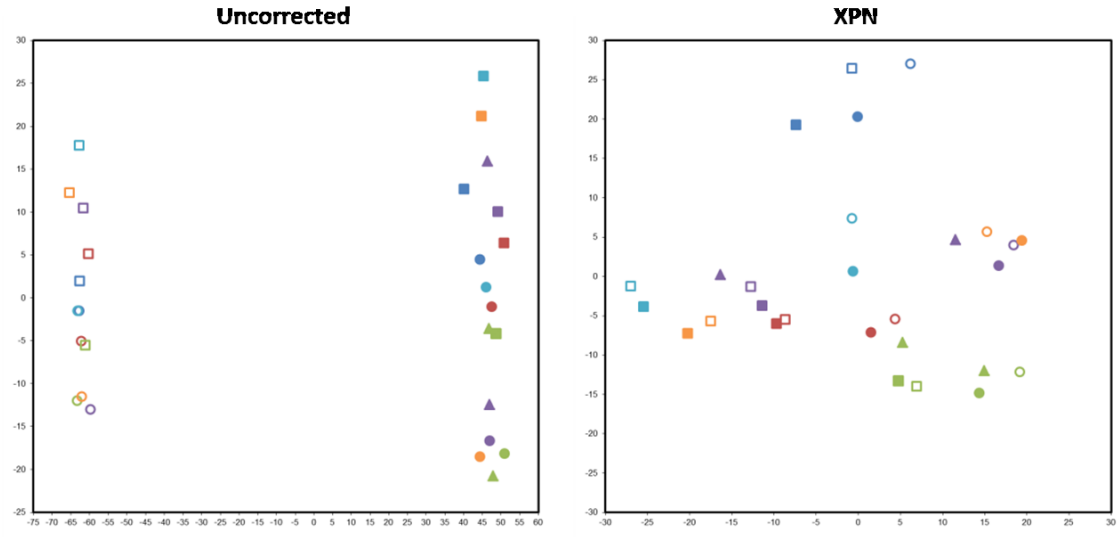


Figure 22. ***Comparison of cross-platform matched samples by multidimensional scaling before and after XPN correction****. Multidimensional scaling (MDS) plots before (uncorrected) and after XPN correction demonstrating the relationship between overlapping samples. Circles = pre-treatment, squares = 3 months post treatment with Letrozole, open symbols = Affymetrix, filled symbols = Illumina, triangles = matched Illumina replicate samples, different colours represent different patients.*

The standard deviation across genes for all pre-treatment or 3 month samples was higher in Affymetrix than Illumina, but was dramatically increased after combining the data. Correction with either ComBat or XPN reduced variation to a level similar to that seen in either dataset independently; further suggesting that gene-wise cross-platform bias is reduced, while true biological variation is maintained (figure 23). When all samples of the combined XPN-corrected dataset were clustered by a published list of genes identified as most changed in response to neoadjuvant endocrine therapy [124, 409] the pre-treatment and 3 month samples clustered together regardless of platform (figure 24).

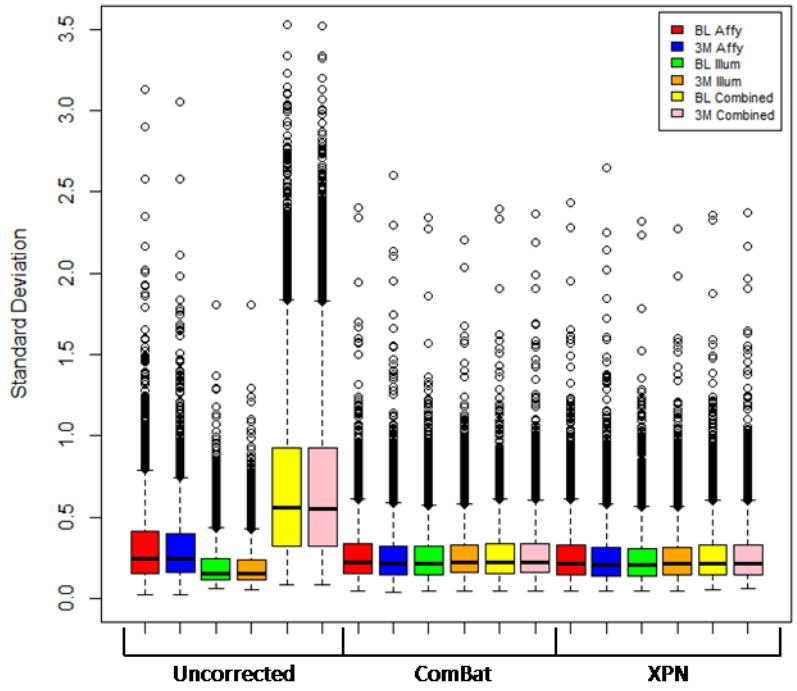


Figure 23. ***Analysis of inter- and intra-platform variance before and after ComBat and XPN correction****. Boxplots of standard deviation for each gene across all samples from the same subgroup (pre-treatment and 3 months) for Affymetrix and Illumina datasets independently and when combined both before and after correction with either ComBat or XPN.*

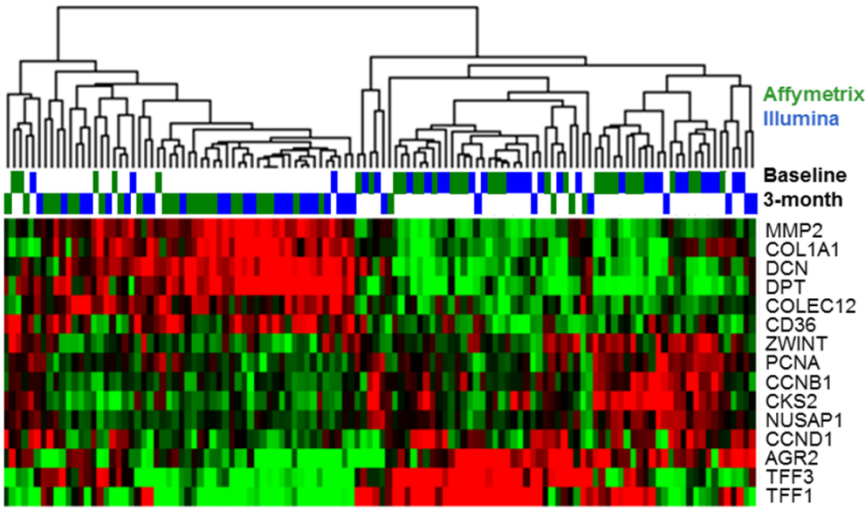


Figure 24. ***Hierarchical clustering and heatmap based on published list of genes identified as most changed between pre-treatment and 3 month samples in patients treated with neoadjuvant Letrozole****. Colour bar indicates the platform the sample was processed on with Affymetrix in green and Illumina in blue.*

Increasing sample number by integration of the Affymetrix and Illumina datasets resulted in the identification of a greater number of significantly differentially expressed genes using pairwise SAM (i.e. the consistency of pre-treatment and 3 month samples from the same patient) at a given false discovery rate (figure 25).

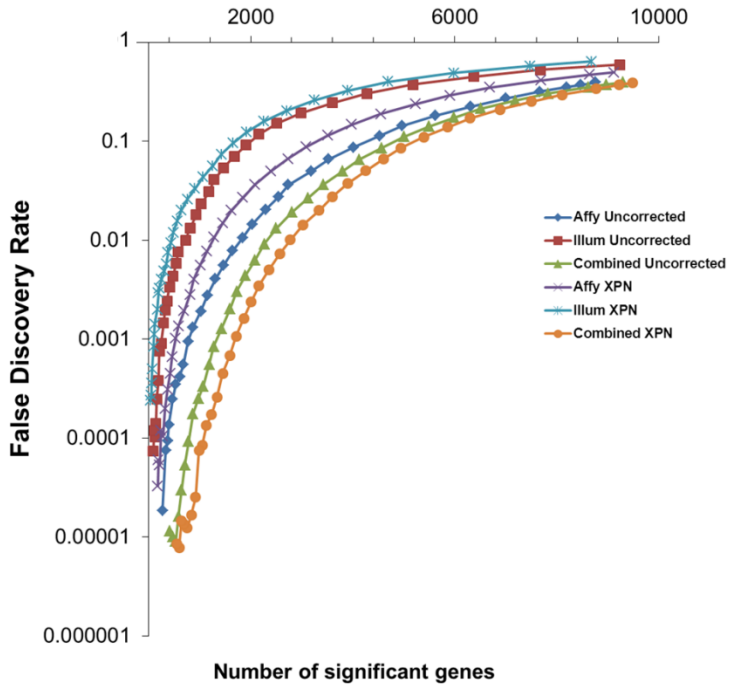


Figure 25. ***Effect of cross-platform integration and correction on differential gene expression analysis****. Plot shows the relationship between the estimated false discovery rate relative to the number of significant differentially expressed genes identified using SAM analysis of Affymetrix and Illumina datasets independently and when combined both before and after XPN correction.*

Interestingly, correction of the combined data by XPN showed only minor improvement compared with uncorrected data in a pairwise SAM analysis with a 93.8% overlap of genes (figure 26A). However, when a non-pairwise SAM method was used (i.e. two unmatched groups: (i) all pre-treatment samples and (ii) all 3 month samples), XPN correction of the integrated data was essential (figures 26B and C). There was an impressive 90% overlap of common differentially expressed genes following XPN correction when comparing the pre-treatment samples from one platform with the 3 month samples from the other. By contrast, the overlap between pre-treatment and 3 month groups in each dataset (Affymetrix or Illumina) independently was only 42.4% (figures 26A and C). Finally, comparing the uncorrected Affymetrix pre-treatment

versus Illumina 3 month samples (and vice versa) with the XPN-corrected equivalent resulted in a very poor overlap (12.1%), indicating the importance of XPN correction for robust differential gene expression of cross-platform integrated datasets.

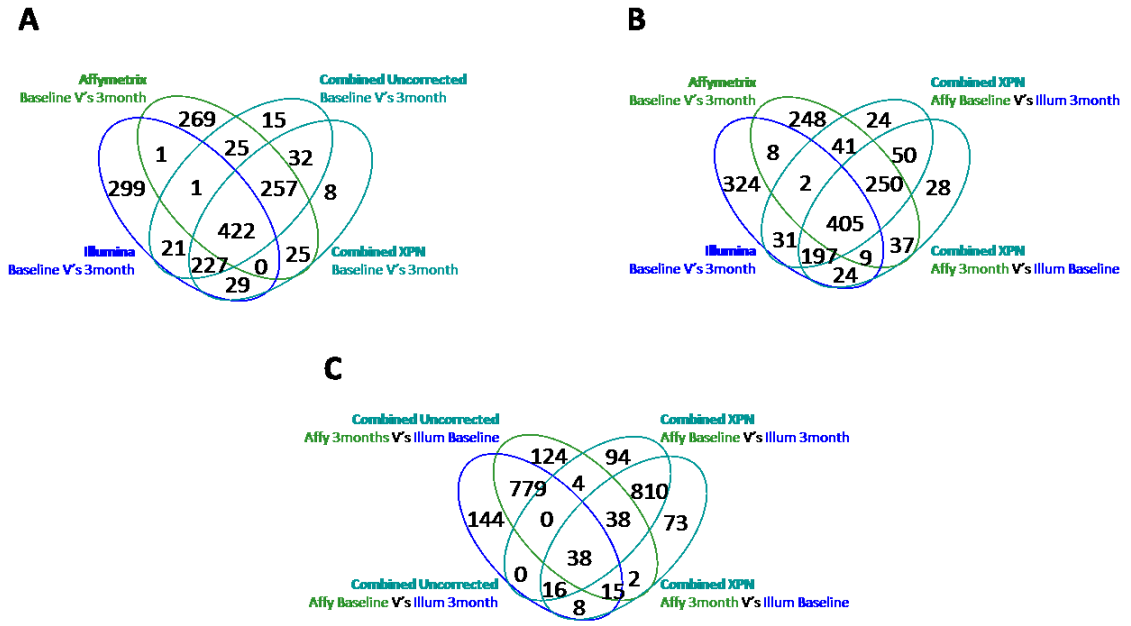


Figure 26. ***Venn diagrams showing the overlaps between the 1000 most significant differentially expressed genes****. A: pairwise SAM analysis and B&C: non–pairwise SAM analysis with Affymetrix (green), Illumina (blue) and combined (teal).*

2.4.3. Published Affymetrix and Illumina Datasets can be Successfully Integrated

Two publicly available non-subtype specific primary breast cancer datasets of comparable size and composition (Naderi *et al.* [389] n=153 on Illumina HumanWG6 v1 and Desmedt *et al.* [388] n=198 on Affymetrix HGU133A) were assigned to molecular subtypes using centroids from the intrinsic gene signatures of Sørlie *et al.* (2003) [25], Parker *et al.* [27], and Hu *et al.* [390]. This was performed on each dataset independently and then both datasets were combined, both before and after XPN correction. Clustering the integrated data before correction resulted in two distinct clusters representing the two datasets, highlighting the platform-specific systematic bias

(figure 27). Following XPN correction the integrated data clustered based on true biological differences, with two clear clusters representing the basal/ERBB2 intrinsic subtype and the luminal subtype for each of the intrinsic centroids (figure 27). Assignment of molecular subtype was highly consistent (Sørlie: 96.6%, Hu: 94.9% and Parker: 96.6%) between uncorrected and XPN-corrected datasets, further suggesting that the XPN correction method does not adversely affect the biological interpretation of the data.

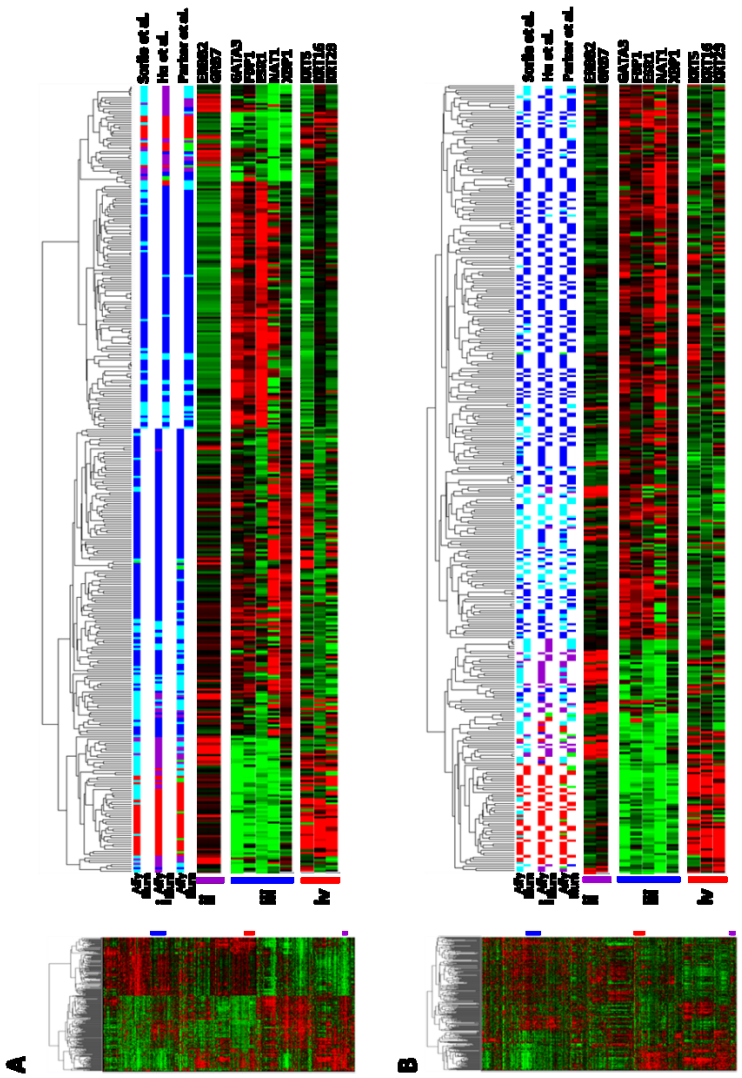


Figure 27. ***Comparison of primary breast tumour gene expression profiles generated on Affymetrix and Illumina platforms****. The Naderi et al. [389] study used Illumina WG6v1 BeadChips, whilst the Desmedt et al. dataset [388] was generated with Affymetrix HG-U133A arrays. A: Before cross-platform correction. B: After XPN correction. Hierarchical clustering of tumours is based upon the 500 most variable genes (thumbnails show all genes). i) Subtypes were assigned by three methods Sorlie et al. (2003) [25], Parker et al. [27] and Hu et al. [390]. Red = basal, purple = ERBB2, blue = luminal A, light blue = luminal B, green = normal-like. Clusters of genes associated with the subtypes are highlighted as follows; ii) ERBB2 gene cluster, iii) luminal gene cluster, iv) basal gene cluster.*

Once again, increasing sample number through integrating datasets results in a greater number of significantly differentially expressed genes, between the Sørlie *et al.* basal and luminal A or the more subtle comparison of luminal A and luminal B subtype samples, at a given FDR (figure 28). Uncorrected integrated data performs poorly in comparison to the integrated data after XPN correction or indeed to either dataset independently.

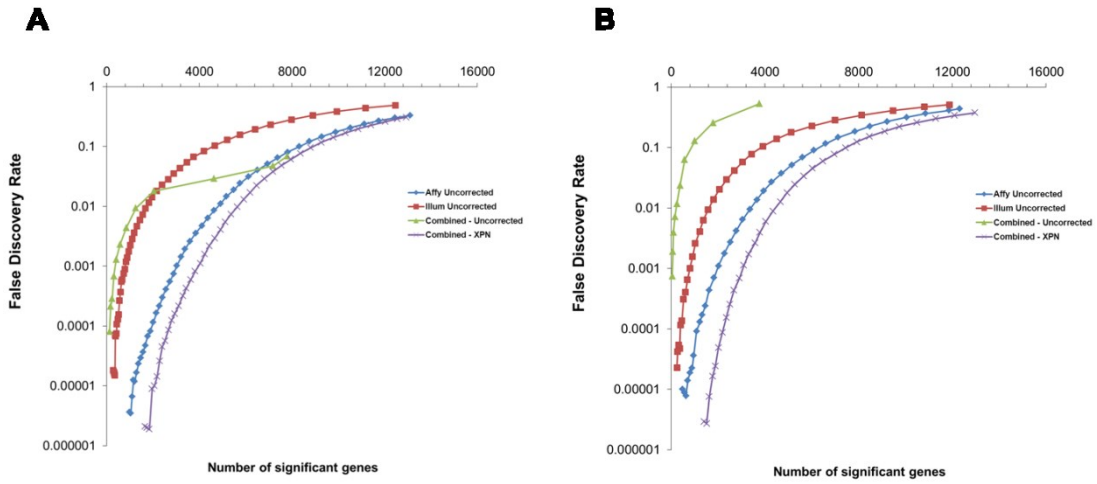


Figure 28. ***Plots showing the relationship between false discovery rate against the number of significant differentially expressed genes identified across a range values of delta using SAM analysis****. Analysis was carried out in Affymetrix (Desmedt) and Illumina (Naderi) datasets independently and when combined both before and after XPN correction to identify genes differentially expressed between the basal and luminal A (A) or luminal A and luminal B subtypes (B).*

2.5. Discussion

The biggest obstacles to the direct comparison of data obtained from different microarray platforms are differences in the sequence and the number of probes that target each transcript. Many studies simply use the most highly or variably expressed probe to represent a gene, despite evidence that some probes hybridise to multiple genes and others have out-dated or incorrect annotation [393, 410-413]. Limiting integration of data to only those genes where the probe sequences are identical, or comparing

measurements simply based upon the official gene symbol would severely restrict our ability to evaluate whether data from different platforms can be directly integrated. For this reason, probes were re-annotated in this study using alternative CDFs [393] for Affymetrix and a validated composite look-up list for Illumina [394] (discussed previously: section 2.2.3).

The microarray quality control (MAQC) project declared that expression values generated on different platforms cannot be directly compared because unique labelling methods and probe sequences will result in variable signals for probes that hybridize to the same target [384]. However, in the interests of making the best use of published data on valuable clinical material, this study was designed to assess whether it would be reasonable to integrate Affymetrix and Illumina data in the interests of improving statistical power and unearthing true biological findings. It has previously been shown that robust classifiers developed using data generated from one platform can accurately predict the phenotype of samples assessed on a different platform [414]. This study demonstrates that it is possible to combine Affymetrix and Illumina gene expression data for meaningful integrative reanalysis. Previous work has shown that for either platform alone (intra-platform), integration of microarray data should only be performed with appropriately similar datasets [369, 370, 376], although exactly where the similarity threshold lies is an important consideration that is still to be determined.

This study revealed that the Distance Weighted Discrimination (DWD) method [397], (which has been used for cross-platform normalisation and cited by more than 50 published studies), may be unreliable in terms of its ability to remove technical noise and preserve biological variability. Perhaps this method is best suited to transformed data such as that generated by two-colour cDNA studies. Relatively strict filter-thresholds were used in the current analyses, including conservative detection p-values to limit the analysis to clearly expressed genes as a previous meta-analysis approach found low or intermediate expressing genes to have poorer inter-platform reproducibility than highly expressed genes [381]. Another recently published comparison of cross-platform normalization methods also found XPN to have the highest inter-platform

concordance [415]. Like this study, this focused on direct adjustment approaches, where the major batch effect (platform used) is clearly identifiable rather than surrogate variable analysis (SVA) approaches [416, 417], which look at latent or unknown variables, such as when samples are processed on different days, in different groups or by different people. Direct integration approaches are only appropriate for small numbers of highly similar datasets specifically selected to answer clearly defined questions, as opposed to recent global survey-based approaches used to identify common tissues or expression profiles across all available datasets [418-420]. Whilst integrating data across platforms increases the number of samples, it also has an impact on the number of genes represented. Genes may be 'lost' at the reannotation stage if not present on both arrays. Therefore integration is a trade-off between increased sample numbers and decreased gene number. Sample numbers are perhaps the biggest factor in the reliability of microarray studies. Ein-Dor *et al.* suggested that thousands of samples are needed to generate a robust gene list for predicting outcome in cancer [378]. The overlap of differentially expressed genes between single and integrated Affymetrix and Illumina datasets was found to be high, although it should be remembered that it has previously been demonstrated that greater biological reliability is seen between studies at the pathway, rather than individual gene level [313].

2.6. Conclusion

This study sought to evaluate whether it is reasonable to combine appropriate Affymetrix and Illumina datasets for reanalysis. Despite fundamental differences in the technology, data from these platforms can legitimately be combined at the normalised and corrected intensity, rather than the fold change level for robust reanalysis, with improved statistical power than the original datasets alone. A validated workflow for integration of Affymetrix and Illumina gene expression data has been presented and cross-platform normalisation (XPN) has been evaluated as the most appropriate correction method to date for the purpose of reducing inter-platform bias allowing for robust gene expression analysis.

3. Understanding and Predicting Response to Endocrine Therapy Early in Breast Cancer Treatment

3.1. Acknowledgments and Contributions

I would like to acknowledge the contribution to this chapter made by Dr Jeremy Thomas (JT) (Consultant Pathologist) who performed all pathological assessments (section 3.4.2.4) and Professor J Michael Dixon (JMD) (Consultant Breast Surgeon) who performed all 3D ultrasound sonography. I would like to acknowledge the contribution made by Anita Dunbier PhD, who carried out the independent validation analysis (section 3.5.4.6).

3.2. Introduction

Third-generation aromatase inhibitors such as letrozole and anastrozole have an established role in the treatment of estrogen receptor alpha positive (ER^+) postmenopausal breast cancer [118, 121-123, 421]. The endocrine effect of these drugs involves specific inhibition of the aromatase enzyme thereby reducing endogenously synthesised estrogen [55, 126, 422, 423]. In hormone sensitive breast cancers estrogen deprivation can lead to cell death, reduced proliferation and reduced tumour progression [424]. However, not all patients respond, with response rates only 50-70% in the neoadjuvant setting and lower in advanced disease [124, 304, 425]. Therefore, there is a clinical need to identify early-on-treatment biomarkers which predict for response/non-response to endocrine therapy and which outperforming currently used parameters for response assessment, in a move towards improved and stratified treatments and ultimately better patient care.

This study involved the generation of a new whole genome gene expression microarray dataset profiling 34 patients with primary breast cancer treated neoadjuvantly with letrozole. Tumour specimens were collected at pre-treatment, 14 days and again at 3 months for each patient (figure 29). The new microarray data was integrated with a pre-

existing published dataset [124] of similar design in order to increase sample numbers and improve statistical power (see section 2) and resulted in generating the largest microarray study of its type comprising a unique series of 89 ER^+ letrozole treated breast cancer patients with dynamic gene expression profiles. Clinical response was determined for each patient using repeated 3D ultrasound measurements taken throughout the treatment window (figure 29).

This work has led to the development of a classifier of response to endocrine therapy in the neoadjuvant setting. The optimal model was based on the expression of 4 genes which could classify response in the letrozole dataset with 96% accuracy. The model was validated in the anastrozole-only arm of an independent clinical trial and found to predict response with 91% accuracy, outperforming previously reported predictive models and currently used clinical and pathological determinants of response.

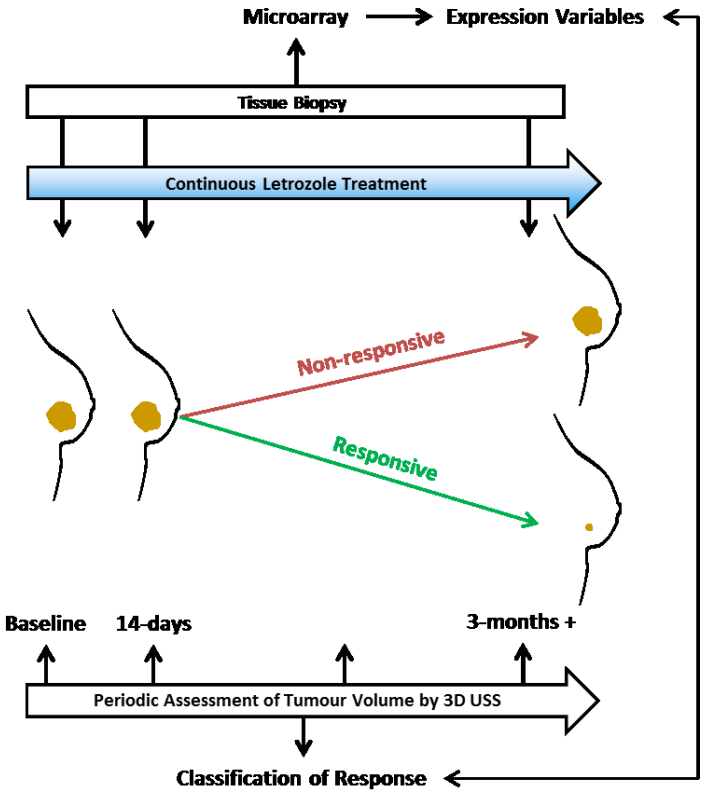


Figure 29. ***Experimental design****. Post-menopausal women with ER$^+$ primary breast tumours treated with continuous neoadjuvant letrozole over approximately 3 months were recruited to the study. Tissue biopsies were taken from the primary tumour at pre-treatment, 14 days and again at approximately 3 months (usually a surgical specimen). RNA was extracted from tissue specimens and used for gene expression microarray analysis. Dynamic clinical response was determined by 3D ultrasound sonography performed by a single operator (JMD) periodically throughout the treatment window.*

3.3. Materials and Methods

3.3.1. Patients

The study population comprised postmenopausal women presenting to the Edinburgh Breast Unit (Western General Hospital) with large primary histologically confirmed invasive breast cancer, immunohistochemically determined to be ER$^+$. A consecutive series of 235 patients were recruited between 2004 and 2011 as part of an on-going letrozole clinical audit. Patients were excluded based on strict predefined criteria: (i) if

the tumour was multifocal or of a certain histological type (mucinous, tubular or lobular), (ii) if the pathologist determined low tumour cellularity or less than 40% malignancy in the pre-treatment biopsy, (iii) if extraction failed to yield sufficient RNA to be suitable for further study, (iv) if follow-up records for clinical assessment of response were unavailable or incomplete, or (iv) if the drug was changed from letrozole to another agent during the treatment window due to an adverse drug reaction. Of the 70 patients who met these criteria, 36 had 3 or more tumour biopsies available for each (comprising pre-treatment, 14 day and approximately 3 month biopsies) and were selected for microarray analysis. All patients gave informed consent to be included in the study which had been approved by the local ethics committee (LREC; 2001/8/80, 2001/8/81 and 06/S1103/65 – see supplementary section 6.4 and 6.5). Patients were treated with a neoadjuvant protocol in which letrozole (Femara, 2.5mg; Novartis Pharma AG, Basel, Switzerland) was given daily for approximately 3 months. Microarray analysis was performed on the 36 patients and analysable results were obtained from 34 patients. Failures were either due to inadequate amplification of total RNA or if the number of probes detected with expression levels above the background (defined by negative control probes) with a P-value less than 0.05 was less than 9000.

3.3.2. Response Assessment

Clinical response was determined using dynamic changes in tumour volumes assessed by repeated measurements taken over the treatment period. Measurements were usually taken around day 45 and again at 3 months and several patients had additional measurements taken throughout the 3 month treatment period. Assessment was based on ultrasound measurements performed by a single sonographer (JMD) (figure 29). Response classifications were determined as described in section 3.4.2.

Pathological response was determined by a pathologist from haematoxylin and eosin (H&E) stained formalin fixed paraffin embedded (FFPE) tissue sections (see section 3.4.2.4).

3.3.3. Tumour Samples

Tumour biopsies were taken with a 14-guage needle: before, approximately 14 days after, and approximately 3 months after commencement of continuous letrozole treatment as described previously [426]. Samples were snap-frozen in liquid nitrogen and frozen sections were taken, H&E stained and the cellularity and percentage presence of cancerous tissue within each specimen was assessed by a pathologist.

3.3.4. RNA Processing and Microarray Hybridisation

Biopsies were homogenised and RNA was extracted using the miRNeasy Mini Kit with RNAse Free DNAse treatment (Qiagen). RNA quantity and quality was verified on a Bioanalyser 2100 with RNA 6000 Nano Kit (Agilent) and Nanodrop 2000c (Thermo Scientific). RNA was reverse transcribed and amplified using the WT-Ovation FFPE System Version 2 (NuGEN), purified using the Qiaquick PCR Purification Kit (Qiagen), biotinylated using the IL Encore Biotin Module (NuGEN), purified using minElute Reaction Cleanup Kit (Qiagen) and quantified once again using the Nanodrop 2000c (Thermo Scientific). Labelled cDNA was then hybridised to Human HT-12v4 whole-genome expression Beadarrays (Illumina) according to the standard protocol for NuGEN amplified samples. Samples in the same response group were assigned random positions across all the Beadarrays with samples from the same patient on the same array. Beadarrays were hybridised and processed in 5 batches and each batch included a replicate Universal Human Reference RNA (UHRR) control sample in order to assess the need for batch effect correction [369, 370, 376]. The Human HT-12v4 whole-genome expression Beadarray covers more than 48000 transcript probes and its annotation is publically available. Data were extracted using the GenomeStudio software (Illumina).

3.3.5. Published Datasets

3.3.5.1. Affymetrix Dataset: Letrozole

Affymetrix gene expression data was generated from primary breast tumour tissue biopsies taken before, approximately 14 days after and again approximately 3 months after commencement of continuous neoadjuvant letrozole treatment in 58 patients (data was available online from 55 patients) as part of a previously described clinical study [124, 284]. Patients were selected from a consecutive series recruited between 2001 and 2003 as part of the aforementioned letrozole clinical audit (see section 3.3.1). RNA was extracted, amplified and labelled as previously described [124] before hybridisation to HGU-133A GeneChips (Affymetrix) according to the standard protocol.

3.3.5.2. Illumina Dataset: Anastrozole

Illumina gene expression data was generated from primary breast tumour tissue biopsies from 44 patients before, and after 2 weeks of continuous anastrozole treatment in the anastrozole-only arm of a multicentre neoadjuvant clinical trial [373]. This study received approval from an institutional review board at each site and was conducted in accordance with the 1964 Declaration of Helsinki and International Conference on Harmonization/Good Clinical Practice guidelines. Written informed consent was obtained from each patient before participation. RNA was extracted, amplified and labelled as previously described [372], before hybridising to HumanWG-6 v2 Expression BeadChips (Illumina) according to the standard protocol. Tumour volume measurements were taken at pre-treatment, 8 weeks and again at 16 weeks and clinical response assessment was determined using the modified UICC/WHO assessment criteria as described previously [372, 373, 427]. Briefly, clinical response was determined by objective status and based on comparison of pre- with on- treatment (16 week) measurements taken as the sum of products of perpendicular diameters of measurable lesions. Measurements were collected using physical examination, calliper, ultrasound or magnetic resonance imaging (MRI); the same method was always used for follow-up measurements for the same patient. A complete response (CR) was characterised by

complete disappearance of all measurable and evaluable disease. A partial response (PR) was a decrease under pre-treatment of greater than or equal to 50% and progressive disease was an increase of 50% above pre-treatment. Stable disease/no response (SD) did not qualify for CR, PR or progressive disease.

3.3.6. Data Processing and Analysis

All data was processed using the R/Bioconductor software and packages [392], and the TM4 Microarray software suite (MeV) [428, 429] unless otherwise stated. Data generated on the Illumina microarray platform (this study, samples from 34 patients) and data generated on the Affymetrix platform (published dataset [124, 284], samples from 55 patients) were each independently pre-processed and re-annotated to Ensembl gene identifiers, then combined and batch corrected (see section 2). UHRR control samples were removed from the Illumina dataset prior to pre-processing. Briefly, Illumina probe profiles were quantile normalised using the *lumi* package (R/Bioconductor) and mapped to Ensembl gene sequences using a composite list comprising mappings from reMOAT [430], Ensembl BioMart [395] and a custom BLAST sequence search of the online Ensembl gene database where there was agreement between at least two of the resources. Where multiple Illumina probes represented an Ensembl gene, the mean expression level was calculated. A custom Chip Definition File (CDF) [393] was used to map the Affymetrix data to Ensembl gene annotations and RMA implemented by the *affy* package (R/Bioconductor) was used for normalisation. The datasets were then filtered using Illumina or Affymetrix probe detection P-values, removing probes that were undetected ($P>0.05$ in the total minus 3 samples). Both datasets were then combined and batch-corrected with cross-platform normalisation (XPN; ArrayMining) [398, 399] to reduce platform associated bias (see section 2). Replicate samples included in both datasets were used to confirm the success of integration as described in section 3.4.1. The combined, corrected Affymetrix and Illumina dataset (89 patients; 267 samples) was used for all further analysis (figure 30). These datasets were considered suitable for integration as both studies were designed with a similar experimental focus and both studies have similar composition in terms of patient/sample numbers and

clinical parameters (see section 3.4.1). R-scripts for data pre-processing are given in supplementary section 6.1)

Differential gene expression analysis was performed using Rank Product (RP) (MeV; TM4 Microarray Software Suite) [428, 429]. Pathway enrichment and functional gene ontology analysis was performed in DAVID Bioinformatics Resources 6.7 [431-433]. Multivariate analysis and regression and classification analysis was carried out using Random Forest (RF) (see section 3.3.6.1) [434] and Classification and Regression Trees (CART) (see section 3.3.6.2) [435, 436] (Salford Predictive Miner, Salford Systems, San Diego, USA) (figure 30). Predictive signatures were assessed using centroid classification and logistic regression (*glm* package: R/Bioconductor). Gene expression heatmaps were generated in MeV using Euclidean distance with complete linkage following gene mean-centering performed in Cluster 3.0 [400]. Multidimensional scaling was performed in R with scaling plots generated in JMP10 (JMP Software, USA). All statistical analyses were carried out in Prism 6 (Graphpad Software, California, USA).

3.3.6.1. Random Forest (RF)

Random Forest (RF) [434] analysis was used to reduce a large list of differentially expressed genes between responsive and non-responsive tumours to a relatively small number comprising the most important factors. RF is a form of dynamically construed nearest neighbour classifier capable of considering interaction effects among variables. It was used to grow a forest of 5000 classification trees each derived from a randomly boot-strapped sample of the original dataset and partially randomly selected predictor variables. Separate models were then combined in an ensemble via a voting or averaging process generating a list of variables ranked based on the contribution of each variable to model accuracy. Measures of contribution to accuracy are defined by the Gini score, which cumulates and standardises a measure of improvement (split of responsive and non-responsive tumours) for every node split by a variable in every tree in the forest.

3.3.6.2. Classification and Regression Trees (CART)

Classification and regression trees (CART) [435, 436] were used to develop decision tree based models to predict response to endocrine therapy. CART models are examples of binary recursive partitioning; binary as parent nodes are always split into exactly 2 child nodes and recursive as the process is repeated by considering each child node as a parent. CART models use exhaustive searches and computer-intensive testing techniques to identify a set of logical variables which can robustly separate heterogeneous data into homogenous segments. CART models are non-parametric, assuming neither a linear nor continuous non-linear relationship between variables and classifiers. CART models begin by attempting to divide the entire dataset into two segments using one variable. In the process every eligible variable is examined for splitting power resulting in partitioning of the data using the best performing predictor. Splitting of the data continues at each node resulting in growth of a maximal tree (a tree which cannot be grown any further) and which can subsequently be pruned. An error rate is defined for the maximal tree and for each sub-tree based on the misclassification rate derived from random internal testing.

In this study, 73 samples were included in the training set, 54 classified as responsive and 19 as non-responsive. Continuous expression variables from 200 genes were used as potential candidates for classification. One disadvantage of CART models is that trees can be unstable and as a result small changes in the data can result in generating very different trees. This is largely due to the fact that if a particular split changes all subsequent splits, which follow after it in the tree, also change. To address this, cross validation was carried out in which the test dataset (n=73) was randomly split into 10 sections, each with a similar distribution of response outcomes. One of the subsets was reserved for testing while the other 10 are combined and used as the test set for model building. The process was then repeated 10 times with a different subset of the data reserved for testing each time.

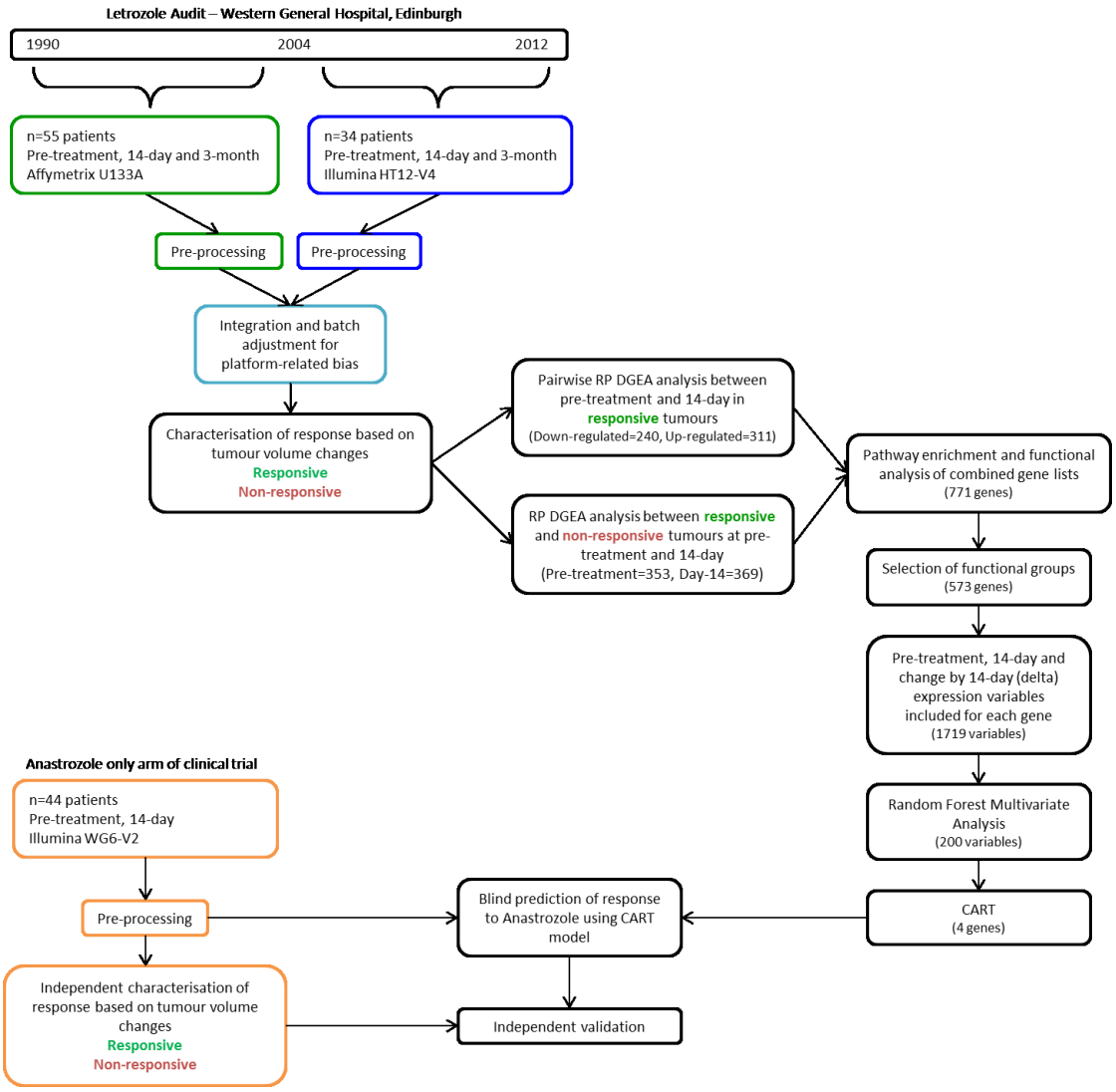


Figure 30. ***Flowchart of data pre-processing, integration, analysis and validation****. Pre-treatment, 14 day and 3 month gene expression data from 89 patients (combined dataset; teal) receiving neoadjuvant letrozole was derived by integrating a new gene expression study (Illumina: dark blue) with a previous study [124] (Affymetrix; dark green). Clinical response was determined by changes in tumour volume and patients were characterised as responsive (green) or non-responsive (red). Rank product differential gene expression analysis and multivariate techniques (RF and CART) and functional analysis were used to build a model to predict response to therapy. The model was independently validated in a gene expression dataset profiling 44 patients receiving neoadjuvant anastrozole (orange).*

3.3.7. Immunohistochemistry

3.3.7.1. Ki67

Immunohistochemical staining of FFPE tissue for Ki67 was performed on a subset of samples derived from the Miller *et al* dataset as part of a previous study [292]. Staining and scoring was performed as described previously [292].

