GENOMIC COPY NUMBER ALTERATIONS IN POORLY DIFFERENTIATED BREAST CANCER

by

Dr Sileida Yunexy Oliveros Delgado

Submitted to the Imperial College for the degree of

MD (Res)

Department of Surgery and Cancer
Faculty of Medicine
Imperial College London
Hammersmith Campus

December 2011

DECLARATION

The development of the present project was based on previous Breast cancer research performed under the appropriate supervision by Prof Leonard and Prof Thomas. Breast cancer cohort data was collected by Dr Mariko Morishita while working in Breast cancer research in the South Wales cancer centre. As part of my project, I cleaned, coded and performed the statistical analysis of the collected data for the characterisation of the cohort according to clinico-pathological parameters and their association with survival as it will be presented in Chapter 2.

Tissue Microarray (TMA) from selected samples was performed at the Pathology Department, Singleton Hospital, Wales under Prof Thomas' supervision. I coordinated TMA staining and central review of all selected cases for array Comparative Genomic Hybridisation (aCGH) working in conjunction with the Diagnostic laboratory at Charing Cross Hospital and Dr Mihir Gudi, a dedicated Breast pathologist. I transcribed the results on the TMA into the Clinical database. I selected samples to be arrayed, extracted DNA and assessed its suitability for aCGH experiments. Under Dr Unger's supervision, I optimised and performed aCGH in all the samples arrayed in the UK. This included all aCGH steps from labelling, hybridisation, washing and scanning of the arrays. Biostatistical analysis of aCGH data was performed by Dr Kristian, Unger.

Although I carried out work on cell lines in the laboratory, the DNA from MCF-7 and Tamoxifen resistant cell lines was provided by Prof Nicholson, from Cardiff University, Wales. I performed aCGH in cell lines and grew cell lines used as positive controls in my experiments.

Finally, I wrote the content of this MD thesis and the papers arising from this breast cancer research working closely with my supervisors and all the team involved in its development.

ACKNOWLEDGEMENT

I would like to show my gratitude to so many people who made this research possible. I am indebted to Prof Leonard and Prof Thomas, my supervisors, who gave me not only the opportunity to fulfil my ambition to do research in breast cancer but also a pertinent project and the means to have it completed. Special thanks goes to my co-supervisor, Dr Kristian Unger whose constant guidance introduced me in the "lab" world and make this project possible. I would also like to express my gratitude to Dr Bill Mathison, my next desk neighbour, for all his encouragement, guidance and friendship. To Marohk, my very supportive ex-lab manager and Dr Mihir for his experienced contribution in the pathological assessment of the samples.

It was an honour to have received the Gunnar Nilsson Cancer Treatment Trust Fellowship which greatly funded this research along with European Commission (TRANSBIG). Without their financial support this research would not have been possible.

To my father, example of excellence and perseverance. The one always pursuing perfection, even after death. From him my strength and ambitions. To my mother, loving heart full of blessings and devotion. Always checking all in place. From her my weakness for those in need, particularly my patients. Just hoping you are as proud of me as I have always been of you both. To my sisters Aury, Maylen, Alicia, Yamile and Yorley. For the madness of a house full of noise and happiness.

For my baby Gabriel and my grown-up Alejandro. To whom I will continue to dedicate my life, the reason to prosper, the essence of simplicity ... what really matters. And for you Alejandro, the one always there, loving, understanding and supporting. Both in a personal and professional level. Since those days of college parties to these sleepless days of writing and nappies. For you my love.

LIST OF ABREVIATIONS

aCGH	Array Comparative Genomic Hybridisation
ADC	Adenocarcinoma
AI	Aromatase Inhibitor
AJCC	American Joint Commission on Cancer
AS	Anti-oestrogene
ATP	Adenosine-5'-triphosphate
BAC	Bacterial Artificial Chromosomes
BC	Breast cancer
BCOU	British Columbia Breast Cancer Outcomes Unit
BCS	Breast Conserving Surgery
BCSS	Breast cancer–specific survival
Bp	Base pair
CAPweb	CGH array Analysis Platform on the Web
cDNA	Complementary deoxyribonuclei acid
CGH	Comparative genomic hybridization
CNA	Copy number alteration
CNS	Copy number signature
CNV	Copy number variation
CSS	Cancer specific survival
CT	Computer Tomography
CTP	Deoxycytidine triphosphate
DBCG	Danish Breast Cancer Cooperative Group
DCIS	Ductal Carcinoma In-situ
DDFS	Distant Disease Free Survival
DFS	Disease Free Survival

DNA	Deoxyribonuclei acid
DOP-PCR	Degenerate oligonucleotide polymerase chain reaction
EFS	Event Free Survival
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ER	Oestrogen receptor
ERE	Oestrogen receptor elements
EtOH	Ethanol
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
FFT	Fresh frozen tissue
FISH	Fluorescence in-situ hybridisation
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GCRMA	GeneChip robust multi-array average
GEO	Gene expression Omnibus
GEP	Gene expression profile
GES	Gene expression signature
GGI	Genomic grade index
GnRH	Gonadotropin-releasing hormone
GTP	Guanosine-5'-triphosphate
HER	Human epidermal growth factor receptor
HER-2	Human epidermal growth factor receptor -2
HGNC	HUGO Gene Nomenclature Committee
HG	Histological grade
HR	Hazard ratio
HUGO	Human Genome Organisation
IGF	Insuline growth factor

IGFR	Insuline growth gactor receptor
IHC	Immunohistochemistry
IQR	Interquartile range
KM	Kaplan Meier
LOH	Loss of heterozygocity
LTED	Long term estrogen deprivation
m	Frequency of mitosis
MAI	Mitotic Activity Index
MAIA	Microarray Image Analysis
MAPKs	Mitogen-activated protein kinases
MBC	Metastatic breast cancer
MINDACT	Microarray in Node Negative Disease May Avoid Chemotherapy
MKPs	MAPK phosphatases
mRNA	Messenger ribonucleid acid
MTE	Multiple testing error
MVA	Multivariate analysis
n	Nuclear pleomorphism
NaCl	Sodium chloride
NPI	Nottingham Prognostic Index
NST	No special type
OS	Overall survival
PAC	P1 artificial chromosomes
PAC	p1 Artificial Chromosome
PCR	Polymerase Chain Reaction
PMT	Photomultiplier tubes
PR	Progesterone receptor
RFS	Relapse Free Survival

RNA	Ribonucleid acid
RPL	Random Prime labelling
RPS	Random primer solution
RT	Radiotherapy
SCF	Supraclavicular fossa
SEER	Surveillance, Epidemiology and End Results
SLNB	Sentinel lymph node biopsy
ssDNA	Single-stranded DNA
t	Tubule
TAILORx	Trial Assigning Individualized Options for Treatment (Rx)
TAMR	Tamoxifen resistance
TAMRG	Tamoxifen resistant group
TAMCG	Tamoxifen control group
TMA	Tissue Microarray
TNBC	Triple negative breast cancer
TNM	Tumour, Nodes and metastasis
TTD	Time-to-death
TTP	Thymidine triphosphate
TTR	Time-to-relapse
UICC	International Union Against Cancer
US	Ultrasound
UVA	Univariate analysis
VAMP	Visualization Analysis of array and other Molecular Profiles
WECCA	Weighted clustering of called aCGH data
WLE	Wide local excision

LIST OF TABLES

Table 2.1- Epidemiological, clinical and pathological parameters collected from clinical files.

Table 2.2- Demographics and description breast cancer patients by HG.

Table 2.3- Distribution according to tumour and nodal staging.

Table2.4- ER,PR and HER-2 status according to HG.

Table 2.5- Adjuvant Treatment received by HG.

Table 2.6- Distribution of number of cases with recurrence BC and death according to HG.

Table 2.7- Cox regression UVA for OS, DDFS and DFS.

Table 2.8- Cox regression MVA OS, DDFS and DFS.

Table 2.9- Cox regression MVA according to tumour groups adjusting for adjuvant therapy for the survival endpoints OS, DDFS and DFS.

Table 3.1- Results of TMA and centrally reviewed IHC cases according to HG.

Table 3.2- Pipetting scheme Multiplex PCR.

Table 3.3- Incorporation of labelled nucleotide in RPEX cell line using different labelling kits

Table 3.4- Preparation of labelled gDNA prior hybridisation.

Table 4.1- Demographics of BC patients with aCGH profile.

Table 4.2- Chromosomal region alteration associated with worse OS and CSS.

Table 4.3- Genes encoded by gained region associated with CSS.

Table 4.4- Copy number alterations significantly associated with nodal status.

Table 4.5- Genes encoded by gained region associated with CSS (chromosome 5:171064773-174710119).

Table 4.6- Classification of DUSP family according to Keyse et al.

Table 5.1- Clinico-pathological characteristics of tamoxifen resistant patients and controls.

Table 5.2- Regions differentially amplified in TAMR in comparison with MCF-7 cell lines showing regions of alteration and HGNC symbols.

Table 5.3- Genelist extracted from Ensembl Biomart database from region Chr7: 94136-6268062 showing HGNC and gene description.

Table 5.4- Probelist extracted GEO for region Chromosome7:94136-6268062.

LIST OF FIGURES

Figure 2.1- Distribution of HG into tumour staging.

Figure 2.2- Distribution of HG into nodal staging categories.

Figure 2.3- Distribution of ER status according to HG.

Figure 2.4- KM Survival estimates for OS (A) and DFS (B) according to HG.

Figure 2.5- KM for OS estimates according to tumour groups.

Figure 3.1- Overview of the project- Flow chart.

Figure 3.2- Patient selection for aCGH.

Figure 3.3- TMA 1 with IHC staining for ER, PR, HER-2 and EGFR.

Figure 3.4- IHC staining for ER and PR.

Figure 3.5- IHC staining for HER-2.

Figure 3.6- Agarose gel image after Multiplex PCR.

Figure 3.7- Cy3 and Cy5 Amersham fluorescent dye emission spectral characteristics.

Figure 3.8- Comparative aCGH profile in RPEX cell lines using different labelling kits.

Figure 3.9- Hybridisation chamber dimensions (SureHyb enabled).

Figure 3.10- Array image taken using MAIA program.

Figure 3.11- Exemplary summary plot for aCGH results.

Figure 4.1- KM survival plots for gain and non-gain regions in RP11-20O22, CTB-54I1, RP11-489P1 and CTB-73D21.

Figure 4.2: Kaplan Meir plot of survival in node-negative and node-positive patients with 5q gain-negative and 5q gain-positive tumours.