3.3.7.2. CD45 Staining

Immunohistochemical staining of FFPE tissue with a mouse monoclonal antibody to CD45 (Cell Signalling Technology Inc.) diluted x100 was performed. Antigen retrieval was carried out using 0.1M sodium citrate/0.1M citric acid pH6 and detection using the EnVision^{TM} kit (Dako, Agilent Technologies) with 1 hour incubation as per the manufacturer's standard protocol. Sections were counterstained using haematoxylin. Scoring was carried out using subjective assessment of two independent scorers as described in section 3.4.4.3.

3.4. Results

3.4.1. Integration of Microarray Data

In order to increase sample number, particularly of non-responsive tumours which were limited in the sample population, data generated on the Illumina microarray platform (this study, samples from 34 patients) and data generated on the Affymetrix platform (published dataset [124, 284], samples from 55 patients) were integrated as discussed previously in sections 2 and 3.3.4. These datasets were considered suitable for integration as both studies have a very similar experimental design including: time-course, drug protocol, response assessment method and tissue sampling schedule. Furthermore, both studies have a similar composition in terms of patient/sample numbers and clinical parameters and indeed statistical analysis using logistic regression (LR) yielded no significant differences (figure 31). Criteria to describe distinct patterns of clinical response based on dynamic tumour volume changes were defined as described in section 3.4.2.1. Briefly, tumours were classified into one of 6 response

pattern groups: (i) 'quick stable' response, (ii) 'slow' response, (iii) 'non-response', (iv) 'intermediate' response, (v) 'acquired response' and (vi) 'acquired non-response', based on strict criteria (see section 3.4.2.1). The proportions of the 6 response groups, in particular the 3 major groups (i-iii), were found to be comparable in both datasets (figure 32), providing further justification for the suitability of these datasets for integration.

Dataset/Platform	Affymetrix	Illumina	Total	P
Number of Patients	55	34	89	
TNM Staging				
Unknown	0	2	2	
Tumour Size				0.134
T1	1 (2%)	5 (16%)	6	
T2	40 (73%)	14 (44%)	54	
T3	4 (7%)	5 (16%)	9	
T4	10 (18%)	8 (24%)	18	
Nodes				0.221
N0	42 (76%)	21 (66%)	63	
N1	13 (24%)	11 (34%)	24	
Metastasis				0.354
M0	51 (93%)	28 (88%)	79	
M1	3 (5%)	2 (6%)	5	
MX	1 (2%)	2 (6%)	3	
Tumour Grade				0.269
Unknown	0	0	0	
1	5	6	11	
2	39	23	62	
3	11	5	16	
Volume at Diagnosis (cm^3)				0.551
Mean	7.5	7.6	7.6	
Range	2.2-36.9	0.7-25.6	0.7-36.9	
ER Status (IHC-Allred Score)				0.234
Unknown	0	0	0	
6	1 (2%)	1 (3%)	2	
7	10 (18%)	5 (15%)	15	
8	44 (80%)	28 (82%)	72	
HER2 Status (IHC)				0.091
Unknown	0	3 (8%)	3	
Negative	48 (87%)	23 (68%)	71	
Positive	7 (13%)	8 (24%)	15	

Figure 31. ***Table comparing the composition of Affymetrix and Illumina datasets****. Affymetrix (dark green) and Illumina (dark blue) datasets were compared in respect of composition of clinical and pathological variables. Logistic regression analysis was performed to compare the two datasets in respect of each variable; P = P-values.*

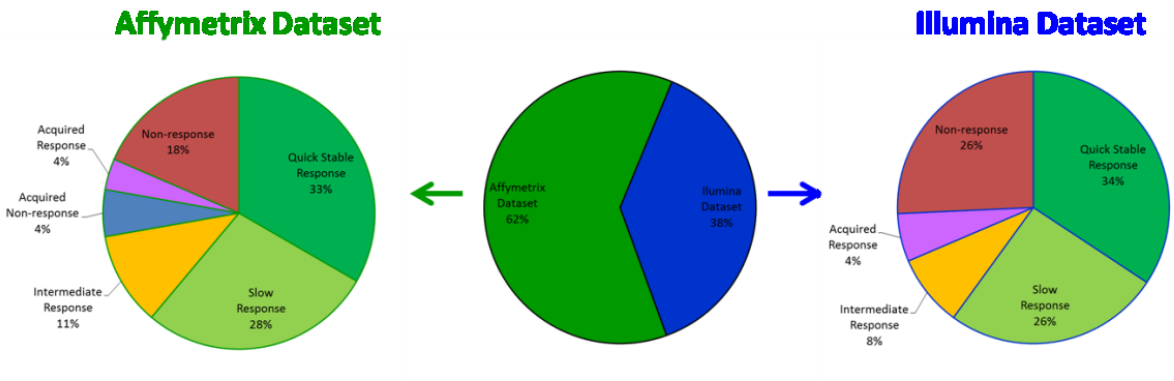


Figure 32. ***Comparison of proportions of clinical response patterns in Affymetrix and Illumina datasets****. The combined dataset comprised 62% of samples from the Affymetrix dataset and 38% from the Illumina dataset (centre). Affymetrix (dark green; left) and Illumina (dark blue; right) datasets were compared in respect of proportions of each clinical response pattern: 'quick stable' (green), 'slow' (light green), 'non-response' (red), 'intermediate' (yellow), 'acquired response' (purple) and 'acquired non-response' (blue).*

Prior to cross-platform integration, batch effects in the Affymetrix data due to sample processing data were minimised using ComBat [396] as discussed previously in section 2.4.2.2 and figure 20. The Illumina data was processed in 5 batches with a Universal Human Reference RNA (UHRR) control sample included in each batch to assess inter-batch variation as in a previous study [370]. An additional replicate UHRR sample was also included in processing batch 4 in order to compare inter and intra-batch variation (figure 33A).

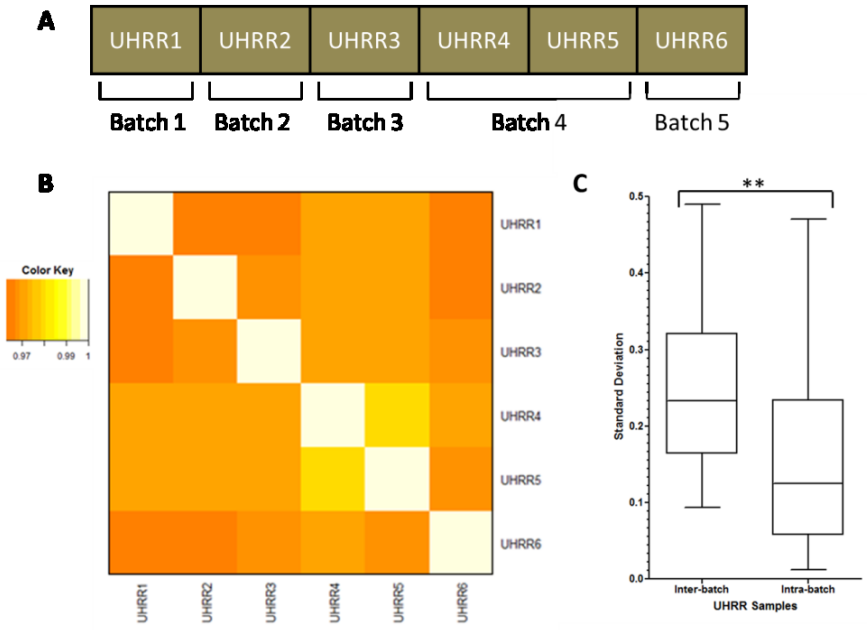


Figure 33. ***Analysis of inter- and intra-batch variation in the Illumina dataset****. A: distribution of replicate UHRR control samples in Illumina dataset processing batches. B: heatmap of Pearson correlations between replicate UHRR samples. R values range from low correlation (orange) to high correlation (white) through shades of yellow reflecting the overall similarity of expression profiles. C: comparison of variance (standard deviation) between inter-batch samples (UHRR1, 2, 3, 4 and 6) and intra-batch samples (UHRR4 and 5). ** = P<0.01.*

Analysis of the UHRR samples revealed that Pearson correlations between intra-batch samples (R=0.99) were marginally higher than inter-batch correlations (R=0.95-0.97) (figure 33B) and the variance between inter-batch UHRR samples was significantly higher (P<0.01) than between intra-batch samples (figure 33C). To correct the marginal batch effect the Illumina data was ComBat [396] corrected to reduce inter-batch variation.

The pre-processed Affymetrix and Illumina datasets were then integrated and XPN [398] corrected to reduce platform-related bias (figure 34). Samples (pre-treatment, 14 day and 3 month) from 6 patients originally included in the Affymetrix dataset were also included as replicates in the Illumina dataset. Pearson correlation analysis based on the expression of all genes revealed high correlation between matched pairs confirming the

success of integration (figure 35). The corrected combined dataset was used for all further analysis.

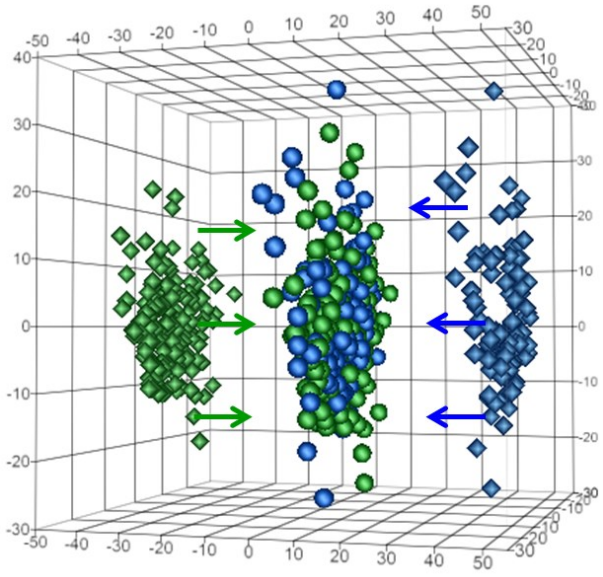


Figure 34. ***Integration and XPN correction of Affymetrix and Illumina datasets****. 3D multidimensional scaling (MDS) plot of all samples from the Affymetrix dataset (dark green) and Illumina dataset (dark blue) based on all genes, before (left and right) and after (centre) XPN correction.*

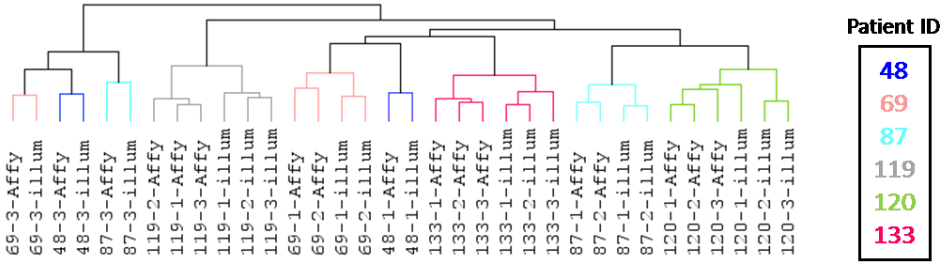


Figure 35. ***Hierarchical clustering of matched samples based on Pearson correlation after XPN correction****. Colour denotes samples from the same patient (as per the patient ID legend), the suffixes on patient ID's (48, 69, 87, 119, 120 and 133) denote as follows: '.1' = pre-treatment, '.2' = 10-14-days, '.3' = 3 months, '.Illum' = Illumina Beadarray data, '.Affy' = Affymetrix GeneChip data.*

3.4.2. Clinical and Pathological Assessment of Response

Clinical response was determined using dynamic changes in tumour volumes assessed by repeated measurements taken over the treatment period. Assessment was based on ultrasound measurements performed by a single sonographer (JMD). Patterns of clinical response were found to be varied and transient in a number of cases. To address this, criteria to describe distinct patterns of clinical response were defined and these were later used as the basis to determine overall clinical response or non-response and to select samples for predictive model development (figure 36).

3.4.2.1. Patterns of Clinical Response

Patterns of clinical response were determined using strict criteria (all percentage reductions relate to changes relative to the pre-treatment measurement). The three major patterns which characterise 82% of tumours in the combined dataset were defined as: (i) 'quick stable' response: a volume reduction of at least 50% by day 45 and at least 70% by 3 months, (ii) 'slow' response: a volume reduction between 0 and 50% by day 45 and at least 70% by 3 months, (iii) 'non-responder': increase in tumour volume or a partial reduction that never exceeds 50% (figure 36).

Three additional clinical response patterns were defined which characterise a small proportion of tumours (18% of the combined dataset) which have a transient or medial response to therapy: (iv) 'intermediate' response: a partial response by day 45 and continued partial response or no further change by 3 months, tumours reduce by at least 50% but not more than 70%, (v) 'acquired response': an increase above initial tumour volume by day 45 then a rapid reduction of at least 70% of initial volume by 3 months, (vi) 'acquired non-response': a decrease in tumour volume of at least 50% by day 45 then rapid increase to greater than the initial volume by 3 months (figure 36).

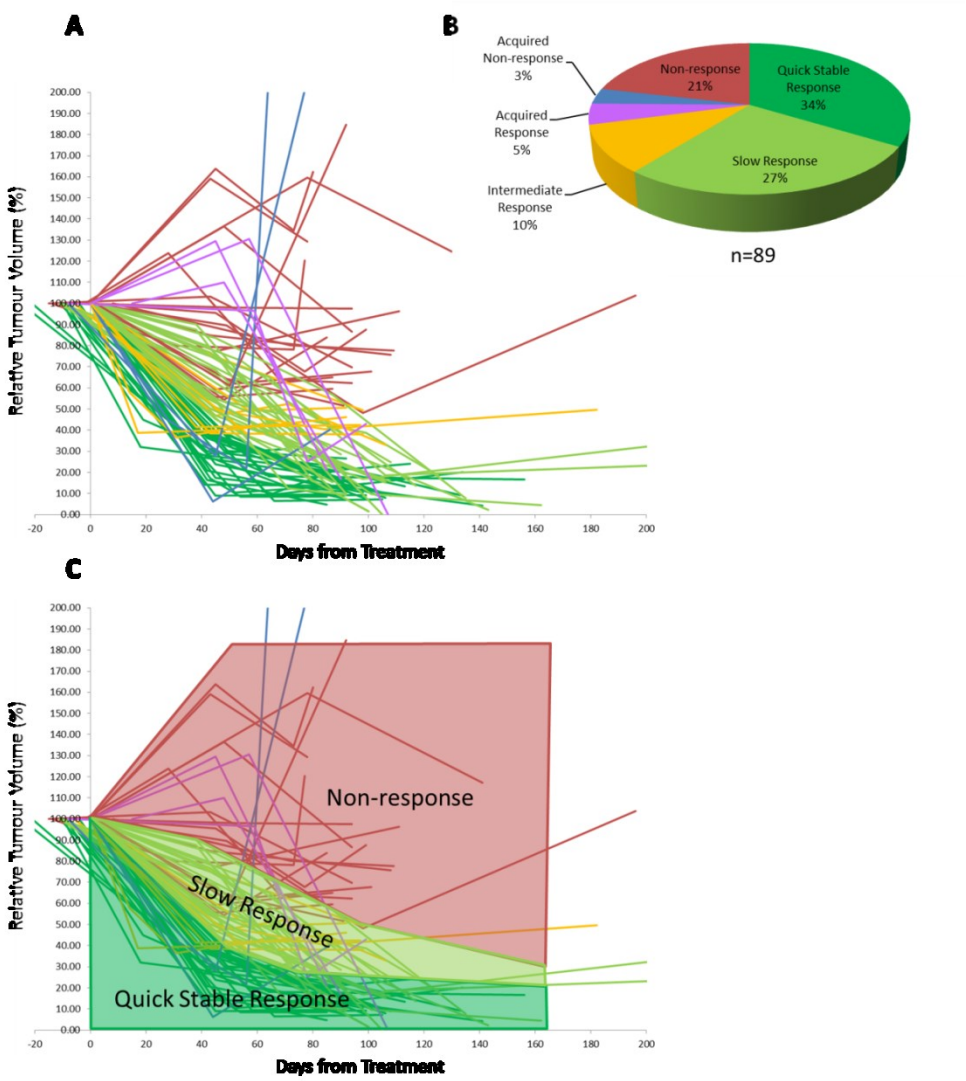


Figure 36. ***Patterns of clinical response to neoadjuvant letrozole therapy****. A: line graph of relative (%) changes in tumour volume assessed periodically over the treatment period. Each line represents one patient. B: Pie chart showing proportions of each clinical response pattern in the combined dataset. A&B: colours denote different clinical response patterns as follows: 'quick stable' (green), 'slow' (light green), 'non-response' (red), 'intermediate' (yellow), 'acquired response' (purple) and 'acquired non-response' (blue).C: For clarity the major clinical response patterns: 'quick stable' (green), 'slow' (light green), 'non-response' (red), are shown as blocks of colour on the graph.*

3.4.2.2. Determination of Overall Clinical Response

Overall response was used as the primary endpoint in this study and refers to the clinical characterisation of response after approximately 3 months of therapy, where in the neoadjuvant setting results can impact significantly on future clinical treatment decisions including the most appropriate surgery to perform. Tumours which respond well to endocrine therapy with significant reduction in volume are better candidates for less invasive breast conserving surgeries. The heterogeneous nature of response as demonstrated by the response pattern groups (see section 3.4.2.1), in particular the 'slow' response group which respond as well as the 'quick stable' responders by 3 months but which fail to respond as quickly, (figure 36) precludes early measurements or early changes in volume from being accurate indications of overall response. Therefore, in-keeping with currently used clinical parameters, overall response in this study was characterised by tumour volume changes at 3 months with the following cut-offs: 'responsive': tumour volume is less than or equal to 30% of initial volume, and 'non-responsive': tumour volume is greater than 50% of initial volume. Assigning these criteria gave a 'responsive' group comprising all tumours with 'quick stable' response and 'slow' response patterns, and a 'non-responsive' group consisting of the tumours with 'non-response' patterns (figure 37).

For the purposes of identifying early predictive molecular biomarkers of overall response, tumours with 'intermediate' response patterns (which comprise 10% of the combined dataset) were excluded, as tumour volume changes by 3 months are ambiguous, falling out with the predefined criteria for either 'responsive' or 'non-responsive' tumours. Additionally tumours with 'acquired response/non-response' patterns, which account for 8% of the dataset, were also excluded. Dramatic changes in clinical response of these atypical tumours are 'acquired' late in treatment suggesting molecular changes which occur out with the window of early molecular response biomarker determination (figure 37). However, response status in these excluded samples is considered in sections 3.4.4.5/6.

Overall Response	Response Pattern
Responsive	Quick Stable Response
	Slow Response
Non-responsive	Non-response
Excluded	Intermediate Response
	Acquired Response
	Acquired Non-response

Figure 37. ***Association between overall response and clinical response patterns****. Tumours classified as 'responsive' comprise all those with 'quick stable' and 'slow' response patterns and those classified as 'non-responsive' comprise all those with 'non-response' patterns. Tumours with 'intermediate' or 'acquired response/non-response' patterns were excluded.*

In the dataset used for later independent validation analysis (see section 3.4.4.7), in which 44 patients were treated with neoadjuvant anastrozole [373], clinical response was determined using the modified UICC/WHO criteria applied to tumour measurements after 16 weeks of continuous therapy (see section 3.3.5.2). To ensure consistency of response, classification patients in the combined dataset (this study) were also assessed for clinical response using the modified UICC/WHO criteria applied to 3 month tumour volume measurements. All patients determined to have a complete response (CR) or partial response (PR) using the modified UICC/WHO criteria were classified as responsive using the overall response criteria used in this study. All patients with stable disease (SD) or progressive disease using the modified UICC/WHO criteria were classified as non-responsive in this study (figure 38). Based on these findings, patients in the validation dataset with clinical responses classified as CR or PR were taken as responsive and patients classified as SD or PD were taken as non-responsive.

Overall Response (This study)	Modified UICC/WHO Criteria			
	Responsive		Non-responsive	
	Complete Response	Partial Response	Stable Disease	Progressive Disease
Responsive	1	53	0	0
Non-responsive	0	0	17	2

Figure 38. ***Association between response criteria used in this study and the modified UICC/WHO criteria****. Patients in this study were assessed for clinical response, based on 3D ultrasound measurements performed after 3 months of therapy, using both the overall response criteria (this study) and the modified UICC/WHO criteria. Responsive = green, non-responsive = red.*

3.4.2.3. Clinical Features of Response

Clinicopathological data were collected for every patient including TNM staging, tumour grade, and tumour volume at diagnosis and immunohistochemical determination of ER and HER2 status and were assessed for prediction of overall response outcome by logistic regression (LR) (figure 39). Of the clinicopathological variables tested, HER2 positivity was found to be significantly associated with non-response to therapy (P=0.006). However, in the non-responsive tumours HER2 over-expression occurs in only 42% and even within the HER2 positive population only 50% of tumours are non-responsive. Other variables were not found to have a statistically significant association with response (figure 39).

Clincal Response	Responsive	Non-responsive	Total	P
Number of Patients	54	19	73	
TNM Staging				
Unknown	1	2	3	
Tumour Size				0.272
T1	4	0	4	
T2	32	10	42	
T3	5	2	7	
T4	12	5	17	
Nodes				0.151
N0	35	15	50	
N1	18	2	20	
Metastasis				0.739
M0	47	16	63	
M1	3	1	4	
MX	3	0	3	
Tumour Grade				0.256
Unknown	0	0	0	
1	8	2	10	
2	39	10	49	
3	8	7	15	
Volume at Diagnosis (cm^3)				0.551
Mean	14.5	9.9	12.2	
Range	1.4-36.9	1.2-34.3	0	
ER Status (IHC-Allred Score)				0.21
Unknown	0	0	0	
6	2	0	2	
7	14	0	14	
8	38	19	57	
HER2 Status (IHC)				0.006
Unknown	3	0	3	
Negative	43	11	54	
Positive	8	8	16	

Figure 39. ***Table showing association of clinicopathological data with clinical response****. Clinicopathological data comprising TNM staging, grade at diagnosis, initial tumour volume and range of volumes across the response groups and immunohistochemical scoring of HER2 and ER were collected for every patient. Logistic regression analysis was performed to assess prediction of overall clinical response; P = P-values.*

3.4.2.4. Pathological Response

With the objective of neoadjuvant endocrine therapy to shrink tumour size such that breast conserving surgery becomes possible, clinical response was taken as the primary endpoint. However, pathological response was also determined at pre-treatment, day 14 and 3 months for each patient by a pathologist (JT) (figure 40A). Pathological response was assessed by changes in tumour grade as determined by microscopic examination of acinar/tubule formation, nuclear atypia/pleomorphism and presence of mitotic cells. Tumour cellularity was also estimated as a percentage of the tissue section. A complete

response, defined as no remaining tumour, occurred in only 5 of 73 patients by 3 months. A partial response was defined as either a decrease by more than 20% in tumour cellularity, improvement of grade or both and progression was characterised by an increase in tumour cellularity of greater than 20%, increased grade or both. Pathological response at 3 months is given in figure 40B. In terms of tumour cellularity, partial response to therapy occurs in the majority (65%) of clinically responding tumours and in approximately one third (39%) of non-responsive tumours. However, approximately half (53%) of the non-responsive tumours do not show any significant changes in cellularity. The majority of tumours do not change in grade by 3 months (58% of responsive and 58% of non-responsive) however a greater proportion of responsive tumours decrease in grade compared with non-responsive tumours (31% and 11% respectively). Conversely, a greater proportion of non-responsive tumours progress with an increase in tumour grade compared with the responsive tumours (32% and 4% respectively).

Pathological response was also determined at day 14 for each patient (figure 41A). In terms of tumour cellularity, 41% of responsive tumours were classified as having a partial response while another 41% were found not to have changed. In the non-responsive tumours, 47% were classified as having a partial response and 37% were found not to have changed. In terms of changes in tumour grade by day 14, the majority 74% of responsive and 89% of non-responsive tumours did not change. As an indication of overall response, early (day 14) pathological changes were not found to be significantly informative in logistic regression (LR) analysis (figure 41B).

A

Clincal Response	Responsive	Non-responsive	Total
Number of Patients	54	19	73
Pathological Response			
Cellularity			
Complete Response	4	1	5
Partial Response	34	7	41
No Change	8	10	18
Progression	6	1	7
Unknown	2	0	2
Grade			
Complete Response	4	1	5
Partial Response	16	1	17
No Change	30	11	41
Progression	2	6	8
Unknown	2	0	2

B

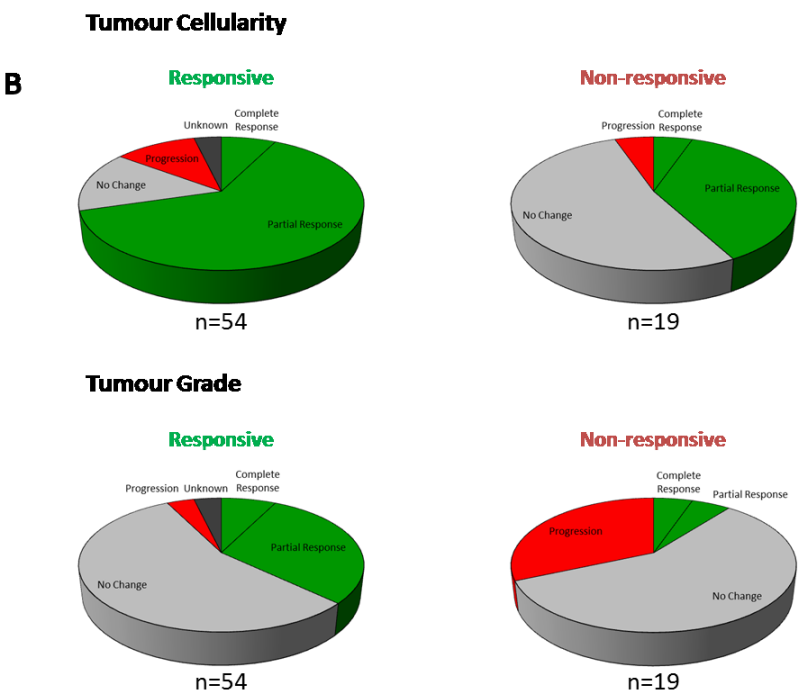
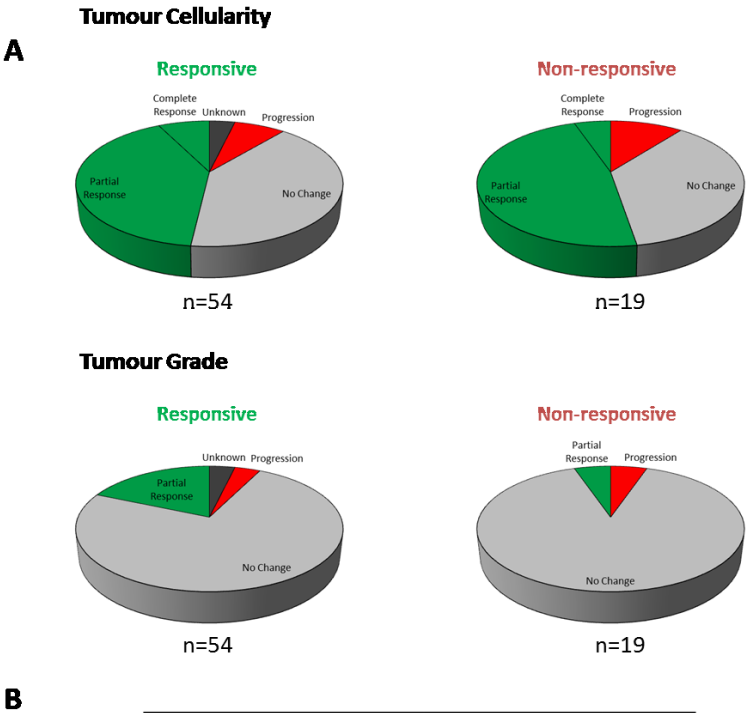


Figure 40. ***Association of pathological response with clinical response after 3 months of therapy****. A: Table comparing pathological response, in respect of changes in tumour cellularity and changes in tumour grade, with overall clinical response. B: Pie charts showing proportion of pathological response classifications associated with clinically responsive (left) and non-responsive (right) tumours in respect of changes in tumour cellularity (top) and changes in tumour grade (bottom). Pathological response classifications are as follows: complete and partial response (green), no change (grey), progression (red) and unknown (black).*



B

Pathological Variable	P-value	Odds Ratio	95% CI
Change in tumour cellularity at 14 days	0.267	0.54	0.18-1.60
Change in tumour grade at 14 days	0.241	3.61	0.42-30.89

Figure 41. ***Association of pathological response with clinical response after 14 days of therapy****. A: Pie charts showing proportion of pathological response classifications associated with clinically responsive (left) and non-responsive (right) tumours in respect of changes in tumour cellularity (top) and changes in tumour grade (bottom). Pathological response classifications are as follows: complete and partial response (green), no change (grey), progression (red) and unknown (black). B: results of logistic regression (LR) analysis to assess prediction of overall clinical response using early pathological changes.*

3.4.2.5. Breast Cancer Molecular Subtypes and Response

All tumours were profiled using the intrinsic subtype classifications proposed by Sørlie *et al* [25] applied to pre-treatment expression (figure 42). Of 73 tumours, 82% were found to be classified as luminal A and the remaining 18% as luminal B. Within the luminal B tumours the majority (62%) were non-responsive. However, within the non-responsive population, luminal B only accounts for 42% of tumours with the majority classified as luminal A.

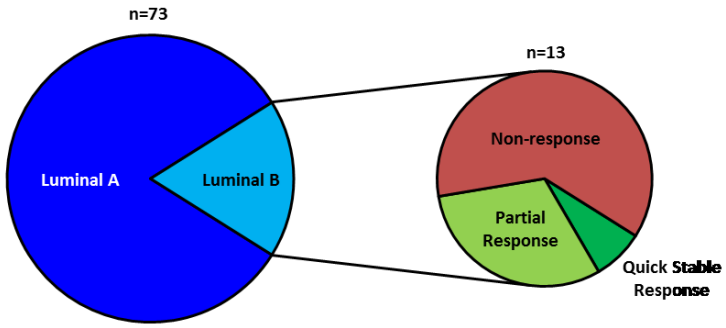


Figure 42. ***Association of luminal B molecular subtype with clinical response****. Left: pie chart showing the proportion of tumours classified as luminal A (dark blue) and luminal B (light blue) using the Sørlie intrinsic subtype classifications applied to pre-treatment expression. Right: pie chart showing the proportion of luminal B tumours belonging to each of three clinical response patterns: 'quick stable' (green), 'slow' (light green) and 'non-response' (red).*

3.4.3. Molecular Profile of Quick Stable Response

Within the 'quick stable' response subgroup, pairwise Rank Product (RP) differential gene expression analysis (FDR=0.01) was used to identify the most changed genes which characterise a good clinical response to therapy. The analysis gave rise to 551 consistent early changed genes between pre-treatment and day 14 (figure 44) and 1063 consistent overall changes between pre-treatment and 3 months (figure 43). Pathway enrichment and biological process gene ontology analysis was carried out separately for down-regulated and up-regulated genes in DAVID Bioinformatics Resources 6.7. Functional analysis revealed genes down-regulated after 3 months of therapy were enriched for processes including proliferation, glycolysis and oxidative phosphorylation. Genes found to be up-regulated at 3 months were enriched for processes including extracellular matrix (ECM) or stromal remodelling (including cellular adhesion and angiogenesis) and inflammatory immune response. Common to both up and down-regulated gene lists, were genes with biological process annotations enriched for cell signalling (G-protein complex receptor signalling and small GTPase mediated signal transduction), transcription (transcription machinery, transcriptional regulation and

transcriptional cofactor activity) and protein processing (protein folding, amino acid metabolism, phosphorylation, glycosylation and acetylation). Several of the genes (78 of the 551 early changed genes and 162 of the 1063 overall changed genes) had unknown biological function. Another set with unknown biological function was annotated as 'metal ion binding' reflecting their known molecular function.

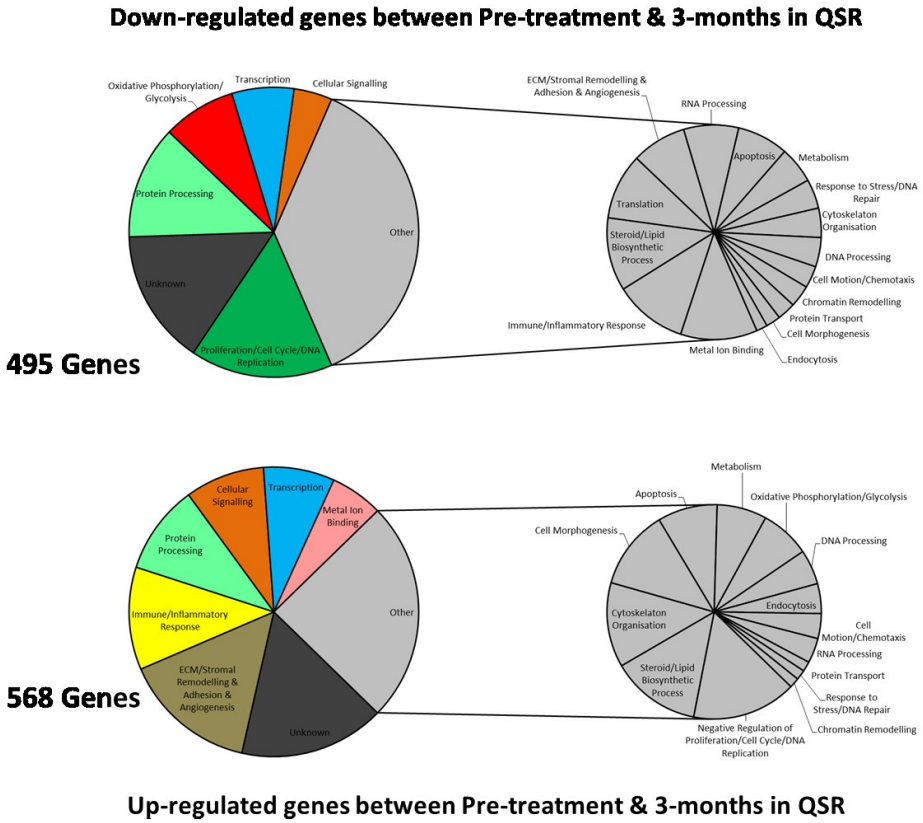


Figure 43. ***Functional analysis of significantly differentially expressed genes between pre-treatment and 3 months in the 'quick stable' response group****. Pie charts showing genes down-regulated (top) and up-regulated (bottom) at 3 months grouped by functional association. Each gene was assigned to only one group which was the most significantly enriched functional process or pathway. Large pie charts (left) comprise functional groups represented by at least 15 genes, minor functional groups are shown in small pie charts (right). Colours assigned to each functional group are consistent throughout all figures in this study.*

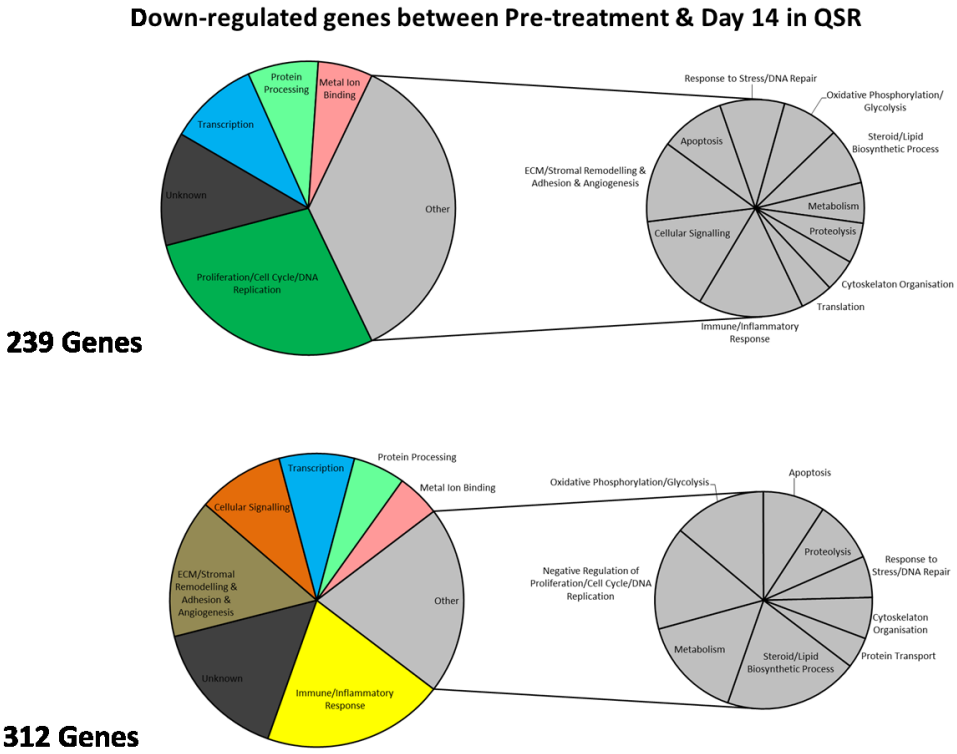


Figure 44. ***Functional analysis of significantly differentially expressed genes between pre-treatment and 14 days in the 'quick stable' response group****. Pie charts showing genes down-regulated (top) and up-regulated (bottom) at 14 days grouped by functional association. Each gene was assigned to only one group which was the most significantly enriched functional process or pathway. Large pie charts (left) comprise functional groups represented by at least 15 genes, minor functional groups are shown in small pie charts (right). Colours assigned to each functional group are consistent throughout all figures in this study.*

3.4.3.1. Down-regulated Genes and Processes

Proliferation associated genes including those involved in the regulation of cell cycle such as cyclins (CCNA2, CCNB1, CCND1 and CCNE2), cyclin-dependent kinases (CDK1, CDK2 and CDK3) and cell-division cycle genes (CDC7, CDC20, CDC23 and CDC25C) were significantly down-regulated at 3 months. In addition, genes involved with DNA replication such as the mini chromosome maintenance genes (MCM2,

MCM3, MCM4, MCM5, MCM6 and MCM7) and replication factor genes (RFC2, RFC3, RFC4 and RFC5) were also significantly down-regulated at 3 months. Other proliferation associated genes found to be down-regulated at 3 months were involved with the M-phase of mitosis including formation of the centromere-kinetochore complex (CENPF, CENPI, CENPN and CENPQ) and formation of mitotic spindle (ASPM, AURKA, TPX2 and SAC3D1).