Figure 4.3- Kapplan Meier survival curves for ER+ subgroup for OS (A) and CSS (B) and for Her-2 negative patients for OS (C) and CSS (D). **Figure 4.4**- Unsupervised and supervised analysis of copy number alterations in relation to BC clinical parameters.

Figure 5.1- Tamoxifen resistant study design.

Figure 5.2- Summary plot of CNA in Chromosome 2 and 17 for MCF-7 and TAMR cell lines.

Figure 5.3- Exemplary comparison of aCGH profile in MCF-7 cell lines analysed using different “calling” platforms: CAPweb (top) and CGHcall (bottom).

Figure 5.4- Summary plots of TAMCG and TAMRG according to survival.
Figure 5.5- Boxplot for the expression of SNX-8 gene.

ABSTRACT

This project was designed to assess clinical and biological parameters of prognosis in high grade breast cancer (BC). The first approach aimed to define Clinico-epidemiological parameters associated with poor survival in a retrospective BC cohort of 1339 non-metastatic patients presenting at Singleton Hospital, South Wales, UK. Median follow-up was 5.4 years (range 0.09-10.14 years). Results of this analysis supported the role of histological grade (HG) as a prognostic factor and a tumour classifier. HG stood as one of the main variables associated with OS, DDFS and DFS with only N3 disease conferring worse prognosis. This analysis contributed with clinical information and survival data required for sample selection for subsequent comparative genomic hybridisation (CGH) studies.

Array Comparative genomic hybridisation (aCGH) was performed in 78 cases (67 HG3 and 11 HG2) cases aiming to identify copy number alterations (CNA) associated with poor survival. aCGH protocol optimisation was required to obtain reproducible results and a new simplified aCGH protocol was described. A region of chromosomal gain in Chromosome 5 (5q35.1 to 5q35.2) was significantly associated with Cancer-specific survival (CSS; FDR<0.2). DUSP-1 and MSX2 genes were among those candidate genes validated In-silico for poorer prognosis in BC.

FINALLY, THE ISSUE OF TAMOXIFEN RESISTANCE WAS ADDRESSED BY ATTEMPTING TO IDENTIFY CNA IN OESTROGEN RECEPTOR (ER) POSITIVE/TAMOXIFEN TREATED PATIENTS ASSOCIATED WITH EARLY RELAPSE/DEATH ≤5 YEARS (TAMRG) COMPARED WITH A SUB-GROUP OF PATIENTS ALIVE AND WELL AFTER 5 YEARS ON FOLLOW-UP (TAMCG). A REGION OF GAIN ON CHROMOSOME 7 (188219-6234052) WAS ASSOCIATED WITH TMACG (P:0.05) WHICH WAS ALSO ASSOCIATED WITH SIGNIFICANT OVER-EXPRESSION OF SNX8 USING IN-SILICO VALIDATION IN LOI ET AL DATASET. RESULTS FROM THIS STUDY CONTRIBUTE TOWARDS THE IDENTIFICATION OF CANDIDATE PROGNOSTIC GENES IN BC. BIOLOGICAL VALIDATION OF THESE RESULTS IS RECOMMENDED. FURTHER RESEARCH

IS NEEDED TO ASSESS THE FUNCTIONAL RELEVANCE OF THESE CANDIDATE GENES.TABLE OF CONTENT

DECLARATION 2

ACKNOWLEDGEMENT 3

LIST OF ABREVIATIONS 4

LIST OF TABLES 8

LIST OF FIGURES 9

ABSTRACT 11

CHAPTER 1 – INTRODUCTION 16

1.1- BREAST CANCER BACKGROUND 16
1.2- PROGNOSTIC CLASSIFICATION 17
1.3- PATHOLOGY 21
1.3.1- Grade as a prognostic marker 21
1.3.2- Molecular and Clinical heterogeneity within Grade 3 tumours 22
1.3.3- Oestrogen receptor and tamoxifen resistance 25
1.4- GENETIC VARIABLES 27
1.4.1- Characterisation using aCGH 27
1.5- OVERVIEW OF THE PROJECT 34

CHAPTER 2 – RETROSPECTIVE COHORT ANALYSIS. 35

2.1- COHORT OVERVIEW – STATISTICAL ANALYSIS 35
2.1.1- Patients characteristics. 35
2.1.2- Results 37
2.1.2- Receptor Status 40
2.1.3- Adjuvant Treatment 42
2.1.4- Follow up 43
2.2- IDENTIFICATION OF CLINICO-PATHOLOGICAL VARIABLES ASSOCIATED WITH OS, DDFS AND DFS IN BC: COX REGRESSION ANALYSIS. 44
2.2.1- Univariate analysis. 44
2.2.2- Identification of Clinico-pathological variables associated with OS, DDFS and DFS in BC. Multivariate analysis. 46

2.3- RELEVANCE OF ER STATUS IN HG3 TUMOURS 47
2.4- TUMOUR CLASSIFICATION USING IHC AS A SURROGATE FOR GEP MOLECULAR SUB-TYPES. 48
2.5- DISCUSSION 50

CHAPTER 3 – MATERIALS AND METHODS. 56

3.1 – PROJECT OVERVIEW. 56
3.1- PATIENTS SELECTION FOR TISSUE MICROARRAY (TMA). 57
3.2 DNA ISOLATION 63
3.2.1- DNA from Cell lines 63
3.2.2- DNA isolation from FFPE tissues 63
3.2.3 – FFPE cell pellets as positive control 64
3.3- MULTIPLEX PCR 64
3.4 - BAC ARRAY 66
3.4.1- Random Prime labelling (RPL) 66
3.4.1.1- BioPrime® DNA Labelling System - Invitrogen 67
3.4.1.2- Bioprime ® Total Genomic labelling System- Invitrogen 68
3.4.1.3 Amersham CyScribe Array CGH genomic DNA labelling kit (GE Healthcare 28909726) 69
3.4.1.4- Comparison of different genomic DNA labelling systems for aCGH 69
3.4.2- Summary of labelling results in BC tumour samples. 72
3.4.3 Initial aCGH experiments in FFPE tissues. 72
3.4.4- Agilent Oligonucleotide Array-based CGH for Genomic DNA protocol. 73
3.4.4.1- Preparation of Labelled Genomic DNA prior hybridization. 73
3.4.4.2- Washing 75
3.4.4.3- Scanning 75
3.6- BIOINFORMATIC ANALYSIS OF CGH 75
3.6.1 Extraction of Fluorescence intensities. 75
3.6.2- Primary data analysis 76
3.6.3 Secondary Analysis 79
3.6.3.1- Clustering of aCGH data 79
3.6.4- In silico validation of candidates genes 80

CHAPTER 4 – MOLECULAR DIVERSITY IN HIGH GRADE BREAST CANCER. ASSOCIATION OF 5Q GAIN WITH SURVIVAL.. 81

4.1- PATIENTS COHORT 81
4.2 – RESULTS 82
4.2.1 – Patients characteristics 82
4.2.2 – Association of Copy Number Changes with survival 83
4.2.3 Association of copy number changes with Nodal Status 87
4.2.4 – In-silico validation of candidate genes. 88
4.2.5- Hierarchical Unsupervised Clustering Analysis 90
4.3- DISCUSSION 91
4.3.1- DUSP-1 gene 93
4.3.2- MSX2 96

CHAPTER 5 – BAC ARRAY ANALYSIS AND RESISTANCE TO TAMOXIFEN.... 99

5.1- AIMS OF THE STUDY 99
5.2- SAMPLE SELECTION 100
5.3- MATERIALS AND METHOD 101
5.4- RESULTS 102
5.4.1- Comparative aCGH profile between MCF-7 and TAMR cell lines. 102
5.3.2 – Array CGH profile in Tamoxifen resistant tumours 104
5.3.3 – Association of gene expression located in Chromosome 7 region with early and late relapse in Tamoxifen-treated patients...................................... 106
5.5- DISCUSSION 108
5.5.1- Cell lines results 108
5.5.2- Sorting-Nexin 8 (SNX-8) over-expression might have a protective effect for ER+ patients treated with tamoxifen 111
5.5.3- Findings differ between CNA between Tamoxifen resistant cell lines and tumours. 113

CHAPTER 6 – CONCLUSION 116

6.1- RESEARCH CONTRIBUTIONS 116
6.1.1- HISTOLOGICAL GRADE APPEARS TO BE THE MOST IMPORTANT BIOLOGICAL PROGNOSTIC PARAMETER ROUTINELY ASSESSED IN BREAST CANCER TUMOURS IN CLINICAL PRACTICE. 116

6.1.2- MOLECULAR HETEROGENEITY OF HIGH GRADE BREAST CANCER – ASSOCIATION OF GAIN OF A CHROMOSOMAL REGION ON 5Q WITH SURVIVAL. 117
6.1.3- DUSP-1 and MSX2 were up-regulated in poorly differentiated breast cancers using In-silico validation. 118
6.1.4- HOXB13 and TOB1 genes are gained in TAMR cell lines. 119
6.1.5- SNX8 gained – allowing Tamoxifen sensitivity? 120
6.2- RESEARCH LIMITATIONS. 121
6.3- RECOMMENDATIONS FOR FUTURE RESEARCH 123

REFERENCES 124

APPENDICES 144

APPENDIX 1 CODING SYSTEM USED FOR THE STATISTICAL ANALYSIS OF THE BREAST CANCER DATABASE 144
APPENDIX 2 – TISSUE MICROARRAY MAP 147
APPENDIX 3 GENOMIC DNA ISOLATION FROM FFPE TISSUE BLOCKS 148
APPENDIX 4- PROTOCOL FOR THE PREPARATION OF FFPE CELL PELLETS. 152
APPENDIX 5.- MULTIPLEX-PCR CHECK OF GENOMIC DNA ISOLATED FROM FFPE TISSUE FOR ITS USABILITY IN ARRAY CGH ANALYSIS 154
APPENDIX 4- PROTOCOL FOR THE PREPARATION OF FFPE CELL PELLETS. 158
APPENDIX 5.- MULTIPLEX-PCR CHECK OF GENOMIC DNA ISOLATED FROM FFPE TISSUE FOR ITS USABILITY IN ARRAY CGH ANALYSIS 160
APPENDIX 6.- SUMMARY OF THE DNA CONCENTRATION 165
APPENDIX 7.- AGAROSE GEL IMAGE AFTER MULTIPLEX PCR OF TUMOUR SAMPLES. 166
APPENDIX 8.- GE HEALTHCARE. AMERSHAM. 167
APPENDIX 9.- RANDOM PRIMER LABELLING 168
APPENDIX 10- SUMMARY OF LABELLING RESULTS IN BC TUMOUR SAMPLES 171
APPENDIX 11.- PREPARATION OF LABELLED GENOMIC DNA FOR ARRAY 172

CHAPTER 1 – INTRODUCTION

1.1- Breast Cancer Background

Breast cancer (BC) is the most common female cancer in the UK accounting for just over 39600 new cases per year in 2008. The incidence rates of BC has doubled over the last three decades, a trend that was apparent even before the introduction of the National Screening program in 1988 and has continued to rise until recent years estimating an increase of 4% between 2007 and 2008 (Office-for-National-Statistics, 2010).