Genes involved with glycolysis (GAPDH and PDHB) and several involved with the mitochondrial generation of energy through oxidative phosphorylation were also found to be significantly down-regulated following 3 months of therapy. The latter includes genes encoding major protein complexes involved in the electron transport chain such as cytochrome c oxidase (COX6C and COX7A2), ATP synthase (ATP5E and ATP5EP2) and NADH dehydrogenase ubiquinone (NDUFA4, NDUFA7, NDUFA8, NDUFB2, NDUFB5 and NDUFB7).

3.4.3.2. Up-regulated Genes and Processes

Genes significantly up-regulated by 3 months were enriched for immune processes which might suggest an inflammatory response to treatment (discussed later: see section 3.4.4.3). Up-regulated genes in this category included those encoding lysosomal acid hydrolases such as proteases (cathepsins: CTSA, CTSB, CTSG, CTSK, CTSL1, CTSS and CTSZ), glycosidases (HEXB and FUCA1) and lipases (LIPA and ACP5), genes encoding the major histocompatibility complex type two (HLA-DMB, HLA-DOB, HLA-DPA1, HLA-DPA2 and HLA-DRA) and genes involved in the complement and coagulation cascade (C1S, C7, VWF, CFD, CFH and CFI). Additionally, genes encoding inflammatory chemokines including C-X-C motif and C-C motif ligands were significantly up-regulated by 3 months suggesting a possible chemokine mediated recruitment of immune cells and induction of inflammatory response.

Genes associated with ECM/stromal remodelling, adhesion and angiogenesis were also significantly up-regulated at 3 months. These include genes involved in fibrosis and production and remodelling of connective and vascular tissue such as elastin (ELN),

biglycan (BGN), collagens (COL1, COL2, COL3, COL5, COL6, COL14, COL15, COL16 and COL21 family of collagen genes) and laminins (LAMA, LAMB and LAMC family of laminin genes) and genes involved in cellular and matrix adhesion (CD36, ANXA9, CLDN5, CLDN8, FBLN2, FBLN5, THBS2 and THBS3).

3.4.3.3. Comparison of 'Quick Stable' and 'Slow' Response Gene Changes

A comparison of early and overall consistent gene changes derived from pairwise RP (FDR=0.01) analysis in 'quick stable' response and 'slow' response tumours revealed significant overlap in both early and overall changed genes (figure 45). However, a greater proportion of genes were significantly up or down-regulated in the 'quick stable response' tumours both at day 14, consistent with the early dramatic changes in volume seen in these tumours, and at 3 months compared with the 'slow response' tumours. Given the clear differences in early clinical response between these two groups of responsive tumours (see section 3.4.2.1), robust prediction of overall response and non-response is likely to rely on identifying molecular biomarkers which have consistent early expression or early changed expression across the 'quick stable' and 'slow' response tumours and have significantly different expression in the non-responsive tumours.

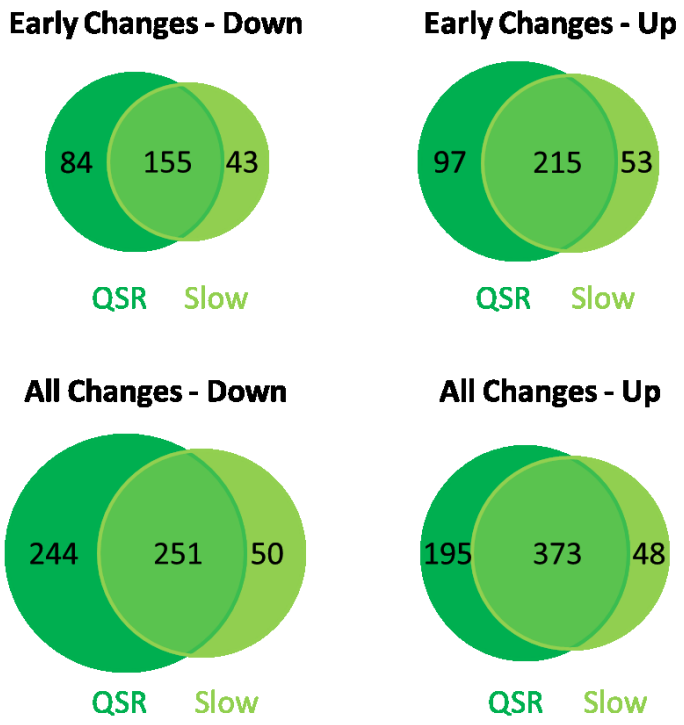


Figure 45. ***Comparison of early and overall gene expression changes in 'quick stable' and 'slow' response tumours****. Venn diagrams comparing early changes (by day 14: top) and overall changes (by 3 months: bottom) in differentially expressed genes both down-regulated (left) and up-regulated (right) in 'quick stable' (green) and 'slow' (light green) response tumours.*

3.4.4. Predicting Response to Therapy

3.4.4.1. Early and Late Gene Changes in Major Functional Groups

The major functional groups identified as significantly up or down-regulated within the 'quick stable' response tumours were analysed further to identify gene groups that change early in treatment (by day 14). Changes were assessed in both responsive and non-responsive tumours to identify unique patterns of change and differences in expression levels associated with response (figure 46). Hypothetically, genes and processes which change early in treatment and/or have early expression profiles differentially associated with clinical response represent promising candidates for early predictive marker exploration.

Proliferation associated genes were found to be significantly down-regulated in both 'quick stable' and 'slow' response tumours with changes occurring as early as day 14. Consistent early down-regulated genes include the cyclins (CCNA2, CCNB1 and CCND1), the mini chromosome maintenance genes (MCM2, MCM4 and MCM6) and the mitotic spindle associated genes (ASPM and AURKA). Within the non-responsive tumours proliferation associated genes were found to be down-regulated in some patients, however to a significantly lesser extent than in responding tumours with several tumours showing no change by day 14 or 3 months (figure 46). The full list of proliferation genes is given in supplementary section 6.3.

Glycolysis and oxidative phosphorylation associated genes were found to be significantly and consistently down-regulated by 3 months in both 'quick stable' and 'slow' response tumours. Some tumours were found to change as early as day 14 however early changes were not consistent across all responsive tumours, with several tumours maintaining high expression levels of these genes at day 14. Within the non-responsive tumours, expression levels and associated changes were found to be more heterogeneous with some tumours having low expression of these genes at pre-treatment and increased expression at day 14 and 3 months while others had down-regulated expression consistent with the responsive tumours (figure 46).

Genes involved with the immune/inflammatory response and ECM/stromal remodelling were found to be significantly co-expressed, with the same tumours having similar expression levels and consistent patterns of changed expression across all differentially expressed genes in these functional categories, suggesting that these biological processes are linked. In the responsive tumours, significant up-regulation of these genes occurs by 3 months in the majority of tumours while only some tumours have increased expression by day 14. Within the non-responsive tumours expression of these genes remained consistently low at day 14 and at 3 months with only a minority of tumours having significantly increased expression by 3 months.

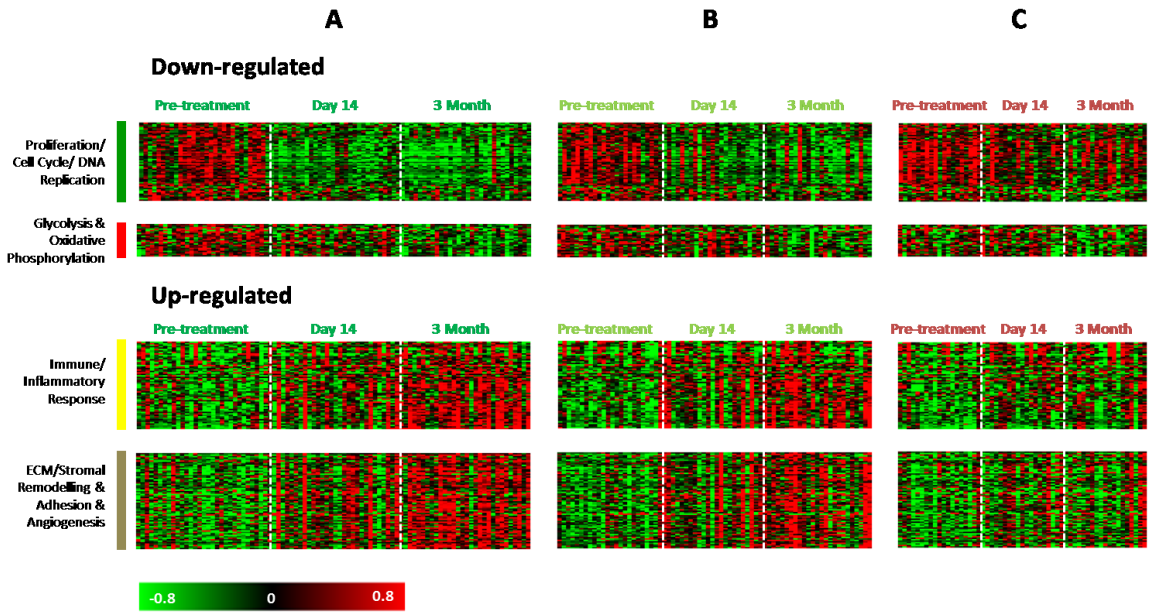


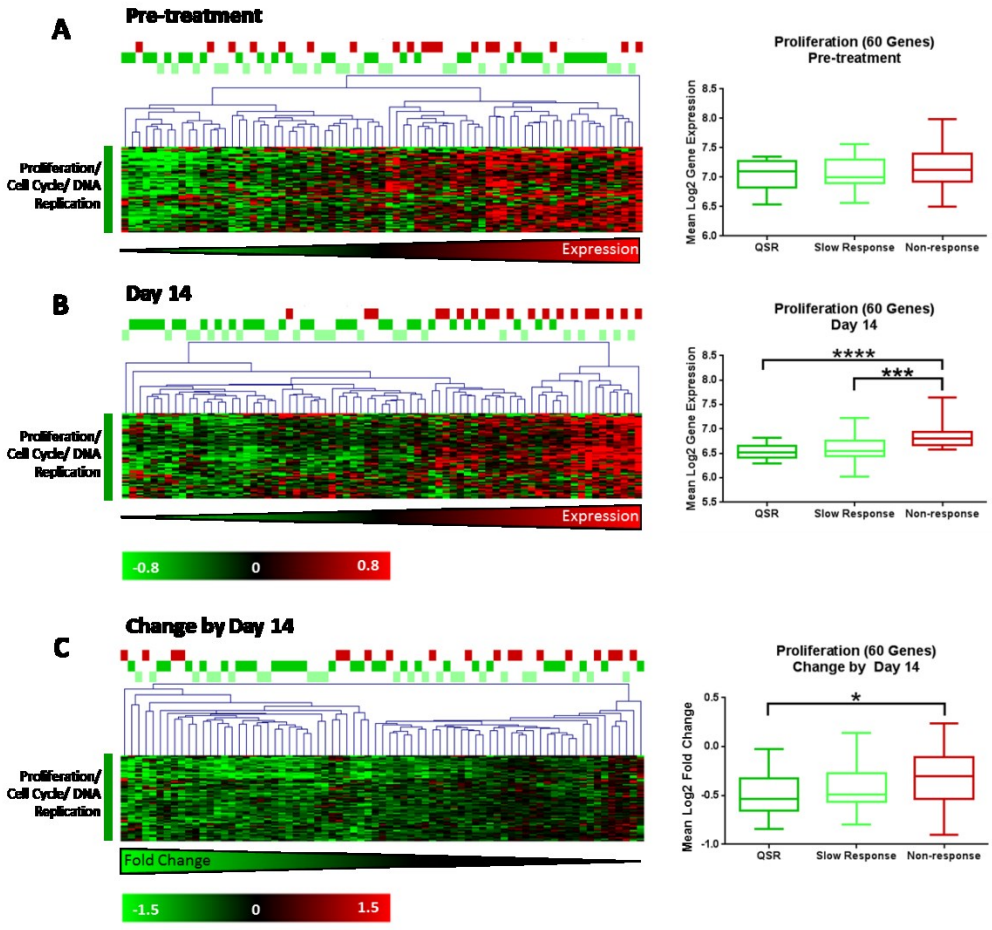
Figure 46. ***Heatmaps showing changes in gene expression over time in major functional groups****. Expression of genes involved in four major function groups: proliferation/cell cycle/DNA replication (green), glycolysis and oxidative phosphorylation (red), immune/inflammatory response (yellow) and EMC/stromal remodelling and adhesion and angiogenesis (brown) identified as differentially down-regulated (top) or up-regulated (bottom) between pre-treatment and day 14 in 'quick stable' response tumours. Samples are in numerical order of patient ID for each time point shown (pre-treatment, day 14 and 3 month) above each heatmap block. Log2 expression levels have been gene mean centred across all response groups shown, A: 'quick stable' (green; left), B: 'slow' (light green; centre) and C: 'non-response' (red; right) tumours. In the heatmaps, colours represent relative differences in expression with red denoting higher expression and green lower expression.*

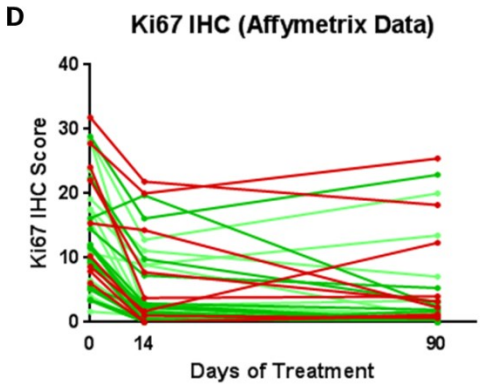
3.4.4.2. Proliferation and Response to Therapy

Further analysis of the proliferation associated gene set (n=60, the full list of proliferation genes is given in supplementary section 6.3.) revealed strong associations with response after 14 days of treatment (figure 47B). While expression of proliferation genes at pre-treatment was poorly associated with response, high expression at day 14 was found to be reasonably correlated, with the majority of non-responsive tumours expressing higher levels of these genes compared with both 'quick stable' ($P<0.0001$)

and ‘slow’ (P=0.0053) response tumours (figures 47A). However, while the latter is true for most of the responsive tumours, some maintained high expression levels of proliferation genes, similar to the non-responsive tumours, at day 14. Of these the majority belong to the ‘slow’ response group which, early in treatment, have been shown (see section 3.4.2.1) to behave clinically like the non-responsive tumours. In addition, analysis of fold changes (change in proliferation gene expression between pre-treatment and day 14) also revealed poor associations with response (figure 47C). Together, these findings suggest that while proliferation associated genes may comprise an important part of a predictive model of response, proliferation alone is not robustly predictive of response, in particular in distinguishing between ‘slow’ response and non-responsive tumours early in treatment. Furthermore, changes in early gene expression and expression levels at pre-treatment may not be as predictive as day 14 expression levels.

Immunohistochemically-assessed expression of the proliferation marker Ki67 was carried out at pre-treatment, 14 days and 3 months for samples in the Affymetrix part of the combined dataset as part of a previous study [292]. Interestingly, expression of Ki67 was not found to be significantly different between responsive and non-responsive tumours at any time-point (figure 47D).


A
Pre-treatment
Proliferation/ Cell Cycle/ DNA Replication
Expression
Proliferation (60 Genes) Pre-treatment
Mean Log2 Gene Expression
QSR
Slow Response
Non-response
B
Day 14
Proliferation/ Cell Cycle/ DNA Replication
Expression
-0.8
0
0.8
Proliferation (60 Genes) Day 14
****
***
Mean Log2 Gene Expression
QSR
Slow Response
Non-response
C
Change by Day 14
Proliferation/ Cell Cycle/ DNA Replication
Fold Change
-1.5
0
1.5
Proliferation (60 Genes) Change by Day 14
*
Mean Log2 Fold Change
QSR
Slow Response
Non-response


D
Ki67 IHC (Affymetrix Data)
Ki67 IHC Score
Days of Treatment
0
14
90

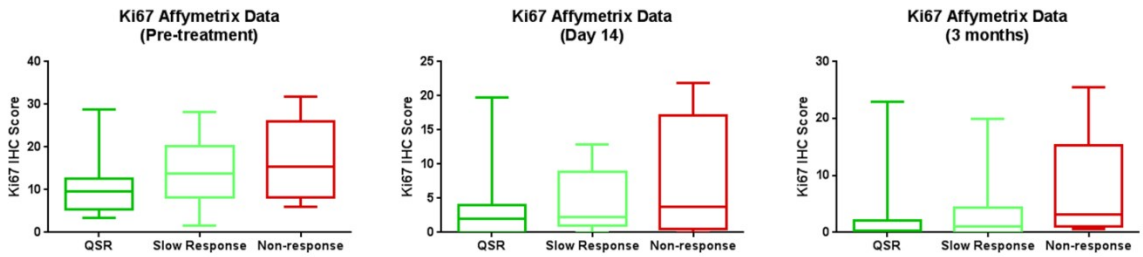

Ki67 Affymetrix Data (Pre-treatment)
Ki67 IHC Score
QSR
Slow Response
Non-response
Ki67 Affymetrix Data (Day 14)
Ki67 IHC Score
QSR
Slow Response
Non-response
Ki67 Affymetrix Data (3 months)
Ki67 IHC Score
QSR
Slow Response
Non-response

Figure 47. ***Association between proliferation and clinical response****. A&B: heatmaps and boxplots showing expression and mean expression respectively of 60 proliferation genes identified as significantly down-regulated in the 'quick stable' response tumours at pre-treatment (A) and 14 days (B). In heatmaps samples were ordered from low (left) to high (right) based on mean expression. Coloured bars above heatmaps correspond to colours in boxplots and represent clinical response patterns: 'quick stable' (green), 'slow', (light green) and 'non-response' (red). In the heatmaps log2 expression was gene mean centered across all samples and heatmap colours represent relative differences in expression with red denoting higher expression and green lower expression. C: heatmap and boxplot showing change in expression (fold change) of proliferation genes by day 14. In the heatmap colours represent log2 fold changes: red = increased expression, black = no change and green = decreased expression. (Continued on page 128) D: line graph and boxplots for each time point of immunohistochemically determined scores for Ki67 over time with assessments made at pre-treatment, 14 days and 3 months. In the line graph each line represents a single patient. Line and boxplot colours denote clinical response patterns: 'quick stable' (green), 'slow' (light green) and 'non-response' (red). A-D: stars above boxplots denote statistical significance: **** = P<0.0001, *** = P<0.001, * = P<0.05. All possible statistical comparisons were conducted between groups; only statistically significant differences are represented in plots.*

3.4.4.3. Inflammation, Stromal Remodelling and Response to Therapy

Multidimensional scaling (MDS) was used to model associations between responsive and non-responsive tumours at pre-treatment, day 14 and 3 months based on the 551 early (by day 14) most changed genes in 'quick stable' response tumours from pairwise RP (FDR=0.01) analysis (figure 48). Clustering of responsive and non-responsive tumours based on all 551 genes was heterogeneous at pre-treatment and day 14, however by 3 months a cluster comprising only responsive tumours and enriched for 'quick stable' response tumours was clearly identifiable (figure 48). Tumours in the responsive-only cluster were those with the most changed gene expression profile between pre-treatment and 3 months. RP analysis (FDR=0.01) and pathway enrichment and biological process gene ontology analysis at 3 months revealed that tumours in this cluster were significantly enriched for high expression of immune/inflammatory response and ECM/stromal remodelling genes compared with all other tumours,

providing further evidence that these processed are linked. These results indicate that a local immune/inflammatory response and ECM/stromal remodelling occurs in some tumours and is significantly associated with a positive response to therapy, however, such a response is not dependent on up-regulation of these processes. Conversely, non-responsive tumours were not found to be up-regulated in respect of these processes. Despite the strong association with a positive response to therapy, these processes have been shown to be significantly up-regulated in only some responsive tumours by 3-months and only in a small subset of those as early as day 14. However, identification of individual genes involved which are consistently up-regulated early in treatment may be possible and these may represent promising candidate markers for discriminating between subsets of responsive and non-responsive tumours.

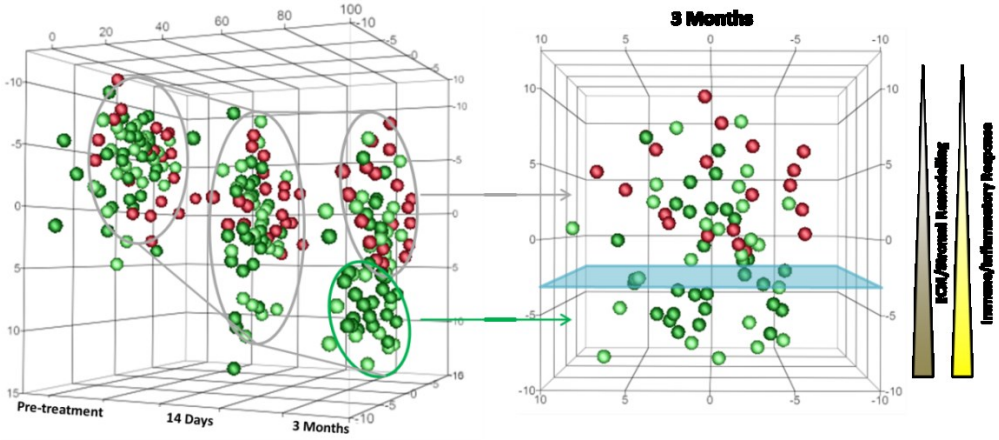


Figure 48. ***Modelling of association between responsive and non-responsive tumours over time reveals an immune/ECM enriched responsive cluster at 3 months****. Left: 3D MDS plot (2D scaling coordinated with third dimension as time) of all tumours at each time point based on the 551 most differentially expressed genes between pre-treatment and 14 days in 'quick stable' response tumours. Right: 2D MDS plot of all tumours at 3 months. Tumours belonging to the immune/ECM enriched cluster are shown beneath the blue plane and enrichment of functional groups are represented by the gradient-coloured triangles (dark = high expression, light = low expression). Both: each sphere represents a single sample at a given time point, spheres are coloured based on clinical response pattern: 'quick stable' (green), 'slow' (light green) and 'non-response' (red).*

Inflammatory activity can be readily assessed in histological specimens by staining for leukocyte infiltration using the leukocyte common antigen: CD45. This molecule is a type 1 transmembrane protein tyrosine phosphatase which is commonly expressed on the cell surface of mature haematopoietic lineage cells with the exclusion of platelets and mature erythrocytes [437]. Presence of leukocyte infiltration was determined by examination of CD45 stained specimens for each patient at pre-treatment and 3 months. Assessments were made using semi-quantitative subjective criteria in which a score between 0 and 3+ was assigned based on intensity and proportion of stained cells (figure 49A). Leukocytes were found to be present either 'diffusely', homogeneously distributed throughout the specimen, as heterogeneous 'aggregates' or a combination of both (figure 49A). Analysis of results revealed that the majority of responsive tumours had no leukocyte infiltration at pre-treatment, including the subset of responsive tumours with significantly increased 3 month expression of immune/inflammatory and ECM related genes (figures 49B, D and E). However, 10 of the 19 non-responsive tumours were found to have leukocyte infiltration at pre-treatment suggesting a possible link between the presence of immune/inflammatory cells and non-response to therapy in some tumours (figures 49B and E). Interestingly, logistic regression (LR) analysis revealed a strong association P=0.007 between leukocyte infiltration and non-response at pre-treatment compared to the responsive group: comprising all 'quick stable' and 'slow' response tumours (figure 49B). Analysis to compare the presence of leukocyte infiltration with proliferation was carried out and the presence of immune/inflammatory cells at pre-treatment was found to be associated with higher expression of proliferation genes at day 14 (figure 49C). The majority of both responsive and non-responsive tumours were found to have no leukocyte infiltration at 3 months. Interestingly, the presence of leukocytes in the subset of responsive tumours with increased expression of immune/inflammatory and ECM related genes at 3 months was found in only 11 of 25 tumours suggesting that immune related cells were not responsible for increased expression of these genes (figures 49D and F).

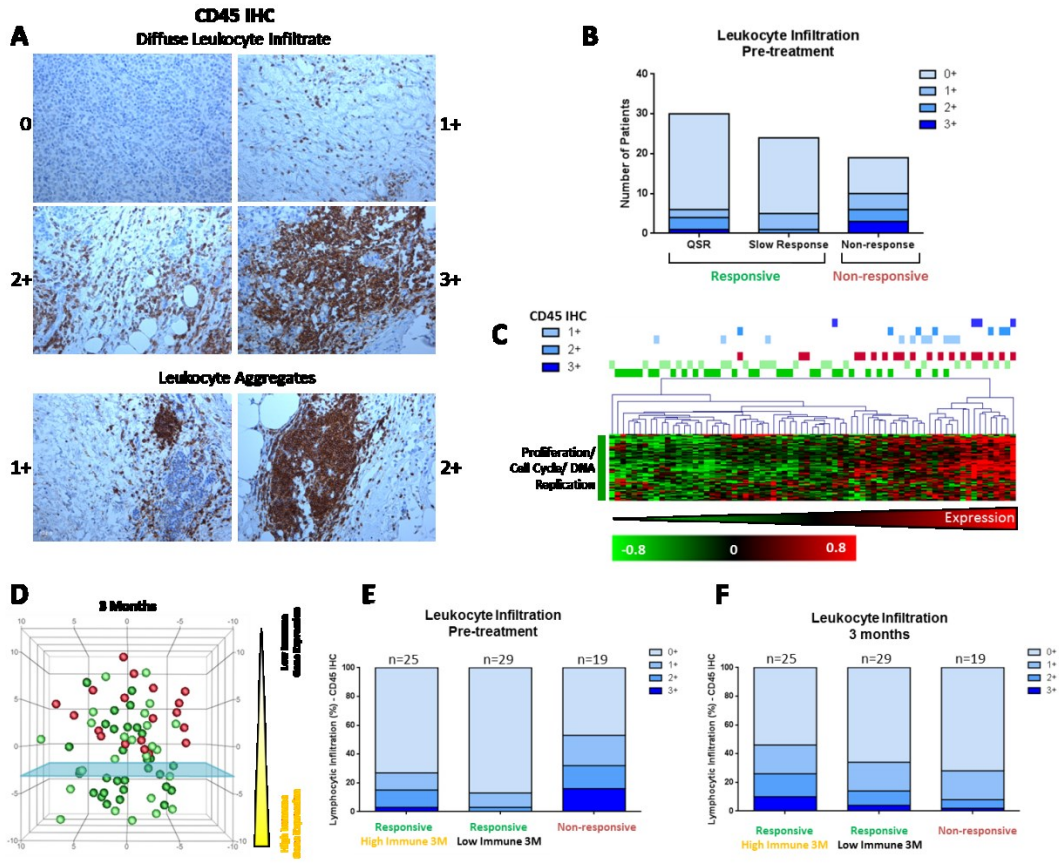


Figure 49. ***Association between leukocyte infiltration and clinical/proliferative response to therapy****. A: tissue sections of pre-treatment and 3 month samples were stained for CD45 and scored between 0 and a maximum of 3+ using semi-quantitative subjective criteria. Leukocytes were 'diffusely' distributed through the tissue (top), 'aggregated' (bottom) or a combination of both. B: bar chart showing IHC score proportions for leukocyte infiltration across 'quick stable', 'slow' and 'non-response' groups. Colours in bars of graph refer to IHC score for leukocyte infiltration as per the legend. C: heatmap showing the association between leukocyte infiltration and expression of the 60 most significantly down-regulated proliferation-associated genes by day 14 in 'quick stable' response tumours across all tumours. Log2 expression was mean centre and ordered from low (left) to high (right) based on mean expression. Colours in the heatmap range from green to red through black and refer to relative expression, red = higher expression and green = lower expression.* *(Continued on page 141)*

Coloured bars above the heatmap represent clinical response patterns: 'quick stable' (green), 'slow', (light green) and 'non-response' (red) and IHC scores for leukocyte infiltration as per the key. D: 2D MDS plot of all tumours at 3 months from figure 48.Tumours belonging to the immune enriched cluster are shown beneath the blue plane and enrichment of the immune functional group is represented by the gradient-coloured triangle (yellow = high expression, black = low expression). Each sphere represents a single sample at 3 months, spheres are coloured based on clinical response pattern: 'quick stable' (green), 'slow' (light green) and 'non-response' (red). E&F: scores for pre-treatment (E) and 3 months (F) were pooled and graphed as a percentage of the total tumours assessed in the group. Groups used were: 'responsive' (green) or 'non-responsive' (red). Responsive tumours were further split into those with either high (yellow) or low (black) immune/inflammatory related gene expression at 3 months, as per figure 49D. Colours in bars of graph refer to IHC score for leukocyte infiltration as per the legend.

3.4.4.4. Differential Gene Expression between Responsive and Non-responsive Tumours

Rank product analysis (FDR 0.01) was also used to identify significantly differentially expressed genes between responsive and non-responsive tumours at pre-treatment (n=353) and day 14 (n=369). Pathway enrichment and biological process gene ontology analysis was then used to determine enriched groups of functionally related genes. At pre-treatment and day 14 non-responsive tumours were found to express higher levels of proliferation associated genes compared with responsive tumours (figure 50). Given the strong association between proliferation and response (see section 3.4.4.2) these early differentially expressed genes represent likely candidates for predictive response marker exploration within this functional group. Responsive tumours were found to expresses higher levels of genes at pre-treatment associated with immune/inflammatory response and also at day 14, coupled with higher expression of ECM/stromal remodelling associated genes. Given the strong association between up-regulation of these functional groups at 3 months and positive response (see section 3.4.4.3), early differentially expressed genes involved in these processes may represent early markers of a positive response in some tumours. Other differentially expressed genes between responsive and non-responsive tumours were found to be involved in biological processes including cell

signalling, transcription, protein processing, metabolism and metal ion binding. A number of genes also have unknown biological function. In addition, a proportion of the genes were designated as belonging to 'minor biological processes' indicating that they were assigned a functional processes annotation which was enriched in the gene list by only 10 genes or less.

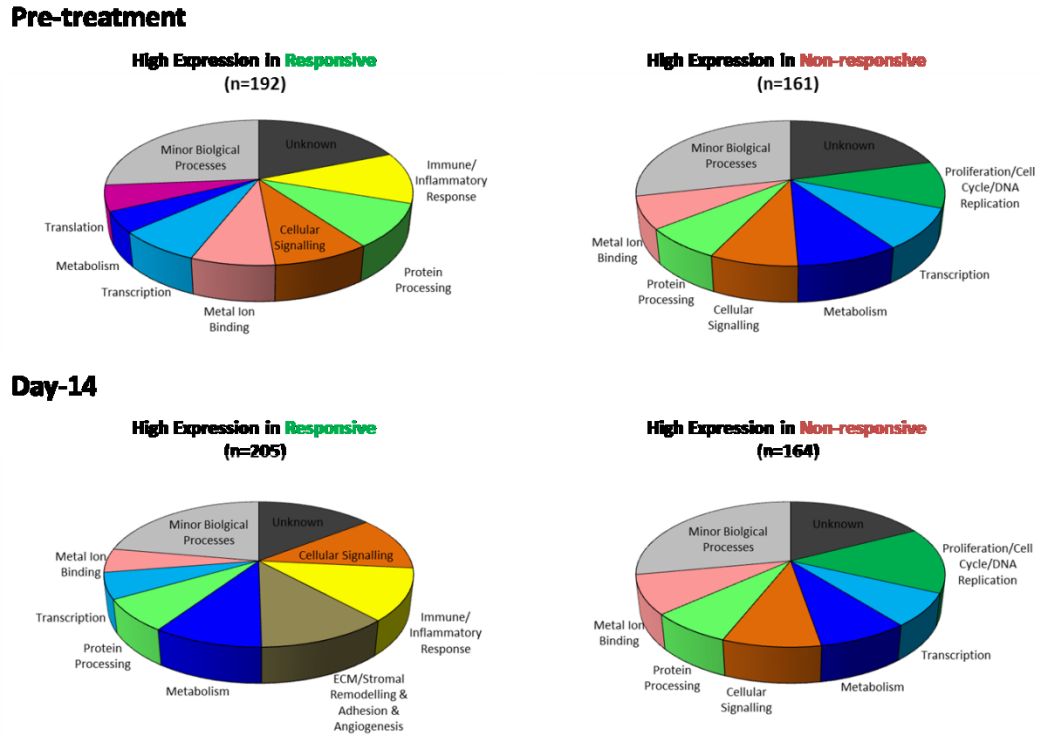


Figure 50. ***Functional groups of genes differentially expressed between responsive and non-responsive tumours at pre-treatment and 14 days****. RP analysis was carried out to identify significantly differentially expressed genes between responsive and non-responsive tumours at pre-treatment (top) and 14 days (bottom). Each gene was assigned to only one group which was the most significantly enriched functional process or pathway. Functional groups shown were represented by at least 15 genes. 'Minor biological processes' comprise processes represented by less than 15 genes are shown as one group (grey). Colours assigned to each functional group are consistent throughout all figures in this study.*

3.4.4.5. A Gene Expression Model for Early Prediction of Response

Lists of genes differentially expressed in the 'quick stable' response tumours and between responsive and non-responsive tumours at pre-treatment and at 14 days were pooled together (n=771) and filtered, removing genes with unknown function or genes designated as 'metal ion binding', giving 573 genes. A list of variables comprising pre-treatment, 14 day expression and delta (change in expression between pre-treatment and 14 days) were included for each of the 573 genes and used for random forest (RF) analysis. This multivariate procedure was used to identify the 200 most informative variables differentiating between responsive and non-responsive tumours. Of these, 82 were pre-treatment variables, 82 were day 14 variables and 36 were delta variables. The 200 most informative variables comprised 131 distinct genes (the full gene list is provided in supplementary section 6.3). Pathway enrichment and biological process gene ontology analysis of these genes was carried out and revealed significant enrichment for proliferation and immune/inflammatory associated genes as well as genes involved with transcription and metabolism. Other biological processes found to be enriched to a lesser extent include apoptosis, protein processing, cellular signalling and ECM/stromal remodelling (figure 51).

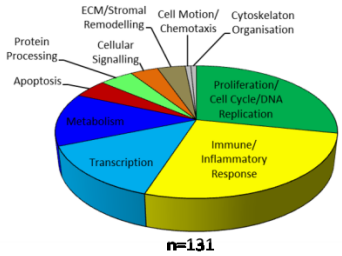


Figure 51. ***Functional enrichment of the 131 most informative genes differentiating between responsive and non-responsive tumours****. RF was used to identify the 200 most informative variables, comprising 131 distinct genes, differentiating between responsive and non-responsive tumours. Each gene was assigned to only one group which was the most significantly enriched functional process or pathway. Colours assigned to each functional group are consistent throughout all figures in this study.*

The 200 most informative variables were then used as the input for classification and regression tree (CART) modelling. The optimal model generated had 8 terminal nodes and was subsequently pruned to reduce the error rate (figure 52).

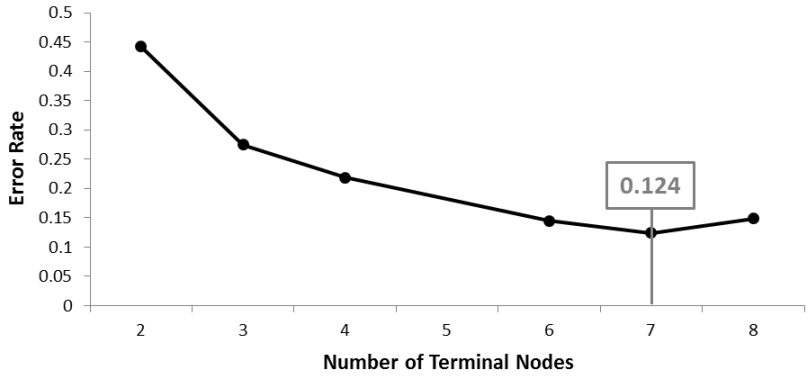


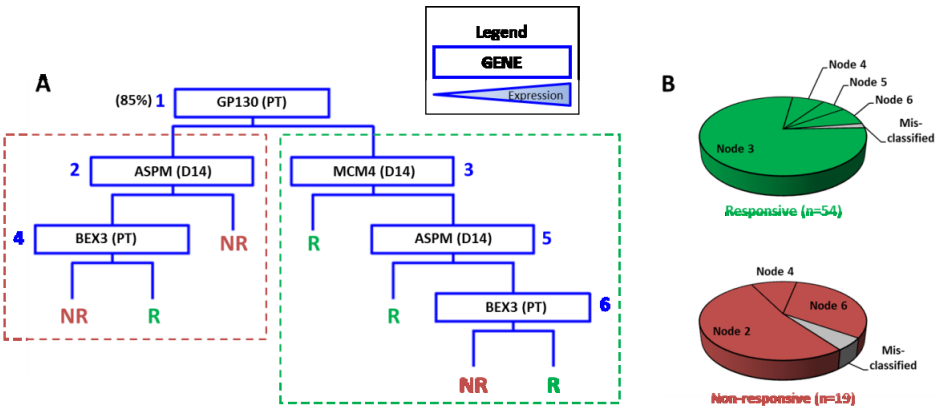
Figure 52. ***Plot showing the association between error rate and the number of terminal nodes in the CART model****. The CART algorithm begins by attempting to divide the entire dataset into two segments using one variable, in the process every eligible variable is examined for splitting power resulting in partitioning of the data using the best performing predictor. Splitting of the data at each node continues resulting in growth of a maximal tree (a tree which cannot be grown any further). An error rate is defined for the maximal tree and for each sub-tree based on the misclassification rate derived from random internal testing. The maximal tree was pruned from 8 terminal nodes to 7 in order to reduce the error rate.*

The final model (figures 53-55) comprised 7 terminal nodes using expression values from 4 genes (figure 53): pre-treatment expression of the immune related gene IL6ST (GP130) and the apoptosis induction related gene NGFRAP1 (BEX3); and day 14 expression of 2 proliferation associated genes: ASPM (mitotic spindle associated) and MCM4 (DNA replication associated). The CART model calculates expression value cut-offs or splitters for each node (figure 54D). These determine the path taken by individual subjects through the classification tree. Subjects propagate down to terminal nodes (leafs) based on their gene expression profile resulting in a classification.

Ensmebl Gene ID	Time Point Measurement	HGNC Symbol	Anonyms	Functional Cateogry
ENSG00000134352	Pre-treatment	IL6ST	GP130, CD130	Immune
ENSG00000066279	Day 14	ASPM	ASP	Proliferation
ENSG00000104738	Day 14	MCM4	CDC21	Proliferation
ENSG00000166681	Pre-treatment	NGFRAP1	BEX3	Apoptosis

Figure 53. ***Table showing details of the 4 genes used in the CART model****. The CART model uses pre-treatment expression of 2 genes: the immune related gene GP130 and the apoptosis related gene BEX3, and day 14 expression of two proliferation associated genes: ASPM and MCM4.*

The 4 gene model uses pre-treatment expression of GP130 as a primary splitter which has prediction accuracy itself of 85% (figure 54A). Tumours with low expression of GP130 are strongly associated with non-response and are further classified by day 14 expression of ASPM and pre-treatment expression of BEX3. Conversely, tumours with high expression of GP130 are strongly associated with positive response and are further classified by day 14 expression of MCM4 and ASPM and pre-treatment expression of BEX3. After primary node splitting, node 2 (ASPM) represents the primary predictive terminal node for non-response, accurately identifying 53% of the non-responsive tumours (figure 54B), whereas, node 3 (MCM4) represents the primary predictive terminal node for positive response, accurately identifying 78% of responsive tumours (figure 54B). Overall, the model was found to have 96% accuracy, 89% sensitivity, 98% specificity, a positive predictive value (PPV) of 89%, a negative predictive value (NPV) of 96% and an area under the receiver operating characteristic AUC (ROC) curve of 0.96 units^2 in the test set (letrozole data), only failing to predict response accurately in 3 of 73 patients (figure 54C).



C

Accuracy	Sensitivity	Specificity	PPV	NPV	AUC (ROC)
0.96	0.89	0.98	0.89	0.96	0.96

D

Node Number	Gene ID	Time Point	Log2 Splitter Value
1	GP130	Pre-treatment	7.20
2	ASPM	Day 14	5.87
3	BEX3	Pre-treatment	7.70
4	MCM4	Day 14	6.88
5	ASPM	Day 14	5.83
6	BEX3	Pre-treatment	8.44

Figure 54. ***The 4 gene CART model****. A: decision tree diagram showing the hierarchical structure of the model and relationship between the constituent components. Each node is uniquely numbered 1 to 6 (blue numbers) and represents the expression of a single gene at pre-treatment (PT) or 14 days (D14). From each node the path through the model is determined by a CART algorithm calculated splitter value based on log2 expression of the given gene. For each node the branch to the right indicates an expression level above the model defined splitter value for the given gene and the branch to the left indicates an expression level below the splitter value (as per the legend). The primary splitter is GP130 which has prediction accuracy itself of 85%; expression levels below the GP130 splitter value (left branch) are associated with non-response (red dotted-line box) and expression levels above the splitter (right branch) are associated with good response (green dotted-line box). Terminal nodes are represented by their association with a particular response group: responsive (green 'R') and non-responsive (red 'NR'). B: pie charts showing the proportion of responsive (green) and non-responsive (red) patients correctly classified at each node in the decision tree. Node numbers correspond to blue node numbers on the decision tree diagram (A). C: classification test statistics calculated for the 4 gene CART model using the training set (letrozole data): accuracy, sensitivity, specificity, PPV, NPV and AUC (ROC). D: table of CART algorithm calculated splitter values for each gene at each node.*

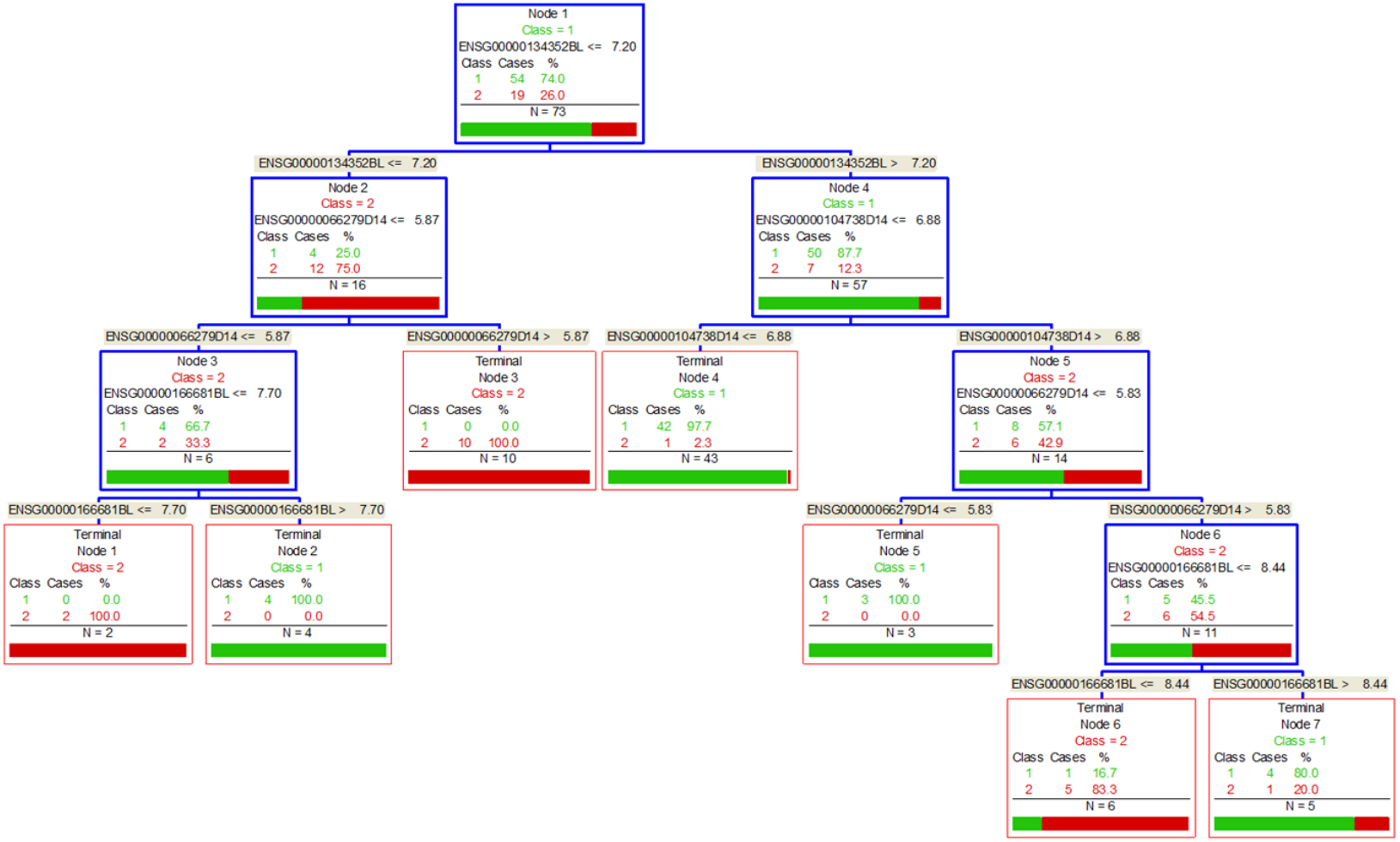

Node 1
Class = 1
ENSG00000134352BL <= 7.20
Class Cases %
1 54 74.0
2 19 26.0
N = 73
ENSG00000134352BL <= 7.20
Node 2
Class = 2
ENSG00000066279D14 <= 5.87
Class Cases %
1 4 25.0
2 12 75.0
N = 16
ENSG00000134352BL > 7.20
Node 4
Class = 1
ENSG00000104738D14 <= 6.88
Class Cases %
1 50 87.7
2 7 12.3
N = 57
ENSG00000066279D14 <= 5.87
Node 3
Class = 2
ENSG00000166681BL <= 7.70
Class Cases %
1 4 66.7
2 2 33.3
N = 6
ENSG00000066279D14 > 5.87
Terminal
Node 3
Class = 2
Class Cases %
1 0 0.0
2 10 100.0
N = 10
ENSG00000104738D14 <= 6.88
Terminal
Node 4
Class = 1
Class Cases %
1 42 97.7
2 1 2.3
N = 43
ENSG00000104738D14 > 6.88
Node 5
Class = 2
ENSG00000066279D14 <= 5.83
Class Cases %
1 8 57.1
2 6 42.9
N = 14
ENSG00000166681BL <= 7.70
Terminal
Node 1
Class = 2
Class Cases %
1 0 0.0
2 2 100.0
N = 2
ENSG00000166681BL > 7.70
Terminal
Node 2
Class = 1
Class Cases %
1 4 100.0
2 0 0.0
N = 4
ENSG00000066279D14 <= 5.83
Terminal
Node 5
Class = 1
Class Cases %
1 3 100.0
2 0 0.0
N = 3
ENSG00000066279D14 > 5.83
Node 6
Class = 2
ENSG00000166681BL <= 8.44
Class Cases %
1 5 45.5
2 6 54.5
N = 11
ENSG00000166681BL <= 8.44
Terminal
Node 6
Class = 2
Class Cases %
1 1 16.7
2 5 83.3
N = 6
ENSG00000166681BL > 8.44
Terminal
Node 7
Class = 1
Class Cases %
1 4 80.0
2 1 20.0
N = 5

Figure 55. ***The 4 gene CART model (full decision tree)****. Full decision tree diagram showing the hierarchical structure of the model and relationship between the constituent components. Each node is numbered, parent nodes are outlined in blue and terminal nodes in red. Each node shows the number of cases and proportion therein of responsive (class 1; green) and non-responsive (class 2; red) patients partitioned by each splitter at each node. Splitter values are given for each node next to Ensembl gene IDs representing the 4 genes in the model suffixed with 'BL' for baseline/pre-treatment expression or 'D14' for 14 day expression.*

Further analysis of the 4 gene CART model using the gene expression profiles of each tumour showed that higher GP130 expression at pre-treatment is strongly associated with a positive response to therapy. In addition, higher day 14 expression of proliferation associated genes ASPM and MCM4 is strongly associated with non-response and higher pre-treatment expression of the apoptosis induction gene BEX3 is associated with a positive response to therapy (figure 56).

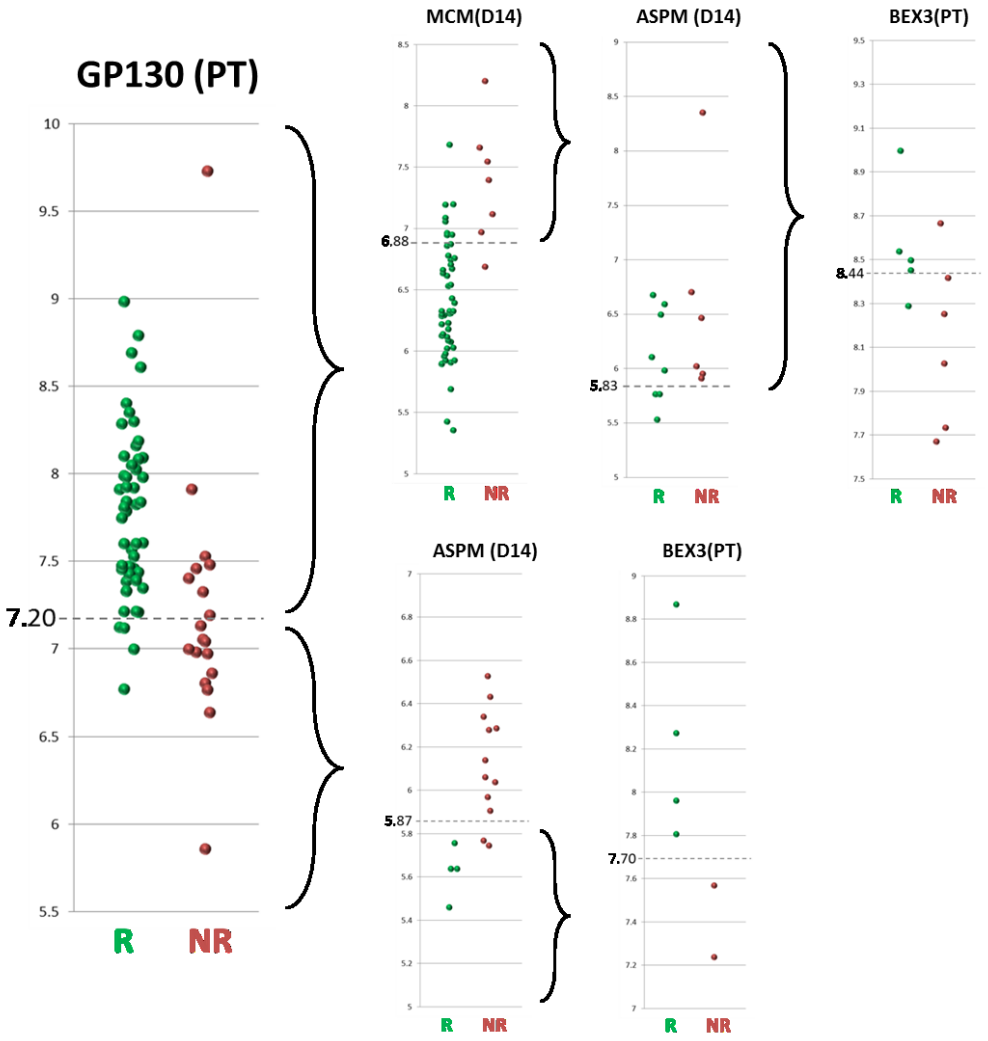


Figure 56. ***Distribution of expression variables and association with partitioning in the 4 gene CART model****. Scatter diagrams showing the distribution of log2 expression for each of the four genes in respect of the responsive (green) and non-responsive (red) patients partitioned at each node in the model. Each graph represents a node in the model. Splitter values are shown for each gene (dotted-lines) and brackets denote branches in the model. Only patients within a given bracket are partitioned at the next node which in the model indicated by the direction of the bracket. Terminal nodes have no brackets. PT=pre-treatment expression and D14= 14 day expression.*

Mean gene expression profiles for each of the 4 genes were compared in responsive ('quick stable' and 'slow' response tumours independently) and non-responsive tumours over time (expression at pre-treatment, day 14 and 3 months) (figure 57). Expression

profiles of ‘quick stable’ and ‘slow’ response tumours were found to be very similar with no significant differences in expression at any of the three time points for all 4 genes. Mean expression of GP130 and BEX3 was found to be higher across the time points in responsive tumours compared with non-responsive whereas the converse was true for the proliferation associated genes: ASPM and MCM4. Individual analysis at each of the time points revealed that GP130 has significantly lower expression at pre-treatment, day 14 and 3 months in non-responsive tumours compared with ‘quick stable’ response tumours ($P=0.009$, 0.001 and 0.0005 respectively) and ‘slow’ response tumours ($P=0.0447$, 0.0124 and 0.0719 respectively). Differences at pre-treatment in expression of ASPM between non-responsive tumours and both subsets of responsive tumours were not significant. Differences at day 14 and 3 months were significant between non-responsive and ‘quick stable’ tumours ($P=0.081$ and 0.008 respectively) however comparisons between non-responsive and ‘slow’ tumours failed to reach significance. Differences in expression of MCM4 between non-responsive and responsive tumours were only significant at 14 days ($P<0.0001$ for both). BEX3 expression was significantly different between non-responsive and ‘quick stable’ response tumours at pre-treatment, 14-days and at 3 months ($P=0.0029$, 0.0217 and 0.001 respectively). Comparison of expression between non-responsive and ‘slow’ response tumours was only significant at pre-treatment ($P=0.0071$) and 3 months ($P=0.0069$).

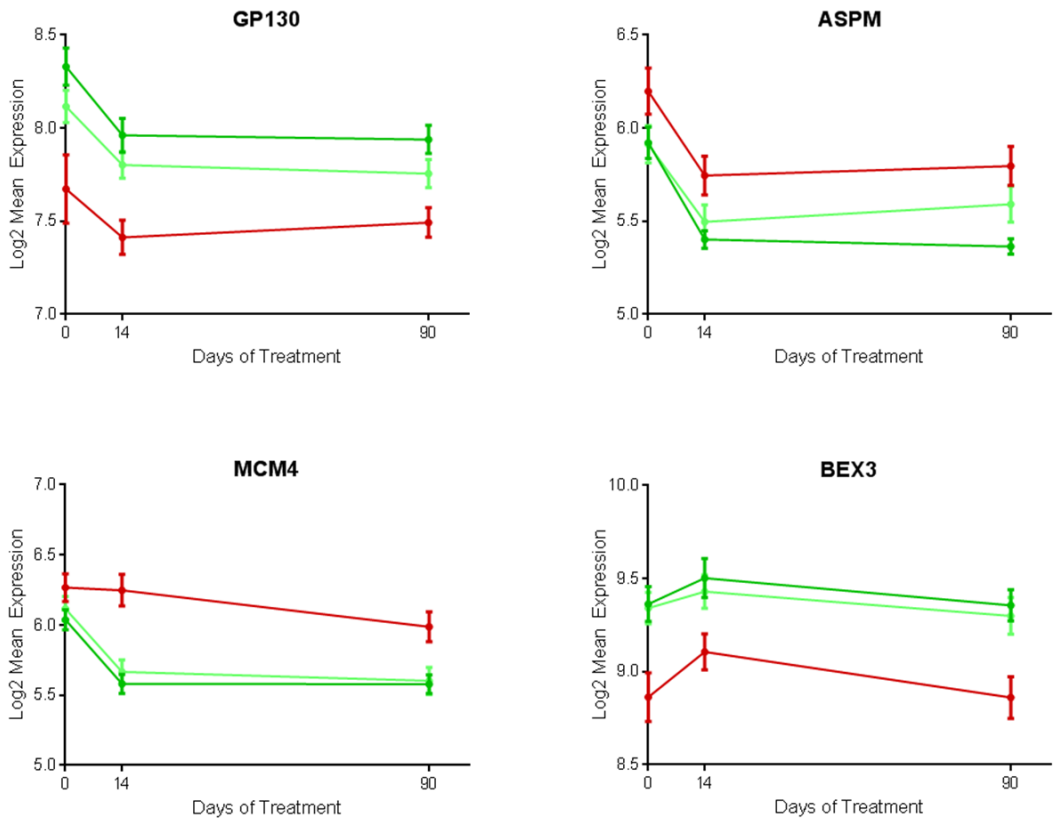


Figure 57. ***Mean gene expression profiles over time in responsive and non-responsive tumours****. Mean log2 gene expression values were calculated for each of the 4 genes at each time-point: pre-treatment (0), day 14 (14) and 3 months (90) for all patients in each response pattern subgroup: 'quick stable' (green), 'slow' (light green) and 'non-response' (red). Error bars denote standard error of the mean.*

The 4 gene CART model was able to classify all tumours, including those with 'intermediate' response patterns (previously excluded: see section 3.4.2.1), as either: 'responsive' characterised by a dramatic reduction in tumour volume by 3 months of therapy or 'non-responsive' characterised by significantly less or no change in tumour volume over 3 months (figure 58A). Progression free survival (PFS) was analysed for all tumours, based on model predicted classifications. A progression event was defined by any or all of the following criteria: increase in tumour volume greater than or equal to 20% above previous ultrasound measurement. Kaplan-Meier survival analysis (figure 58B) revealed significantly improved progression free survival in tumours classified by

the model as 'responsive' compared with those classified as 'non-responsive' (P<0.0001).The logrank hazard ratio was calculated to be 0.196 (95% CI: 0.037 to 0.286), taking the reciprocal (1/0.196 = 5.516) suggests that the estimated hazard (disease progression event) function is around 5.5 times more for patients classified as non-responsive by the model.

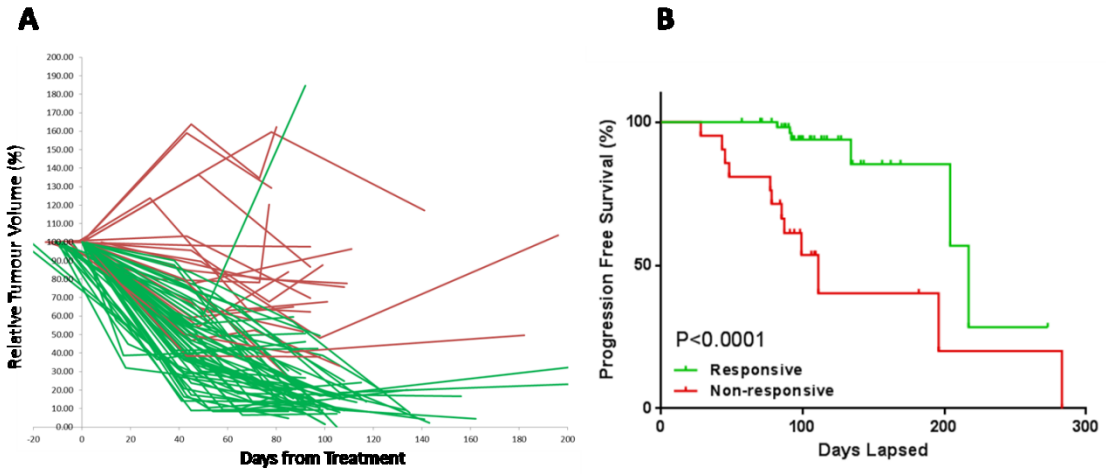


Figure 58. ***Predicted clinical response profiles and progression free survival****. A: line graph of relative (%) changes in tumour volume assessed periodically over the treatment period. Each line represents one patient. Colours indicated predicted overall clinical response: responsive (green), non-responsive (red). B: curve showing Kaplan-Meier progression free survival analysis based on response classifications predicted using the 4 gene CART model.*

3.4.4.6. Acquired Response/Non-response

The 4 gene CART model was applied to assess response in tumours with transient patterns of clinical response defined as 'acquired response' or 'acquired non-response', which had previously been excluded from the analysis (see sections 3.4.2.1 and 3.4.2.2). Interestingly, the model predicts all but one of the tumours to be non-responsive. The responsive tumour was classified as an early non-responder and despite have increasing in tumour volume by 30% above pre-treatment at day 50, by 3 months of therapy had a complete clinical and pathological response with disappearance of all identifiable tumour, suggesting that this tumour may be more akin to the 'slow' response tumours

rather than those with acquired response. The remaining 'acquired response' tumours may be inherently non-responsive to endocrine therapy and potentially undergo late molecular changes, out with the 14 day window for prediction, which result in eventual clinical response. The tumours with patterns of clinical response defined as 'acquired non-response' are characterised by a reduction in tumour volume from pre-treatment by around day 45 followed by a dramatic increase in volume by 3 months. Based on overall clinical response (response by 3 months) these tumours are accurately identified as non-responsive using the 4 gene CART model potentially suggesting that despite an early clinical response to therapy, the molecular profiles of these tumours at pre-treatment and 14 days are characteristic and predictive of overall non-response to therapy (figure 59).

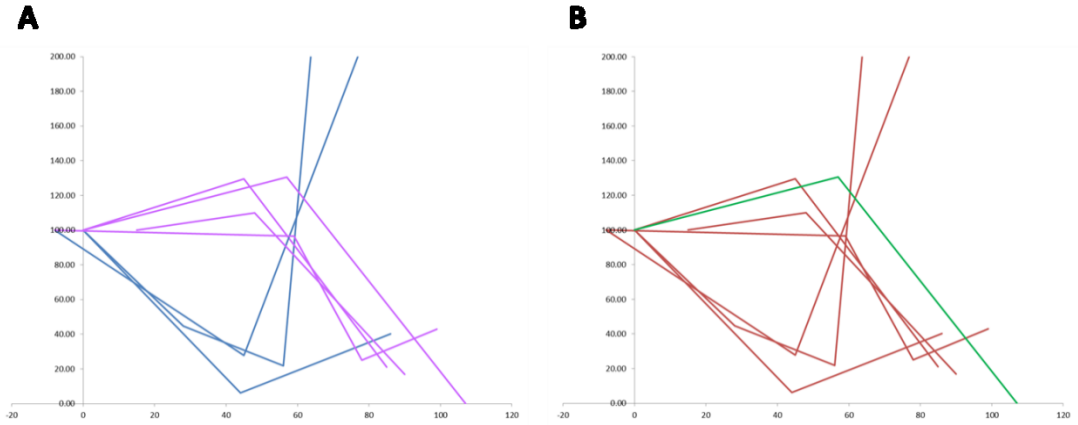


Figure 59. ***Profile and prediction of tumours with 'acquire response/non-response'.*** *A&B: line graphs of tumours with 'acquired response/non-response' patterns of clinical response showing relative (%) changes in tumour volume assessed periodically over the treatment period. Each line represents one patient. A: coloured by clinical response pattern: 'acquired response' (purple) and 'acquired non-response' (blue). B: coloured by predicted overall clinical response using the 4 gene CART model: responsive (green) and non-responsive (red).*

3.4.4.7. Independent Validation of Predictive Model

In order to independently validate the model, data was derived from the anastrozole-only arm of an independent clinical study in which patients (n=44) with primary breast cancer were given neoadjuvant anastrozole for 16 weeks. Tumour tissue core biopsies were

taken at pre-treatment and 14 days and response was assessed by changes in tumour volume using the modified UICC/WHO criteria. For the purposes of validation, patients IDs were coded for blind prediction with clinical response data held by an independent collaborator.

As the CART model is based on model-calculated expression cut-off values of the 4 genes it was necessary to XPN correct the training (letrozole data) and test set (anastrozole data) to remove batch and platform effects between the datasets (see sections 2 and 3.4.1) (figure 60).

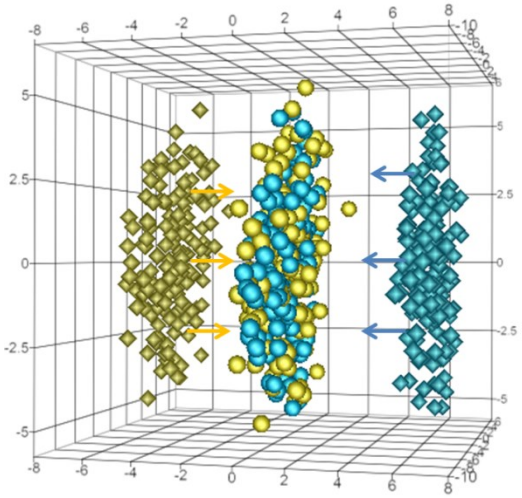


Figure 60. ***Integration and XPN correction of letrozole and anastrozole datasets****. 3D multidimensional scaling (MDS) plot of all samples from the letrozole dataset (Affymetrix and Illumina combined dataset; this study) (yellow) and anastrozole dataset (blue) based on all genes, before (left and right) and after (centre) XPN correction.*

The XPN-corrected letrozole training set was then used to re-compute expression cut-off values (splitters) for each of the genes which were then applied to predict response in the anastrozole test set. Prediction in the letrozole training set maintained 96% accuracy following XPN-correction. Each tumour in the anastrozole test set was classified as either responsive or non-responsive using the 4 gene CART model and results were sent to the collaborators who reported the prediction accuracy of the model to be 91%, correctly identifying 28 of 29 (97%) responsive tumours and 12 of 15 non-responsive

tumours (80%) (figure 61A). Sensitivity and specificity of the model in the test dataset were calculated to be 80% and 97% respectively and PPV and NPV were found to be 92% and 90% respectively (figure 61B). Interestingly, prediction accuracy based on expression of GP130 alone was found to be 82%.

A

Model	4 Gene CART Model	
Dataset	Letrozole	Anastrozole
Training/Test	Training	Test
Responsive	53/54 (98%)	28/29 (97%)
Non-responsive	17/19 (89%)	12/15 (80%)
Total	70/73 (96%)	41/44 (91%)

B

Accuracy	Sensitivity	Specificity	PPV	NPV
0.91	0.80	0.97	0.92	0.90

Figure 61. ***Table of 4 gene CART model prediction results in training and test datasets****. A: total prediction accuracy and accuracy within both response groups: responsive (green) and non-responsive (red) were determined within the training dataset (letrozole; yellow) and the test dataset (anastrozole: blue). B: classification test statistics calculated for the 4 gene CART model using the test data (anastrozole data): accuracy, sensitivity, specificity, PPV and NPV.*

Gene expression profiles for each of the 4 genes were also compared in all clinically responsive and non-responsive tumours at time-points relevant to the 4 gene CART model in the anastrozole test dataset (figure 62). Results were similar to those found in the letrozole training dataset with pre-treatment expression of GP130 and BEX3 found to be significantly higher in responsive tumours compared with non-responsive tumours (P=0.0029 and 0.029 respectively). In addition, day 14 expression of ASPM and MCM4 was found to be significantly higher in non-responsive compared with responsive tumours (P=0.0154 and 0.0092 respectively), consistent with the findings from the letrozole training set.

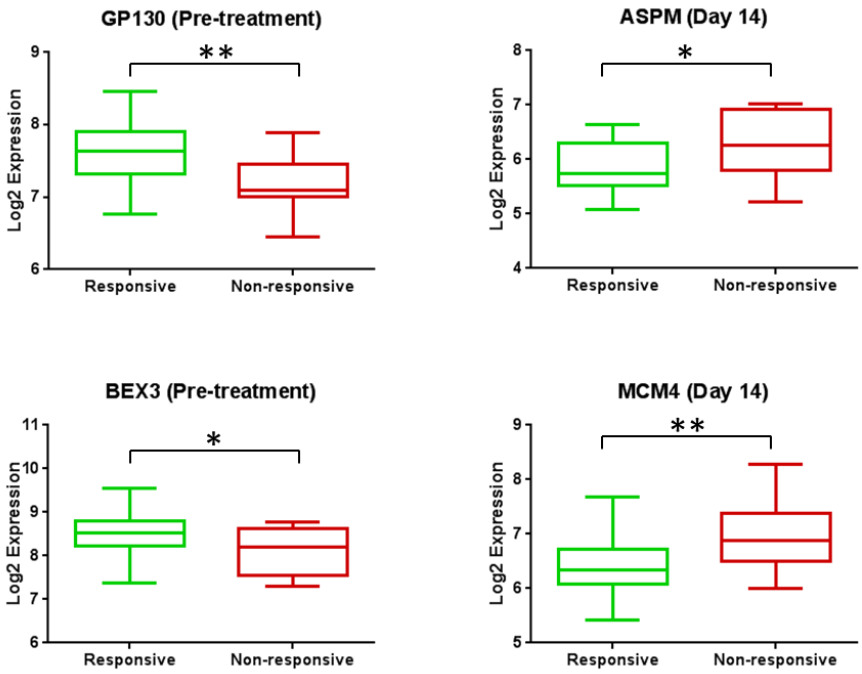


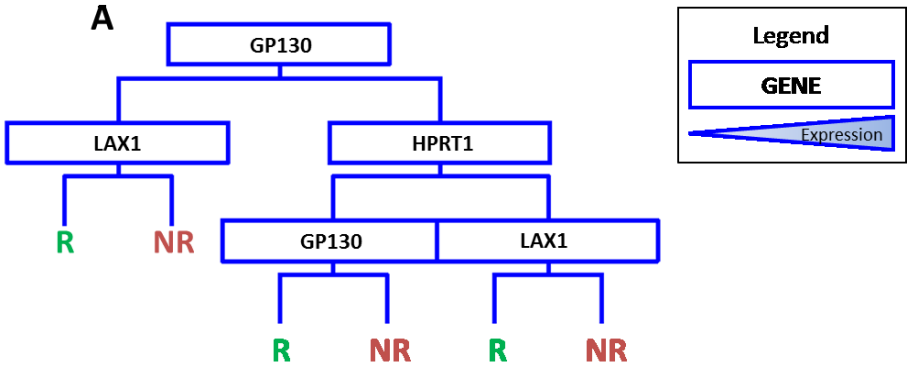
Figure 62. ***Comparing expression of the 4 CART model genes between responsive and non-responsive tumours in the anastrozole test set at relevant time points.*** *Boxplots comparing log2 gene expression of the 4 genes at either pre-treatment (GP130 and BEX3) or 14 days (ASPM and MCM4) between clinically responsive (green) or non-responsive (red) tumours.* * = P<0.05, * = P<0.01.*

3.4.4.8. The Clinical Need for a 14 Day Biopsy

The 4 gene CART model is dependent on expression of 2 genes at pre-treatment and 2 genes at 14 days and therefore as well as the routinely performed pre-treatment biopsy also requires a 14 day biopsy to be performed. Clinically, a predictive model based solely on pre-treatment gene expression would be preferable, dispensing with the need for this second biopsy, if suitably high prediction accuracy could be achieved. To assess this, 82 pre-treatment variables derived from the random forest list of the 200 most informative variables differentiation between pre-treatment and 14 days were used as input to CART. The optimal CART model had 6 terminal nodes comprising pre-treatment expression of 3 genes (figure 63). As with the 4 gene CART model (see section 3.4.4.5) the primary splitter is GP130. The other genes included in the model are

the immune cell signalling associated gene LAX1 and the HPRT1 gene which plays an essential role in purine metabolism via the purine salvage pathway where it catalyses conversion of hypoxanthine and guanine to inosine monophosphate and guanosine monophosphate respectively [438].

A



B

Ensmebl Gene ID	HGNC Symbol	Anonyms	Functional Cateogry
ENSG00000134352	IL6ST	GP130, CD130	Immune
ENSG00000122188	LAX1	LAX	Immune
ENSG00000166681	HPRT1	HPRT/HGPRT	Metabolism

Figure 63. ***The 3 gene pre-treatment only CART model.*** *A: decision tree diagram showing the hierarchical structure of the model and relationship between the constituent components. Each node represents the expression of a single gene at pre-treatment. From each node the path through the model is determined by a CART algorithm calculated splitter value based on log2 expression of the given gene. For each node the branch to the right indicates an expression level above the model defined splitter value for the given gene and the branch to the left indicates an expression level below the splitter value (as per the legend). Terminal nodes are represented by their association with a particular response group: responsive (green 'R') and non-responsive (red 'NR'). B: Table showing details and functional association of the 3 genes used in the pre-treatment only CART model.*

Predication accuracy with the 3 gene pre-treatment CART model was 93% in the training set (letrozole data) however was shown to perform less well (84% accuracy) than the 4 gene pre-treatment and 14-day CART model (91% accuracy) in the test set (anastrozole data) (figure 64A). Interestingly, the PPV was found to be significantly

lower in both training (76%) and test (63%) datasets compared with the pre-treatment and 14 day model (PPV = 89% and 92% respectively).

A

Model	Pre-treatment Only	
Dataset	Letrozole	Anastrozole
Training/Test	Training	Test
Responsive	49/54 (91%)	22/29 (76%)
Non-responsive	19/19 (100%)	12/15 (80%)
Total	68/73 (93%)	37/44 (84%)

B **Letrozole Data**

Accuracy	Sensitivity	Specificity	PPV	NPV	AUC (ROC)
0.93	1.00	0.89	0.76	1.00	0.97

C **Anastrozole Data**

Accuracy	Sensitivity	Specificity	PPV	NPV
0.84	0.80	0.76	0.63	0.88

Figure 64. ***Table of prediction results for 3 gene pre-treatment only CART model in training and test datasets****. Total prediction accuracy and accuracy within both response groups: responsive (green) and non-responsive (red) were determined within the training dataset (letrozole; yellow) and the test dataset (anastrozole: blue). B&C: classification test statistics calculated for the pre-treatment only model in training (B) and test (C) datasets: accuracy, sensitivity, specificity, PPV, NPV and AUC (ROC).*

3.4.4.9. Improving on Previous Models

A study by Harvell *et al* in 2008 published a gene expression signature comprising 50 genes [368], later refined to 25 pre-treatment genes [367], derived from gene expression microarray analysis of pre- and on-treatment samples from 6 patients treated neoadjuvantly with either: exemestane or exemestane plus tamoxifen for 4 months. Of the 6 patients, 3 were defined as responsive and 3 as non-responsive based on clinical response assessed by physical exam, mammography, ultrasound or MRI. Clinical response categories used were: complete response (disappearance of all detectable disease), partial response (longest dimension decrease by ≥ 30%), marginal response (longest dimension decreased by 20-29%), stable disease (no significant change in tumour size) and progressive disease (increase in tumour volume ≥ 20%). In their study patients with complete, partial or marginal response were classed as 'responsive' and

those with stable or progressive disease were defined as 'non-responsive'. In the combined letrozole dataset (this study), 20 of the genes in the 25 gene signature (figure 65) were represented. Pearson correlation analysis was carried out between the 20 pre-treatment genes and centroids in which the expression of each gene was defined as either -1 (lower expression) or 1 (higher expression) based on the association between expression and clinical response reported in the original study (figures 65B and C). Samples from the current study were then ranked in order of Pearson correlation to the most highly correlated centroid (responsive or non-responsive). The analysis revealed poor association between expression pattern of the 20 genes and clinical response. Prediction accuracy was calculated to be only 64%.

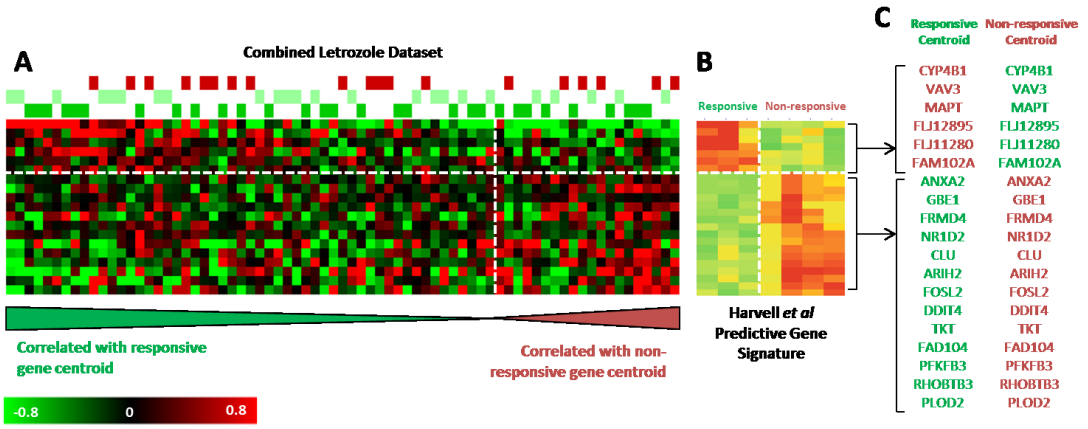


Figure 65. ***Assessment of a previously reported predictive signature by Harvell et al (2008).*** *A-C: heatmap of the combined letrozole dataset (this study) showing expression of 20 genes from a 25 gene signature identified as being predictive of clinical response to neoadjuvant endocrine therapy in a published study [367]. Samples were ranked by highest Pearson correlation to either responsive (left to right; as per green triangle under heatmap) or non-responsive (right to left; as per red triangle under heatmap) centroids (C). C: genes in centroids were assigned either -1 (green; lower expression) or 1 (red; higher expression) based on the association between expression and clinical response in the original study (B). A: coloured bars above the heatmap represent clinical response patterns: 'quick stable' (green), 'slow', (light green) and 'non-response' (red). Log2 expression was gene mean centered across all samples and heatmap colours represent relative differences in expression with red denoting higher expression and green lower expression. B: heatmap adapted from Harvell et al (2008) [367] showing the association between relative expression of the 20 genes and clinical response. Colours range from red to green through yellow with red denoting higher expression and green denoting lower expression.*

The most closely related to the current study was conducted by Miller *et al* [124] in 2009. They published a gene expression signature reported to differentiate between responsive and non-responsive tumours comprising 205 variables: 69 pre-treatment, 45 day 14 and 91 change be day 14 variables. In the original study the signature was derived from the Affymetrix-only part of our combined letrozole dataset and therefore could not be independently tested in this combined dataset. Instead it was tested in the Illumina-only part of the dataset. Testing was carried out using logistic regression (LR) analysis and centroid classification (CC). Centroids were calculated for clinically

responsive and non-responsive patients based on expression profiles of the 205 variables in the Affymetrix-only part of the letrozole data, as in the original study. The predictive accuracy of the 205 expression variable signature was found to be 91% with CC and 100% with LR analysis in the original training set (Affymetrix data) but only 53% with CC and 43% with LR analysis when applied to the Illumina-only part of the dataset (figure 66).

A

Dataset	Letrozole Affymetrix	Letrozole Illumina	Letrozole Affymetrix	Letrozole Illumina
Training/Test	Training	Test	Training	Test
Classification Type	Centroid Classification		Logistic Regression	
Responsive	2/33 (94%)	12/21 (43%)	33/33 (100%)	13/21 (38%)
Non-responsive	2/10 (80%)	2/9 (78%)	10/10 (100%	4/9 (56%)
Total	4/43 (91%)	14/30 (53%)	43/43 (100%)	17/30 (43%)

B

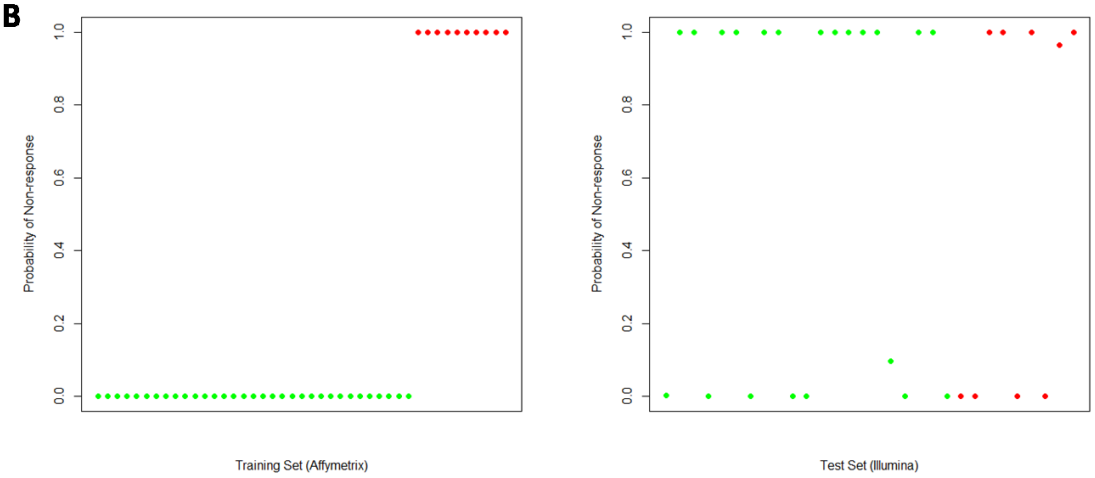


Figure 66. ***Validation of a previously reported predictive signature by Miller et al (2009) based on expression of 205 variables****. A: table of prediction results using the 205 variable model in training (Affymetrix; dark green) and test (Illumina; dark blue) datasets. Total prediction accuracy and accuracy within both response groups: responsive (green) and non-responsive (red) were determined from CC validation and LR analysis. B: logistic regression modelling of the relationship between responsive (green) and non-responsive (red) samples and probability of non-response (constrained between 0.0 and 1.0) for the training set (Affymetrix; left) and test set (Illumina; right).*

3.5. Discussion

In hormone sensitive breast cancer, aromatase inhibitors such as letrozole and anastrozole have an established therapeutic role, particularly in post-menopausal women [118, 121-123, 421]. However, a number of patients respond less well to therapy and therefore, in a move towards better patient care and treatment, there is an important clinical need to identify early-on-treatment biomarkers which predict for response/non-response to endocrine therapy and which outperform currently used parameters for response assessment and are translational with the potential for routine clinical use.