By contrast, BC mortality is declining. Five year relative survival increased by 0.9% for patients diagnosed between 2001 and 2006 and followed up to 2007 in comparison to those diagnosed during 2000- 2005 and followed up to 2005 (Walters S, 2009). This significant achievement is likely to be related to the success of early detection programme and to improvements in treatment particularly adjuvant systemic therapies such as tamoxifen. Clinical factors associated with an increased risk of developing BC include age, family history, obesity, exposure to exogenous female hormones, previous history of breast irradiation, alcohol intake and contra-lateral BC.

BC is a heterogeneous disease and several factors determine the prognosis and therapeutic options used in specific cases. Initial diagnosis is based on clinical, radiological and histological evaluation of the primary tumour in order to confirm diagnosis and assess extent of the disease. Ideally all cases should be discussed in a multidisciplinary meeting that would agree with diagnosis and propose, accordingly, the clinical management. Current routine clinical management relies on an accurate staging of the anatomical and pathological extent of the disease which will often, but not always, requires further investigations such as blood tests and imaging studies like ultrasound (US), Computer Tomography (CT) and bone scan.

1.2- Prognostic Classification

The most widely used method for classifying the extent of anatomical cancer spread for solid tumours is the TNM classification that looks into local tumour volume (T), loco-regional nodal spread (N) and distant metastasis (M). This method was initially developed by Dr Pierre Denoix at the Institute Gustave-Roussy over several years (Denoix and Baclesse, 1955, Baclesse and Denoix, 1957). However, it was not until 1958 that the first TNM recommendations for the staging of BC and Larynx were published as recognition for staging standardization by a special Committee from the Research Commission of the International Union Against Cancer (UICC, 1958). Since then, there have been several revised versions of the TNM classification published which have helped clinicians in assessment of prognosis, planning of treatment, evaluation of treatment response and standardising information record and exchange between centres (Gospodarowicz et al., 1998).

However useful and standardized, the TNM staging is not a perfect system as it does not include non-anatomical prognostic factors that influence prognosis or even influence treatment decisions. Specifically in BC, factors such as age, histological grade (HG), Oestrogen receptor (ER), progesterone receptor (PR) or Human Epidermal Growth factor receptor (HER-2) status are well known to influence outcome and are not currently included in the TNM classification. As more non-anatomical factors become available in different solid tumours, there are good practical arguments to add these to TNM. HG has been incorporated in the TNM staging of tumours like prostate, soft tissue and bone. Similarly, age was included in the TNM staging of thyroid cancer and serum markers in the staging of testicular cancer. The list of non-anatomical prognostic factors for BC has increased considerably. On one hand, it is appealing to incorporate all prognostic factors in the initial clinical assessment of patients. However, this carries the risk of overloading the TNM with variables that although might increase its prognostic accuracy, can convert the TNM in a very complex tool to be applicable in clinical practice, losing in this way the reason for which it was conceived. On these basis some believe that the TNM should remain as it is (Sobin, 2001), a simple classification system to be use

upfront to help clinicians with the treatment decision process and a system upon which other prognostic systems are built.

Another useful treatment guideline used in clinical practice is St Gallen International Expert Consensus (Goldhirsch et al., 2001). This guideline has been developed by a group of expert who have met at St Gallen, Switzerland, since 1978 in order to reach a consensus about the clinical management of early BC based on updated evidence. At their most recent meeting, the panel of experts highlighted the relative indications for chemo-endocrine or endocrine therapy alone based not only on tumour size, nodal status and ER/PR status but also on HG, proliferation (either assessed by Ki-67 labelling index or pathological description of frequency of mitosis), peri-tumoural vascular invasion and patient preference. HG was considered, by the majority of the panel, to be “sufficient indication for adjuvant chemotherapy, although genomic grade could be considered as an adjunct”, if available. More importantly, the panel supported for the first time the use of validated multigene tests to assist clinical decisions in cases where the use of a particular therapy is still uncertain after consideration of conventional clinico-pathological markers (Goldhirsch et al., 2009).

The Nottingham prognostic Index (NPI) is another clinically useful prognostic index. It was developed from a retrospective analysis of 387 BC patients with operable disease. Only 3 variables remain statistically significant in the multivariate analysis (MVA), namely tumour size, nodal status and tumour grade, which were subsequently used for the development of an index to classify patients into three categories according to the chance of dying from BC at 15 years. Patients in the good prognostic group (29%) had 80%, moderate group (54%) 42% and poor prognostic patients (17%) had 13% 15-year survival (Galea et al., 1992). The index emphasises the importance of HG in the prognosis of BC patients.

One of the most accurate tools for assessing prognosis in patients with BC is Adjuvant! Online (Ravdin et al., 2001). This is an evidence-based computerised online program developed on 10-year-observed overall survival (OS) of Stage I and II women age between 36 to 69 years diagnosed between 1988 and 1992 and recorded in the Surveillance, Epidemiology and End Results (SEER) registry in the United States. The programme also estimates Breast cancer–specific survival

(BCSS) for untreated patients which was projected from observed BCSS in the same population but adjusted for the frequency and efficacy of the systemic therapy. The other endpoint estimated is Event Free Survival (EFS) which was indirectly derived from SEER mortality data. Adjuvant Online also provides the user with an estimate of efficacy of adjuvant therapies such as tamoxifen and different chemotherapy regimens. This information is particularly important at the time of discussing therapeutic options in clinic and is known to influence the patients willingness to accept such therapies (Ravdin et al., 1998).

Adjuvant! Online includes several clinico-pathological characteristics such as age, co-morbidities, HG, tumour size and nodal status and treatment-related factors including different options for adjuvant hormone therapy and chemotherapy. It is widely available, free access for online users and results can be printed out to illustrate the discussion of therapeutic options in clinical practice.

Adjuvant Online has also been validated in different populations. A validation study performed at the British Columbia Breast Cancer Outcomes Unit (BCOU) confirm that 10-year predicted and observed outcomes were within 1% for OS, breast cancer specific survival (BCSS) and EFS (all $p > 0.05$). Adjuvant! overestimated OS, BCSS, and EFS in women younger than age 35 years (predicted-observed = 8.6%, 9.6%, and 13.6%, respectively; all $p < 0.001$) or with lymphatic or vascular invasion (Olivotto et al., 2005). More recently a large European retrospective validation study also confirm that Adjuvant! performed reliably with a difference between predicted and observed outcomes (OS and BCSS) within 2% for most relevant clinico-pathological subgroups and similarly to Olivotto et al the study outcomes were overestimated by approximately 4% in young patient (less than 40 years) (Mook et al., 2009). However, a recent study from Oxford shows that the UK outcomes may be significantly inferior for OS, BCSS and EFS by 5.54, 4.53 and 3.51 % respectively across the majority of the analysed categories (Campbell et al., 2009). Discrepancies are likely to be due to a higher mortality rate in the UK in comparison to the US. However, confirmation of these results including a representative number of centres across the UK is required.

Among all prognostic factors in BC, nodal status has traditionally been regarded as the most important. It is estimated that approximately 30% and 40% of the clinically node-negative patients have occult positive axillary nodes (Fisher et al., 1985). Therefore, it is routine clinical practice to perform axillary surgery at the time of the surgical treatment of the primary tumour as it will allow pathological staging and, in those cases with lymph node (LN) involvement, axillary dissection also provides effective local control in early and advanced BC (Bembenek and Schlag, 2000). There is also consensus on the prognostic significance of the number of positive nodes; the higher the number of lymph nodes involved, the worse the prognosis (Fisher et al., 1983).

Although universally accepted and validated, the use of the current guidelines for the prognosis of BC might be challenged as a consequence of a change in the spectrum of BC due to the implementation of screening and increase awareness of BC in the population (Cady, 1997). The tumours in screened women are smaller and more likely to require Breast Conserving Surgery (BCS) than they would have been in symptomatic patients (Advisory-Committee-on-Breast-Cancer-Screening, 2006). Screened cancers also have fewer LN metastases. This may affect the relative statistical weighting of features applied to pathology indicators of survival in the current guidelines as the NPI, formulated 20 years ago on the basis of symptomatic practice and might not be optimal for a population in the context of screening (Anderson et al., 2000).

Tumour size is also considered a prognostic factor. It is widely accepted that the bigger the tumour the higher the chance of relapse. However, it has also become obvious that BC is a heterogeneous disease with variable prognosis even in cases of similar tumour size. As an example, the reported relapse free survival (RFS) at 8 years for a tumour less than 1 cm node-negative without systemic therapy ranges from 81% (Fisher et al., 2001) to 100% (Tinnemans et al., 1989). Therefore, predicting prognosis in BC is complex and even very early tumours can metastasise despite systemic therapy. Discrepancies in outcome in a population of patients with the same TNM staging are likely to be related to the biology of the tumour, which dictates its aggressiveness (Bonnefoi, 2007).

1.3- Pathology

1.3.1- Grade as a prognostic marker

Histological grade of the tumour is the most commonly used pathological prognostic variable that reflects tumour biology. The system to assess the degree of tumour differentiation was initially described by Scarff, Bloom & Richardson (Bloom and Richardson, 1957) and subsequently modified in order to make it more objective and reproducible (Elston and Ellis, 1991). The system currently in used in the UK, endorsed by the Royal College of Pathologist (NHSBSP, 2005), involves a semi-quantitative evaluation of three morphological features tubule/acinar/glandular formation (t), nuclear pleomorphism (n) and frequency of mitoses (m). The resulting scores for each feature are added together and assigned to grade 1, 2 and 3. HG1 includes tnm scores 3-5, HG2 tnm scores of 6 and 7, and HG3 tnm scores of 8 and 9 (NHSBSP, 2005).

Despite previous efforts to standardise reporting systems for HG, there is still significant inter-observer discrepancies in reproducibility among pathology departments. Boiesen reported only 31% complete agreement on HG with a moderate reproducibility among different departments (overall mean Kappa 0.54) (Boiesen et al., 2000). These results were comparable to those reported by Frierson (Frierson et al., 1995). Further analysis revealed that the concordance on the analysis on tubular formation was better than for nuclear pleomorphism and mitotic count which are usually affected by fixation artefacts, necrosis, severe inflammation or fibrosis (Boiesen et al., 2000). Although other authors have found the reproducibility to be acceptable when specified guidelines are in use with complete agreement ranging from 83.3% to 87 % (Robbins et al., 1995, Dalton et al., 1994). However, the poor reproducibility of HG assessment found in some studies is being given as the reason for not including the HG into the new TNM classification despite of its recognised prognostic significance (Singletary and Greene, 2003).

Unfortunately, HG is not always useful as prognostic information due to the fact that 30-60% of tumours are classified as HG2, even when it is assessed by a single pathologist. In a publication from the Bordet group, Sotiriou et al found that HG1 and HG3 BC had distinct gene expression profiles. This group identified a Genomic grade signature which was the base to develop a Genomic Grade Index (GGI). GGI was later used to separate HG2 patients into two groups with similar expression profile and risk of recurrence to HG1 or HG3 (Sotiriou et al., 2006). It seems that some of the discrepancies issues in the HG assessment might be overcome with the use of GGI when available and it has even stronger prognostic significance in the MVA than LN status or tumour size (Sotiriou et al., 2006).

As discussed previously, as more BC are diagnosed at an early stage the prognostic factors used to classify patients at diagnosis will migrate from the assessment of disease burden only, as in the TNM staging, to those tools that allow assessment of tumour biology. There is evidence to suggest HG3 disease is not the result of dedifferentiation from HG1 disease but they represented a separate entity with different molecular origins (Roylance et al., 1999). HG3 tumours are also more aggressive and common in symptomatic BC patients accounting for about half of all cases with published ratios for grades 1, 2 and 3 being approximately 2:3:5 (Elston and Ellis, 1991). In addition, the proportion of cases with HG3 is significantly higher in patients ≤ 35 years of age compared to older patients, a fact that contributes to the poorer prognosis in young patients (Kroman et al., 2000).

1.3.2- Molecular and Clinical heterogeneity within Grade 3 tumours

Heterogeneity is not only limited to tumours of different histological grades. Significant differences have been found within poorly differentiated tumours in terms of clinical behaviour and survival. Different methods can be applied to sub-classify HG3 cases into subgroups such as Immuno-histochemistry (IHC) profile, gene expression profile (GEP) and more recently Array Comparative Genomic Hybridisation (aCGH). Recent publications used both GEP and CNA (Copy Number alterations) to predict BC outcome (Chin et al., 2006, Bergamaschi et al., 2006,

Haverty et al., 2008). Some of these have successfully identified subgroups of worse prognosis within HG3 tumours.

The majority of grade 3 cases are ER negative. So starting from ER, a group of ER+/HG3 tumours have a worse prognosis than ER+ with other histological grading but better prognosis than the rest, ER-/HG3 tumours.

It is recognised that ER+ tumours, regardless of the grading, express genes commonly found in “luminal” epithelial cells (Perou et al., 2000). These luminal tumours were sub-classified further using GEP into Luminal A, usually well differentiated tumours with low proliferation rate and a group with high proliferation rate called Luminal B (Sorlie et al., 2001). Patients with Luminal B tumours have consistently worse prognosis than Luminal A though slightly better (Sorlie et al., 2001), outcome to basal-like or HER-2 tumours (Rakha et al., 2008).

Three sub-groups were found within ER negative tumours: “basal-like”, “HER-2” and another, even more heterogeneous group, “normal- breast-like” tumours (Perou et al., 1999). This “breast tumour intrinsic” classification has been validated in other studies (Sorlie et al., 2003, Hu et al., 2006). However, GEP for BC remains a research tool. Complementary deoxyribonucleic acid (cDNA) arrays are expensive, requires both good quality Ribonucleic acid (RNA) that could only be obtained from frozen samples and the expertise of a research group for the analysis and interpretation of results (Rakha et al., 2008).

The knowledge that some BC express markers found in myoepithelial/basal cells preceded the identification of the “basal-like” group by Perou et al, although its clinical significance was not clear. Using pathological criteria, current definition of basal-like tumours include those tumours that express cytokeratins 5/6 , 17, 14 (Gusterson et al., 1982, Wetzels et al., 1991), vimentin (Gould et al., 1990), lack expression of ER/PR and HER-2 [triple negative breast cancer, TNBC] (Perou et al., 1999, Sorlie et al., 2001, Hu et al., 2006), EGFR expression (Reis-Filho et al., 2005a), high HG, central scar, significant necrosis and pushing borders (Fulford et al., 2006).

Although a comprehensive definition of “basal-like” tumours using IHC is still awaited, multiple attempts have been made to correlate “basal-like” tumours defined by GEP with that defined by pathological criteria. Niesel found that using a battery of IHC including Cyt 5/6 and EGFR, in addition to the commonly requested ER, PR and HER-2, it will provide a specificity for the detection of basal- like tumours (defined by GEP) of 100% with a sensitivity of 75% (Nielsen et al., 2004). It is interesting that using unsupervised clustering analysis of TMA results from a panel of IHC staining also classify BC into groups closely comparable to those obtained from cDNA (Abd El-Rehim et al., 2005, Makretsov et al., 2004, Jacquemier et al., 2005, Zhang et al., 2003).

It is been reported that “basal-like” tumours have the worse prognosis among all the groups even below HER-2 positive tumours, whose prognosis had dramatically improved after the advent of trastuzumab therapy. This poor prognosis is extended to TNBC, greatly represented by “basal-like” tumours, in view of the lack of efficacious therapeutic options, probably a similar picture to those HER-2 positive patients 20 years ago. Patients within the “normal-breast-like” also have TNBC disease. However, they lack expression of basal cytokeratins, do experience a lower response to neo-adjuvant chemotherapy (Rouzier et al., 2005) and have a clinical outcome more similar to the no–TNBC group than they have to the basal-like group (Foulkes et al., 2009).

HG3 BC have also been characterised according to their aCGH profile (Natrajan, 2009) following classification into 3 groups (luminal, basal-like and HER-2) using IHC definitions by Nielsen et al (Nielsen et al., 2004). Patterns of aCGH profile using oligonucleotide arrays had been previously described (Hicks et al., 2006) but Natrajan et at found that Basal-like tumours are more likely to have an aCGH profile called “sawtooth” (segments of gain alternating with deletions involving all the chromosomes) whereas HER-2 tumours more commonly have a “firestorm” profile characterised for a group of narrow peaks of amplifications confined to a single chromosome arm. Luminal tumours had a more heterogeneous profile being the

Luminal B tumours characterised by high level gene amplification (Natrajan, 2009, Bergamaschi et al., 2006).

It is important to emphasise that all the previously defined groups, either by GEP and/or IHC, are more commonly classified as poorly differentiated tumours, except for the Luminal A group. High grade tumours are not only a heterogeneous group pathologically, biologically and genetically but also clinically and further research is required to identify potentially therapeutic options to improve its poor prognosis.

1.3.3- Oestrogen receptor and tamoxifen resistance

Oestrogens induce both differentiation and proliferation of normal breast tissue. Specifically, oestradiol triggers a chain of intracellular events in breast cells after binding ER that ultimately stimulates the transcription of oestrogen-dependent genes.

ER is recognised as a prognostic and predictive factor in BC. The presence and density of ER expression is also the best predictor of response to endocrine therapy together with PR and HER-2 (Rastelli and Crispino, 2008). For ER positive disease, adjuvant tamoxifen reduces the risk of BC recurrence by 11.8% and BC mortality by 9.2% at 15 years and is the standard of care in premenopausal women (Wa et al., 2005, EBCTCG, 2005) and some postmenopausal patients. Unfortunately, even in ER positive patients, there will be at least 30 % of therapeutic failure to tamoxifen (*de novo* resistance) with the majority of patients eventually developing resistance over time (acquired resistance) (Riggins et al., 2007).

At least three different mechanism of action have been described for tamoxifen including both genomic and non-genomic pathways of controlling replication of oestrogen-dependent genes. Similarly, tamoxifen resistance has been extensively investigated and several mechanisms have been postulated to explain the development of resistance including changes in ER, tamoxifen metabolism, alteration of regulatory proteins and through the interaction with growth factor signalling pathways (Bender and Nahta, 2008). There is significant evidence both in

vivo and in vitro to confirm that the majority of tamoxifen resistant cells continue to express ER (Johnston et al., 1995) and approximately 20% of tamoxifen resistant cells will respond to an aromatase inhibitor or fulvestrant indicating that the oestrogen pathway still has effect in cell proliferation (Osborne et al., 2002, Howell et al., 2002).

In addition, it has been proven in vivo that inappropriate activation of intracellular growth factor signalling pathway including Epidermal Growth Factor (EGF) and its receptor (EGFR) (Nicholson et al., 2002) , Insulin Growth Factor (IGF) (Guvakova and Surmacz, 1997), Insulin-Growth Factor receptor (IGFR) (Stephen et al., 2001) and HER family receptors can also occur in tamoxifen resistant tumours. It is of particular interest the "cross-talk" between HER-2 and ER intracellular pathway. HER-2 amplification occurs in tamoxifen resistant cells even in those originally HER-2 negative (Nicholson et al., 2004, Gutierrez et al., 2005). Despite the extensive scientific work directed to understand mechanism of tamoxifen resistance there is still no useful therapeutic intervention applicable in clinical practice that would reverse such resistance once developed.

Studies in the ER pathway have been facilitated due to the availability of endocrine-sensitive wild type MCF-7 breast cancer cell. A number of different sub-lines of the MCF-7 are available including tamoxifen resistant (TAMR) (Knowlden et al., 2003). However, it has been described that MCF-7 cells develop rapid genetic changes in culture producing MCF-7 sub-lines that differ both at the genomic and phenotypic levels (Nugoli et al., 2003). Some of the tumour properties are best studied in cell culture such as anchorage–independence which reflects the tumour cells metastatic potential in vivo (Mori et al., 2009). However, findings from cell cultures are also difficult to translate to an in vivo tumour setting despite of some evidence that support that both cell lines and BC tumours have similar genomic profile (Naylor et al., 2005).

As adjuvant tamoxifen therapy is usually recommended for 5 years, it would be ideal to predict what patients will benefit from it and are at low risk of relapse. Han et al found that there are significant differences in the deoxyribonucleic acid (DNA) CVA

between patients who relapse within 5 years of diagnosis compared with patients in the non-recurrence group (Han et al., 2006). Further studies using DNA array technology are required to elucidate further genetic alterations involved in hormone resistant process as a step prior predicting for tamoxifen response.