3.5.1. Currently Used Measures and Biomarkers of Response

3.5.1.1. ER as a Biomarker

The most widely used biomarker to determine suitability of hormone therapy in breast cancer is estrogen receptor alpha (ER). However, within the ER^+ population response rates to anti-estrogen therapies are only 50-70% in the neoadjuvant setting and lower in advanced disease [124, 304, 425]. As a result, there is a need to elucidate biomarkers which have the power to accurately identify those patients within the ER-positive population who will respond less well to endocrine therapy.

3.5.1.2. Early Clinical Response

Analysis of clinical response based on dynamic changes in 3D ultrasound measurements over the treatment period revealed heterogeneous patterns of response. Within the population of patients who responded well to therapy there were two subgroups; the first defined as ‘quick stable’ response patients were characterised by significantly reduced tumour volumes by around day 45 of treatment and continued positive response thereafter. The second responsive group denoted as ‘slow’ response patients responded as well as the ‘quick stable’ patients by 3 months of therapy with significant reductions in tumour volume. However, by around day 45 they showed little change akin to the subgroup of ‘non-response’ patients. The latter group was characterised by little or no change or an increase in tumour volume by around day 45 and at 3 months. The

similarity in clinical response between the 'non-response' and 'slow' response tumours by day 45 precludes clinical response, based on early changes in tumour volume or size, from being an accurate predictor of overall response to therapy. Three additional subgroups were defined to characterise distinct patterns of clinical response. The 'intermediate' response subgroup clearly gained some benefit from the therapy however tumour volume reductions were not as dramatic as with the 'quick stable' and 'slow' response tumours. The 'acquired response/non-response' tumours represented only a small proportion of the patients in the study and these were characterised by sudden and dramatic changes or reversals in response which occur late in treatment and involve molecular changes that occur out-with the window of early molecular marker detection. For the objective of designing a molecular classifier of response only 'quick stable' and 'slow' response tumours (collectively referred to as the 'responsive' tumours) and the 'non-responsive' tumours were used, thus creating two distinct well-separated groups, each with clearly defined clinical response criteria.

3.5.1.3. Early Pathological Response

Pathological response by 14 days was found to be poorly associated with clinical response. Approximately half of both the clinically responsive and non-responsive tumours have a reduction in tumour cellularity by 14 days while the remaining half of both groups either do not change or have an increase in cellularity. In terms of changes in grade, the majority of both responsive and non-responsive patients do not change by 14 days. Indeed, logistic regression analysis shows that neither changes in tumour cellularity nor grade were found to be significantly predictive of clinical response. Together these findings suggest that early pathological changes are poor indictors of overall clinical response to endocrine therapy.

3.5.1.4. Molecular Subtyping and Response

With the advent of molecular subtyping, ER^+ breast cancer was subdivided into either luminal A or the more aggressive luminal B form (see section 1.1.3). Luminal B breast cancer is often characterised by loss of progesterone receptor (PR) expression which is

thought to represent a surrogate for a more aggressive disease phenotype which is less dependent on estrogen signalling and has been linked to reduced ER expression, spread to lymph nodes, higher proliferation and increased expression of EGFR and HER2 [439]. Indeed, luminal B tumours have been linked to poor prognosis and lack of response to hormone therapy [439, 440]. In this study, tumours were profiled based on the Sørlie classifications for intrinsic subtypes [25]. Of the 73 patients 18% were classified as luminal B and of these 62% were non-responsive. Indeed, within the non-responsive population only 42% of tumours were classified as luminal B. While the luminal B molecular profile is associated with non-response, it is not accurately predictive of response as the majority of non-responsive tumours were classified as luminal A and 38% of luminal B tumours responded well to endocrine therapy.

3.5.1.5. HER2 as a Biomarker

Over-expression of epidermal growth factor receptors, particularly HER2, have also been linked endocrine therapy resistance in breast cancer [182]. Indeed, HER2 expression has been reported to be a prognostic factor and predictor of response to HER2 targeted therapy [441]. However, the ability of HER2 expression to predict response to neoadjuvant endocrine therapy is controversial [442]. In this study, HER2 protein expression was determined immunohistochemically (see section 1.1.3) as part of routine histology at diagnosis for the majority of patients. Of the 70 patients tested, 23% were determined to be $HER2^{+}$, 62% of which were clinically non-responsive. However, within the non-responsive population only 42% of patients were determined to be HER2-positive. In logistic regression analysis HER2 positivity was found to be significantly associated with non-response however is not robustly predictive of response as 38% of HER2-positive patients responded well to endocrine therapy. Instead HER2 signalling might represent a mechanism, active in a subset of non-responsive patients, which drives proliferation in a non-estrogen dependent manner. Such patients who gain only partial or no benefit from endocrine therapy may represent promising candidates for combination or alternative therapy targeted against HER2.

3.5.2. Proposed Predictors of Response

3.5.2.1. Ki67

Uncontrolled proliferation is regarded as a hallmark of cancer and a number of studies have linked proliferation to lack of response to endocrine therapy [50, 285, 292, 443, 444]. Immunohistochemical assessment of the proportion of cells staining for the nuclear antigen Ki67 has become a widely used methodology for comparing proliferation between tumours [444]. Potential uses suggested include: prognosis, prediction of response or resistance to chemotherapy or endocrine therapy and estimation of residual risk. Ki67 has been shown to independently improve the prediction of response to chemotherapy and prognosis in a group of breast cancer patients [445] and together with HER2 status to predict pathological complete response in HER2-positive patients receiving neoadjuvant chemotherapy [446]. Interestingly, one study reported that post-treatment (chemotherapy) levels of Ki67 were more predictive of prognosis than pre-treatment levels [447]. With regard to endocrine therapy, a recent study of metastatic breast cancer reported that low Ki67 expression was associated with higher clinical benefit and longer time to progression on first-line endocrine therapy and longer survival after metastatic recurrence [448]. Another study investigated the association between early proliferation changes and clinical response to neoadjuvant letrozole. Proliferation was determined by both Ki67 immunohistochemistry and the gene expression grade index (GGI); a 97 gene algorithm designed to measure proliferation (see section 1.3.2.4.2). The study reported that low proliferation, determined by either Ki67 or the GGI, after 10-14 days of neoadjuvant letrozole was associated with a good clinical response in high grade but not low grade tumours [443]. In the current study, Ki67 was immunohistochemically determined at baseline, 14 days and 3 months in a subset of patients (n=55) as part of a previous study [292]. However, no significant associations between Ki67 expression levels or changes in expression and clinical response were found in our cohort. The potential for Ki67 as a surrogate of proliferation and predictor of response and prognosis has been demonstrated in a number of studies. However, with regard to endocrine therapy in the neoadjuvant setting, while

some studies have shown early changes in Ki67 to be positively and significantly correlated with both clinical and pathological response, they have also eluded to discordancy occurring in a number of cases, rendering it a poor predictor of overall response [292, 303]. Furthermore, recent studies have suggested that enormous variation in the analytical practice of Ki67 measurement dramatically limits its value in this context [444, 449].

3.5.2.2. Proliferation Gene Expression

Hierarchical cluster analysis based on the most significantly down-regulated proliferation associated genes (n=60) over 3 months of therapy in the 'quick stable' response tumours revealed poor association with response at pre-treatment or based on changed expression by 14 days. However, expression of proliferation genes at day 14 was found to be strongly associated with response, with the majority of non-responsive tumours expressing relatively high levels of proliferation genes. Despite the strong association, a number of responsive tumours also had high day 14 expression suggesting that while proliferation may constitute an important component of a predictive model, when considered independently, it is a poor predictor of overall response. Interestingly, these findings suggest that on- or post-treatment biomarkers may have more predictive power than at pre-treatment, a view supported by a recent review of biomarkers in breast cancer in preoperative studies [449].

3.5.2.3. Immune Signature and Leukocyte Infiltration

The immune system has been shown to have both tumour supressing and tumour promoting roles [450]. Cytokines such as interleukins and interferons play a critical role in controlling the immune system. They are mainly secreted by lymphocytes and macrophages and work by modulating the function of target cells in a paracrine or autocrine fashion [451]. Several studies have shown that breast tumours are heavily infiltrated by leukocytes including T-cells, B-cells and tumour-associated macrophages (TAMs) [228, 452-456]. A recent study of leukocyte composition in breast cancer reported that CD8 expressing T-cells comprise the major population of infiltrating

leukocytes in both chemotherapy treated and naïve patients. Interestingly, granzyme B expression was found to be low in the chemotherapy naïve cohort and in two thirds of the treated cohort suggesting negligible cytotoxic activity [457]. Indeed, a number of studies have suggested that despite the presence of lymphocytes, T-cell mediated anti-tumour responses are impaired in breast cancer, potentially due to the action of inhibitory cytokines in the tumour microenvironment [451-453]. Despite the complexity surrounding the role of infiltrating leukocytes in mediating tumour behaviour, the presence of lymphocytes has been reported as an independent predictor of good response to cytotoxic chemotherapy in breast cancer [348] and infiltration of CD8 expressing T-cells in particular was found to be an indicator of good clinical outcome in patients with ER^- breast cancer [458]. Contrary to this, a recent study reported that infiltrating leukocytes were associated with poor response to endocrine therapy in breast cancer. In addition, they suggested that infiltrating dendritic cells, which have been previously linked to breast tumorigenesis by a mechanism involving polarisation of CD4 expressing T-cells [459], could be involved in mediating a poor response to endocrine therapy in highly proliferative breast tumours. Interestingly, immunohistochemical staining for CD45 in this study also revealed a strong association between the presence of infiltrating leukocytes in pre-treatment tumour tissue sections and high expression of proliferation associated genes after 14 days of letrozole treatment. However, it should be noted that CD45 is expressed on the surface of most myeloid cells, lymphocytes and dendritic cells and therefore the hematopoietic lineage of the infiltrating leukocytes identified in this study remains unclear. Despite an association with high proliferation, the presence of infiltrating leukocytes was found to have a poor association with clinical response.

Analysis of gene expression in ‘quick stable’ response tumours revealed a subset characterised by dramatic and consistent up-regulation of genes involved with immune/inflammatory and ECM/stromal remodelling at 3 months. However, infiltrating leukocytes were only found in approximately half of these tumours suggesting that these gene changes were largely leukocyte-independent and may instead be a feature of the cancer cells or other infiltrating cell types such as fibroblasts. In agreement with this, a

study in 2007 reported an immune response gene expression signature which identified a good prognosis groups in ER^- breast cancer which was independent of leukocyte infiltration [36].

3.5.2.4. Predictive Gene Expression Signatures

The study by Harvell *et al* in 2008 [367, 368] represents one of the first focused on predicting response to neoadjuvant endocrine therapy using pre- and on-treatment matched biopsies in which the endpoint used was clinical response after extended therapy (4 months). They published a 50 gene signature which was later refined to a 25 gene signature reported to predict clinical response to therapy after 4 months of treatment using pre-treatment gene expression levels. In this study both the 50 and 25 gene signature were validated in the combined letrozole data using hierarchical clustering and were found to have no association with clinical response. The original study was based on expression profiles of only 6 patients: 3 responsive and 3 non-responsive. Despite differences in drug protocol and experimental design, including assessment of clinical response, which exist between the original study and this, the limited sample number is likely to impact significantly on reproducibility of their findings (as discussed in sections 1.3.2.4.5, 1.3.3.1 and 3.5.5.1).

The most closely related study to this using matched pre- and on-treatment biopsies and also using a clinical response endpoint (clinical response following 3 months of therapy) was published by Miller *et al* in 2009 [124]. They developed a gene signature based on 205 early gene expression variables (including pre-treatment, 14 days or change by 14 days) which was reported to distinguish between clinically responsive and non-responsive tumours. At the time of publication their dataset represented the largest of its type, based on expression profiles of 58 patients, significantly larger than previous efforts by Harvell *et al* [367]. At the time no other sufficiently large datasets existed for validation purposes. Therefore, the first validation of their findings was carried out as part of this study; however, their original dataset comprises the Affymetrix component of what is now the combine letrozole dataset used in this study and for that reason it was only appropriate to validate their findings in the illumina-only part dataset. As expected

their signature was found to be highly predictive in the training set (original data: Affymetrix component) in both centroid classification (CC) and logistic regression (LR) analyses despite differences in the criteria for clinical response used in this study (see sections 3.3.5.1, 3.4.2.1/2 and 3.5.5.2). However, in the test set (Illumina component) prediction of clinical response using their gene expression signature was poor. This is likely due to the analysis approaches used in the original study which led to findings which were largely dataset specific and not reproducible in independent datasets. Importantly, this study demonstrated the presence of bias in the Affymetrix dataset associated with variation in processing date (see sections 2.4.2.2 and 3.4.1) which was overlooked in the original study. The existence of such bias has been shown in several studies [369, 370, 376] to significantly impact on gene expression analysis and may be a contributory factor to the lack of reproducibility in the Miller et al [124] study.

3.5.3. The 4 Gene CART Model for Predicting Response

This study has reported a model which has been shown to accurately and robustly predict clinical response to aromatase inhibitors in the neoadjuvant setting. The model has been shown to out-perform currently used clinical and pathological parameters for response assessment and previously reported gene expression based models designed for the same purpose.

Based on the expression of only 4 genes, the model presented in this study has significantly greater translational potential compared to predictive signatures from other studies comprising several genes, the assaying of which would be impractical in the clinical setting largely due to prohibitive costs and the time consuming nature of such assays. Furthermore, the 4 genes have been shown to be functionally relevant, involved in processes central to endocrine therapy response in breast cancer: immune signalling, proliferation and apoptosis. Indeed, all 4 genes have been linked to cancer in numerous studies (see section 3.5.4). Expression levels of the 4 genes were found to be significantly different between all responsive and all non-responsive tumours over time. Importantly, despite differences in early clinical response between ‘quick stable’ and ‘slow’ response tumours, expression levels of the 4 genes were not significantly

different between these groups allowing for both to be distinguished at molecular level from the non-responsive tumours by day 14 of treatment.

Interestingly, all but one of the tumours characterised by 'acquired response/non-response' patterns of clinical response were predicted to be non-responsive. In the 'acquired response' tumours a decrease in tumour volume may involve molecular changes which occur out-with the day 14 window for early prediction and as a result these tumours are predicted to be non-responsive. In the 'acquired non-response tumours' it is possible that, despite an initial reduction in tumour volume before dramatic regrowth, early gene expression profiles remain indicative of overall clinical response. The mechanisms underlying 'acquired response/non-response' to therapy remain unclear. Furthermore, with only 8% of patients in the study having transient clinical responses the performance of the 4 gene model within this subgroup is difficult to ascertain. However, given the aberrant nature of clinical response in these patients they may nonetheless gain clinical benefit from alternative or combination therapy and thus a prediction of 'non-responsive' may be appropriate for this group.

Unlike previous models reported to predict response to endocrine therapy in the neoadjuvant setting, the model presented here has been independently tested under controlled blind conditions in a cohort of patients treated with neoadjuvant anastrozole and found to have 91% accuracy of clinical response prediction. This provides further evidence for the predictive power of the model. Indeed, compared to the previously published models which have now been validated as part of this study, the 4 gene model was shown to vastly out-perform these other models in independent validation analysis. In addition, the 4 gene model has been shown to predict response in patients treated with both letrozole and anastrozole, potentially suggesting a broader role for the model in predicting neoadjuvant clinical response to multiple single and combined anti-estrogen therapies.

3.5.4. Genes in the 4 Gene CART Model

3.5.4.1. Glycoprotein 130 (GP130)

GP130 is a transmembrane protein belonging to the class of type 1 cytokine receptors, essential for signal transduction of cytokines in the interleukin 6 (IL6) family including: IL6, IL11, oncostatin M (OSM), leukaemia inhibitory factor (LIF) and ciliary neurotropic factor (CNTF) [460]. GP130 is known to mediate numerous homeostatic functions including the immune/inflammatory response, bone metabolism and haematopoiesis [461].

GP130 has no intrinsic tyrosine kinase activity; instead it is phosphorylated on tyrosine residues after complexing with the cytokine ligand and other ligand-specific binding subunits, leading to activation of the JAK family of tyrosine kinases (JAK1, JAK2 and TYK2) which in turn stimulate multiple pathways including MAPK signalling, PI3K signalling and STAT signalling leading to transcriptional regulation of down-stream genes [460, 462]. An important target of GP130 is STAT3, a member of the STAT family of transcription factors [463, 464]. Following GP130 activation, STAT3 becomes phosphorylated on a tyrosine residue and homo- or heterodimerizes before translocating to the nucleus where it induces transcription of numerous genes involved in growth regulation, differentiation and cell cycle [463]. STAT3 is an oncogene which is constitutively activated in numerous solid tumours including breast cancer where it may act as a central mediator of malignancy. Its constitutive activation has been shown to involve epidermal growth factor (EGF) signalling and signalling through GP130 [465-470].

The cytokine ligands of GP130 have been reported to have numerous biological effects on breast cancer cells. IL6 and OSM have been shown to have both growth promoting and inhibiting effects [451, 471, 472]. They have also been shown to have a role in disrupting cell adhesion and thus increasing cell motility [473, 474]. Interestingly, IL6 has been identified as a modulator of reductive 17β-hydroxysteroid dehydrogenase type 1 activity which catalyses the inter-conversion of estrone and the biologically more

potent estradiol following the aromatisation of androgens [475]. GP130 has also been reported to activate the estrogen receptor (ESR1) in breast cancer cells [476].

GP130 has also been shown to physically and functionally interact with EGFR family members. In a study of prostate cancer GP130 was found to form an IL6-dependent complex with epidermal growth factor receptor 2 (HER2) which is required for full IL6 activity [477], a feature also found in breast cancer cells [473]. A study investigating OSM signalling in breast cancer cells also identified physical associations between both GP130 and oncostatin M receptor β and HER2 and epidermal growth factor receptor 3 (HER3) [478]. Although the molecular mechanisms and biological significance of interactions between GP130 and EGFR remain poorly understood, a recent study suggested that GP130 signalling is essential for signalling down-stream of EGFR in breast cancer cells and that blocked GP130 signalling is linked with inhibition of invasion, metastasis and angiogenesis [479].

GP130 has been reported to be expressed in varying degrees in the majority of cell lines and primary breast cancer tumours [480, 481]. A study of a cohort of patients with tumours of varying type (breast, gastrointestinal, uterine, ovarian, renal and bladder) and stage, some of whom received chemotherapy and radiotherapy, revealed that blood serum levels of GP130 were significantly elevated in cancer patients (both treated and untreated) compared with healthy control subjects. Furthermore, serum GP130 levels were found to increase significantly with tumour stage [482]. One study reported GP130 expression as a positive prognostic factor for overall survival and disease free survival in advance stage breast cancer [481]. Indeed, GP130 has also been included in the recent 12-gene EndoPredcit assay, where increased expression is correlated with a lower risk of recurrence following endocrine therapy [345-348]. In the predictive model presented here, GP130 represents the primary splitter, partitioning the tumours into either responsive or non-responsive with 85% accuracy in the training set and 81% accuracy in the test set. GP130 expression is highly correlated with response to endocrine therapy and increased expression is strongly associated with significantly improved drug efficacy.

3.5.4.2. Abnormal spindle-like microcephaly-associated protein (ASPM)

Studies in mice and embryonic neuroblasts have shown the protein product of ASPM to be essential for mitotic spindle regulation during mitosis [483, 484]. In glioblastoma cell lines ASPM gene expression was found to be significantly down-regulated in response to ionizing radiation and subsequent investigations revealed that silencing of the gene impaired DNA double-stranded break repair suggesting an important role for ASPM in efficient non-homologous end joining [485]. Another study of gene expression in a panel of glioblastomas and astrocytomas showed that increase expression of ASPM is positively and strongly correlated with proliferation and tumour malignancy [486, 487]. Indeed several studies have implicated ASPM in DNA repair, DNA replication and cell cycle arrest [488, 489]. High concentrations of the APSM protein product have been found to localise at the spindle poles between the prophase and telophase stages of mitosis [490, 491]. In a gene expression study of 97 primary breast cancers, ASPM was identified as being significantly deregulated in relation to outcome in aggressive tumours and its expression was found to correlate significantly with CCNB2 which was identified as a potential biomarker of unfavourable prognosis over short-term follow up in aggressive breast cancer [492, 493]. Interestingly, another study of breast cancer cell lines showed that knockdown of the BRCA1 gene (germline mutations in which have been shown to confer increased risk for breast cancer [494, 495]) results in down-regulation of cell cycle and proliferation associated genes including ASPM and attenuation of the mitotic spindle checkpoint in mitosis [496].

3.5.4.3. Minichromosome Maintenance Complex Component 4 (MCM4)

MCM4 encodes a highly conserved protein which is essential for DNA replication. During the initiation step of DNA replication at the origin of replication, it forms a heterohexameric protein complex with other members of the minichromosome maintenance (MCM) protein family (MCM2-7). The complex is a major component of the pre-replication complex together with other proteins including: members of the origin recognition complex family, CDC6 and CDT1. The MCM complex has ATPase activity and functions as the key component of the replicative helicase that unwinds

double-stranded DNA and drives progression of the replication fork [497-499]. Aberrant fork progression can lead to stalled forks or fork collapse resulting in incomplete DNA replication and potentially introduction of double stranded breaks [498]. Additional chromatin bound MCM complexes also occupy dormant origin regions which can be activated to complete DNA replication in the event of stalled forks [500-502].

A recent study identified MCM4 as a potential biomarker for diagnosis of cervical squamous cell carcinomas after confirming significantly higher expression in tumour compared to control tissue [503]. Another study identified MCM4 as a potential therapeutic target in non-small cell lung cancer (NSCLC) after reporting its essential role in proliferation of NSCLC cell lines and its high correlation of expression with proliferation markers Ki67 and cyclin-E [504]. A study of cutaneous melanoma also reported high correlation of expression between MCM4 and Ki67 [505]. A genetic screen for chromosome instability in mice identified a mutation in the MCM4 gene resulting in destabilisation of the MCM complex which was found to cause exclusively mammary adenocarcinomas in 80% of homozygous female test mice suggesting that such mutations in MCM4 may increase risk of breast cancer in humans [506]. Other studies in mice have identified hypomorphic alleles of MCM4 which can give rise to tumorigenesis resulting from reduced MCM4 protein levels which promotes chromosome instability through a reduction in the number of dormant MCM complex which are available to activate in the event of replicative stress [507-510]. Another group identified a dominantly acting hypomorphic MCM4 allele, the protein product of which is viable and can incorporate into the MCM complex resulting in oncogenic chromosomal abnormalities during DNA replication [511].

3.5.4.4. Nerve Growth Factor Receptor Associated Protein 1 (NGFRAP1/BEX3)

BEX3, previously identified as ‘p75NTR-associated cell death executor’ (NADE), is a member of the brain expressed X-linked (BEX) family of proteins and although its functions remain incompletely understood, it has been implicated in several studies with the induction of apoptosis [512-515]. The p75 neurotrophin receptor (P75NTR) is a

member of the tumour necrosis factor (TNF) family which can mediate both cell survival and cell death in response to nerve growth factor (NGF). It is thought that BEX3 mediates apoptosis in response to NGF by interacting with the cell death domain of P75NTR. In neural cells, the interaction of BEX3 and P75NTR in response to NGF was shown to induce both caspase 2 and caspase 3 pro-apoptotic activities and resulted in fragmentation of DNA [512-514]. Indeed, in the brain, loss of function of BEX3 has been linked to development of tuberous sclerosis complex (TSC) and other tumour pathology [516].

Studies using various human cancer cell lines and tissues revealed that BEX3 was more widely expressed than other members of the BEX family of proteins such as BEX1 and BEX2 whose expression is predominantly limited to the brain [517]. Interestingly, BEX3 was found to be highly expressed in endocrine-related organs. In breast cancer cell lines, forced expression of BEX3 was found to potently supress cellular growth by an unknown mechanism which was independent of NGF signalling [518]. One group investigated the role of BEX3 in alternative apoptosis pathways. They identified BEX3 as a binding partner of the second mitochondria-derived activator of caspase (SMAC) which has been implicated in the activation of apoptosis in response to stress. The interaction between BEX3 and SMAC was found to promote TNF-related apoptosis-inducing ligand (TRAIL)-induced apoptosis in breast cancer cells by blocking the X-linked inhibitor of apoptosis protein (XIAP)-mediated ubiquitination of SMAC and its subsequent degradation [515]. As well as a role in induction of apoptosis potentially via discrete mechanisms, BEX3 has also been directly linked in at least one study to the cellular growth cycle, in particular mitochondrial DNA replication. In the study, silencing of the BEX3 gene was reported to result in significantly reduced growth of teratocarcinoma cells [519].

Interestingly, in prostate cancer, BEX3 has been identified as a potential biomarker for diagnosis with significantly higher expression found in malignant compared to benign control prostate tissue [520]. With regard to response to anticancer therapy, a study of ovarian cancer showed significant down-regulation of BEX3 gene expression following

chemotherapy treatment [521]. Furthermore, a study of primary breast cancers revealed a novel subgroup of ERα-positive tumours with improved outcome after tamoxifen treatment which were characterised by over-expression of the BEX3 homologue: BEX2 [522]. In breast cancer, NGF inhibits C2-induced apoptosis through binding of P75NTR and subsequent NF-κB activation [523, 524]. The study reports that BEX2 expression is necessary for the inhibition of C2-induced apoptosis and that high expression of BEX2 is associated with enhanced efficacy of tamoxifen [522]. In this scenario, tamoxifen treatment reduces levels of estrogen such that BEX2 expression levels reduce reversing the inhibition of C2-induced apoptosis. Interestingly, BEX3 has also been shown to interact with P75NTR and modulate NGF signalling through activation of NF-κB [512-514, 516, 522].

3.5.5. Improving on Previous Work

3.5.5.1. Increased Sample Number and Definition of Response

One of the key advantages of this study over other previously published studies is the increased number of patients (n=89) with available pre- and on-treatment biopsies and clearly defined clinical responses based on 3 months of therapy. Previous studies have been limited by patient number. The study by Harvell *et al* [367, 368] was based on pre- and on-treatment samples from only 6 patients receiving neoadjuvant exemestane with or without tamoxifen while another study by Massarweh *et al* [525] analysed pre- and on-treatment samples from only 12 patients receiving neoadjuvant anastrozole plus fulvestrant with or without gefitinib. Studies such as these, severely limited by sample number, lack statistical significance and fail to account for the clinical and molecular heterogeneity known to exist in endocrine therapy treated breast cancer. As a result reproducibility of findings is often poor and therefore validation in an independent dataset unlikely at best.

The largest study at the start of this thesis in 2009 was that of Miller *et al* [124], in which pre- and on-treatment samples were analysed from 55 patients receiving neoadjuvant letrozole, 15 of which were classified as non-responsive to treatment using

a more relaxed criteria then the current study (only 10/15 would be classified as non-responsive using the more defined criteria in sections 3.4.2.1/2). The limited number of readily available matched pre- and on-treatment tumour tissue samples from patients with complete clinical follow-up restricts the generation of new large datasets. However, in order to significantly increase patient numbers and hence statistical significance in this study, a novel data integration approach was used to combine datasets. Data from the previous Miller *et al* [124] study and new data generated as part of this study were directly integrated using a previously described process pipeline which included XPN correction for inter-platform bias (see sections 2 and 3.4.1). This approach led to creation of the largest dataset of its type, comprising 89 patients, each with pre- and on-treatment biopsies and full clinical response follow up over 3 months. Importantly, this combined dataset comprises the largest proportion of well-defined clinically non-responsive patients (n=19).

3.5.5.2. Improved Model Design

3.5.5.2.1. Data Integration and Reproducibility of Findings

This study has demonstrated that the reproducibility of findings from previous studies by Harvell *et al* [367, 368] and Miller *et al* [124] to predict clinical response to endocrine therapy is poor. Several factors may account for this, including limited sample number and the analysis approaches used potentially leading to dataset specific findings. It is postulated that the novel data integration approaches designed and implemented in this study may have additional advantages over and above increasing sample size and statistical significance. As part of the integration pipeline, during the pre-processing of individual datasets, probes are strictly filtered and re-annotated. In so doing only those with well-designed probe sequences for the gene of interest and a low chance of non-specific hybridisation are retained. While a number of genes are lost at this step in the pipeline, those which remain are likely to have a consistent relative expression across datasets and consistent absolute expression following XPN correction in biologically similar groups of samples. It is hypothesised that predictive signatures derived from this

refined list of accurately measured and consistently expressed genes across different microarray platforms may have improved reproducibility in independent datasets. The exact nature and magnitude of the contribution of data integration and XPN correction to reproducibility remains to be elucidated and is worthy of future work.

3.5.5.2.2. CART Algorithm

The use of a classification and regression tree (CART) algorithm represents a conceptual improvement in model design in this study compared with previous studies. The study of Miller *et al* [124] identified 205 variables which, when considered together, were reported to discriminate between responsive and non-responsive tumours. Application of these variables with centroid based classification yielded poor reproducibility of prediction accuracy in an independent dataset. This may be due to the molecular heterogeneity which has been reported to exist within the non-responsive population; by considering all non-responsive tumours as one group, accuracy of prediction is likely to be highly dependent on the composition of the training set. In this study, the issue of heterogeneity within the non-responsive tumours was addressed by increasing sample number in order to better represent the population, and importantly was accounted for by the CART algorithm. Rather than splitting data into only two groups, the algorithm employs exhaustive searchers and computer-intensive testing techniques to identify variables which can partition the data into homogeneous segments, each associated with either positive response or non-response to therapy.

3.5.6. Clinical Application

3.5.6.1. Clinical Need for an Early On-treatment Biopsy

Another important finding of this study has been identifying the clinical need for an early on-treatment biopsy, as well as the pre-treatment biopsy routinely performed at diagnosis, for substantially improved prediction accuracy of overall response. Proliferation-associated gene expression in particular, which has been shown in numerous studies to be strongly associated with clinical response to therapy and which

comprises 2 of the 4 genes in the 4 gene CART model, has dramatically improved predictive power at day 14 on-treatment compared with pre-treatment.

3.5.6.2. Treating Non-responsive Patients

The majority of non-responsive patients have been shown to exhibit both poor clinical and pathological response to therapy compared with the responsive group. While some seem to gain little benefit from endocrine therapy and thus should be considered for alternative treatment, many gain partial benefit and may be suitable candidates for combination therapy. Reduced response to endocrine therapy in such tumours might suggest that proliferation and tumour development is less or only partially dependent on estrogen. Consistent with this hypothesis, a subset of non-responsive tumours in this study (42%) were identified as $HER2^+$, suggesting an alternative mechanism involved in tumorigenesis. Consistent with evidence from a recent study [269], patients who are both ER^+ and $HER2^+$ at diagnosis and are predicted to be non-responsive to endocrine therapy alone may benefit from aromatase inhibitor therapy combined with anti-HER2 therapy such as trastuzumab, pertuzumab or lapatinib. Other groups have studied alternative drugs, combination therapy protocols and treatment strategies (see section 1.2.4) for potential roles in reversing endocrine therapy resistance with mixed success, doubtless due to the molecular and mechanistic heterogeneity of non-response. Rather than a 'one approach fits all', further study is required to elucidated the individual mechanisms which underlie non-response in distinct subsets of non-responsive patients and to stratify such patients in a move towards personalised medicine whereby the most appropriate alternative or combined therapy is identified for each patient.

4. Conclusions and Future Work

This study has led to the development and validation of novel data integration approaches to increase sample size, and as a result statistical significance, for robust microarray data meta-analysis with potentially improved reproducibility and elucidation of platform-independent findings. Using this approach, 4 novel biomarkers for accurate and robust prediction of response to neoadjuvant aromatase inhibitor therapy by two weeks of treatment were identified. Using pre-treatment expression levels of the immune signalling and apoptosis related genes: GP130 and BEX3 respectively, and 14 day expression levels of the proliferation-associated genes: ASPM and MCM4, accuracy of prediction was found to be 96% and 91% in training (n=73) and validation (n=44) datasets. In particular, GP130 was identified as a key biomarker in the model, pre-treatment expression of which was shown to have 85% and 81% prediction accuracy alone in training and validation datasets respectively. Importantly, the 4 gene model has been shown to significantly outperform currently used clinical and pathological parameters for assessing response and previously reported gene expression based models for the purpose.

4.1. Applicability to Other Settings and Therapies

4.1.1. Other Endocrine Therapies

The 4 gene model has been shown to accurately predict response in patients treated with two different aromatase inhibitors (AIs): letrozole and anastrozole, potentially suggesting a broader role for predicting response to other AIs such as exemestane or even other anti-estrogen therapies such as tamoxifen or fulvestrant. Future work will elucidated the potential applicability of the model to predict response to other single and combined anti-estrogen agents.

4.1.2. The Adjuvant Setting

The 4 gene model was designed to predict response in the neoadjuvant setting. However, adjuvant treatment represents the major clinical setting for endocrine therapy and

therefore it will be interesting to assess the applicability of the model to prediction of long-term outcome to adjuvant endocrine therapy. To that end a cohort of patients (n=50) with available pre-treatment and 14 day biopsy material, treated with 14 days of either neoadjuvant letrozole or anastrozole followed by surgery and adjuvant endocrine therapy (tamoxifen or AI) with long-term follow up have been identified. Future work will involve measuring expression of the 4 genes in these samples and determining association with long-term outcome.

4.2. Translational Potential

Based on expression of only 4 genes, the model has great potential for use in the clinical setting or as a commercially available test. Future work will focus on determining the applicability of the model to other endocrine therapies and the adjuvant setting, as well as establishing a clinical protocol for optimum patient care and the most appropriate assay and delivery thereof for robust, accurate and reproducible prediction.

4.2.1. Technology

4.2.1.1. Gene Expression Based Technology: qRT-PCR

Given the restrictive associated costs and unnecessary coverage, commercial gene expression microarray is not the most suitable technology to assay expression of only 4 genes for clinical application. Instead, tests such as Oncotype DX®, based on expression of 21 genes, employ qRT-PCR. However, complications can arise from analytical variation in qRT-PCR, which are likely to impact on predictive accuracy of the model across different sites. In the case of Oncotype DX®, these complications have been largely circumvented by providing the test at only one centralised location where measures of gene expression are derived using the same technology and strictly adhered to protocol. In the case of the 4 gene model it seems logical that a similar setup would be required should the test become commercially available. As a first step, expression levels of the 4 genes in the training set would have to be re-analysed using a chosen RT-PCR technology and controlled methodology in order to establish new splitter values which could be applied to samples from new patients to determine clinical response.

4.2.1.2. Immunohistochemistry

Alternatively, currently used biomarkers such as ER, PR and HER2 are routinely assayed at diagnosis by assessing protein level using IHC rather than gene expression. Thus, it will be advantageous to determine if variation at protein level of the 4 genes could be used as an alternative to gene expression to distinguish responsive from non-responsive patients. If a protein-based test yielded positive results, there may be potential for employing the model in the clinical setting as part of routine histology. However, in order to cut down on analytical variation across sites, there would doubtless be a need to develop standard comprehensive protocols for analytical assessment (including selection of reliable validated antibodies), interpretation and scoring, akin to current efforts to standardise the analytical practice for Ki67 assessment [444].

4.2.2. Collecting, Processing and Storing Tissue

4.2.2.1. FFPE Tissue

The move to qRT-PCR or IHC has another important advantage in that sufficient RNA for qRT-PCR or protein for IHC can be obtained from FFPE tissue, which unlike fresh frozen tissue is prepared routinely at diagnosis for every patient, can be transported easily in the event of a centralised qRT-PCR based test, is easier to work with and can be stored long-term at room temperature without degrading, unlike fresh frozen tissue which requires to be stored at -80^{o}C.

4.2.2.2. Final Needle Aspiration

This study has demonstrated the clinical need for a 14 day biopsy for improved prediction of clinical response to endocrine therapy compared with pre-treatment gene expression alone. Given the importance of a 14 day sample to prediction and considering the cost and discomfort for the patient associated with taking a core biopsy there is a clinical need, for improved patient care, to determine if RNA sufficient for qRT-PCR and be derived from a final needle aspiration (FNA). Future work will also include

expression analysis, in respect of the 4 genes, of patients treated with neoadjuvant letrozole with core biopsies taken at diagnosis and FNAs taken at day 14 of treatment.

4.3. Understanding the Mechanisms underlying GP130 Expression

GP130 has been identified as the key gene involved with predicting clinical responsiveness to endocrine therapy in this study and has an important role in determining risk of recurrence in the EndoPredcit assay, yet little is known about the biology underlying its relevance. A growing number of studies, including this, now exist which provide compelling evidence that high expression of GP130 is associated with both a good clinical response to endocrine therapy and better prognosis [345, 347, 481]. Indeed, the findings of this study indicate that withdrawal of estrogen has a more beneficial effect, in terms of clinical response, in tumours that have higher levels of GP130 at pre-treatment.

While future work will help to elucidate the underlying mechanisms, hypothetically if high GP130 expression is indicative of increased GP130-mediated signalling either through the JAK/STAT, MAPK or PI3K pathways then one possible conclusion could be that disruption of the estrogen signalling pathway has a greater downstream effect when GP130 signalling is active due to crosstalk between the pathways. Increased expression of GP130 could indicate active cytokine signalling via one of its primary upstream activators: IL6 which has been shown to increase the activity of the aromatase enzyme and thus local estrogen production [526]. While crosstalk between the GP130-mediated and estrogen signalling pathways remains poorly understood, in breast tumours the principle mechanism of STAT3, the downstream effector of IL6 signalling, activation has been shown to be via the IL6/GP130/JAK/STAT pathway and is associated with regulation of proliferation [527, 528]. While the expression of STAT3 in breast cancer has been reported as constitutive, studies have shown that estrogen can induce phosphorylation of STAT3 and that STAT-dependent transcription is itself dependent on estrogen-bound ESR1 and intact MAPK and PI3K pathways [529]. Withdrawal of estrogen may result in disruption of the estrogen signalling pathway and disruption of the IL6/GP130/JAK/STAT pathway through altered activity of STAT3,

both having a concerted effect on downstream gene expression. Further work has to be done to confirm the presence of GP130-mediate signal transduction and its association with the estrogen signalling pathway which remains poorly understood.

4.4. Investigating Mechanisms of Non-response

Future work will also focus on gaining a better understanding of the mechanisms underlying non-response in a move towards better treatments, including new, alternative or combined therapies, to improve outcome of patients predicted to respond less well to endocrine therapy alone.

4.4.1. HER2 Signalling

Initial work will focus on establishing the role of HER2 signalling as a driver of proliferation in non-responsive tumours which over-express the HER2 gene, with a view to conducting a clinical trial to determine benefit from combined endocrine therapy and HER2 targeted therapy in the neoadjuvant setting.

4.4.2. Immune and Stromal Cell Infiltration

This study has also identified a subgroup of clinically responsive patients which following 3 months of endocrine therapy are characterised by increased expression of immune/inflammatory and ECM/stromal rearrangement associated genes. Future work will investigate the underlying mechanisms leading to increased expression of immune and stromal gene signatures in responsive tumours and the implications for clinical response. This will include establishing the extent to which infiltrating leukocytes and stromal cells such as fibroblasts, identified in some tumours, are involved and their importance to clinical response. Furthermore, a cohort of patients with pre-treatment and 14 day biopsies taken before surgery following no-intervening treatment will also be assessed by gene expression to determine the extent to which early increases in immune/inflammatory and ECM/stromal gene expression are a result of an inflammatory reaction to the core biopsy rather than response to the drug.

5. References

1. CRUK, *Cancer cases and deaths in the UK.* available at: http://publications.cancerresearchuk.org/cancerstats/statsincidence/dtcasesdeaths.html (accessed 16th June 2012), 2012.
2. ONS *Breast Cancer in England, 2010*. Office for National Statistics, 2012.
3. Smith, I.E., *Aromatase inhibitors--extending the benefits of adjuvant therapy beyond tamoxifen.* Breast, 2004. **13 Suppl 1**: p. S3-9.
4. Stingl, J. and C. Caldas, *Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis.* Nat Rev Cancer, 2007. **7**(10): p. 791-9.
5. Sims, A.H., et al., *Origins of breast cancer subtypes and therapeutic implications.* Nat Clin Pract Oncol, 2007. **4**(9): p. 516-25.
6. Malhotra, G.K., et al., *Histological, molecular and functional subtypes of breast cancers.* Cancer Biol Ther, 2010. **10**(10): p. 955-60.
7. Turashvili, G., et al., *Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis.* BMC Cancer, 2007. **7**: p. 55.
8. Yen, M.F., et al., *Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening.* Eur J Cancer, 2003. **39**(12): p. 1746-54.
9. Li, C.I., D.J. Uribe, and J.R. Daling, *Clinical characteristics of different histologic types of breast cancer.* Br J Cancer, 2005. **93**(9): p. 1046-52.
10. Bloom, H.J. and W.W. Richardson, *Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years.* Br J Cancer, 1957. **11**(3): p. 359-77.
11. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.* Histopathology, 2002. **41**(3A): p. 154-61.
12. Genestie, C., et al., *Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems.* Anticancer Res, 1998. **18**(1B): p. 571-6.
13. Lester, S.C., et al., *Protocol for the examination of specimens from patients with invasive carcinoma of the breast.* Arch Pathol Lab Med, 2009. **133**(10): p. 1515-38.
14. Escobar, P.F., et al., *The 2003 revised TNM staging system for breast cancer: results of stage re-classification on survival and future comparisons among stage groups.* Ann Surg Oncol, 2007. **14**(1): p. 143-7.
15. Harris, L., et al., *American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer.* J Clin Oncol, 2007. **25**(33): p. 5287-312.
16. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Combinatorial biomarker expression in breast cancer.* Breast Cancer Res Treat, 2010. **120**(2): p. 293-308.

17. Payne, S.J., et al., *Predictive markers in breast cancer--the present.* Histopathology, 2008. **52**(1): p. 82-90.
18. Hortobagyi, G.N., *Toward individualized breast cancer therapy: translating biological concepts to the bedside.* Oncologist, 2012. **17**(4): p. 577-84.
19. Allred, D.C., et al., *Prognostic and predictive factors in breast cancer by immunohistochemical analysis.* Mod Pathol, 1998. **11**(2): p. 155-68.
20. Gown, A.M., et al., *High concordance between immunohistochemistry and fluorescence in situ hybridization testing for HER2 status in breast cancer requires a normalized IHC scoring system.* Mod Pathol, 2008. **21**(10): p. 1271-7.
21. Tvrdik, D., et al., *Comparison of the IHC, FISH, SISH and qPCR methods for the molecular diagnosis of breast cancer.* Mol Med Rep, 2012. **6**(2): p. 439-43.
22. Slamon, D.J., et al., *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.* Science, 1987. **235**(4785): p. 177-82.
23. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
24. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
25. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
26. Sorlie, T., *Molecular portraits of breast cancer: tumour subtypes as distinct disease entities.* Eur J Cancer, 2004. **40**(18): p. 2667-75.
27. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.
28. Mohammadizadeh, F., et al., *Expression of basal and luminal cytokeratins in breast cancer and their correlation with clinicopathological prognostic variables.* Indian J Med Sci, 2009. **63**(4): p. 152-62.
29. Prat, A., et al., *Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer.* Breast Cancer Res, 2010. **12**(5): p. R68.
30. Perou, C.M., *Molecular stratification of triple-negative breast cancers.* Oncologist, 2011. **16 Suppl 1**: p. 61-70.
31. Collins, L.C., et al., *Androgen receptor expression in breast cancer in relation to molecular phenotype: results from the Nurses' Health Study.* Mod Pathol, 2011. **24**(7): p. 924-31.
32. Farmer, P., et al., *Identification of molecular apocrine breast tumours by microarray analysis.* Oncogene, 2005. **24**(29): p. 4660-71.
33. Lim, E., et al., *Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.* Nat Med, 2009. **15**(8): p. 907-13.
34. Prat, A. and C.M. Perou, *Mammary development meets cancer genomics.* Nat Med, 2009. **15**(8): p. 842-4.
35. Lehmann, B.D., et al., *Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.* J Clin Invest, 2011. **121**(7): p. 2750-67.

36. Teschendorff, A.E., et al., *An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.* Genome Biol, 2007. **8**(8): p. R157.
37. Das, P., G.M. Siegers, and L.M. Postovit, *Illuminating luminal B: QSOX1 as a subtype-specific biomarker.* Breast Cancer Res, 2013. **15**(3): p. 104.
38. Katchman, B.A., et al., *Expression of quiescin sulfhydryl oxidase 1 is associated with a highly invasive phenotype and correlates with a poor prognosis in Luminal B breast cancer.* Breast Cancer Res, 2013. **15**(2): p. R28.
39. Holland, D.G., et al., *ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium.* EMBO Mol Med, 2011. **3**(3): p. 167-80.
40. Yanagawa, M., et al., *Luminal A and luminal B (HER2 negative) subtypes of breast cancer consist of a mixture of tumors with different genotype.* BMC Res Notes, 2012. **5**: p. 376.
41. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.
42. Stebbing, J. and P. Ellis, *An overview of drug development for metastatic breast cancer.* Br J Nurs, 2012. **21**(4): p. S18-22.
43. Poortmans, P., *Evidence based radiation oncology: breast cancer.* Radiother Oncol, 2007. **84**(1): p. 84-101.
44. ACS, *Surgery for Breast Cancer.* available at: http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-treating-surgery (accessed 20th June 2012), 2012.
45. Breakthrough, *The Best Treatment: Your guide to breast cancer treatment in England & Wales.* available at: http://www.breakthrough.org.uk/breast_cancer/treatment/ (accessed on 20th June 2012), 2009.
46. Yu, J.M. and Y.S. Wang, *Breast cancer radiotherapy: controversies and prospectives.* Chin Med J (Engl), 2008. **121**(20): p. 1957-9.
47. Kaufmann, M., et al., *Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006.* Ann Oncol, 2007. **18**(12): p. 1927-34.
48. McArthur, H.L. and C.A. Hudis, *Breast cancer chemotherapy.* Cancer J, 2007. **13**(3): p. 141-7.
49. Kelly, C.M. and G.N. Hortobagyi, *Adjuvant chemotherapy in early-stage breast cancer: what, when, and for whom?* Surg Oncol Clin N Am, 2010. **19**(3): p. 649-68.
50. Larionov, A.A. and W.R. Miller, *Challenges in defining predictive markers for response to endocrine therapy in breast cancer.* Future Oncol, 2009. **5**(9): p. 1415-28.
51. Collignon, J. and G. Jerusalem, *[Will targeted therapies replace chemotherapy?].* Rev Med Liege, 2012. **67 Spec No**: p. 29-36.
52. McArthur, H.L. and C.A. Hudis, *Trastuzumab: a picky partner?* Clin Cancer Res, 2009. **15**(20): p. 6311-3.

53. Herbst, R.S., et al., *Clinical Cancer Advances 2005: major research advances in cancer treatment, prevention, and screening--a report from the American Society of Clinical Oncology.* J Clin Oncol, 2006. **24**(1): p. 190-205.
54. Sabine, V.S., et al., *Gene expression profiling of response to mTOR inhibitor everolimus in pre-operatively treated post-menopausal women with oestrogen receptor-positive breast cancer.* Breast Cancer Res Treat, 2010. **122**(2): p. 419-28.
55. Johnston, S.R. and M. Dowsett, *Aromatase inhibitors for breast cancer: lessons from the laboratory.* Nat Rev Cancer, 2003. **3**(11): p. 821-31.
56. Smuk, M. and J. Schwers, *Aromatization of androstenedione by human adult liver in vitro.* J Clin Endocrinol Metab, 1977. **45**(5): p. 1009-12.
57. Simpson, E.R. and M. Dowsett, *Aromatase and its inhibitors: significance for breast cancer therapy.* Recent Prog Horm Res, 2002. **57**: p. 317-38.
58. Larionov, A.A., et al., *Aromatase in skeletal muscle.* J Steroid Biochem Mol Biol, 2003. **84**(4): p. 485-92.
59. Perel, E. and D.W. Killinger, *The interconversion and aromatization of androgens by human adipose tissue.* J Steroid Biochem, 1979. **10**(6): p. 623-7.
60. Longcope, C., et al., *Aromatization of androgens by muscle and adipose tissue in vivo.* J Clin Endocrinol Metab, 1978. **46**(1): p. 146-52.
61. Sasano, H., et al., *Aromatase in human bone tissue.* J Bone Miner Res, 1997. **12**(9): p. 1416-23.
62. Miller, W.R. and A.P. Forrest, *Oestradiol synthesis by a human breast carcinoma.* Lancet, 1974. **2**(7885): p. 866-8.
63. Pasqualini, J.R., et al., *Concentrations of estrone, estradiol, and estrone sulfate and evaluation of sulfatase and aromatase activities in pre- and postmenopausal breast cancer patients.* J Clin Endocrinol Metab, 1996. **81**(4): p. 1460-4.
64. Key, T., et al., *Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies.* J Natl Cancer Inst, 2002. **94**(8): p. 606-16.
65. Miller, W.R. and J. O'Neill, *The importance of local synthesis of estrogen within the breast.* Steroids, 1987. **50**(4-6): p. 537-48.
66. Bulun, S.E., et al., *Distribution of aromatase P450 transcripts and adipose fibroblasts in the human breast.* J Clin Endocrinol Metab, 1996. **81**(3): p. 1273-7.
67. Harada, N., T. Utsumi, and Y. Takagi, *Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and promoters, and switching of tissue-specific exons 1 in carcinogenesis.* Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11312-6.
68. Schrey, M.P. and K.V. Patel, *Prostaglandin E2 production and metabolism in human breast cancer cells and breast fibroblasts. Regulation by inflammatory mediators.* Br J Cancer, 1995. **72**(6): p. 1412-9.
69. Horwitz, K.B., et al., *Nuclear receptor coactivators and corepressors.* Mol Endocrinol, 1996. **10**(10): p. 1167-77.
70. Kumar, V., et al., *Functional domains of the human estrogen receptor.* Cell, 1987. **51**(6): p. 941-51.
71. Beato, M., *Gene regulation by steroid hormones.* Cell, 1989. **56**(3): p. 335-44.

72. Tsai, M.J. and B.W. O'Malley, *Molecular mechanisms of action of steroid/thyroid receptor superfamily members.* Annu Rev Biochem, 1994. **63**: p. 451-86.
73. Kato, S., et al., *Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase.* Science, 1995. **270**(5241): p. 1491-4.
74. Wakeling, A.E., *Similarities and distinctions in the mode of action of different classes of antioestrogens.* Endocr Relat Cancer, 2000. **7**(1): p. 17-28.
75. Reid, G., et al., *Cyclic, proteasome-mediated turnover of unliganded and liganded ERalpha on responsive promoters is an integral feature of estrogen signaling.* Mol Cell, 2003. **11**(3): p. 695-707.
76. Bourdeau, V., et al., *Genome-wide identification of high-affinity estrogen response elements in human and mouse.* Mol Endocrinol, 2004. **18**(6): p. 1411-27.
77. Klein-Hitpass, L., et al., *A 13 bp palindrome is a functional estrogen responsive element and interacts specifically with estrogen receptor.* Nucleic Acids Res, 1988. **16**(2): p. 647-63.
78. Kamalakaran, S., S.K. Radhakrishnan, and W.T. Beck, *Identification of estrogen-responsive genes using a genome-wide analysis of promoter elements for transcription factor binding sites.* J Biol Chem, 2005. **280**(22): p. 21491-7.
79. Gruber, C.J., et al., *Anatomy of the estrogen response element.* Trends Endocrinol Metab, 2004. **15**(2): p. 73-8.
80. Scholz, A., M. Truss, and M. Beato, *Hormone-induced recruitment of Sp1 mediates estrogen activation of the rabbit uteroglobin gene in endometrial epithelium.* J Biol Chem, 1998. **273**(8): p. 4360-6.
81. Klinge, C.M., *Estrogen receptor interaction with estrogen response elements.* Nucleic Acids Res, 2001. **29**(14): p. 2905-19.
82. Brown, M. and P.A. Sharp, *Human estrogen receptor forms multiple protein-DNA complexes.* J Biol Chem, 1990. **265**(19): p. 11238-43.
83. Lees, J.A., S.E. Fawell, and M.G. Parker, *Identification of two transactivation domains in the mouse oestrogen receptor.* Nucleic Acids Res, 1989. **17**(14): p. 5477-88.
84. Kumar, V. and P. Chambon, *The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer.* Cell, 1988. **55**(1): p. 145-56.
85. Musgrove, E.A. and R.L. Sutherland, *Biological determinants of endocrine resistance in breast cancer.* Nat Rev Cancer, 2009. **9**(9): p. 631-43.
86. Clarke, R.B., et al., *Dissociation between steroid receptor expression and cell proliferation in the human breast.* Cancer Res, 1997. **57**(22): p. 4987-91.
87. Prall, O.W., et al., *Estrogen-induced activation of Cdk4 and Cdk2 during G1-S phase progression is accompanied by increased cyclin D1 expression and decreased cyclin-dependent kinase inhibitor association with cyclin E-Cdk2.* J Biol Chem, 1997. **272**(16): p. 10882-94.
88. Foster, J.S., et al., *Multifaceted regulation of cell cycle progression by estrogen: regulation of Cdk inhibitors and Cdc25A independent of cyclin D1-Cdk4 function.* Mol Cell Biol, 2001. **21**(3): p. 794-810.

89. Sabbah, M., et al., *Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11217-22.
90. Liu, M.M., et al., *Opposing action of estrogen receptors alpha and beta on cyclin D1 gene expression.* J Biol Chem, 2002. **277**(27): p. 24353-60.
91. Cicatiello, L., et al., *Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter.* Mol Cell Biol, 2004. **24**(16): p. 7260-74.
92. Safe, S., *Transcriptional activation of genes by 17 beta-estradiol through estrogen receptor-Sp1 interactions.* Vitam Horm, 2001. **62**: p. 231-52.
93. Webb, P., et al., *Tamoxifen activation of the estrogen receptor/AP-1 pathway: potential origin for the cell-specific estrogen-like effects of antiestrogens.* Mol Endocrinol, 1995. **9**(4): p. 443-56.
94. Neuman, E., et al., *Cyclin D1 stimulation of estrogen receptor transcriptional activity independent of cdk4.* Mol Cell Biol, 1997. **17**(9): p. 5338-47.
95. Meyerson, M. and E. Harlow, *Identification of G1 kinase activity for cdk6, a novel cyclin D partner.* Mol Cell Biol, 1994. **14**(3): p. 2077-86.
96. Caldon, C.E., et al., *Cell cycle control in breast cancer cells.* J Cell Biochem, 2006. **97**(2): p. 261-74.
97. Dubik, D. and R.P. Shiu, *Mechanism of estrogen activation of c-myc oncogene expression.* Oncogene, 1992. **7**(8): p. 1587-94.
98. Galaktionov, K., X. Chen, and D. Beach, *Cdc25 cell-cycle phosphatase as a target of c-myc.* Nature, 1996. **382**(6591): p. 511-7.
99. Hermeking, H., et al., *Identification of CDK4 as a target of c-MYC.* Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2229-34.
100. Lacy, S. and P. Whyte, *Identification of a p130 domain mediating interactions with cyclin A/cdk 2 and cyclin E/cdk 2 complexes.* Oncogene, 1997. **14**(20): p. 2395-406.
101. Prall, O.W., et al., *c-Myc or cyclin D1 mimics estrogen effects on cyclin E-Cdk2 activation and cell cycle reentry.* Mol Cell Biol, 1998. **18**(8): p. 4499-508.
102. Planas-Silva, M.D. and R.A. Weinberg, *Estrogen-dependent cyclin E-cdk2 activation through p21 redistribution.* Mol Cell Biol, 1997. **17**(7): p. 4059-69.
103. Foster, J.S., et al., *Estrogens down-regulate p27Kip1 in breast cancer cells through Skp2 and through nuclear export mediated by the ERK pathway.* J Biol Chem, 2003. **278**(42): p. 41355-66.
104. Castoria, G., et al., *Role of atypical protein kinase C in estradiol-triggered G1/S progression of MCF-7 cells.* Mol Cell Biol, 2004. **24**(17): p. 7643-53.
105. Schick, P., et al., *Preoperative chemotherapy followed by mastectomy for locally advanced breast cancer.* J Surg Oncol, 1983. **22**(4): p. 278-82.
106. Kaufmann, M., et al., *Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update.* J Clin Oncol, 2006. **24**(12): p. 1940-9.

107. Semiglazov, V.F., et al., *Phase 2 randomized trial of primary endocrine therapy versus chemotherapy in postmenopausal patients with estrogen receptor-positive breast cancer.* Cancer, 2007. **110**(2): p. 244-54.
108. Perloff, M. and G.J. Lesnick, *Chemotherapy before and after mastectomy in stage III breast cancer.* Arch Surg, 1982. **117**(7): p. 879-81.
109. Colleoni, M., et al., *Influence of endocrine-related factors on response to perioperative chemotherapy for patients with node-negative breast cancer.* J Clin Oncol, 2001. **19**(21): p. 4141-9.
110. Gianni, L., et al., *Feasibility and tolerability of sequential doxorubicin/paclitaxel followed by cyclophosphamide, methotrexate, and fluorouracil and its effects on tumor response as preoperative therapy.* Clin Cancer Res, 2005. **11**(24 Pt 1): p. 8715-21.
111. Dowsett, M., et al., *Proliferation and apoptosis as markers of benefit in neoadjuvant endocrine therapy of breast cancer.* Clin Cancer Res, 2006. **12**(3 Pt 2): p. 1024s-1030s.
112. Doisneau-Sixou, S.F., et al., *Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells.* Endocr Relat Cancer, 2003. **10**(2): p. 179-86.
113. Shiau, A.K., et al., *The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen.* Cell, 1998. **95**(7): p. 927-37.
114. Klinge, C.M., et al., *Estrogen response element sequence impacts the conformation and transcriptional activity of estrogen receptor alpha.* Mol Cell Endocrinol, 2001. **174**(1-2): p. 151-66.
115. Long, X. and K.P. Nephew, *Fulvestrant (ICI 182,780)-dependent interacting proteins mediate immobilization and degradation of estrogen receptor-alpha.* J Biol Chem, 2006. **281**(14): p. 9607-15.
116. Howell, A., et al., *Comparison of fulvestrant versus tamoxifen for the treatment of advanced breast cancer in postmenopausal women previously untreated with endocrine therapy: a multinational, double-blind, randomized trial.* J Clin Oncol, 2004. **22**(9): p. 1605-13.
117. Buzdar, A.U., *Fulvestrant--a novel estrogen receptor antagonist for the treatment of advanced breast cancer.* Drugs Today (Barc), 2008. **44**(9): p. 679-92.
118. Goss, P.E., et al., *A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer.* N Engl J Med, 2003. **349**(19): p. 1793-802.
119. Buzdar, A.U., *Aromatase inhibitors: changing the face of endocrine therapy for breast cancer.* Breast Dis, 2005. **24**: p. 107-17.
120. Kao, Y.C., et al., *Binding characteristics of seven inhibitors of human aromatase: a site-directed mutagenesis study.* Cancer Res, 1996. **56**(15): p. 3451-60.
121. Carpenter, R. and W.R. Miller, *Role of aromatase inhibitors in breast cancer.* Br J Cancer, 2005. **93 Suppl 1**: p. S1-5.

122. Coombes, R.C., et al., *A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer.* N Engl J Med, 2004. **350**(11): p. 1081-92.
123. Smith, I.E. and M. Dowsett, *Aromatase inhibitors in breast cancer.* N Engl J Med, 2003. **348**(24): p. 2431-42.
124. Miller, W.R., et al., *Gene expression profiles differentiating between breast cancers clinically responsive or resistant to letrozole.* J Clin Oncol, 2009. **27**(9): p. 1382-7.
125. Howell, A. and M. Dowsett, *Endocrinology and hormone therapy in breast cancer: aromatase inhibitors versus antioestrogens.* Breast Cancer Res, 2004. **6**(6): p. 269-74.
126. Miller, W.R. and J. Jackson, *The therapeutic potential of aromatase inhibitors.* Expert Opin Investig Drugs, 2003. **12**(3): p. 337-51.
127. Early Breast Cancer Trialists' Collaborative, G., *Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.* Lancet, 2005. **365**(9472): p. 1687-717.
128. Brodie, A., L. Macedo, and G. Sabnis, *Aromatase resistance mechanisms in model systems in vivo.* J Steroid Biochem Mol Biol, 2010. **118**(4-5): p. 283-7.
129. Bartsch, R., et al., *Ovarian function suppression and fulvestrant as endocrine therapy in premenopausal women with metastatic breast cancer.* Eur J Cancer, 2012. **48**(13): p. 1932-8.
130. Doughty, J.C., *When to start an aromatase inhibitor: now or later?* J Surg Oncol, 2011. **103**(7): p. 730-8.
131. Smith, I.E., et al., *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, 2005. **23**(22): p. 5108-16.
132. Ellis, M.J., et al., *Estrogen-independent proliferation is present in estrogen-receptor HER2-positive primary breast cancer after neoadjuvant letrozole.* J Clin Oncol, 2006. **24**(19): p. 3019-25.
133. Burcombe, R., et al., *Evaluation of Ki-67 proliferation and apoptotic index before, during and after neoadjuvant chemotherapy for primary breast cancer.* Breast Cancer Res, 2006. **8**(3): p. R31.
134. Guarneri, V. and P.F. Conte, *The curability of breast cancer and the treatment of advanced disease.* Eur J Nucl Med Mol Imaging, 2004. **31 Suppl 1**: p. S149-61.
135. Gonzalez-Angulo, A.M., F. Morales-Vasquez, and G.N. Hortobagyi, *Overview of resistance to systemic therapy in patients with breast cancer.* Adv Exp Med Biol, 2007. **608**: p. 1-22.
136. Dowsett, M., et al., *Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen.* J Clin Oncol, 2010. **28**(3): p. 509-18.
137. Mustacchi, G., et al., *Neo-adjuvant exemestane in elderly patients with breast cancer: a phase II, multicentre, open-label, Italian study.* Ann Oncol, 2009. **20**(4): p. 655-9.
138. Mouridsen, H., et al., *Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a*

phase III study of the International Letrozole Breast Cancer Group. J Clin Oncol, 2001. **19**(10): p. 2596-606.
139. Takatsuka, Y., et al., *[A comparison of two doses of tamoxifen in patients with advanced breast cancer: 20 mg/day versus 40 mg/day].* Gan To Kagaku Ryoho, 1989. **16**(5): p. 2093-7.
140. Harbeck, N. and A. Rody, *Lost in translation? Estrogen receptor status and endocrine responsiveness in breast cancer.* J Clin Oncol, 2012. **30**(7): p. 686-9.
141. Iwamoto, T., et al., *Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry.* J Clin Oncol, 2012. **30**(7): p. 729-34.
142. Dunnwald, L.K., M.A. Rossing, and C.I. Li, *Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients.* Breast Cancer Res, 2007. **9**(1): p. R6.
143. Carracedo, A., et al., *Are ER+PR+ and ER+PR- breast tumors genetically different? A CGH array study.* Cancer Genet, 2012. **205**(4): p. 138-46.
144. Whitman, M., et al., *Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate.* Nature, 1988. **332**(6165): p. 644-6.
145. Anderson, K.E., et al., *Translocation of PDK-1 to the plasma membrane is important in allowing PDK-1 to activate protein kinase B.* Curr Biol, 1998. **8**(12): p. 684-91.
146. Franke, T.F., et al., *The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase.* Cell, 1995. **81**(5): p. 727-36.
147. Maehama, T. and J.E. Dixon, *The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate.* J Biol Chem, 1998. **273**(22): p. 13375-8.
148. Gewinner, C., et al., *Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling.* Cancer Cell, 2009. **16**(2): p. 115-25.
149. Campbell, R.A., et al., *Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance.* J Biol Chem, 2001. **276**(13): p. 9817-24.
150. Yamnik, R.L., et al., *S6 kinase 1 regulates estrogen receptor alpha in control of breast cancer cell proliferation.* J Biol Chem, 2009. **284**(10): p. 6361-9.
151. Coso, O.A., et al., *The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway.* Cell, 1995. **81**(7): p. 1137-46.
152. Minden, A., et al., *Selective activation of the JNK signaling cascade and c-Jun transcriptional activity by the small GTPases Rac and Cdc42Hs.* Cell, 1995. **81**(7): p. 1147-57.
153. Petz, L.N., et al., *Estrogen receptor alpha and activating protein-1 mediate estrogen responsiveness of the progesterone receptor gene in MCF-7 breast cancer cells.* Endocrinology, 2002. **143**(12): p. 4583-91.

154. Shou, J., et al., *Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer.* J Natl Cancer Inst, 2004. **96**(12): p. 926-35.
155. Miller, T.W., et al., *Loss of Phosphatase and Tensin homologue deleted on chromosome 10 engages ErbB3 and insulin-like growth factor-I receptor signaling to promote antiestrogen resistance in breast cancer.* Cancer Res, 2009. **69**(10): p. 4192-201.
156. Miller, T.W., J.M. Balko, and C.L. Arteaga, *Phosphatidylinositol 3-kinase and antiestrogen resistance in breast cancer.* J Clin Oncol, 2011. **29**(33): p. 4452-61.
157. Goetz, M.P., et al., *The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen.* Breast Cancer Res Treat, 2007. **101**(1): p. 113-21.
158. Hoskins, J.M., L.A. Carey, and H.L. McLeod, *CYP2D6 and tamoxifen: DNA matters in breast cancer.* Nat Rev Cancer, 2009. **9**(8): p. 576-86.
159. Colomer, R., et al., *A single-nucleotide polymorphism in the aromatase gene is associated with the efficacy of the aromatase inhibitor letrozole in advanced breast carcinoma.* Clin Cancer Res, 2008. **14**(3): p. 811-6.
160. Garcia-Casado, Z., et al., *A polymorphism at the 3'-UTR region of the aromatase gene defines a subgroup of postmenopausal breast cancer patients with poor response to neoadjuvant letrozole.* BMC Cancer, 2010. **10**: p. 36.
161. Barone, I., L. Brusco, and S.A. Fuqua, *Estrogen receptor mutations and changes in downstream gene expression and signaling.* Clin Cancer Res, 2010. **16**(10): p. 2702-8.
162. Johnston, S.R., et al., *Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer.* Cancer Res, 1995. **55**(15): p. 3331-8.
163. Gutierrez, M.C., et al., *Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase.* J Clin Oncol, 2005. **23**(11): p. 2469-76.
164. Clarke, R., et al., *The properties of high-dimensional data spaces: implications for exploring gene and protein expression data.* Nat Rev Cancer, 2008. **8**(1): p. 37-49.
165. Herynk, M.H. and S.A. Fuqua, *Estrogen receptor mutations in human disease.* Endocr Rev, 2004. **25**(6): p. 869-98.
166. Riggins, R.B., et al., *Pathways to tamoxifen resistance.* Cancer Lett, 2007. **256**(1): p. 1-24.
167. Wang, D., et al., *Induction of estrogen receptor alpha-36 expression by bone morphogenetic protein 2 in breast cancer cell lines.* Mol Med Rep, 2012. **6**(3): p. 591-6.
168. Shi, L., et al., *Expression of ER-{alpha}36, a novel variant of estrogen receptor {alpha}, and resistance to tamoxifen treatment in breast cancer.* J Clin Oncol, 2009. **27**(21): p. 3423-9.
169. Riggins, R.B., et al., *ERRgamma mediates tamoxifen resistance in novel models of invasive lobular breast cancer.* Cancer Res, 2008. **68**(21): p. 8908-17.
170. Paech, K., et al., *Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites.* Science, 1997. **277**(5331): p. 1508-10.

171. Kushner, P.J., et al., *Estrogen receptor pathways to AP-1.* J Steroid Biochem Mol Biol, 2000. **74**(5): p. 311-7.
172. Speirs, V., et al., *Increased expression of estrogen receptor beta mRNA in tamoxifen-resistant breast cancer patients.* Cancer Res, 1999. **59**(21): p. 5421-4.
173. Madeira, M., et al., *Estrogen receptor alpha/beta ratio and estrogen receptor beta as predictors of endocrine therapy responsiveness--a randomized neoadjuvant trial comparison between anastrozole and tamoxifen for the treatment of postmenopausal breast cancer.* BMC Cancer, 2013. **13**(1): p. 425.
174. Gururaj, A.E., et al., *Novel mechanisms of resistance to endocrine therapy: genomic and nongenomic considerations.* Clin Cancer Res, 2006. **12**(3 Pt 2): p. 1001s-1007s.
175. Jordan, V.C., *Selective estrogen receptor modulation: concept and consequences in cancer.* Cancer Cell, 2004. **5**(3): p. 207-13.
176. Knowlden, J.M., et al., *Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells.* Endocrinology, 2003. **144**(3): p. 1032-44.
177. McClelland, R.A., et al., *Enhanced epidermal growth factor receptor signaling in MCF7 breast cancer cells after long-term culture in the presence of the pure antiestrogen ICI 182,780 (Faslodex).* Endocrinology, 2001. **142**(7): p. 2776-88.
178. Ross-Innes, C.S., et al., *Differential oestrogen receptor binding is associated with clinical outcome in breast cancer.* Nature, 2012. **481**(7381): p. 389-93.
179. Carroll, J.S., et al., *Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1.* Cell, 2005. **122**(1): p. 33-43.
180. Lupien, M., et al., *FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription.* Cell, 2008. **132**(6): p. 958-70.
181. Hurtado, A., et al., *FOXA1 is a key determinant of estrogen receptor function and endocrine response.* Nat Genet, 2011. **43**(1): p. 27-33.
182. Lupien, M., et al., *Growth factor stimulation induces a distinct ER(alpha) cistrome underlying breast cancer endocrine resistance.* Genes Dev, 2010. **24**(19): p. 2219-27.
183. Sainsbury, J.R., et al., *Epidermal-growth-factor receptors and oestrogen receptors in human breast cancer.* Lancet, 1985. **1**(8425): p. 364-6.
184. Koga, M., E.A. Musgrove, and R.L. Sutherland, *Modulation of the growth-inhibitory effects of progestins and the antiestrogen hydroxyclomiphene on human breast cancer cells by epidermal growth factor and insulin.* Cancer Res, 1989. **49**(1): p. 112-6.
185. Song, R.X., et al., *Estrogen utilization of IGF-1-R and EGF-R to signal in breast cancer cells.* J Steroid Biochem Mol Biol, 2010. **118**(4-5): p. 219-30.
186. Pontiggia, O., et al., *The tumor microenvironment modulates tamoxifen resistance in breast cancer: a role for soluble stromal factors and fibronectin through beta1 integrin.* Breast Cancer Res Treat, 2012. **133**(2): p. 459-71.
187. Lindberg, K., et al., *Estrogen receptor beta represses Akt signaling in breast cancer cells via downregulation of HER2/HER3 and upregulation of PTEN: implications for tamoxifen sensitivity.* Breast Cancer Res, 2011. **13**(2): p. R43.

188. Hurtado, A., et al., *Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen.* Nature, 2008. **456**(7222): p. 663-6.
189. Beauchemin, D., C. Lacombe, and C. Van Themsche, *PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype.* Mol Cancer, 2011. **10**: p. 148.
190. Zuo, T., et al., *FOXP3 is an X-linked breast cancer suppressor gene and an important repressor of the HER-2/ErbB2 oncogene.* Cell, 2007. **129**(7): p. 1275-86.
191. Hua, G., et al., *A negative feedback regulatory loop associates the tyrosine kinase receptor ERBB2 and the transcription factor GATA4 in breast cancer cells.* Mol Cancer Res, 2009. **7**(3): p. 402-14.
192. O'Brien, C.S., et al., *Breast cancer stem cells and their role in resistance to endocrine therapy.* Horm Cancer, 2011. **2**(2): p. 91-103.
193. Farnie, G., et al., *Novel cell culture technique for primary ductal carcinoma in situ: role of Notch and epidermal growth factor receptor signaling pathways.* J Natl Cancer Inst, 2007. **99**(8): p. 616-27.
194. Park, S.Y., et al., *Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer.* Clin Cancer Res, 2010. **16**(3): p. 876-87.
195. Pece, S., et al., *Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content.* Cell, 2010. **140**(1): p. 62-73.
196. Sheridan, C., et al., *CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis.* Breast Cancer Res, 2006. **8**(5): p. R59.
197. Ouhtit, A., et al., *In vivo evidence for the role of CD44s in promoting breast cancer metastasis to the liver.* Am J Pathol, 2007. **171**(6): p. 2033-9.
198. Creighton, C.J., et al., *Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13820-5.
199. Sutherland, R.L. and E.A. Musgrove, *Cyclins and breast cancer.* J Mammary Gland Biol Neoplasia, 2004. **9**(1): p. 95-104.
200. Ishii, Y., S. Waxman, and D. Germain, *Tamoxifen stimulates the growth of cyclin D1-overexpressing breast cancer cells by promoting the activation of signal transducer and activator of transcription 3.* Cancer Res, 2008. **68**(3): p. 852-60.
201. Akli, S., et al., *Low-molecular-weight cyclin E can bypass letrozole-induced G1 arrest in human breast cancer cells and tumors.* Clin Cancer Res, 2010. **16**(4): p. 1179-90.
202. Caldon, C.E., et al., *Cyclin E2 overexpression is associated with endocrine resistance but not insensitivity to CDK2 inhibition in human breast cancer cells.* Mol Cancer Ther, 2012. **11**(7): p. 1488-99.
203. Venditti, M., et al., *C-myc gene expression alone is sufficient to confer resistance to antiestrogen in human breast cancer cells.* Int J Cancer, 2002. **99**(1): p. 35-42.
204. Cariou, S., et al., *Down-regulation of p21WAF1/CIP1 or p27Kip1 abrogates antiestrogen-mediated cell cycle arrest in human breast cancer cells.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9042-6.

205. Ellis, M.J., et al., *Whole-genome analysis informs breast cancer response to aromatase inhibition.* Nature, 2012. **486**(7403): p. 353-60.
206. Riggins, R.B., et al., *Antiestrogens, aromatase inhibitors, and apoptosis in breast cancer.* Vitam Horm, 2005. **71**: p. 201-37.
207. Mandlekar, S. and A.N. Kong, *Mechanisms of tamoxifen-induced apoptosis.* Apoptosis, 2001. **6**(6): p. 469-77.
208. Yde, C.W., et al., *Breast cancer cells with acquired antiestrogen resistance are sensitized to cisplatin-induced cell death.* Mol Cancer Ther, 2007. **6**(6): p. 1869-76.
209. Frasor, J., et al., *Positive cross-talk between estrogen receptor and NF-kappaB in breast cancer.* Cancer Res, 2009. **69**(23): p. 8918-25.
210. Yde, C.W., et al., *NFkappaB signaling is important for growth of antiestrogen resistant breast cancer cells.* Breast Cancer Res Treat, 2012. **135**(1): p. 67-78.
211. Shen, H.M. and V. Tergaonkar, *NFkappaB signaling in carcinogenesis and as a potential molecular target for cancer therapy.* Apoptosis, 2009. **14**(4): p. 348-63.
212. Madrid, L.V., et al., *Akt stimulates the transactivation potential of the RelA/p65 Subunit of NF-kappa B through utilization of the Ikappa B kinase and activation of the mitogen-activated protein kinase p38.* J Biol Chem, 2001. **276**(22): p. 18934-40.
213. Nehra, R., et al., *BCL2 and CASP8 regulation by NF-kappaB differentially affect mitochondrial function and cell fate in antiestrogen-sensitive and -resistant breast cancer cells.* FASEB J, 2010. **24**(6): p. 2040-55.
214. Zhu, Y., et al., *Expression patterns among interferon regulatory factor-1, human X-box binding protein-1, nuclear factor kappa B, nucleophosmin, estrogen receptor-alpha and progesterone receptor proteins in breast cancer tissue microarrays.* Int J Oncol, 2006. **28**(1): p. 67-76.
215. Gu, Z., et al., *Association of interferon regulatory factor-1, nucleophosmin, nuclear factor-kappaB, and cyclic AMP response element binding with acquired resistance to Faslodex (ICI 182,780).* Cancer Res, 2002. **62**(12): p. 3428-37.
216. Zhou, Y., et al., *Enhanced NF kappa B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer.* BMC Cancer, 2007. **7**: p. 59.
217. deGraffenried, L.A., et al., *NF-kappa B inhibition markedly enhances sensitivity of resistant breast cancer tumor cells to tamoxifen.* Ann Oncol, 2004. **15**(6): p. 885-90.
218. Zhou, Y., et al., *The NFkappaB pathway and endocrine-resistant breast cancer.* Endocr Relat Cancer, 2005. **12 Suppl 1**: p. S37-46.
219. Pradhan, M., et al., *Proinflammatory cytokines enhance estrogen-dependent expression of the multidrug transporter gene ABCG2 through estrogen receptor and NF{kappa}B cooperativity at adjacent response elements.* J Biol Chem, 2010. **285**(41): p. 31100-6.
220. Nettles, K.W., et al., *CBP Is a dosage-dependent regulator of nuclear factor-kappaB suppression by the estrogen receptor.* Mol Endocrinol, 2008. **22**(2): p. 263-72.

221. Cutrupi, S., et al., *Targeting of the adaptor protein Tab2 as a novel approach to revert tamoxifen resistance in breast cancer cells.* Oncogene, 2012. **31**(40): p. 4353-61.
222. Takaesu, G., et al., *TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway.* Mol Cell, 2000. **5**(4): p. 649-58.
223. Viatour, P., et al., *Phosphorylation of NF-kappaB and IkappaB proteins: implications in cancer and inflammation.* Trends Biochem Sci, 2005. **30**(1): p. 43-52.
224. Bruunsgaard, H., M. Pedersen, and B.K. Pedersen, *Aging and proinflammatory cytokines.* Curr Opin Hematol, 2001. **8**(3): p. 131-6.
225. Pfeilschifter, J., et al., *Changes in proinflammatory cytokine activity after menopause.* Endocr Rev, 2002. **23**(1): p. 90-119.
226. Baumgarten, S.C. and J. Frasor, *Minireview: Inflammation: an instigator of more aggressive estrogen receptor (ER) positive breast cancers.* Mol Endocrinol, 2012. **26**(3): p. 360-71.
227. Kelly, P.M., et al., *Macrophages in human breast disease: a quantitative immunohistochemical study.* Br J Cancer, 1988. **57**(2): p. 174-7.
228. Leek, R.D., et al., *Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma.* Cancer Res, 1996. **56**(20): p. 4625-9.
229. Hagemann, T., et al., *Macrophages induce invasiveness of epithelial cancer cells via NF-kappa B and JNK.* J Immunol, 2005. **175**(2): p. 1197-205.
230. Paik, S., et al., *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med, 2004. **351**(27): p. 2817-26.
231. Sheen-Chen, S.M., et al., *Serum concentration of tumor necrosis factor in patients with breast cancer.* Breast Cancer Res Treat, 1997. **43**(3): p. 211-5.
232. Miles, D.W., et al., *Expression of tumour necrosis factor (TNF alpha) and its receptors in benign and malignant breast tissue.* Int J Cancer, 1994. **56**(6): p. 777-82.
233. Rivas, M.A., et al., *TNF alpha acting on TNFR1 promotes breast cancer growth via p42/P44 MAPK, JNK, Akt and NF-kappa B-dependent pathways.* Exp Cell Res, 2008. **314**(3): p. 509-29.
234. Rubio, M.F., et al., *TNF-alpha enhances estrogen-induced cell proliferation of estrogen-dependent breast tumor cells through a complex containing nuclear factor-kappa B.* Oncogene, 2006. **25**(9): p. 1367-77.
235. Buss, H., et al., *Constitutive and interleukin-1-inducible phosphorylation of p65 NF-{kappa}B at serine 536 is mediated by multiple protein kinases including I{kappa}B kinase (IKK)-{alpha}, IKK{beta}, IKK{epsilon}, TRAF family member-associated (TANK)-binding kinase 1 (TBK1), and an unknown kinase and couples p65 to TATA-binding protein-associated factor II31-mediated interleukin-8 transcription.* J Biol Chem, 2004. **279**(53): p. 55633-43.
236. Song, Y.J., et al., *IL-1 receptor-associated kinase 1 is critical for latent membrane protein 1-induced p65/RelA serine 536 phosphorylation and NF-kappaB activation.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2689-94.