1.4- Genetic Variables

1.4.1- Characterisation using aCGH

Comparative genomic hybridisation has emerged as a very powerful molecular cytogenetic method for the detection of copy number alteration. CGH allows wide genome analysis DNA for the detection and mapping of unbalanced genomic alterations such as amplification, deletion, chromosomal loss/duplication and unbalanced chromosomal translocations. As the normal chromosomal content is diploid, BAC (Bacterial Artificial Chromosomes) array results might reveal a homozygous loss (loss of 2 copies), heterozygous loss (loss of 1 copy or loss of heterozygosity or LOH), gain (more than 2 copies) or amplification (usually defined as >5 copies) (van Beers and Nederlof, 2006). However, this is not the appropriate technique for the detection of balanced translocations or inversions.

In a hallmark publication, CGH was first described as a technique in which tumour DNA and normal genomic reference DNA were differentially labelled with two different fluorophores and simultaneously hybridised to normal metaphase chromosomes in the presence of Cot-1 DNA. The amount of DNA bound is proportional to the relative quantity of DNA present and can be quantified by the measurement of green-to-red fluorescence ratios along the length of chromosomes. A software computer then calculates the foreground and background signal intensities and generates a copy number karyotype for each tumour sample (Kallioniemi et al., 1992). Although conventional CGH became a very powerful technique, this method required of high quality metaphase chromosomes usually difficult to obtain in solid tumours, is time consuming and has a low detection resolution ranging from <20 Mb (Pinkel et al., 1998) down to 3Mb as its best (Kirchhoff et al., 1999). In addition, only a limited number of cells can be analysed by conventional aCGH which might not represent the main tumour cell type or even the

overall tumour cell population as BC might contain many different cell types (Cingoz et al., 2003).

Conventional arrays have been replaced by aCGH in which DNA sequences serving as hybridisation targets are attached to a hydrophilic polymer-coated glass slide (Pinkel et al., 1998, Solinas-Toldo et al., 1997). This target DNA sequences consist on large-insert clones with known chromosomal location called BAC and P1 artificial chromosomes (PAC). BAC clones have an average size of 100-150 Kb, sufficient to provide high binding specificity (Pinkel et al., 1998).

As approximately 3500 BAC clones are required for the construction of an array of 1 Mb resolution, it would be expensive and laborious to rely solely on large-scale bacterial cultures. Instead, BAC clones are currently produced using techniques such as degenerate oligonucleotide polymerase chain reaction (DOP-PCR) among others (Fiegler et al., 2003). The resolution of aCGH is dictated by the density of mapped sequences (distance of DNA sequences on the chromosome) and the size of the cloned DNA target (Fiegler et al., 2007). In other words, the smaller the size of the DNA target sequences and the closer they are on the chromosome the higher the array resolution (Bejjani and Shaffer, 2006). Tiling pathways arrays offer a much higher resolution than aCGH in view of overlapping sequences in the BAC or PAC clones (Redon et al., 2005, Ichimura et al., 2006, Ishkanian et al., 2004). It is estimated that 30,000 clones are required for the construction of tiling arrays (Fiegler et al., 2007) which will provide a resolution of approximately 100 Kb. Further spatial resolution is achieved only by oligonucleotide arrays with a maximum resolution of > 10Kb. However, the appropriate platform used depend on several factors including tissue and platform related issues which are dictated by reproducibility, availability, costs, expertise and type of aberration for detection (Tan and Reis-Filho, 2008). An interesting approach is to screen the whole-genome with a low resolution platform (1Mb) with subsequent fine mapping of a region of interest (Garcia et al., 2005).

Array CGH results will depend to great extent on the genetic input material. Most of the research in BAC array has been performed in fresh frozen tissue (FFT) that will provide with the best quality DNA. However, Formalin Fixed Paraffin Embedded

(FFPE) tissue samples historically represent the method of choice for tissue preservation and a vast amount of human tumour material is kept in paraffin blocks. Furthermore, invaluable clinical information might be available for correlation with biological findings in FFPE tumours, increasing the opportunity for successful translational research. Optimisation of DNA isolation protocols has provided us with the required technology for the extraction of DNA from FFPE with excellent results but the DNA quality is still dependent on the hypoxic period of the tissue before fixation, time of fixation, the use of buffered formalin (Legrand et al., 2002) and the duration of storage among others. Improvements in sampling handling have been universally standardised recently in an attempt to achieve better genetic material preservation. Nevertheless, a proportion of archival DNA samples will be unsuitable for aCGH as it may have been stored in inappropriate conditions.

Array CGH results also depend on the purity of the input DNA. As a general rule, aCGH should be performed on samples containing ≥ 70% of tumour cells in order to minimise normal cell contamination and increase detection sensitivity (Garnis et al., 2005, van Beers and Nederlof, 2006). In another words, normal cell contamination will increase the proportion of normal DNA input in aCGH experiments which, after hybridization with normal reference DNA, bring the green-to-red ratio closer to 1 and by doing so reducing the probability of copy number variation (CNV) detection (Weiss et al., 1999). In those samples with low tumours cell content (<75%) laser capture microdissection (LCM) should be considered in order to overcome tissue heterogeneity (Devries et al., 2005, Garnis et al., 2005). However, the main disadvantage of LCM is the low amount of DNA retrieved after dissection, particularly for small sections. If the DNA yield drops below the minimal requirements for aCGH then whole genomic amplification (WGA) is required prior aCGH experiments. However, WGA will usually introduce amplification errors and until better amplification techniques are developed to overcome this difficulty, unamplified DNA remains the gold standard for aCGH experiments (Arriola et al., 2007). However, new bio-informatic corrections can be made to minimise the impact of tissue heterogeneity on the genomic profile in tiling array being able to detect CNV with only 30% of tumour cell content (Garnis et al., 2005).

As aCGH is expensive and time consuming, it is of paramount importance to perform DNA quality control on FFPE samples in order to predict aCGH success. Multiplex Polymerase Chain Reaction (PCR) is a quick and reliable method of assessing DNA quality and its suitability for aCGH where PCR is performed using four primer sets that produce 100, 200, 300 and 400 bp fragments in the housekeeping Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene. The higher the content of cross-linked DNA the less abundant PCR amplification products using multiplex PCR which might even disappear in decreasing order. This means that 400 bp amplification products will decrease or disappear first, followed by the 300 bp, 200 bp and finally 100 bp as the content of DNA cross links increases (van Beers et al., 2006). According to the authors, samples of ≥200 bp multiplex PCR fragment are likely to give good aCGH profiles with highly reproducible hybridisation.

Biological validation of aCGH findings is usually performed by Fluorescent in-situ hybridisation (FISH) in which a previously known or suspected regions of alteration is interrogated using a specific probe. The underlying mechanism is similar to aCGH in the sense that it is a FISH experiment performed on thousands of loci simultaneously on interphase or metaphase chromosomes (Bejjani and Shaffer, 2006). Other techniques are also available for aCGH validation including quantitative real time polymerase chain reaction (qRT-PCR).

As mentioned before, aCGH is a powerful method for the genome-wide detection of CNV, particularly in solid tumours that usually exhibit complex genetic abnormalities previously impossible to decode. This has facilitated the identification of possible therapeutic targets as it is able to pinpoint the location of cancer related genes and might translate into clinical practice as is the case of HER-2Neu and c-Myc among others (Reis-Filho et al., 2005b, King et al., 1985, Slamon et al., 1987) .

Other types of arrays using cDNAs have been used to obtain gene expression profiles. Pollack described the first study using array based genome wide profile using cDNA microarrays (Pollack et al., 1999). cDNA microarrays have revolutionised the molecular classification of BC (Perou et al., 2000). Perou at al demonstrated that each BC tumour has a distinctive expression pattern and that in

general BC tumours can be allocated into at least 4 groups according to certain phenotypic characteristics namely ER+/luminal-like, basal-like, Erb-B2+ and normal breast (Perou et al., 2000). Further sub-classification of the luminal type into A and B was later developed according to the level of expression of ER receptor, which also correlates with HG (Sorlie et al., 2001).

Other investigators have concentrated on the development of prognostic signatures in order to improve risk stratification over standard predictors. Three gene expression-based prognostic breast cancer tests have been licensed for use namely Oncotype DX® (Genomic Health, California), H/I (Avaria DX®, California) and Mammaprint (Agendia®, Huntington Beach, California, USA).

Oncotype DX® was developed from 250 candidate gene set measured in 447 patients, within the NSABP-B20 trial. Using 21 selected genes, Paik et al classified ER+, LN negative patients into 3 different groups according to their 10 year risk of distant recurrence. Patients within the high risk group experienced a large benefit from chemotherapy (relative risk, 0.26; 95% CI, 0.13 - 0.53; absolute decrease in 10-year distant recurrence rate: mean, 27.6%). It is also a useful tool to identify low risk patients (Recurrence score <18) that had minimal, if any, benefit from adjuvant chemotherapy avoiding unnecessary treatment in patients with low probability of recurrence even when standard prognostic factors initially classified those patients as high risk (relative risk, 1.31; 95% CI, 0.46 - 3.78; absolute decrease in distant recurrence rate at 10 years: mean, -1.1%) (Paik et al., 2006). Although Oncotype DX® is a useful tool in the decision making process and has the advantage of being performed in FFPE tissue, it has significant cost implications. The test itself costs thousands of dollars, although this could be set against the significant costs of unnecessary chemotherapy treatments or even with the medical cost of those untreated patients that will develop metastatic disease in the future (Sims et al., 2006). A validation clinical Study is ongoing (Trial Assigning Individualized Options for Treatment Rx, TAILORx Study).

Mammaprint® is another prognostic signature initially developed in N–ve patients using FFT. However, validation study included a mixed population of LN+ and ER-ve patients. This signature classified patients using a 70-gene signature into good and poor group according to their risk of 5 and 10 year response rate [RR] (van 't Veer et

al., 2002). MINDACT (Microarray in Node Negative Disease May Avoid Chemotherapy), a clinical trial to validate Mammaprint® is currently ongoing.

Homeobox 13 (HOXB13) and interleukin 17B receptor (IL17BR), the two gene ratio, has been developed in ER+ patients. High levels of HOXB13 predict recurrence in patients treated with adjuvant tamoxifen (Ma et al., 2004). This ratio has been validated in more than 850 patients Stage I and II, N- ER+ BC (Ma et al., 2006).