237. Hardy, D.B., et al., *Progesterone receptor inhibits aromatase and inflammatory response pathways in breast cancer cells via ligand-dependent and ligand-independent mechanisms.* Mol Endocrinol, 2008. **22**(8): p. 1812-24.
238. Kobayashi, S., et al., *Mechanisms of progesterone receptor inhibition of inflammatory responses in cellular models of breast cancer.* Mol Endocrinol, 2010. **24**(12): p. 2292-302.
239. Sikora, M.J., et al., *Mechanisms of estrogen-independent breast cancer growth driven by low estrogen concentrations are unique versus complete estrogen deprivation.* Breast Cancer Res Treat, 2012. **134**(3): p. 1027-39.
240. Chan, C.M., et al., *Molecular changes associated with the acquisition of oestrogen hypersensitivity in MCF-7 breast cancer cells on long-term oestrogen deprivation.* J Steroid Biochem Mol Biol, 2002. **81**(4-5): p. 333-41.
241. Martin, L.A., et al., *Enhanced estrogen receptor (ER) alpha, ERBB2, and MAPK signal transduction pathways operate during the adaptation of MCF-7 cells to long term estrogen deprivation.* J Biol Chem, 2003. **278**(33): p. 30458-68.
242. Herynk, M.H., et al., *A hypersensitive estrogen receptor alpha mutation that alters dynamic protein interactions.* Breast Cancer Res Treat, 2010. **122**(2): p. 381-93.
243. Thiery, J.P., et al., *Epithelial-mesenchymal transitions in development and disease.* Cell, 2009. **139**(5): p. 871-90.
244. Voulgari, A. and A. Pintzas, *Epithelial-mesenchymal transition in cancer metastasis: mechanisms, markers and strategies to overcome drug resistance in the clinic.* Biochim Biophys Acta, 2009. **1796**(2): p. 75-90.
245. Beavon, I.R., *The E-cadherin-catenin complex in tumour metastasis: structure, function and regulation.* Eur J Cancer, 2000. **36**(13 Spec No): p. 1607-20.
246. Faleiro-Rodrigues, C., et al., *Association of E-cadherin and beta-catenin immunoexpression with clinicopathologic features in primary ovarian carcinomas.* Hum Pathol, 2004. **35**(6): p. 663-9.
247. Hirohashi, S. and Y. Kanai, *Cell adhesion system and human cancer morphogenesis.* Cancer Sci, 2003. **94**(7): p. 575-81.
248. Nagafuchi, A., *Molecular architecture of adherens junctions.* Curr Opin Cell Biol, 2001. **13**(5): p. 600-3.
249. Hiscox, S., et al., *Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of beta-catenin phosphorylation.* Int J Cancer, 2006. **118**(2): p. 290-301.
250. Dhasarathy, A., M. Kajita, and P.A. Wade, *The transcription factor snail mediates epithelial to mesenchymal transitions by repression of estrogen receptor-alpha.* Mol Endocrinol, 2007. **21**(12): p. 2907-18.
251. Trimarchi, M.P., M. Mouangsavanh, and T.H. Huang, *Cancer epigenetics: a perspective on the role of DNA methylation in acquired endocrine resistance.* Chin J Cancer, 2011. **30**(11): p. 749-56.
252. Esteller, M., *Epigenetic changes in cancer.* F1000 Biol Rep, 2011. **3**: p. 9.
253. Hatziapostolou, M. and D. Iliopoulos, *Epigenetic aberrations during oncogenesis.* Cell Mol Life Sci, 2011. **68**(10): p. 1681-702.

254. Szyf, M., *Epigenetics, DNA methylation, and chromatin modifying drugs.* Annu Rev Pharmacol Toxicol, 2009. **49**: p. 243-63.
255. Fan, M., et al., *Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant.* Cancer Res, 2006. **66**(24): p. 11954-66.
256. van Agthoven, T., et al., *Induction of estrogen independence of ZR-75-1 human breast cancer cells by epigenetic alterations.* Mol Endocrinol, 1994. **8**(11): p. 1474-83.
257. Chang, H.G., et al., *Tamoxifen-resistant breast cancers show less frequent methylation of the estrogen receptor beta but not the estrogen receptor alpha gene.* J Mol Med (Berl), 2005. **83**(2): p. 132-9.
258. van Agthoven, T., et al., *CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer.* Br J Cancer, 2009. **101**(11): p. 1824-32.
259. Girault, I., I. Bieche, and R. Lidereau, *Role of estrogen receptor alpha transcriptional coregulators in tamoxifen resistance in breast cancer.* Maturitas, 2006. **54**(4): p. 342-51.
260. Stone, A., et al., *Tamoxifen-induced epigenetic silencing of oestrogen-regulated genes in anti-hormone resistant breast cancer.* PLoS One, 2012. **7**(7): p. e40466.
261. Cook, K.L., A.N. Shajahan, and R. Clarke, *Autophagy and endocrine resistance in breast cancer.* Expert Rev Anticancer Ther, 2011. **11**(8): p. 1283-94.
262. Clarke, R., *Cannibalism, cell survival, and endocrine resistance in breast cancer.* Breast Cancer Res, 2011. **13**(4): p. 311.
263. Gonzalez-Malerva, L., et al., *High-throughput ectopic expression screen for tamoxifen resistance identifies an atypical kinase that blocks autophagy.* Proc Natl Acad Sci U S A, 2011. **108**(5): p. 2058-63.
264. Miller, W.R. and A. Larionov, *Changes in expression of oestrogen regulated and proliferation genes with neoadjuvant treatment highlight heterogeneity of clinical resistance to the aromatase inhibitor, letrozole.* Breast Cancer Res, 2010. **12**(4): p. R52.
265. Long, B.J., et al., *Therapeutic strategies using the aromatase inhibitor letrozole and tamoxifen in a breast cancer model.* J Natl Cancer Inst, 2004. **96**(6): p. 456-65.
266. Dowsett, M., et al., *Impact of tamoxifen on the pharmacokinetics and endocrine effects of the aromatase inhibitor letrozole in postmenopausal women with breast cancer.* Clin Cancer Res, 1999. **5**(9): p. 2338-43.
267. Jelovac, D., et al., *Additive antitumor effect of aromatase inhibitor letrozole and antiestrogen fulvestrant in a postmenopausal breast cancer model.* Cancer Res, 2005. **65**(12): p. 5439-44.
268. Fedele, P., et al., *Targeted agents to reverse resistance to endocrine therapy in metastatic breast cancer: where are we now and where are we going?* Crit Rev Oncol Hematol, 2012. **84**(2): p. 243-51.
269. Nahta, R. and R.M. O'Regan, *Therapeutic implications of estrogen receptor signaling in HER2-positive breast cancers.* Breast Cancer Res Treat, 2012. **135**(1): p. 39-48.

270. Barnadas, A., et al., *An overview of letrozole in postmenopausal women with hormone-responsive breast cancer.* Adv Ther, 2011. **28**(12): p. 1045-58.
271. Barnett, C.M., *Everolimus: targeted therapy on the horizon for the treatment of breast cancer.* Pharmacotherapy, 2012. **32**(4): p. 383-96.
272. Cavazzoni, A., et al., *Overcoming acquired resistance to letrozole by targeting the PI3K/AKT/mTOR pathway in breast cancer cell clones.* Cancer Lett, 2012. **323**(1): p. 77-87.
273. Villarreal-Garza, C., et al., *mTOR inhibitors in the management of hormone receptor-positive breast cancer: the latest evidence and future directions.* Ann Oncol, 2012. **23**(10): p. 2526-35.
274. Johnston, S.R., *BOLERO-2 - will this change practice in advanced breast cancer?* Breast Cancer Res, 2012. **14**(3): p. 311.
275. Baselga, J., et al., *Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer.* N Engl J Med, 2012. **366**(6): p. 520-9.
276. Bobustuc, G.C., et al., *MGMT inhibition restores ERalpha functional sensitivity to antiestrogen therapy.* Mol Med, 2012. **18**: p. 913-29.
277. Agrawal, A., J.F. Robertson, and K.L. Cheung, *Clinical relevance of "withdrawal therapy" as a form of hormonal manipulation for breast cancer.* World J Surg Oncol, 2011. **9**: p. 101.
278. Sabnis, G., et al., *Sensitivity to the aromatase inhibitor letrozole is prolonged after a "break" in treatment.* Mol Cancer Ther, 2010. **9**(1): p. 46-56.
279. Lewis-Wambi, J.S. and V.C. Jordan, *Estrogen regulation of apoptosis: how can one hormone stimulate and inhibit?* Breast Cancer Res, 2009. **11**(3): p. 206.
280. Ariazi, E.A., et al., *Estrogen induces apoptosis in estrogen deprivation-resistant breast cancer through stress responses as identified by global gene expression across time.* Proc Natl Acad Sci U S A, 2011. **108**(47): p. 18879-86.
281. Sun, X.Z., D. Zhou, and S. Chen, *Autocrine and paracrine actions of breast tumor aromatase. A three-dimensional cell culture study involving aromatase transfected MCF-7 and T-47D cells.* J Steroid Biochem Mol Biol, 1997. **63**(1-3): p. 29-36.
282. Masri, S., et al., *Genome-wide analysis of aromatase inhibitor-resistant, tamoxifen-resistant, and long-term estrogen-deprived cells reveals a role for estrogen receptor.* Cancer Res, 2008. **68**(12): p. 4910-8.
283. Taylor, K.J., et al., *Dynamic changes in gene expression in vivo predict prognosis of tamoxifen-treated patients with breast cancer.* Breast Cancer Res, 2010. **12**(3): p. R39.
284. Miller, W.R., et al., *Changes in breast cancer transcriptional profiles after treatment with the aromatase inhibitor, letrozole.* Pharmacogenet Genomics, 2007. **17**(10): p. 813-26.
285. Miller, W.R., et al., *Aromatase inhibitors--gene discovery.* J Steroid Biochem Mol Biol, 2007. **106**(1-5): p. 130-42.
286. Dixon, J.M., et al., *Lessons from the use of aromatase inhibitors in the neoadjuvant setting.* Endocr Relat Cancer, 1999. **6**(2): p. 227-30.

287. Miller, W.R., et al., *Pathological features of breast cancer response following neoadjuvant treatment with either letrozole or tamoxifen.* Eur J Cancer, 2003. **39**(4): p. 462-8.
288. Miller, W.R., et al., *Biological and clinical effects of aromatase inhibitors in neoadjuvant therapy.* J Steroid Biochem Mol Biol, 2001. **79**(1-5): p. 103-7.
289. Robbins, P., et al., *Histological grading of breast carcinomas: a study of interobserver agreement.* Hum Pathol, 1995. **26**(8): p. 873-9.
290. Miller, W.R., *Aromatase inhibitors: prediction of response and nature of resistance.* Expert Opin Pharmacother, 2010. **11**(11): p. 1873-87.
291. Ma, C.X., C.G. Sanchez, and M.J. Ellis, *Predicting endocrine therapy responsiveness in breast cancer.* Oncology (Williston Park), 2009. **23**(2): p. 133-42.
292. Miller, W.R., et al., *Proliferation, steroid receptors and clinical/pathological response in breast cancer treated with letrozole.* Br J Cancer, 2006. **94**(7): p. 1051-6.
293. van Dierendonck, J.H., et al., *Nuclear distribution of the Ki-67 antigen during the cell cycle: comparison with growth fraction in human breast cancer cells.* Cancer Res, 1989. **49**(11): p. 2999-3006.
294. Gerdes, J., et al., *Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67.* J Immunol, 1984. **133**(4): p. 1710-5.
295. du Manoir, S., et al., *Ki-67 labeling in postmitotic cells defines different Ki-67 pathways within the 2c compartment.* Cytometry, 1991. **12**(5): p. 455-63.
296. Haerslev, T., G.K. Jacobsen, and K. Zedeler, *Correlation of growth fraction by Ki-67 and proliferating cell nuclear antigen (PCNA) immunohistochemistry with histopathological parameters and prognosis in primary breast carcinomas.* Breast Cancer Res Treat, 1996. **37**(2): p. 101-13.
297. Gonzalez, M.A., et al., *Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer.* J Clin Oncol, 2003. **21**(23): p. 4306-13.
298. Faneyte, I.F., et al., *Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome.* Br J Cancer, 2003. **88**(3): p. 406-12.
299. Petit, T., et al., *Comparative value of tumour grade, hormonal receptors, Ki-67, HER-2 and topoisomerase II alpha status as predictive markers in breast cancer patients treated with neoadjuvant anthracycline-based chemotherapy.* Eur J Cancer, 2004. **40**(2): p. 205-11.
300. Harper-Wynne, C.L., et al., *Comparison of the systemic and intratumoral effects of tamoxifen and the aromatase inhibitor vorozole in postmenopausal patients with primary breast cancer.* J Clin Oncol, 2002. **20**(4): p. 1026-35.
301. Chang, J., et al., *Prediction of clinical outcome from primary tamoxifen by expression of biologic markers in breast cancer patients.* Clin Cancer Res, 2000. **6**(2): p. 616-21.

302. Makris, A., et al., *Changes in hormone receptors and proliferation markers in tamoxifen treated breast cancer patients and the relationship with response.* Breast Cancer Res Treat, 1998. **48**(1): p. 11-20.
303. Tao, Y., et al., *Clinical and biomarker endpoint analysis in neoadjuvant endocrine therapy trials.* J Steroid Biochem Mol Biol, 2005. **95**(1-5): p. 91-5.
304. Ellis, M.J., et al., *Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial.* J Clin Oncol, 2001. **19**(18): p. 3808-16.
305. Elledge, R.M., et al., *Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study.* Int J Cancer, 2000. **89**(2): p. 111-7.
306. Ravdin, P.M., et al., *Prognostic significance of progesterone receptor levels in estrogen receptor-positive patients with metastatic breast cancer treated with tamoxifen: results of a prospective Southwest Oncology Group study.* J Clin Oncol, 1992. **10**(8): p. 1284-91.
307. Wright, C., et al., *Relationship between c-erbB-2 protein product expression and response to endocrine therapy in advanced breast cancer.* Br J Cancer, 1992. **65**(1): p. 118-21.
308. Lal, P., L.K. Tan, and B. Chen, *Correlation of HER-2 status with estrogen and progesterone receptors and histologic features in 3,655 invasive breast carcinomas.* Am J Clin Pathol, 2005. **123**(4): p. 541-6.
309. Ma, X.J., et al., *Gene expression profiles of human breast cancer progression.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5974-9.
310. Dowsett, M., et al., *Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status.* Ann Oncol, 2006. **17**(5): p. 818-26.
311. Weigelt, B., F.L. Baehner, and J.S. Reis-Filho, *The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade.* J Pathol, 2010. **220**(2): p. 263-80.
312. Sims, A.H., et al., *High-throughput genomic technology in research and clinical management of breast cancer. Exploiting the potential of gene expression profiling: is it ready for the clinic?* Breast Cancer Res, 2006. **8**(5): p. 214.
313. Sims, A.H., *Bioinformatics and breast cancer: what can high-throughput genomic approaches actually tell us?* J Clin Pathol, 2009. **62**(10): p. 879-85.
314. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
315. Sparano, J.A. and S. Paik, *Development of the 21-gene assay and its application in clinical practice and clinical trials.* J Clin Oncol, 2008. **26**(5): p. 721-8.
316. Buyse, M., et al., *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst, 2006. **98**(17): p. 1183-92.
317. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.

318. Bueno-de-Mesquita, J.M., et al., *Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER).* Lancet Oncol, 2007. **8**(12): p. 1079-87.
319. Bueno-de-Mesquita, J.M., et al., *Validation of 70-gene prognosis signature in node-negative breast cancer.* Breast Cancer Res Treat, 2009. **117**(3): p. 483-95.
320. Cardoso, F., et al., *Clinical application of the 70-gene profile: the MINDACT trial.* J Clin Oncol, 2008. **26**(5): p. 729-35.
321. Straver, M.E., et al., *The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer.* Breast Cancer Res Treat, 2010. **119**(3): p. 551-8.
322. Simon, R., *Development and validation of therapeutically relevant multi-gene biomarker classifiers.* J Natl Cancer Inst, 2005. **97**(12): p. 866-7.
323. Simon, R., *Roadmap for developing and validating therapeutically relevant genomic classifiers.* J Clin Oncol, 2005. **23**(29): p. 7332-41.
324. Simon, R., et al., *Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification.* J Natl Cancer Inst, 2003. **95**(1): p. 14-8.
325. Paik, S., *Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen.* Oncologist, 2007. **12**(6): p. 631-5.
326. Goldstein, L.J., et al., *Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features.* J Clin Oncol, 2008. **26**(25): p. 4063-71.
327. Habel, L.A., et al., *A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients.* Breast Cancer Res, 2006. **8**(3): p. R25.
328. Paik, S., et al., *Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.* J Clin Oncol, 2006. **24**(23): p. 3726-34.
329. Mina, L., et al., *Predicting response to primary chemotherapy: gene expression profiling of paraffin-embedded core biopsy tissue.* Breast Cancer Res Treat, 2007. **103**(2): p. 197-208.
330. Sotiriou, C., et al., *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* J Natl Cancer Inst, 2006. **98**(4): p. 262-72.
331. Loi, S., et al., *Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade.* J Clin Oncol, 2007. **25**(10): p. 1239-46.
332. Metzger Filho, O., M. Ignatiadis, and C. Sotiriou, *Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis.* Crit Rev Oncol Hematol, 2011. **77**(1): p. 20-9.
333. Tordai, A., et al., *Evaluation of biological pathways involved in chemotherapy response in breast cancer.* Breast Cancer Res, 2008. **10**(2): p. R37.
334. Toussaint, J., et al., *Improvement of the clinical applicability of the Genomic Grade Index through a qRT-PCR test performed on frozen and formalin-fixed paraffin-embedded tissues.* BMC Genomics, 2009. **10**: p. 424.

335. Desmedt, C., et al., *Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes.* Clin Cancer Res, 2008. **14**(16): p. 5158-65.
336. Wirapati, P., et al., *Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures.* Breast Cancer Res, 2008. **10**(4): p. R65.
337. Goetz, M.P., et al., *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, 2006. **12**(7 Pt 1): p. 2080-7.
338. Ma, X.J., et al., *The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer.* J Clin Oncol, 2006. **24**(28): p. 4611-9.
339. Reid, J.F., et al., *Limits of predictive models using microarray data for breast cancer clinical treatment outcome.* J Natl Cancer Inst, 2005. **97**(12): p. 927-30.
340. Reya, T., et al., *Stem cells, cancer, and cancer stem cells.* Nature, 2001. **414**(6859): p. 105-11.
341. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3983-8.
342. Liu, R., et al., *The prognostic role of a gene signature from tumorigenic breast-cancer cells.* N Engl J Med, 2007. **356**(3): p. 217-26.
343. Finak, G., et al., *Stromal gene expression predicts clinical outcome in breast cancer.* Nat Med, 2008. **14**(5): p. 518-27.
344. Symmans, W.F., et al., *Genomic index of sensitivity to endocrine therapy for breast cancer.* J Clin Oncol, 2010. **28**(27): p. 4111-9.
345. Dubsky, P., et al., *EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer.* Ann Oncol, 2013. **24**(3): p. 640-7.
346. Kronenwett, R., et al., *Decentral gene expression analysis: analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test.* BMC Cancer, 2012. **12**: p. 456.
347. Filipits, M., et al., *A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors.* Clin Cancer Res, 2011. **17**(18): p. 6012-20.
348. Denkert, C., et al., *Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.* J Clin Oncol, 2010. **28**(1): p. 105-13.
349. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13550-5.
350. Holm, K., et al., *Characterisation of amplification patterns and target genes at chromosome 11q13 in CCND1-amplified sporadic and familial breast tumours.* Breast Cancer Res Treat, 2012. **133**(2): p. 583-94.
351. Loi, S., et al., *PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10208-13.

352. Creighton, C.J., et al., *Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors.* Cancer Res, 2006. **66**(7): p. 3903-11.
353. Amundson, S.A., et al., *Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen.* Cancer Res, 2008. **68**(2): p. 415-24.
354. Reis-Filho, J.S., C. Westbury, and J.Y. Pierga, *The impact of expression profiling on prognostic and predictive testing in breast cancer.* J Clin Pathol, 2006. **59**(3): p. 225-31.
355. Ahmed, A.A. and J.D. Brenton, *Microarrays and breast cancer clinical studies: forgetting what we have not yet learnt.* Breast Cancer Res, 2005. **7**(3): p. 96-9.
356. Brenton, J.D., et al., *Molecular classification and molecular forecasting of breast cancer: ready for clinical application?* J Clin Oncol, 2005. **23**(29): p. 7350-60.
357. Thomassen, M., et al., *Comparison of gene sets for expression profiling: prediction of metastasis from low-malignant breast cancer.* Clin Cancer Res, 2007. **13**(18 Pt 1): p. 5355-60.
358. Haibe-Kains, B., et al., *Comparison of prognostic gene expression signatures for breast cancer.* BMC Genomics, 2008. **9**: p. 394.
359. Haibe-Kains, B., et al., *A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all?* Bioinformatics, 2008. **24**(19): p. 2200-8.
360. Fan, C., et al., *Concordance among gene-expression-based predictors for breast cancer.* N Engl J Med, 2006. **355**(6): p. 560-9.
361. Natrajan, R., et al., *Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype?* Genes Chromosomes Cancer, 2009. **48**(4): p. 351-65.
362. Natrajan, R., et al., *Tiling path genomic profiling of grade 3 invasive ductal breast cancers.* Clin Cancer Res, 2009. **15**(8): p. 2711-22.
363. Reyal, F., et al., *A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer.* Breast Cancer Res, 2008. **10**(6): p. R93.
364. Teschendorff, A.E. and C. Caldas, *A robust classifier of high predictive value to identify good prognosis patients in ER-negative breast cancer.* Breast Cancer Res, 2008. **10**(4): p. R73.
365. Jansen, M.P., et al., *Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling.* J Clin Oncol, 2005. **23**(4): p. 732-40.
366. Mackay, A., et al., *Molecular response to aromatase inhibitor treatment in primary breast cancer.* Breast Cancer Res, 2007. **9**(3): p. R37.
367. Harvell, D.M., et al., *Estrogen regulated gene expression in response to neoadjuvant endocrine therapy of breast cancers: tamoxifen agonist effects dominate in the presence of an aromatase inhibitor.* Breast Cancer Res Treat, 2008. **112**(3): p. 489-501.

368. Harvell, D.M., et al., *Molecular signatures of neoadjuvant endocrine therapy for breast cancer: characteristics of response or intrinsic resistance.* Breast Cancer Res Treat, 2008. **112**(3): p. 475-88.
369. Sims, A.H., et al., *The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets - improving meta-analysis and prediction of prognosis.* BMC Med Genomics, 2008. **1**: p. 42.
370. Kitchen, R.R., et al., *Correcting for intra-experiment variation in Illumina BeadChip data is necessary to generate robust gene-expression profiles.* BMC Genomics, 2010. **11**: p. 134.
371. Shi, L., et al., *The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.* Nat Biotechnol, 2010. **28**(8): p. 827-38.
372. Ghazoui, Z., et al., *Close and stable relationship between proliferation and a hypoxia metagene in aromatase inhibitor-treated ER-positive breast cancer.* Clin Cancer Res, 2011. **17**(9): p. 3005-12.
373. Smith, I.E., et al., *A phase II placebo-controlled trial of neoadjuvant anastrozole alone or with gefitinib in early breast cancer.* J Clin Oncol, 2007. **25**(25): p. 3816-22.
374. Tseng, G.C., D. Ghosh, and E. Feingold, *Comprehensive literature review and statistical considerations for microarray meta-analysis.* Nucleic Acids Res, 2012. **40**(9): p. 3785-99.
375. Lin, C.Y., et al., *Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells.* Genome Biol, 2004. **5**(9): p. R66.
376. Kitchen, R.R., et al., *Relative impact of key sources of systematic noise in Affymetrix and Illumina gene-expression microarray experiments.* BMC Genomics, 2011. **12**: p. 589.
377. Sims, A.H. and J.M. Bartlett, *Approaches towards expression profiling the response to treatment.* Breast Cancer Res, 2008. **10**(6): p. 115.
378. Ein-Dor, L., O. Zuk, and E. Domany, *Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5923-8.
379. Ong, K.R., et al., *Biomarkers of dietary energy restriction in women at increased risk of breast cancer.* Cancer Prev Res (Phila Pa), 2009. **2**(8): p. 720-31.
380. Kendall, A., et al., *Impact of estrogen deprivation on gene expression profiles of normal postmenopausal breast tissue in vivo.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(4): p. 855-63.
381. Culhane, A.C. and J. Quackenbush, *Confounding effects in "A six-gene signature predicting breast cancer lung metastasis".* Cancer Res, 2009. **69**(18): p. 7480-5.
382. Zhang, Z., et al., *Cross-platform expression microarray performance in a mouse model of mitochondrial disease therapy.* Mol Genet Metab. **99**(3): p. 309-18.
383. Barnes, M., et al., *Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms.* Nucleic Acids Res, 2005. **33**(18): p. 5914-23.

384. Consortium, M., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-61.
385. Shen, R., D. Ghosh, and A.M. Chinnaiyan, *Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data.* BMC Genomics, 2004. **5**(1): p. 94.
386. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nat Biotechnol, 1996. **14**(13): p. 1675-80.
387. Gunderson, K.L., et al., *Decoding randomly ordered DNA arrays.* Genome Res, 2004. **14**(5): p. 870-7.
388. Desmedt, C., et al., *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series.* Clin Cancer Res, 2007. **13**(11): p. 3207-14.
389. Naderi, A., et al., *A gene-expression signature to predict survival in breast cancer across independent data sets.* Oncogene, 2007. **26**(10): p. 1507-16.
390. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC Genomics, 2006. **7**: p. 96.
391. Mackay, A., et al., *Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement.* J Natl Cancer Inst, 2011. **103**(8): p. 662-73.
392. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.
393. Dai, M., et al., *Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.* Nucleic Acids Res, 2005. **33**(20): p. e175.
394. Barbosa-Morais, N.L., et al., *A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data.* Nucleic Acids Res, 2009.
395. Kinsella, R.J., et al., *Ensembl BioMarts: a hub for data retrieval across taxonomic space.* Database (Oxford), 2011. **2011**: p. bar030.
396. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
397. Benito, M., et al., *Adjustment of systematic microarray data biases.* Bioinformatics, 2004. **20**(1): p. 105-14.
398. Shabalin, A.A., et al., *Merging two gene-expression studies via cross-platform normalization.* Bioinformatics, 2008. **24**(9): p. 1154-60.
399. Glaab, E., J.M. Garibaldi, and N. Krasnogor, *ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization.* BMC Bioinformatics, 2009. **10**: p. 358.
400. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
401. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5116-21.
402. Snedecor, G.W. and W.G. Cochran, *Statistical Methods (8th Edition)*. 1989: Iowa State University Press.

403. Neter, J., W. Wasserman, and M.H. Kutner, *Applied Linear Statistical Models, Regression, Analysis of Variance, and Experimental Design, (2nd Edition)*. 1985: Homewood.
404. Oberg, A.L. and D.W. Mahoney, *Linear mixed effects models.* Methods Mol Biol, 2007. **404**: p. 213-34.
405. Tichopad, A., et al., *Design and optimization of reverse-transcription quantitative PCR experiments.* Clin Chem, 2009. **55**(10): p. 1816-23.
406. Kitchen, R.R., M. Kubista, and A. Tichopad, *Statistical aspects of quantitative real-time PCR experiment design.* Methods, 2010. **50**(4): p. 231-6.
407. Lindstrom, M.L. and D.M. Bates, *Nonlinear mixed effects models for repeated measures data.* Biometrics, 1990. **46**(3): p. 673-87.
408. Laird, N.M. and J.H. Ware, *Random-effects models for longitudinal data.* Biometrics, 1982. **38**(4): p. 963-74.
409. Miller, W.R., et al., *Genes Discriminating between Breast Cancers Responsive or Resistant to the Aromatase Inhibitor, Letrozole.* EJCMO, 2010. **2010;2:(2)**.
410. Leong, H.S., et al., *ADAPT: a database of affymetrix probesets and transcripts.* Bioinformatics, 2005. **21**(10): p. 2552-3.
411. Okoniewski, M.J. and C.J. Miller, *Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations.* BMC Bioinformatics, 2006. **7**: p. 276.
412. Lu, X. and X. Zhang, *The effect of GeneChip gene definitions on the microarray study of cancers.* Bioessays, 2006. **28**(7): p. 739-46.
413. Sandberg, R. and O. Larsson, *Improved precision and accuracy for microarrays using updated probe set definitions.* BMC Bioinformatics, 2007. **8**: p. 48.
414. Fan, X., et al., *Consistency of predictive signature genes and classifiers generated using different microarray platforms.* Pharmacogenomics J, 2010. **10**(4): p. 247-57.
415. Rudy, J. and F. Valafar, *Empirical comparison of cross-platform normalization methods for gene expression data.* BMC Bioinformatics, 2011. **12**: p. 467.
416. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
417. Teschendorff, A.E., J. Zhuang, and M. Widschwendter, *Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies.* Bioinformatics, 2011. **27**(11): p. 1496-505.
418. McCall, M.N., et al., *The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes.* Nucleic Acids Res, 2011. **39**(Database issue): p. D1011-5.
419. Engreitz, J.M., et al., *ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression.* Bioinformatics, 2011. **27**(23): p. 3317-8.
420. Engreitz, J.M., et al., *Content-based microarray search using differential expression profiles.* BMC Bioinformatics, 2010. **11**: p. 603.
421. Winer, E.P., et al., *American Society of Clinical Oncology technology assessment working group update: use of aromatase inhibitors in the adjuvant setting.* J Clin Oncol, 2003. **21**(13): p. 2597-9.

422. Miller, W.R. and J.M. Dixon, *Local endocrine effects of aromatase inhibitors within the breast.* J Steroid Biochem Mol Biol, 2001. **79**(1-5): p. 93-102.
423. Geisler, J., et al., *Influence of letrozole and anastrozole on total body aromatization and plasma estrogen levels in postmenopausal breast cancer patients evaluated in a randomized, cross-over study.* J Clin Oncol, 2002. **20**(3): p. 751-7.
424. Miller, W.R., *Biological rationale for endocrine therapy in breast cancer.* Best Pract Res Clin Endocrinol Metab, 2004. **18**(1): p. 1-32.
425. Eiermann, W., et al., *Preoperative treatment of postmenopausal breast cancer patients with letrozole: A randomized double-blind multicenter study.* Ann Oncol, 2001. **12**(11): p. 1527-32.
426. Iqbal, S., et al., *MIB-1 assessments in breast cancers.* Breast, 2002. **11**(3): p. 252-6.
427. Green, S. and G.R. Weiss, *Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria.* Invest New Drugs, 1992. **10**(4): p. 239-53.
428. Saeed, A.I., et al., *TM4 microarray software suite.* Methods Enzymol, 2006. **411**: p. 134-93.
429. Saeed, A.I., et al., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-8.
430. Barbosa-Morais, N.L., et al., *A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data.* Nucleic Acids Res, 2010. **38**(3): p. e17.
431. Huang da, W., et al., *The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists.* Genome Biol, 2007. **8**(9): p. R183.
432. Huang da, W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W169-75.
433. Sherman, B.T., et al., *DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis.* BMC Bioinformatics, 2007. **8**: p. 426.
434. Breiman, L., *Random forests.* Machine Learning, 2001. **45**(1): p. 5-32.
435. Breiman, L., et al., *Classification and Regression Trees*. 1984: Pacific Grove: Wadsworth.
436. Steinberg, D. and P. Colla, *CART: Tree-Structured Non-Parametric Data Analysis*. 1995, San Diego: CA: Salford Systems.
437. Dahlke, M.H., et al., *The biology of CD45 and its use as a therapeutic target.* Leuk Lymphoma, 2004. **45**(2): p. 229-36.
438. Sapag, A., E. Frischling, and H. Laborde, *Hypoxanthine-guanine phosphoribosyltransferase deficiency in a patient with a Madrid II mutation.* Joint Bone Spine, 2013. **80**(1): p. 93-5.
439. Creighton, C.J., *The molecular profile of luminal B breast cancer.* Biologics, 2012. **6**: p. 289-97.

440. Karn, T., et al., *Gene expression profiling of luminal B breast cancers reveals NHERF1 as a new marker of endocrine resistance.* Breast Cancer Res Treat, 2011. **130**(2): p. 409-20.
441. Weigel, M.T. and M. Dowsett, *Current and emerging biomarkers in breast cancer: prognosis and prediction.* Endocr Relat Cancer, 2010. **17**(4): p. R245-62.
442. Takei, H., et al., *Neoadjuvant endocrine therapy of breast cancer: which patients would benefit and what are the advantages?* Breast Cancer, 2011. **18**(2): p. 85-91.
443. Bedard, P.L., et al., *Low residual proliferation after short-term letrozole therapy is an early predictive marker of response in high proliferative ER-positive breast cancer.* Endocr Relat Cancer, 2011. **18**(6): p. 721-30.
444. Dowsett, M., et al., *Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group.* J Natl Cancer Inst, 2011. **103**(22): p. 1656-64.
445. Fasching, P.A., et al., *Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment.* BMC Cancer, 2011. **11**: p. 486.
446. Zhang, G.C., et al., *Pre-treatment hormonal receptor status and Ki67 index predict pathologic complete response to neoadjuvant trastuzumab/taxanes but not disease-free survival in HER2-positive breast cancer patients.* Med Oncol, 2012. **29**(5): p. 3222-31.
447. Tanei, T., et al., *Prognostic significance of Ki67 index after neoadjuvant chemotherapy in breast cancer.* Eur J Surg Oncol, 2011. **37**(2): p. 155-61.
448. Delpech, Y., et al., *Ki67 expression in the primary tumor predicts for clinical benefit and time to progression on first-line endocrine therapy in estrogen receptor-positive metastatic breast cancer.* Breast Cancer Res Treat, 2012. **135**(2): p. 619-27.
449. Sheri, A. and M. Dowsett, *Developments in Ki67 and other biomarkers for treatment decision making in breast cancer.* Ann Oncol, 2012. **23 Suppl 10**: p. x219-27.
450. Schreiber, R.D., L.J. Old, and M.J. Smyth, *Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion.* Science, 2011. **331**(6024): p. 1565-70.
451. Knupfer, H. and R. Preiss, *Significance of interleukin-6 (IL-6) in breast cancer (review).* Breast Cancer Res Treat, 2007. **102**(2): p. 129-35.
452. Stewart, T.H. and G.H. Heppner, *Immunological enhancement of breast cancer.* Parasitology, 1997. **115 Suppl**: p. S141-53.
453. Hadden, J.W., *The immunology and immunotherapy of breast cancer: an update.* Int J Immunopharmacol, 1999. **21**(2): p. 79-101.
454. Mantovani, A., et al., *The origin and function of tumor-associated macrophages.* Immunol Today, 1992. **13**(7): p. 265-70.
455. Crowther, M., et al., *Microenvironmental influence on macrophage regulation of angiogenesis in wounds and malignant tumors.* J Leukoc Biol, 2001. **70**(4): p. 478-90.
456. van Netten, J.P., et al., *Macrophage-tumor cell associations: a factor in metastasis of breast cancer?* J Leukoc Biol, 1993. **54**(4): p. 360-2.

457. Ruffell, B., et al., *Leukocyte composition of human breast cancer.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2796-801.
458. Mahmoud, S.M., et al., *Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer.* J Clin Oncol, 2011. **29**(15): p. 1949-55.
459. Aspord, C., et al., *Breast cancer instructs dendritic cells to prime interleukin 13-secreting CD4+ T cells that facilitate tumor development.* J Exp Med, 2007. **204**(5): p. 1037-47.
460. Taga, T. and T. Kishimoto, *Gp130 and the interleukin-6 family of cytokines.* Annu Rev Immunol, 1997. **15**: p. 797-819.
461. Peters, M., A.M. Muller, and S. Rose-John, *Interleukin-6 and soluble interleukin-6 receptor: direct stimulation of gp130 and hematopoiesis.* Blood, 1998. **92**(10): p. 3495-504.
462. Hong, D.S., L.S. Angelo, and R. Kurzrock, *Interleukin-6 and its receptor in cancer: implications for translational therapeutics.* Cancer, 2007. **110**(9): p. 1911-28.
463. Hirano, T., K. Ishihara, and M. Hibi, *Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors.* Oncogene, 2000. **19**(21): p. 2548-56.
464. Schindler, C.W., *Series introduction. JAK-STAT signaling in human disease.* J Clin Invest, 2002. **109**(9): p. 1133-7.
465. Bromberg, J.F., et al., *Stat3 as an oncogene.* Cell, 1999. **98**(3): p. 295-303.
466. Bromberg, J., *Stat proteins and oncogenesis.* J Clin Invest, 2002. **109**(9): p. 1139-42.
467. Li, L. and P.E. Shaw, *Autocrine-mediated activation of STAT3 correlates with cell proliferation in breast carcinoma lines.* J Biol Chem, 2002. **277**(20): p. 17397-405.
468. Garcia, R., et al., *Constitutive activation of Stat3 by the Src and JAK tyrosine kinases participates in growth regulation of human breast carcinoma cells.* Oncogene, 2001. **20**(20): p. 2499-513.
469. Burke, W.M., et al., *Inhibition of constitutively active Stat3 suppresses growth of human ovarian and breast cancer cells.* Oncogene, 2001. **20**(55): p. 7925-34.
470. Ren, Z. and T.S. Schaefer, *ErbB-2 activates Stat3 alpha in a Src- and JAK2-dependent manner.* J Biol Chem, 2002. **277**(41): p. 38486-93.
471. Zhang, F., et al., *Delineating an oncostatin M-activated STAT3 signaling pathway that coordinates the expression of genes involved in cell cycle regulation and extracellular matrix deposition of MCF-7 cells.* Oncogene, 2003. **22**(6): p. 894-905.
472. Schafer, Z.T. and J.S. Brugge, *IL-6 involvement in epithelial cancers.* J Clin Invest, 2007. **117**(12): p. 3660-3.
473. Badache, A. and N.E. Hynes, *Interleukin 6 inhibits proliferation and, in cooperation with an epidermal growth factor receptor autocrine loop, increases migration of T47D breast cancer cells.* Cancer Res, 2001. **61**(1): p. 383-91.
474. Tamm, I., et al., *Cell-adhesion-disrupting action of interleukin 6 in human ductal breast carcinoma cells.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3329-33.

475. Speirs, V., et al., *Interactive effects of interleukin-6, 17 beta-estradiol and progesterone on growth and 17 beta-hydroxysteroid dehydrogenase activity in human breast carcinoma cells.* J Steroid Biochem Mol Biol, 1993. **46**(1): p. 11-5.
476. Speirs, V., et al., *Direct activation of oestrogen receptor-alpha by interleukin-6 in primary cultures of breast cancer epithelial cells.* Br J Cancer, 2000. **82**(7): p. 1312-6.
477. Qiu, Y., L. Ravi, and H.J. Kung, *Requirement of ErbB2 for signalling by interleukin-6 in prostate carcinoma cells.* Nature, 1998. **393**(6680): p. 83-5.
478. Grant, S.L., et al., *An unexpected biochemical and functional interaction between gp130 and the EGF receptor family in breast cancer cells.* Oncogene, 2002. **21**(3): p. 460-74.
479. Selander, K.S., et al., *Inhibition of gp130 signaling in breast cancer blocks constitutive activation of Stat3 and inhibits in vivo malignancy.* Cancer Res, 2004. **64**(19): p. 6924-33.
480. Douglas, A.M., et al., *Expression and function of members of the cytokine receptor superfamily on breast cancer cells.* Oncogene, 1997. **14**(6): p. 661-9.
481. Karczewska, A., et al., *Expression of interleukin-6, interleukin-6 receptor, and glycoprotein 130 correlates with good prognoses for patients with breast carcinoma.* Cancer, 2000. **88**(9): p. 2061-71.
482. Kovacs, E., *Investigation of interleukin-6 (IL-6), soluble IL-6 receptor (sIL-6R) and soluble gp130 (sgp130) in sera of cancer patients.* Biomed Pharmacother, 2001. **55**(7): p. 391-6.
483. Fish, J.L., et al., *Aspm specifically maintains symmetric proliferative divisions of neuroepithelial cells.* Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10438-43.
484. Bond, J., et al., *ASPM is a major determinant of cerebral cortical size.* Nat Genet, 2002. **32**(2): p. 316-20.
485. Kato, T.A., et al., *ASPM influences DNA double-strand break repair and represents a potential target for radiotherapy.* Int J Radiat Biol, 2011. **87**(12): p. 1189-95.
486. Hagemann, C., et al., *Expression analysis of the autosomal recessive primary microcephaly genes MCPH1 (microcephalin) and MCPH5 (ASPM, abnormal spindle-like, microcephaly associated) in human malignant gliomas.* Oncol Rep, 2008. **20**(2): p. 301-8.
487. Horvath, S., et al., *Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target.* Proc Natl Acad Sci U S A, 2006. **103**(46): p. 17402-7.
488. van der Voet, M., et al., *NuMA-related LIN-5, ASPM-1, calmodulin and dynein promote meiotic spindle rotation independently of cortical LIN-5/GPR/Galpha.* Nat Cell Biol, 2009. **11**(3): p. 269-77.
489. Higgins, J., et al., *Human ASPM participates in spindle organisation, spindle orientation and cytokinesis.* BMC Cell Biol, 2010. **11**: p. 85.
490. Rhoads, A. and H. Kenguele, *Expression of IQ-motif genes in human cells and ASPM domain structure.* Ethn Dis, 2005. **15**(4 Suppl 5): p. S5-88-91.