In an interesting publication, Fan et al applied the prediction derived from 5 different gene-expression models to a dataset of 295 BC patients (Fan et al., 2006). The different signatures or predictors used were the intrinsic subtype classification (Perou et al., 1999), 70-gene profile (van 't Veer et al., 2002, van de Vijver et al., 2002) , wound response (Chang et al), recurrence score (Paik et al., 2004) and the two-gene ratio (Ma et al., 2004). Even though there was little overlapping in terms of the genes involved, 4 out of the 5 predictors used show high agreement with clinical outcome, possibly because they all were focusing on the same biological pathways. Only the H/I ratio failed to identify significant differences in outcome in this study but questions remains as to whether its performance was assessed in the right population. More relevant for our research, all 4 models, to varying degrees, showed significant correlation with HG ($p<0,001$). However, the signatures seem to provide prognostic information beyond that offered by the HG alone (Fan et al., 2006).

It is important to remember that cDNA arrays measure specific mRNA whereas CGH measures DNA relative CNV and they do not necessarily have a close relationship, as occurs in gene silencing due to post-translational changes (van Beers and Nederlof, 2006). However, in practice, several published studies have demonstrated a good correlation (Pollack et al., 2002, Hyman et al., 2002) and their prognostic significance improve when Gene expression signature (GES) and copy number signature (CNS) are used together (Bergamaschi et al., 2006, Chin et al., 2006, Haverty et al., 2008).

However, cDNA arrays might have several disadvantages in comparison with aCGH: it represents only the expressed genes, lack of intron sequences, low signal-to-noise

ratio, variable signal intensities due to different length of the cDNA targets and less achievable resolution (Davies et al., 2005).

Although the available data for aCGH in BC is not as mature as the expression array data, significant advances have been made. In relation to hereditary cancer, Wessels studied patients with BRCA-1 mutation against a group of patients with unknown BRCA status. He found that BRCA-1 mutation carriers exhibit specific somatic mutations including 3p, 3q and 5q. (Wessels et al., 2002). Similarly, another publication confirmed distinct genomic profiles in both BRCA-1 and BRCA-2 mutation when compared to sporadic BC cases (Jonsson et al., 2005). It has also been possible to establish a profile pattern among synchronous tumours (Nyante et al., 2004) and help to differentiate new primary lesions from recurrences (Wa et al., 2005).

There seems to be consistency among several publications that have described the most common genetic alterations in BC even though some of these papers used low resolution technologies (including conventional aCGH). The frequently reported regions of gain include 1q, 8q, 11q, 17q and 20q being the most frequently lost regions 6q, 8p, 9p, 13q, 16q, 17p and 18 q (Kallioniemi et al., 1994, Tirkkonen et al., 1998, Rennstam et al., 2003) .

Some researchers have found significant correlation between a high level gain at 8q and 20q13 with early recurrence from BC defined as occurred in less than 5 years and association with nodal involvement with loss of 13q (Cingoz et al., 2003). Other authors also described common amplicons in BC such as 1q32.1, 8q24.3, and 16p13.3 (Naylor et al., 2005).

As more molecular classifications using expression analysis in BC are being developed and clinical trials validated further research is needed into the classification of BC using DNA arrays in order to elucidate molecular signature for prognosis with possible therapeutic implications.

1.5- Overview of the Project

Two main studies were designed based in the use of aCGH in BC patients in this project. Patients selection for both studies were based on the results of a descriptive analysis of clinico-pathological parameters of a BC cohort with long term follow-up treated in a single centre in the UK. Therefore, in order to increase the clinical relevance of the results, MVA for OS, Distant disease free survival (DDFS) and Disease-Free survival (DFS) was performed and is presented in Chapter 2. Relevance of receptor status in grade 3 cases is also discussed along with the Kaplan-Meier (KM) survival analysis among subgroups classified using IHC and HG.

For the first study, it was decided to include BC patients with poorly differentiated tumours, mainly HG3, as HG was one of the main prognostic factors in the MVA and probably the only one that reflects tumour biology rather than disease burden. The aim of this study was the identification of CNA significantly associated with poor survival in a sub-group of patients with poorly differentiated tumours. Design and results are presented in Chapter 4.

The second study aimed to research the relevant clinical problem of tamoxifen resistance. For this, aCGH profiles were obtained from both tamoxifen resistant cell lines and FFPE tissue obtained from patients clinically defined as tamoxifen resistant. Comparison of those CNA obtained among tamoxifen resistance cell lines and tumours were also performed. Results of this study are presented in Chapter 5. Description of Materials and Methods used for both studies are presented in Chapter 3.

Finally, Chapter 6 includes a short discussion on the conclusion, limitations and recommendations for future research.

CHAPTER 2 – Retrospective cohort analysis.

2.1- Cohort Overview – Statistical analysis

2.1.1- Patients characteristics.

A breast cancer database was built retrospectively in order to obtain epidemiological, clinical and histo-pathological data from patients seen at the BC services in the Oncology Department at Singleton Hospital, Swansea, South Wales, UK. Data was collected using a database designed by a statistician in "Access®" computer software. Data was collected by a medically qualified single observer over two-year-period (2004-2006) and included patient's records from 1996 to 2002. Information collected from clinical records consisted in 5 different sections as in table 2.1.

Table 2.1.- Epidemiological, clinical and pathological parameters collected from clinical files.

Patient	Hospital number, date of birth, sex, menopausal status, hormone replacement therapy, family history, date of surgery, side of surgery, type of breast surgery, type of axillary surgery, current status*, cancer related death and date of death.
Pathology	Pathology specimen number, histological type, presence of ductal carcinoma in-situ, tumour size, multifocality, number of positive nodes, total number of nodes removed, histological grade , ER, PR, HER-2 status and NPI score.
Adjuvant therapy	Adjuvant radiotherapy: irradiated sites, fractionation and start/completion date.
	Adjuvant chemotherapy regimen, number of cycles, start/completion date.
	Adjuvant hormones: drug used, duration, start/completion date.
Neoadjuvant Therapy	Neoadjuvant treatment given, neoadjuvant chemotherapy start/completion date, neoadjuvant hormone therapy start/completion date, neoadjuvant radiotherapy start/completion date.
Metastatic Disease	Metastatic location, first, second and third line treatment administered including dates, modality, start/completion date and response.

* Defined as status at last available observation.

Data from a total of 1489 non-metastatic patients with invasive BC from the database were obtained. Patients with Ductal carcinoma-in-situ (DCIS) only or metastatic patients at diagnosis were excluded. As the main histo-pathological variable in consideration in the present study was HG, it was agreed to include only 1339 patients for whom HG was available. Database was exported into Excel, revised and corrected for inconsistencies when possible. Dichotomous variables were

numerically coded according to a coding system (appendix 1). Continuous variables were transformed into categorical using widely accepted classifications when possible as in the case of TNM staging system (Singletary and Greene, 2003) or NPI (Galea et al., 1992).

ER results were reported using the quick (Allred) score as recommended in the UK (NHSBSP, 2005). An ER negative result was considered so if Quick (Allred) score was reported ≤2 being therefore positive if scored ≥3 to 8. Patients were divided into 4 categories according to ER, PR and HER-2 receptor status: ER positive (ER+) patients were classified as ER+/HG1 (broadly representative of the "Luminal A" group classified on molecular profiling) or ER+/HG3 and/or HER-2+ tumours (broadly representative of the "Luminal B" group). ER negative (ER-) patients were allocated into the HER-2 group if HER-2 positive (HER-2+) and triple negative (TN) if ER, PR and HER-2 were negative.

Statistical variables were generated as followed:

- Age: was categorized into 3 groups (≤50, 50-70 and ≥70 years) as an approximation of the ages considered as thresholds between premenopausal and postmenopausal women (Kroman et al., 2000) and elderly population.
- Follow-up period: Time (in years) between date of diagnosis (DODx) and date last seen either on follow-up clinic or as in-patient and (Date last seen-DODx / 365.25).
- DFS: Time from DODx to date of relapse of BC. This definition includes local, regional or distant relapse as well as patients with a second primary BC and death from any cause. This definition has been used in previous clinical trials (Coombes et al., 2004).
- DDFS: Time from DODx to date of distant BC relapse as well as patients with a second primary BC and death from any cause.
- OS: Time from DODx to DOD from any cause (BC, non-BC and unknown).
- Tamoxifen-resistant group (TAMRG): defined as ER+ cases treated after surgery/RT with tamoxifen only and who died from BC within the first 5 years of follow-up; patients who received neo-adjuvant or adjuvant chemotherapy were excluded. These patients also had an evaluable aCGH profile.

- Tamoxifen-control group (TAMCG): defined as ER+ cases treated with tamoxifen only who were alive and well after 5 years of follow-up; patients who received neo-adjuvant or adjuvant chemotherapy were excluded. These patients also had an evaluable aCGH profile.

Data was then transferred using commercially available software (Stat/Transfer©, Circle System, version 8; http://www.stattransfer.com) into the appropriate format and subsequently analysed using Stata (Data Analysis and Statistical Software©, StataCorp LP, version 10 SE; http://www.stata.com). Categorical variables were analysed using Chi-square test. Normally distributed numerical variables were analysed using t-tests whereas non-parametric methods including Mann-Whitney test were used for non-normally distributed data. Cox regression models were built using a step forward approach to identify factors associated with OS, DDFS and DFS. Only factors found significantly associated with the outcomes in the Univariate analysis (UVA) were included in the MVA. KM plots and Log-rank test were used to compare survival between groups. The level of significance for p-value was set 0.05. Two-sided p-values were used.

2.1.2- Results

General patient's characteristics are summarised in Table 2.2. In our cohort, the median age was 59.69 years (range 18.5-91.43). Over half of patients were between 50-70 years old and in this subgroup the distribution of HG is 24.7%, 44% and 31% for HG 1, 2 and 3 respectively. By contrast, 134 (51.3%) of cancers in <50 y old patients had HG3 disease with only 34 (13%) patients in the young group having HG1 tumours. Adenocarcinoma was the most common histological type (1209, 90%), of which 86.5% were Invasive Ductal Carcinoma (IDC), 9.18% invasive lobular and 2.68% had mixed histology. Special histological types were rare (only 5 cases).