491. Kouprina, N., et al., *The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein.* Hum Mol Genet, 2005. **14**(15): p. 2155-65.
492. Parris, T.Z., et al., *Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma.* Clin Cancer Res, 2010. **16**(15): p. 3860-74.
493. Shubbar, E., et al., *Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome.* BMC Cancer, 2013. **13**: p. 1.
494. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas.* Science, 1994. **266**(5182): p. 120-2.
495. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
496. Bae, I., et al., *BRCA1 regulates gene expression for orderly mitotic progression.* Cell Cycle, 2005. **4**(11): p. 1641-66.
497. Allen, C., et al., *More forks on the road to replication stress recovery.* J Mol Cell Biol, 2011. **3**(1): p. 4-12.
498. Bochman, M.L. and A. Schwacha, *The Mcm complex: unwinding the mechanism of a replicative helicase.* Microbiol Mol Biol Rev, 2009. **73**(4): p. 652-83.
499. Truong, L.N. and X. Wu, *Prevention of DNA re-replication in eukaryotic cells.* J Mol Cell Biol, 2011. **3**(1): p. 13-22.
500. Woodward, A.M., et al., *Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress.* J Cell Biol, 2006. **173**(5): p. 673-83.
501. Ge, X.Q., D.A. Jackson, and J.J. Blow, *Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress.* Genes Dev, 2007. **21**(24): p. 3331-41.
502. Ibarra, A., E. Schwob, and J. Mendez, *Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication.* Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8956-61.
503. Gan, N., et al., *Increase of Mcm3 and Mcm4 expression in cervical squamous cell carcinomas.* Eur J Gynaecol Oncol, 2010. **31**(3): p. 291-4.
504. Kikuchi, J., et al., *Minichromosome maintenance (MCM) protein 4 as a marker for proliferation and its clinical and clinicopathological significance in non-small cell lung cancer.* Lung Cancer, 2011. **72**(2): p. 229-37.
505. Ladstein, R.G., et al., *Ki-67 expression is superior to mitotic count and novel proliferation markers PHH3, MCM4 and mitosin as a prognostic factor in thick cutaneous melanoma.* BMC Cancer, 2010. **10**: p. 140.
506. Shima, N., T.R. Buske, and J.C. Schimenti, *Genetic screen for chromosome instability in mice: Mcm4 and breast cancer.* Cell Cycle, 2007. **6**(10): p. 1135-40.
507. Shima, N., et al., *A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice.* Nat Genet, 2007. **39**(1): p. 93-8.
508. Chuang, C.H., et al., *Incremental genetic perturbations to MCM2-7 expression and subcellular distribution reveal exquisite sensitivity of mice to DNA replication stress.* PLoS Genet, 2010. **6**(9): p. e1001110.

509. Kawabata, T., et al., *Stalled fork rescue via dormant replication origins in unchallenged S phase promotes proper chromosome segregation and tumor suppression.* Mol Cell, 2011. **41**(5): p. 543-53.
510. Kunnev, D., et al., *DNA damage response and tumorigenesis in Mcm2-deficient mice.* Oncogene, 2010. **29**(25): p. 3630-8.
511. Bagley, B.N., et al., *A dominantly acting murine allele of Mcm4 causes chromosomal abnormalities and promotes tumorigenesis.* PLoS Genet, 2012. **8**(11): p. e1003034.
512. Mukai, J., P. Suvant, and T.A. Sato, *Nerve growth factor-dependent regulation of NADE-induced apoptosis.* Vitam Horm, 2003. **66**: p. 385-402.
513. Mukai, J., et al., *NADE, a p75NTR-associated cell death executor, is involved in signal transduction mediated by the common neurotrophin receptor p75NTR.* J Biol Chem, 2000. **275**(23): p. 17566-70.
514. Mukai, J., et al., *Structure-function analysis of NADE: identification of regions that mediate nerve growth factor-induced apoptosis.* J Biol Chem, 2002. **277**(16): p. 13973-82.
515. Yoon, K., H.D. Jang, and S.Y. Lee, *Direct interaction of Smac with NADE promotes TRAIL-induced apoptosis.* Biochem Biophys Res Commun, 2004. **319**(2): p. 649-54.
516. Yasui, S., et al., *The TSC1 gene product hamartin interacts with NADE.* Mol Cell Neurosci, 2007. **35**(1): p. 100-8.
517. Alvarez, E., et al., *Characterization of the Bex gene family in humans, mice, and rats.* Gene, 2005. **357**(1): p. 18-28.
518. Tong, X., et al., *NADE (p75NTR-associated cell death executor) suppresses cellular growth in vivo.* Int J Oncol, 2003. **22**(6): p. 1357-62.
519. Kim, A.J., C.S. Lee, and D. Schlessinger, *Bex3 associates with replicating mitochondria and is involved in possible growth control of F9 teratocarcinoma cells.* Gene, 2004. **343**(1): p. 79-89.
520. Romanuik, T.L., et al., *Novel biomarkers for prostate cancer including noncoding transcripts.* Am J Pathol, 2009. **175**(6): p. 2264-76.
521. L'Esperance, S., et al., *Gene expression profiling of paired ovarian tumors obtained prior to and following adjuvant chemotherapy: molecular signatures of chemoresistant tumors.* Int J Oncol, 2006. **29**(1): p. 5-24.
522. Naderi, A., et al., *BEX2 is overexpressed in a subset of primary breast cancers and mediates nerve growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines.* Cancer Res, 2007. **67**(14): p. 6725-36.
523. El Yazidi-Belkoura, I., et al., *Tumor necrosis factor receptor-associated death domain protein is involved in the neurotrophin receptor-mediated antiapoptotic activity of nerve growth factor in breast cancer cells.* J Biol Chem, 2003. **278**(19): p. 16952-6.
524. Descamps, S., et al., *Nerve growth factor stimulates proliferation and survival of human breast cancer cells through two distinct signaling pathways.* J Biol Chem, 2001. **276**(21): p. 17864-70.

525. Massarweh, S., et al., *A phase II neoadjuvant trial of anastrozole, fulvestrant, and gefitinib in patients with newly diagnosed estrogen receptor positive breast cancer.* Breast Cancer Res Treat, 2011. **129**(3): p. 819-27.
526. Purohit, A., S.P. Newman, and M.J. Reed, *The role of cytokines in regulating estrogen synthesis: implications for the etiology of breast cancer.* Breast Cancer Res, 2002. **4**(2): p. 65-9.
527. Berishaj, M., et al., *Stat3 is tyrosine-phosphorylated through the interleukin-6/glycoprotein 130/Janus kinase pathway in breast cancer.* Breast Cancer Res, 2007. **9**(3): p. R32.
528. Lieblein, J.C., et al., *STAT3 can be activated through paracrine signaling in breast epithelial cells.* BMC Cancer, 2008. **8**: p. 302.
529. Bjornstrom, L. and M. Sjoberg, *Signal transducers and activators of transcription as downstream targets of nongenomic estrogen receptor actions.* Mol Endocrinol, 2002. **16**(10): p. 2202-14.

6. Supplementary

6.1. R-Scripts

```
#Cross-Platform Integration Data Pre-processing and Analysis Illumina and Affymetrix
#Arran Turnbull

##############################################################################################
#Pre-processing of Illumina Data
##############################################################################################

#QC data in GenomeStudio removing samples with less than 9000 detected genes below a detection P value of 0.05
#Export Probe Profile from Genome Studio

#Load required programs in R
library(lumi)

#Load data using lumiR function - do not convert IDs
filename="Illum_SPPraw.txt"
raw.lumi <- lumiR(filename, convertNuID = FALSE)

#Log2 Transformation (Log2)
#For other transformation methods change "log2" for: "vst", "cubicRoot"
lumi.T=lumiT(raw.lumi, method=c("log2"), verbose=TRUE)

#Quantile Normalisation (QN)
#For other normalisation methods change "quantile" for: "rsn", "ssn", "loess", "vsn", "rankinvariant"
lumi.N=lumiN(lumi.T, method=c("quantile"), verbose=TRUE)

#Perform quality control
lumi.N.Q=lumiQ(lumi.N)
summary(lumi.N.Q, 'QC')

#Create an expression matrix of the normalised and filtered data
data.processed <- exprs(lumi.N.Q)

#Write data to directory as a .txt file (make sure that row names and column names are displayed correctly in file)
write.table(data.processed,file="Illum_QN.txt", append=FALSE, quote=FALSE, sep="\t", row.names=TRUE, col.names=TRUE)

#Filter Data
#In Excel, use detection P values generated from GenomeStudio - remove probes which are not detected with a P-value of under 0.05
in the total number of samples minus 3

#Reannotate Data
#In Excel reannotate each probe to an Ensembl (ENSM) gene ID. Use the 50mer probe sequences exported from GenomeStudio to
look up mappings in the following resources
#reMOAT: http://remoat.sysbiol.cam.ac.uk/search.php
#BioMART: http://www.ensembl.org/biomart (Ensembl Human release 68 July 2012)
#BLAST: http://www.ensembl.org/Homo_sapiens/blastview
#Only retain probes where mappings to a an Ensembl gene agree with at least two of the resources

#Combine Multiple Probes
#In Excel ,where a single ENSM gene is represtend by multiple probes - calculate an average

####################################
#Optional Batch Correction
####################################

#Load filtered, reannotated ENSM data
data.processed1=read.delim("Illum_QN_filt_ENSM.txt",header=TRUE, sep="\t", row.names=1)

#Analysis for the presence of batch effects
#Create multidimensional scaling (MDS) plot
dataMatrix <- data.processed
```

```
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
write.table(z, "L3_bg_QN_mds.txt", quote=F, sep="\t")

#Correct with ComBat if necessary

#Batch correction using ComBat
#ComBat - Download ComBat R script from http://statistics.byu.edu/johnson/ComBat/Download.html and place the file in the
directory
#Create the Sample Information File and save as a .txt file to the directory
source('Combat.R')
ComBat('Illum_QN_filt_ENSM.txt','Sample Information File (Combat).txt',skip=1)

#Load ComBat corrected data
data.combat=read.delim("Illum_QN_filt_ENSM-ComBat.txt",header=TRUE, sep="\t", row.names=1)

#Assess correction of data
#Create multidimensional scaling (MDS) plot
dataMatrix <- data.combat
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
write.table(z, "Illum_QN_filt_ENSM-ComBat_mds.txt", quote=F, sep="\t")

#Illumina data ready for integration

##############################################################################################
#Pre-processing of Affymetrix Data
##############################################################################################

#Place CEL files in the working directory

#Load required programs
library(affy)

#Download the alternate_cdf from "http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp"
#Place the alternate_cdf file in the root of the working directory
#In this analysis I used alternate_cdf for ENSG version 15
#Read .CEL files into R and load the required alternate_cdf file
data<-ReadAffy(cdfname = "HGU133A_HS_ENSG")

#Apply the Robust Multichip Average (RMA) Algorithm
#.CEL files which contain expression values for each probe are combined by this alogirthm to give an expression for each probe set
(transcript represented on the chip)
#RMA Background Adjustment
#Quantile Normalisation
#Summary of Expression by medianpolish method
#Log2 Transformation

data.processed=rma(data)
write.exprs(data.processed, file="Affy-RMA.txt")
```

```
#mas5 detection calls
calls<-mas5calls(data)
write.exprs(calls, file="Affy-mas5DetCalls.txt")

#Filter RMA data in Excel based on mas5 detection calls: Remove all are A or M in the total number of the samples
#Present: p < 0.04
#Marginal: p between 0.04 and 0.06
#Absent: p > 0.06

####################################
#Optional Batch Correction
####################################

#Load the filtered and normalised Affyemtrix file
data.filtered=read.delim("Affy-RMA_filt.txt",header=TRUE, sep="\t", row.names=1)

#Analysis for the presence of batch effects
#Create multidimensional scaling (MDS) plot
dataMatrix <- data.filtered
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
write.table(z, "Affy-RMA_filt_mds.txt", quote=F, sep="\t")

#Correct with ComBat if necessary

#Batch correction using ComBat
#ComBat - Download ComBat R script from http://statistics.byu.edu/johnson/ComBat/Download.html and place the file in the
directory
#Create the Sample Information File and save as a .txt file to the directory
source('Combat.R')
ComBat('Affy-RMA_filt_forComBat.txt','Sample Information File (Combat).txt',skip=1)

#Load ComBat corrected data
data.combat=read.delim("Affy-RMA_filt_ComBat.txt",header=TRUE, sep="\t", row.names=1)

#Assess correction of data
#Create multidimensional scaling (MDS) plot
dataMatrix <- data.combat
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
write.table(z, "Affy-RMA_filt_ComBat_mds.txt", quote=F, sep="\t")

#Affy data ready for integration

#############################################################################################
#L2 Affymetrix and L3 Illumina Combined Data
#############################################################################################

#Combine Illumina and Affymetrix datasets in Excel - retaining only ENSM genes where data is available from both platforms.
#Rename the file to "AffyIllumCombined.txt"
```

```
#Load Affy and Illumina combined data
data.combined=read.delim("AffyIllumCombined.txt",header=TRUE, sep="\t", row.names=1)

#View a box plot of expression for the combined dta
boxplot(data.combined, main="Amplitude of Expression Plot - Combined Data")

#Create multidimensional scaling (MDS) plot
dataMatrix <- data.combined
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
write.table(z, "AffyIllumCombined_mds.txt", quote=F, sep="\t")

#############################################################################################
#Optional Cross Platform Correction Methods
#############################################################################################

#################################################################
#Mean Centering
#################################################################

#Perform mean centering of data in Cluster 3
#http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm

#################################################################
#ComBat Correction
#################################################################

#Batch correction using ComBat
#ComBat - Download ComBat R script from http://statistics.byu.edu/johnson/ComBat/Download.html and place the file in the
directory
#Create the Sample Information File and save as a .txt file to the directory
source('Combat.R')
ComBat('AffyIllumCombined.txt','Sample Information File (Combat).txt',skip=1

#################################################################
#XPN correction
#################################################################

#Perform XPN online at ArrayMining
#http://arraymining.org/

#################################################################
#DWD-standard correction
#################################################################

#Perform DWD-standard in downloaded DWD software
#https://genome.unc.edu/pubsup/dwd/

#############################################################################################
#Additional R Codes
#############################################################################################

#################################################################
#Code for Similarity/Dissimilairy Heatmaps
#################################################################

#Drawing similarity/dissimilairy heatmap in R using Gplots
#Determine correlation between columns of dataset
data.cor <- cor(data.1, method="pearson")
```

```
library(gplots)
breaks <- seq(from = 0.90, to = 1, length = 20)
colors = "heat.colors"
heatmap.2(data.cor, dendrogram="none", col=colors, Rowv=F, Colv=F, scale="none", key=TRUE, symkey=FALSE,
density.info="none", trace="none", cexRow=1, cexCol=1, breaks=breaks)
write.table(data.cor, "Data_cor.txt", quote=F, sep="\t")

#################################################################
#Differential gene expression analysis
#################################################################

#Differential gene expression analysis was performed in MeV
#http://www.tm4.org/mev.html

#################################################################
#Expression heatmaps
#################################################################

#Expression heatmaps were created in MeV
#http://www.tm4.org/mev.html

#################################################################
#Multidimensional Scaling
#################################################################

dataMatrix <- data.1
d1=dist(t(scale(dataMatrix)))
dN=dimnames(dataMatrix)[[2]]
nS=length(dN)
d1M=as.matrix(d1)
dimnames(d1M)=list(dN, dN)
d10.2.mds=cmdscale(d1M, k=2)
x=d10.2.mds[,1]
y=d10.2.mds[,2]
plot(x,y)
z=cbind(x,y)
```

6.2. Proliferation Set Gene List

ENSM Gene ID	Office Gene Symbol	Gene Name
ENSG00000168274	HIST1H2AB	histone cluster 1, H2ae; histone cluster 1, H2ab
ENSG00000168274	HIST1H2AE	histone cluster 1, H2ae; histone cluster 1, H2ab
ENSG00000167325	RRM1	ribonucleotide reductase M1
ENSG00000117724	CENPF	centromere protein F, 350/400ka (mitosin)
ENSG00000056736	IL17RB	interleukin 17 receptor B
ENSG00000122952	ZWINT	ZW10 interactor
ENSG00000138346	DNA2	DNA replication helicase 2 homolog (yeast)
ENSG00000111602	TIMELESS	timeless homolog (Drosophila)
ENSG00000126803	HSPA2	heat shock 70kDa protein 2
ENSG00000102384	CENPI	centromere protein I
ENSG00000164104	HMGB2	high-mobility group box 2
ENSG00000104738	MCM4	minichromosome maintenance complex component 4
ENSG00000164109	MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)
ENSG00000100479	POLE2	polymerase (DNA directed), epsilon 2 (p59 subunit)
ENSG00000145386	CCNA2	cyclin A2
ENSG00000101773	RBBP8	retinoblastoma binding protein 8
ENSG00000131747	TOP2A	topoisomerase (DNA) II alpha 170kDa
ENSG00000164032	H2AFZ	H2A histone family, member Z
ENSG00000213066	FGFR1OP	FGFR1 oncogene partner
ENSG00000136518	ACTL6A	actin-like 6A
ENSG00000163808	KIF15	kinesin family member 15
ENSG00000173473	SMARCC1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1
ENSG00000126787	DLGAP5	discs, large (Drosophila) homolog-associated protein 5
ENSG00000076003	MCM6	minichromosome maintenance complex component 6
ENSG00000112312	GMNN	geminin, DNA replication inhibitor
ENSG00000087586	AURKA	aurora kinase A; aurora kinase A pseudogene 1
ENSG00000087586	AURKAPS1	aurora kinase A; aurora kinase A pseudogene 1
ENSG00000073111	MCM2	minichromosome maintenance complex component 2
ENSG00000138175	ARL3	ADP-ribosylation factor-like 3
ENSG00000111666	CHPT1	choline phosphotransferase 1
ENSG00000166451	CENPN	centromere protein N
ENSG00000154473	BUB3	budding uninhibited by benzimidazoles 3 homolog (yeast)
ENSG00000066279	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)
ENSG00000134057	CCNB1	cyclin B1
ENSG00000170312	CDK1	cell division cycle 2, G1 to S and G2 to M
ENSG00000138160	KIF11	kinesin family member 11
ENSG00000197061	HIST1H4C	histone

ENSG00000197061	HIST1H4D	histone
ENSG00000197061	HIST1H4F	histone
ENSG00000197061	HIST1H4H	histone
ENSG00000197061	HIST1H4I	histone
ENSG00000197061	HIST1H4J	histone
ENSG00000197061	HIST2H4A	histone
ENSG00000197061	HIST1H4A	histone
ENSG00000197061	HIST1H4B	histone
ENSG00000197061	HIST2H4B	histone
ENSG00000197061	HIST1H4E	histone
ENSG00000197061	HIST1H4K	histone
ENSG00000197061	HIST1H4L	histone
ENSG00000197061	HIST4H4	histone
ENSG00000088325	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)
ENSG00000131153	GINS2	GINS complex subunit 2 (Psf2 homolog)
ENSG00000110092	CCND1	cyclin D1
ENSG00000140443	IGF1R	insulin-like growth factor 1 receptor
ENSG00000132646	PCNA	proliferating cell nuclear antigen
ENSG00000137804	NUSAP1	nucleolar and spindle associated protein 1
ENSG00000106462	EZH2	enhancer of zeste homolog 2 (Drosophila)
ENSG00000198087	CD2AP	CD2-associated protein
ENSG00000188486	H2AFX	H2A histone family, member X
ENSG00000149554	CHEK1	CHK1 checkpoint homolog (S. pombe)
ENSG00000138180	CEP55	centrosomal protein 55kDa
ENSG00000080986	NDC80	NDC80 homolog, kinetochore complex component (S. cerevisiae)
ENSG00000031691	CENPQ	centromere protein Q
ENSG00000117399	CDC20	cell division cycle 20 homolog (S. cerevisiae)
ENSG00000109084	TMEM97	transmembrane protein 97
ENSG00000100526	CDKN3	cyclin-dependent kinase inhibitor 3
ENSG00000162702	ZNF281	zinc finger protein 281
ENSG00000163918	RFC4	replication factor C (activator 1) 4, 37kDa
ENSG00000109805	NCAPG	non-SMC condensin I complex, subunit G
ENSG00000168061	SAC3D1	SAC3 domain containing 1
ENSG00000171848	RRM2	ribonucleotide reductase M2 polypeptide
ENSG00000196074	SYCP2	synaptonemal complex protein 2
ENSG00000213005	PTTG3P	pituitary tumor-transforming 3
ENSG00000198901	PRC1	protein regulator of cytokinesis 1
ENSG00000175063	UBE2C	ubiquitin-conjugating enzyme E2C

6.3. 131 Genes Input to CART Gene List

ENSM Gene ID	Offical Gene Symbol	Gene Name
ENSG00000196139	AKR1C3	aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)
ENSG00000169744	LDB2	LIM domain binding 2
ENSG00000111404	RERGL	RERG/RAS-like
ENSG00000138755	CXCL9	chemokine (C-X-C motif) ligand 9
ENSG00000122952	ZWINT	ZW10 interactor
ENSG00000115414	FN1	fibronectin 1
ENSG00000179041	RRS1	RRS1 ribosome biogenesis regulator homolog (S. cerevisiae)
ENSG00000213088	DARC	Duffy blood group, chemokine receptor
ENSG00000113387	SUB1	SUB1 homolog (S. cerevisiae)
ENSG00000133740	E2F5	E2F transcription factor 5, p130-binding
ENSG00000112936	C7	complement component 7
ENSG00000131747	TOP2A	topoisomerase (DNA) II alpha 170kDa
ENSG00000129226	CD68	CD68 molecule
ENSG00000124831	LRRFIP1	leucine rich repeat (in FLII) interacting protein 1
ENSG00000139629	GALNT6	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)
ENSG00000175745	NR2F1	nuclear receptor subfamily 2, group F, member 1
ENSG00000112742	TTK	TTK protein kinase
ENSG00000163808	KIF15	kinesin family member 15
ENSG00000039068	CDH1	cadherin 1, type 1, E-cadherin (epithelial)
ENSG00000129596	CDO1	cysteine dioxygenase, type I
ENSG00000170476	MGC29506	hypothetical protein MGC29506
ENSG00000146278	PNRC1	proline-rich nuclear receptor coactivator 1
ENSG00000087586	AURKA	aurora kinase A; aurora kinase A pseudogene 1
ENSG00000087586	AURKAPS1	aurora kinase A; aurora kinase A pseudogene 1
ENSG00000168329	CX3CR1	chemokine (C-X3-C motif) receptor 1
ENSG00000111666	CHPT1	choline phosphotransferase 1
ENSG00000156234	CXCL13	chemokine (C-X-C motif) ligand 13
ENSG00000158488	CD1E	CD1e molecule
ENSG00000107562	CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)
ENSG00000186918	ZNF395	zinc finger protein 395
ENSG00000122861	PLAU	plasminogen activator, urokinase
ENSG00000175445	LPL	lipoprotein lipase
ENSG00000097046	CDC7	cell division cycle 7 homolog (S. cerevisiae)
ENSG00000066279	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)
ENSG00000164949	GEM	GTP binding protein overexpressed in skeletal muscle
ENSG00000170458	CD14	CD14 molecule

ENSG00000109436	TBC1D9	TBC1 domain family, member 9 (with GRAM domain)
ENSG00000170312	CDK1	cell division cycle 2, G1 to S and G2 to M
ENSG00000074410	CA12	carbonic anhydrase XII
ENSG00000067113	PPAP2A	phosphatidic acid phosphatase type 2A
ENSG00000216490	IFI30	interferon, gamma-inducible protein 30
ENSG00000148154	UGCG	UDP-glucose ceramide glucosyltransferase
ENSG00000074416	MGLL	monoglyceride lipase
ENSG00000137947	GTF2B	general transcription factor IIB
ENSG00000131153	GINS2	GINS complex subunit 2 (Psf2 homolog)
ENSG00000137804	NUSAP1	nucleolar and spindle associated protein 1
ENSG00000060718	COL11A1	collagen, type XI, alpha 1
ENSG00000164162	ANAPC10P1	anaphase promoting complex subunit 10; anaphase promoting complex subunit 10 pseudogene
ENSG00000164162	ANAPC10	anaphase promoting complex subunit 10; anaphase promoting complex subunit 10 pseudogene
ENSG00000129055	ANAPC13	anaphase promoting complex subunit 13
ENSG00000108691	CCL2	chemokine (C-C motif) ligand 2
ENSG00000153563	CD8A	CD8a molecule
ENSG00000160712	IL6R	interleukin 6 receptor
ENSG00000189221	MAOA	monoamine oxidase A
ENSG00000138180	CEP55	centrosomal protein 55kDa
ENSG00000176887	SOX11	SRY (sex determining region Y)-box 11
ENSG00000177565	TBL1XR1	transducin (beta)-like 1 X-linked receptor 1
ENSG00000158402	CDC25C	cell division cycle 25 homolog C (S. pombe)
ENSG00000211445	GPX3	glutathione peroxidase 3 (plasma)
ENSG00000123384	LRP1	low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor)
ENSG00000128245	YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide
ENSG00000157404	LOC652799	similar to Mast/stem cell growth factor receptor precursor (SCFR)
ENSG00000157404	LOC653882	similar to Mast/stem cell growth factor receptor precursor (SCFR)
ENSG00000157404	KIT	similar to Mast/stem cell growth factor receptor precursor (SCFR)
ENSG00000165304	MELK	maternal embryonic leucine zipper kinase
ENSG00000171848	RRM2	ribonucleotide reductase M2 polypeptide
ENSG00000158473	CD1D	CD1d molecule
ENSG00000198901	PRC1	protein regulator of cytokinesis 1
ENSG00000175054	LOC648152	ataxia telangiectasia and Rad3 related; similar to ataxia telangiectasia and Rad3 related protein
ENSG00000175054	LOC651921	ataxia telangiectasia and Rad3 related; similar to ataxia telangiectasia and Rad3 related protein
ENSG00000175054	ATR	ataxia telangiectasia and Rad3 related; similar to ataxia telangiectasia and Rad3 related protein
ENSG00000091831	ESR1	estrogen receptor 1
ENSG00000117724	CENPF	centromere protein F, 350/400ka (mitosin)
ENSG00000151726	ACSL1	acyl-CoA synthetase long-chain family member 1

ENSG00000137563	GGH	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)
ENSG00000151725	MLF1IP	MLF1 interacting protein
ENSG00000163739	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
ENSG00000134352	IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)
ENSG00000138346	DNA2	DNA replication helicase 2 homolog (yeast)
ENSG00000113558	SKP1	S-phase kinase-associated protein 1
ENSG00000116478	HDAC1	histone deacetylase 1
ENSG00000104738	MCM4	minichromosome maintenance complex component 4
ENSG00000147257	GPC3	glypican 3
ENSG00000164109	MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)
ENSG00000169032	MAP2K1	mitogen-activated protein kinase kinase 1
ENSG00000145386	CCNA2	cyclin A2
ENSG00000164754	RAD21	RAD21 homolog (S. pombe)
ENSG00000123689	G0S2	G0/G1switch 2
ENSG00000101773	RBBP8	retinoblastoma binding protein 8
ENSG00000094880	CDC23	cell division cycle 23 homolog (S. cerevisiae)
ENSG00000115641	FHL2	four and a half LIM domains 2
ENSG00000123124	WWP1	WW domain containing E3 ubiquitin protein ligase 1
ENSG00000197905	TEAD4	TEA domain family member 4
ENSG00000090920	LOC100133944	Fc fragment of IgG binding protein; similar to IgGFc-binding protein precursor (FcgammaBP) (Fcgamma-binding protein antigen)
ENSG00000090920	FCGBP	Fc fragment of IgG binding protein; similar to IgGFc-binding protein precursor (FcgammaBP) (Fcgamma-binding protein antigen)
ENSG00000131471	AOC3	amine oxidase, copper containing 3 (vascular adhesion protein 1)
ENSG00000105664	COMP	cartilage oligomeric matrix protein
ENSG00000172201	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
ENSG00000006007	GDE1	glycerophosphodiester phosphodiesterase 1
ENSG00000168040	FADD	Fas (TNFRSF6)-associated via death domain
ENSG00000085415	SEH1L	SEH1-like (S. cerevisiae)
ENSG00000104267	CA2	carbonic anhydrase II
ENSG00000137033	IL33	interleukin 33
ENSG00000166483	WEE1	WEE1 homolog (S. pombe)
ENSG00000155792	DEPDC6	DEP domain containing 6
ENSG00000137573	SULF1	sulfatase 1
ENSG00000162604	TM2D1	TM2 domain containing 1
ENSG00000073111	MCM2	minichromosome maintenance complex component 2
ENSG00000135218	CD36	CD36 molecule (thrombospondin receptor)
ENSG00000154473	BUB3	budding uninhibited by benzimidazoles 3 homolog (yeast)
ENSG00000130303	BST2	NPC-A-7; bone marrow stromal cell antigen 2
ENSG00000132341	RAN	RAN, member RAS oncogene family
ENSG00000134057	CCNB1	cyclin B1

ENSG00000172216	CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
ENSG00000205336	GPR56	G protein-coupled receptor 56
ENSG00000067225	LOC652797	similar to Pyruvate kinase
ENSG00000067225	PKM2	similar to Pyruvate kinase
ENSG00000123349	PFDN5	prefoldin subunit 5
ENSG00000172724	CCL19	chemokine (C-C motif) ligand 19
ENSG00000187288	CIDEC	cell death-inducing DFFA-like effector c
ENSG00000160752	FDPS	farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase)
ENSG00000131788	PIAS3	protein inhibitor of activated STAT, 3
ENSG00000166681	NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1
ENSG00000088325	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)
ENSG00000110092	CCND1	cyclin D1
ENSG00000010610	CD4	CD4 molecule
ENSG00000132646	PCNA	proliferating cell nuclear antigen
ENSG00000164128	NPY1R	neuropeptide Y receptor Y1
ENSG00000142684	ZNF593	zinc finger protein 593
ENSG00000149554	CHEK1	CHK1 checkpoint homolog (S. pombe)
ENSG00000108700	CCL8	chemokine (C-C motif) ligand 8
ENSG00000134480	CCNH	cyclin H
ENSG00000197766	CFD	complement factor D (adipsin)
ENSG00000117525	F3	coagulation factor III (thromboplastin, tissue factor)
ENSG00000117399	CDC20	cell division cycle 20 homolog (S. cerevisiae)
ENSG00000161570	CCL5	chemokine (C-C motif) ligand 5
ENSG00000109805	NCAPG	non-SMC condensin I complex, subunit G
ENSG00000167286	CD3D	CD3d molecule, delta (CD3-TCR complex)
ENSG00000139687	RB1	retinoblastoma 1

6.4. LREC Approval

Lothian NHS Board

Lothian Research Ethics Committee
Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9000
Fax 0131 536 9346
www.nhslothian.scot.nhs.uk

NHS Lothian

06 MAR 2002

Mr J Michael Dixon
Consultant Surgeon
Edin Breast Unit
Western General Hosp Edinburgh EH4 2XU

Date 05 March 2002
Your Ref
Our Ref LREC//2001/8/81

Enquiries to Joyce Clearie
Extension 89022
Direct Line 0131 536 9022
Email joyce.clearie@nhs.scot.nhs.uk

Dear Mr Dixon,

STUDY TO DETERMINE WHETHER SHORT TERM CHANGES DURING PRIMARY ENDOCRINE TREATMENT PREDICT FOR RESPONSE

Thank you for submitting the above protocol for ethical approval. The Medicine/Clinical Oncology II Research Ethics Sub-Committee has considered this study and has granted ethical approval. An official Certificate of Ethical Review is enclosed together with a list of Sub-Committee members.

Under the terms of the Scottish Office Home and Health Department Guidelines on Local Research Ethics Committees this decision has been notified to the NHS body under the auspices of which the research is intended to take place. It is that NHS body which has the responsibility of deciding whether or not the research should go ahead taking account of the advice of the Research Ethics Sub-Committee and from whom you must obtain management approval before any work on the study can proceed.

Details of the Lothian Research Ethics Committee and its documentation can be found on http://www.nhslothian.scot.nhs.uk/nhs_lothian/about_lothian_health/lrec/index.html

Yours sincerely

Joyce Clearie

JOYCE CLEARIE
Committee Administrator

Headquarters
Deaconess House 148 Pleasance Edinburgh EH8 9RS

Chair Brian Cavanagh
Chief Executive James Barbour O.B.E.
Lothian NHS Board is the common name of Lothian Health Board

Lothian NHS Board

Lothian Research Ethics Committee
Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9000
Fax 0131 536 9346
www.nhslothian.scot.nhs.uk

NHS Lothian

Mr J Michael Dixon
Academic Office
Paderewski Building
Western General Hospital
Edinburgh EH4 2XU

Date 25 January 2002
Your Ref
Our Ref LREC//2001/8/80

Enquiries to Joyce Clearie
Extension 89022
Direct Line 0131 536 9022
Email joyce.clearie@nhs.scot.nhs.uk

Dear Mr Dixon,

OBTAINING BREAST TISSUE AT DIAGNOSIS BY CORE BIOPSY IN PATIENTS WITH BREAST CANCER

Thank you for submitting the above protocol for ethical approval. The Medicine/Clinical Oncology II Research Ethics Sub-Committee has considered this study and has granted ethical approval. An official Certificate of Ethical Review is enclosed together with a list of Sub-Committee members.

Under the terms of the Scottish Office Home and Health Department Guidelines on Local Research Ethics Committees this decision has been notified to the NHS body under the auspices of which the research is intended to take place. It is that NHS body which has the responsibility of deciding whether or not the research should go ahead taking account of the advice of the Research Ethics Sub-Committee and from whom you must obtain management approval before any work on the study can proceed.

Yours sincerely

Joyce Clearie

JOYCE CLEARIE
Committee Administrator

Headquarters
Deaconess House 148 Pleasance Edinburgh EH8 9RS

Chair Brian Cavanagh
Chief Executive James Barbour O.B.E.

Lothian NHS Board

Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9000
Fax 0131 536 9009
www.nhslothian.scot.nhs.uk

NHS Lothian

13 February 2007

Mr J. M. Dixon
Consultant Surgeon and Senior Lecturer in Surgery
Lothian University Hospitals Division
Edinburgh Breast Unit
Western General Hospital
Crewe Road South, Edinburgh
EH4 2XR

Dear Mr Dixon

Full title of study: **Edinburgh Breast Unit and Edinburgh Cancer Research Centre tissue and body fluids collection with linkage to pathological and clinical data**
REC reference number: **06/S1103/65**

Thank you for your letter of 5 February 2007, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered on behalf of the Committee by the Chair, Dr Christine West.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application		06 December 2006
Investigator CV		
Protocol	1.1	04 December 2006
Covering Letter	2 - with changes	05 February 2007
Covering Letter	1 - with original submission	06 December 2006
Letter of invitation to participant	1.1	04 December 2006
GP/Consultant Information Sheets	1.0	04 December 2006
Participant Information Sheet	1.1	04 December 2006
Participant Consent Form	1.2	05 February 2007
Response to Request for Further Information		

Lothian NHS Board

Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9000
Fax 0131 536 9009
www.nhslothian.scot.nhs.uk

NHS Lothian

13 February 2007

Mr J. M. Dixon
Consultant Surgeon and Senior Lecturer in Surgery
Lothian University Hospitals Division
Edinburgh Breast Unit
Western General Hospital
Crewe Road South, Edinburgh
EH4 2XR

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

REC Reference Number 06/S1103/65 Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely

pp **Chair**
Lothian Local Research Ethics Committee 03
Email: elizabeth.harden@lhb.scot.nhs.uk

Enclosures: *Standard approval conditions*
Site approval form

Copy to: Professor Heather Cubie
Lothian University Hospitals Division
Research & Development Office
Queen's Medical Research Institute
47 Little France Crescent, Edinburgh
EH16 4TJ

R&D Department for NHS Lothian

6.5. R&D Approval

University Hospitals Division

Queen's Medical Research Institute
47 Little France Crescent, Edinburgh, EH16 4TJ

HAC/SM/approval/Dixon/2d/3/3a/3b/3e

7th February 2007

Mr J Mike Dixon
Consultant Surgeon and Senior Lecturer in Surgery
Edinburgh Breast Unit
Western General Hospital
Crewe Road South
Edinburgh
EH4 2XR

NHS Lothian

RESEARCH & DEVELOPMENT OFFICE
Room E1.12

Tel: 0131 242 3330
Fax: 0131 242 3343
Email: R&DOffice@lothian.scot.nhs.uk

Director: *Professor Heather A Cuble*

R&D Governance Manager *Dr Tina McLelland*

PA to Professor Cuble & Dr McLelland: Miss Jill Dobbie

Commercial Research Manager: *Dr Douglas Young*

Research Manager Capacity & Capability: *Dr Janet Hanley*

Research Governance Co-ordinator: *Mrs Susan Shepherd*

Information & Knowledge Manager *Miss Heather Cooper*

AHP Research & Development Facilitator: *Dr Colette Fulton*

Accountant: *Ms Sheevaun McIntyre*

Assistant Accountant: *Mr Neil McLean*

Trial Support Officer: *Ms Dorothy Alden*

Office Manager: *Mrs Glynis Cleland*

Administrative Assistant: *Ms Sandra Muir*

St John's – Administrator: *Mrs Anne Addison*

Dear Mr Dixon

MREC No: N/A
CRF No: Not yet known
LREC No: 06/S1103/65
R&D ID No: 2007/W/BIU/01
Title of Research Edinburgh Breast Unit and Edinburgh Cancer Research Centre tissue and body fluids collection with linkage to pathological and clinical data.

Protocol No/Acronym: N/A

The above project has undergone an assessment of risk to NHS Lothian and review of resource and financial implications. I am satisfied that all the necessary arrangements have been set in place and that all Departments contributing to the project have been informed.

As this is a single site project involving patients, NHS Lothian agrees to act as Sponsor for the study.

Use of Tissue or Samples

- The study involves the use of patient tissue or samples. You must be familiar with NHS Lothian's Tissue Policy and abide by its conditions and also with all regulations in place at the time. Approval is subject to the prevailing legal requirements.
- Approval for the use of tissue is restricted to the protocol associated with this application, but may include additional collaborators within University of Edinburgh. Collaborators who are not named in the original protocol require to be notified to local REC.
- If material is to be transferred to academic collaborators outwith University of Edinburgh or to any commercial entity then a Material Transfer Agreement must be obtained from the R&D Office and signed by all relevant parties prior to transfer of the material. Such collaborations must be fully discussed with the R&D Office and management approval is only effective once this is in place.
- I note that the use of excess diagnostic specimens has been discussed and approved by the Pathologist with delegated authority.
- I note that additional samples will be taken for the study and that this will be done with the patient's explicit consent.
- Where a prospective collection of material is intended, explicit consent must be obtained for storage and use in possible future research projects. A model consent forms along the lines recommended in the MRC Guidelines 'Human Tissue and Biological Samples for Use in Research' should be used and a copy lodged with the R&D Office.

On behalf of the Chief Executive and Medical Director, I am happy to grant management approval from NHS Lothian to allow the project to commence, subject to the approval of the appropriate Research Ethics Committee(s) having also been obtained. You should note that any substantial amendments must be notified to the relevant Research Ethics Committee and to R&D Management with approval being granted from both before the amendments are made.

Please note that under Section A, Q35, NHS Lothian provides indemnity for negligence for NHS and Honorary clinical staff for research associated with their clinical duties. It is not empowered to provide non-negligent indemnity cover for patients. NHS Lothian does not provide indemnity against negligence for healthy volunteer studies. This is the personal responsibility of both NHS and honorary employees and is usually arranged with a medical defence organisation or through the University of Edinburgh.

This letter of approval is your assurance that NHS Lothian is satisfied with your study. As Chief Investigator or local Principal Investigator, you should be fully committed to your responsibilities within the Research Governance Framework for Health and Community Care, an extract of which is attached to this letter.

Yours sincerely

Heather A Cubie

Professor Heather A Cubie
R&D Director

Enc
- Research Governance Certificate ☑ (to be signed and returned)
- NRR authorisation ☑ (to be signed and returned)
- Tissue Policy (if applicable) ☐
- MTA (if applicable) ☐ (to be signed and returned by the recipient)

Copies Administrators, Research Ethics Committee
CRF