Table 2.2- Demographics and description breast cancer patients by HG

Variable	HG1	HG2	HG3
	No (%)	No (%)	No (%)
Total Number of patients	284 (21.1%)	589 (43.9%)	466 (34.8%)
Age - Median (range)	59.20 (33.9-91.4)	61.00 (28.8-88.3)	57.25 (18.5-86.9)
≤50	34 (13.0)	93 (35.6)	134 (51.3)
50-70	179 (24.7)	319 (44)	228 (31.3)
≥70	71 (20.4)	177 (50.7)	101 (28.9)
Unknown	0	0	3 (100)
Histological type			
Invasive Ductal Carcinoma	225 (21.4)	431(41.1)	394 (37.5)
Invasive Lobular carcinoma	27 (22)	87 (70.7)	9 (7.3)
Mixed	4 (11.1)	22 (61.1)	10 (27.8)
Other	2 (40)	0	3 (60)
Unknown	26 (20.8)	49 (39.2)	50 (40)
TNM-T staging*			
T1 (≤ 2 cm)	219 (29.6)	325 (44)	196 (26.4)
T2 (>2 - ≤5 cm)	48 (10.2)	206 (43.8)	216 (46)
T3 (>5 cm)	2 (10)	9 (45)	9 (45)
T4	8 (13.3)	25 (41.7)	27 (45)
Unknown	7 (14.3)	24 (49)	18 (36.7)
TNM N staging**			
No	217 (27.4)	336 (42.5)	238 (30.1)
N1 (1-3 nodes)	53 (15.8)	154 (45.8)	129 (38.4)
N2 (4-9 nodes)	10 (8)	58 (46.4)	57 (45.6)
N3 (≥10 nodes)	3 (3.5)	41 (48.2)	41 (48.2)
Unknown	1 (50)	0	1 (50)
Surgery -Breast			
Mastectomy	72 (12.1)	264 (44.4)	258 (43.4)
Wide Local Excision	205 (29.3)	303 (43.3)	192 (27.4)
Biopsy only	0	0	1 (100)
Unknown	7 (15.9)	22 (50)	15 (34.1)
Surgery -Axilla			
Axillary Clearance	181 (19.2)	447 (44.3)	382 (37.8)
Axillary Sampling	81 (35.5)	92 (40.1)	56 (24.4)
Unknown	22 (22)	50 (50)	28 (28)

* TNM Tumour staging by Singletary et al al 2003.
** TNM Nodal Staging by Singletary et al 2003.

Taking into account 1290 patients where HG, tumour size and nodal staging was available, 738 (57.2%) of patients had T1/T2 node negative BC. As expected, there was a significant association ($p<0.001$) between tumour size and nodal staging. A higher proportion of patients with T2/T3 tumours have node-positive disease when comparing to T1 tumours (Table 2.3).

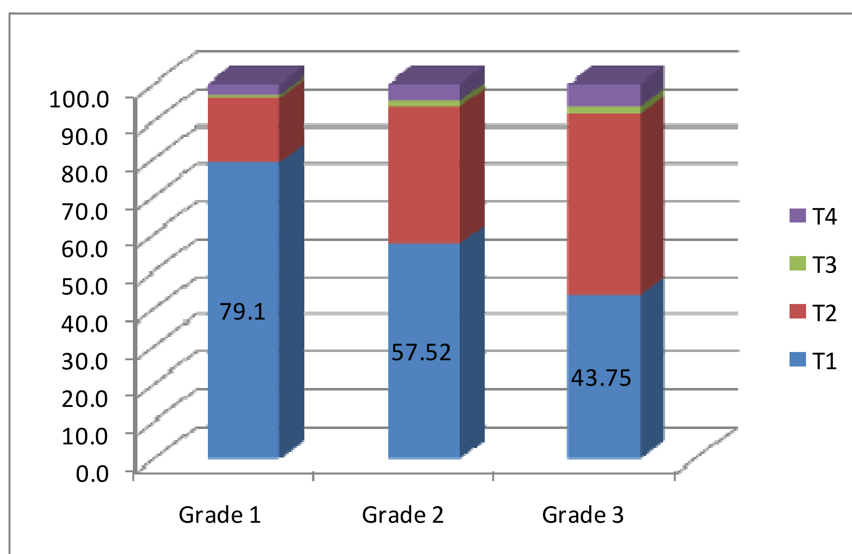
Table 2.3 - Distribution according to tumour and nodal staging*

	Nodal Staging				
	No	N1	N2	N3	Total
Tumour Staging	No (%)	No (%)	No (%)	No (%)	No (%)
T1 (≤ 2cm)	521 (70.4%)	156 (21.1%)	44 (5.9%)	19 (2.6%)	740 (56.9%)
T2 (>2 - ≤ 5 cm)	217 (46.2%)	142 (30.2%)	59 (12.6%)	52 (11.1%)	470 (36%)
T3 (>5 cm)	6 (30.0%)	7 (35.0%)	3 (15%)	4 (20%)	20 (1.7%)
T4	19 (31.7%)	19 (31.7%)	15 (25%)	7 (11.7%)	60 (5.4%)
Total	763	324	121	82	1290

* TNM classification system according to Singletary et al 2003.

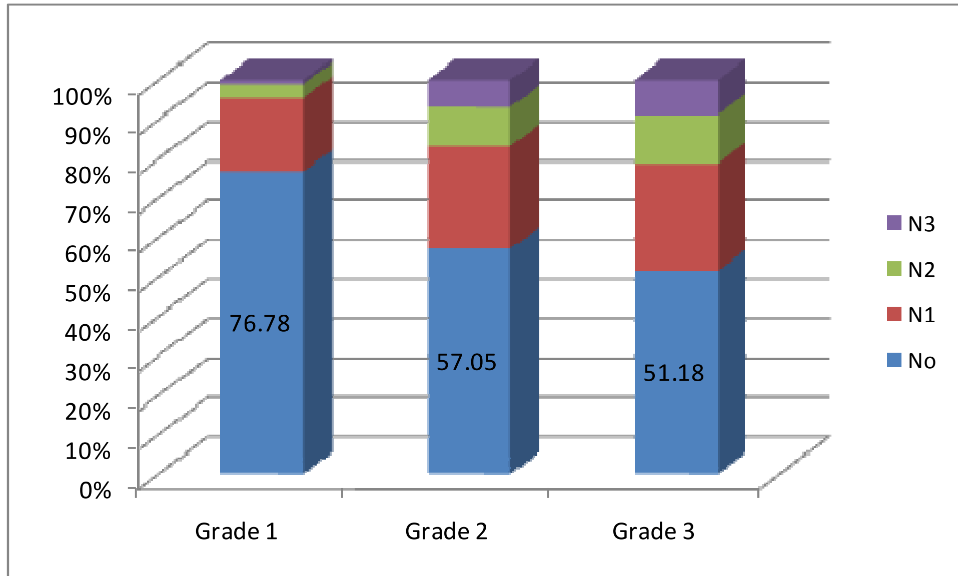
There was also a highly significant association between tumour size and HG ($p<0.001$). Only 26.5% of patients with tumours ≤2 cm had HG3 as opposed to T4 disease were 45% of tumours were poorly differentiated. In another words, 79.1% of those well differentiated tumours were lesions <2 cm with an increment in tumour size/staging in HG2 and 3 cases (Figure 2.1).

Figure 2.1 Distribution of HG into tumour staging



A similar scenario was seen when we looked at the relationship between nodal status and HG ($p<0.0001$) as patients had a 3.1 fold increased risk of node positivity in HG3 disease (Odds ratio 3.12; $p<0.0001$; CI: 2.2-4.3). 76.68% of the HG1 patients had node negative disease. By contrast, 48.82 % of all HG3 cases had node positive disease; a similar picture was seen for HG2 cases (Figure 2.2).

Figure 2.2- Distribution of HG into nodal staging categories.



Wide local excision (WLE) was the main surgical procedure in our cohort, although closely followed by mastectomy (54% and 45% respectively). A significant association ($p<0.001$) was found between the surgery method and tumour size with 70% of T1 patients having BCS as opposed to T2 or even T3 patients were BCS rate was 35 and 10% respectively. Axillary clearance was performed in 75% of cases.

2.1.2- Receptor Status

Information regarding hormone receptor status was available on 835 (62.3%) patients, 217 (16.2%) and 227 (16.95%) for ER, PR and HER-2 respectively. Of those for which data on ER status were available, 624/835 (74.7%) were positive. The proportion of HG1 and HG3 among ER+ cases was similar (HG1:145/624; 23.2% and HG3:164/624; 26.3%). By contrast, the vast majority of ER- patients had HG3 disease (184/211; 87.2%). The positivity rate for HER-2 in the study was 17.2% (39/227). HER-2 testing was either requested by clinicians in view of high risk disease when considering Trastuzumab therapy or subsequently performed for research purposes in the current study. Hence the majority of HER-2 tests were performed in HG3 cases (176/227) with only 49/227 and 2/227 tested cases in HG2

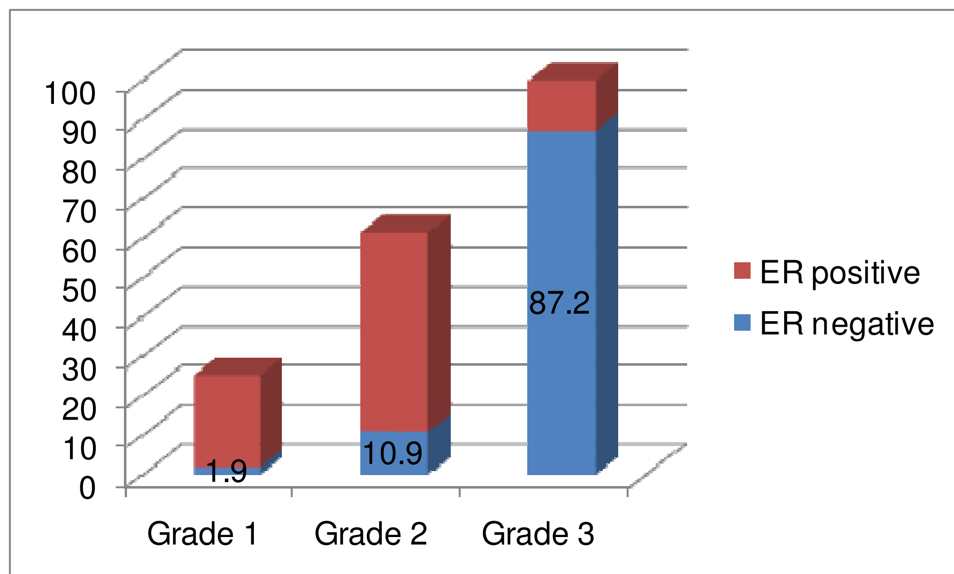
and HG1 respectively. Distribution of ER, PR and HER-2 results according to HG is shown in table 2.4.

Table 2.4 .- ER, PR and HER-2 status according to HG

	Grade 1 No (%)	Grade 2 No (%)	Grade 3 No (%)
ER			
Negative - No (%)	4 (1.9)	23 (10.9)	184 (87.2)
Positive - No (%)	145 (23.2)	315 (50.5)	164 (26.3)
Unknown - No	135 (26.8)	251(49.8)	118 (23.4)
PR			
Negative - No (%)	0	9 (6.0)	140 (93.9)
Positive - No (%)	5 (7.4)	27 (39.7)	36 (52.9)
Unknown - No	279 (24.9)	553 (49.3)	290 (25.8)
HER-2			
Negative - No (%)	2 (1.2)	41 (25.5)	128 (79.5)
Indetermined - No (%)	0	3 (23.1)	10 (76.9)
Positive - No (%)	0	5 (12.8)	38 (97.4)
Unknown - No	282 (25.4)	540 (4.9)	290 (26.1)

There was a significant proportion of HG3 cases with ER- disease in comparison with the ER+ rate in well differentiated tumours (Figure 2.3).

Figure 2.3- Distribution of ER status according to HG



2.1.3- Adjuvant Treatment

A summary of the adjuvant treatment according to HG is shown in Table 2.5.

2.1.3.1- Chemotherapy (CT)

64.15% (859/1339) of patients did not receive adjuvant CT. The majority of those who did not receive CT had N- disease (71.83%). A high proportion of patients treated with CT were HG3 (249/427; 58.3%), N+ disease (67.2%) or NPI >3.5 (90%). Amongst the 427 patients who received CT, 303 patients (71%) received a non-anthracycline containing regimen such as CMF and 90 (21%) anthracycline-containing regimen.

Table 2.5.- Adjuvant treatment received by HG

Variable	Grade 1 No (%)	Grade 2 No (%)	Grade 3 No (%)
Adjuvant Chemotherapy			
No chemotherapy	242 (28.2)	425 (49.5)	192 (22.4)
First generation	29 (9.6)	106 (34.9)	168 (55.4)
Second generation	2 (2.2)	30 (33.3)	58 (64.4)
Third generation	0	11 (32.3)	23 (67.6)
Unknown	11 (20.7)	17 (32.1)	25 (47.2)
Adjuvant Radiotherapy			
No radiotherapy	50 (22.9)	96 (44.0)	72 (33.03)
Breast/Chest wall	190 (24.0)	370 (46.7)	231 (29.2)
Breast/Chest wall+SCF	15 (10.0)	49 (32.9)	85 (57.1)
Breast/Chest wall+SCF+Axilla	18 (15)	56 (46.7)	46 (38.3)
Not specified	0	0	2 (100)
Unknown	11 (23.9)	18 (39.1)	30 (36.9)
Adjuvant Hormones			
No adjuvant hormones	6 (6.3)	20 (21.1)	69 (72.6)
Tamoxifen	259 (22.7)	528 (46.3)	354 (31.0)
AI	5 (15.1)	17 (51.5)	11 (33.3)
Other	3 (25)	6 (50)	3 (25)
unknown	11 (19.0)	18 (31.0)	29 (50.0)

2.1.3.2- Radiotherapy (RT)

1062 (79.31%) patients received adjuvant radiotherapy (RT). Amongst those, 791/1062 (74.4%) received RT to the breast/chest wall (CW) only.

2.1.3.3- Hormones

88.9% (1186/1281) of patients were prescribed anti-hormonal treatment, mainly Tamoxifen. Only 2.5% of patients received adjuvant Aromatase Inhibitor (AI) as initial adjuvant treatment.

2.1.4- Follow up

Median follow-up time in our cohort was 5.42 years (Interquartile range, IQR 3.96-6.77). A total of 284 cases developed recurrent disease during the follow-up period. A higher proportion of HG3 cases developed early BC relapse after diagnosis compared to HG1 patients. In addition, the time between relapse and death is shorter in HG3 cases (table 2.6). Analysis of the relapse pattern showed that HG3 cases had more both local and distant relapses than HG1; from 60 patients who developed local relapse only 3 had HG1 versus 41 patients with HG3. Similarly, among 187 patients with distant recurrence, only 9 were HG1 tumours. KM survival estimates for OS, and DFS in HG cases are shown in Figure 2.4.

Table 2.6.- Distribution of number of cases with recurrence BC and death according to HG

Variable	Grade 1 No (%)	Grade 2 No (%)	Grade 3 No (%)
Recurrence			
No (%)	12 (4.2)	90 (31.7)	148 (52.1)
Median TTR* years (IQR)	3.61 (2.40-5.0)	2.1 (1.4-4.6)	1.61 (1.1-2.9)
Death			
Cancer-related death	8 (4.3)	57 (30.8)	120 (64.9)
Non-cancer related death	5 (19.2)	8 (30.8)	13 (50)
Median TTD**	5.64 (3.6-7.02)	3.3 (2.21-5.3)	2.26 (1.57-3.27)

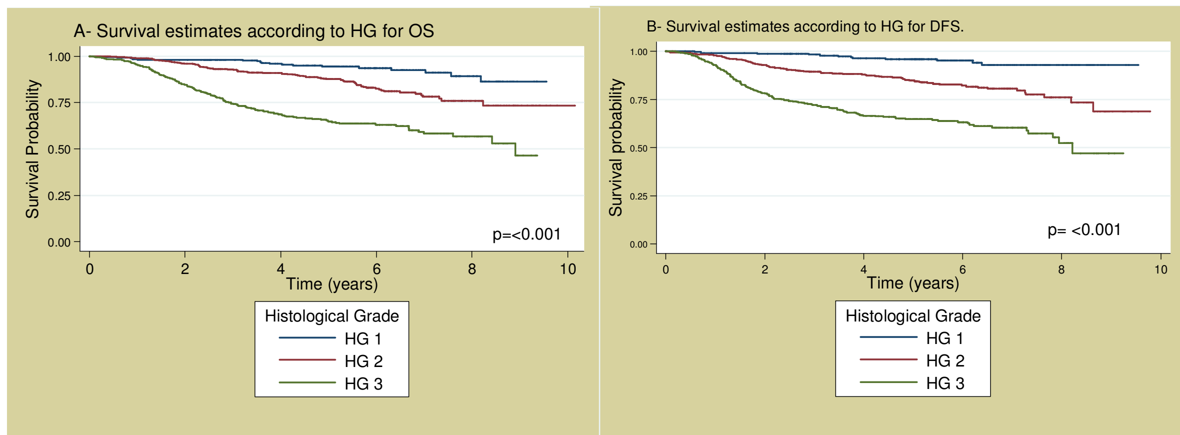
* TTR - time-to-recurrence

** TTD - time-to-death

A total of 1,216 patients had information regarding current status and follow up at the time of completion of the data collection. From those, 277 patients (22.78 %) died during follow-up. 59.20% of those who died were HG3 as opposed to HG1 (20 patients, 7.2%). The proportion of fatal cases among those with HG3 tumours was significantly higher than in any other HG ($p<0.0001$). Analysis of cancer-related death revealed that almost 40% (38.46%) of fatal cases among grade 1 died from

causes other than cancer. By contrast, about 90% of deaths among HG2 and HG3 cases were due to cancer (p<0.011).

Figure 2.4- KM survival estimates for OS (A) and DFS (B) according to HG.



2.2- Identification of Clinico-pathological variables associated with OS, DDFS and DFS in BC: Cox regression analysis.

2.2.1- Univariate analysis.

Table 2.7 shows a summary of the UVA. In HG3 tumours the risk of dying from any cause was 6.6 times higher than patients with HG1 disease (HR:6.64, 95%CI 4.07-10.82). Nodal status analysed as nodal status positive and negative or as in the TNM nodal staging showed significant association with OS (p<0.001). The risk of dying almost doubled from N- to N1 disease (HR:1.76; 95%CI 1.29-2.43) and from N1 to N2 disease (HR:4.23; 95%CI:3.01-5.94). Although there was still an increment in the risk of dying from N2 to N3 disease, the magnitude of this was smaller with partial overlapping in the CI (HR:5.89; 95%CI:4.09-8.49).

When compared with tumour staging and nodal staging, HG showed the strongest association with OS, even more so than N3 disease. However, the factor most strongly associated with OS in the UVA was the NPI which is not surprising as it takes into consideration three of the most significantly associated variables with OS, namely HG, nodal status and albeit with a lesser contribution to the equation, tumour size.

Table 2.7- Cox regression UVA for OS, DDFS and DFS

Variable	OS			DDFS			DFS		
	HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Age									
≤50	1		< 0.001	1		0.024	1		0.020
50-70	0.95	0.67 - 1.34		0.88	0.61 - 1.28		0.91	0.66 - 1.27	
≥70	1.99	1.40 - 2.84		1.35	0.90 - 2.05		1.40	0.97 - 2.00	
Histological Grade									
1	1		< 0.001	1		< 0.001	1		< 0.001
2	2.45	1.47 - 4.07		4.4	2.20 - 8.80		3.4	2.17 - 7.26	
3	6.64	4.07 - 10.82		10.21	5.16 - 20.18		10.00	5.54 - 18.01	
TNM-T Staging*									
T1	1		< 0.001	1		< 0.001	1		< 0.001
T2	2.52	1.91 - 3.33		3.50	2.50 - 4.89		3.58	2.66 - 4.81	
T3	4.99	2.31 - 10.81		8.94	3.83 - 20.86		7.31	3.36 - 15.96	
T4	5.03	3.28 - 7.71		7.50	4.52 - 12.56		6.78	4.24 - 10.51	
Nodal status									
Negative	1		< 0.001	1		< 0.001	1		< 0.001
Positive	2.8	2.16 - 3.61		3.77	2.76 - 5.14		3.52	2.69 - 4.60	
TNM-N Staging**									
Negative	1		< 0.001	1		<0.001	1		< 0.001
N1	1.76	1.29 - 2.43		2.31	1.59 - 3.36		2.23	1.62 - 3.08	
N2	4.23	3.01 - 5.94		5.98	4.02 - 8.90		5.47	3.88 - 7.70	
N3	5.89	4.09 - 8.49		8.27	5.39 - 12.69		7.09	4.86 - 10.40	
NPI***									
≤3.4	1		< 0.001	1		< 0.001	1		< 0.001
3.4 - ≤5.4	5.78	3.56 - 9.40		5.80	3.28 - 10.25		5.47	3.35 - 8.94	
> 5.4	16.64	10.15 - 27.28		21.28	12.01 -37.70		18.82	11.50 - 30.79	
ER									
Negative	1		< 0.001	1		0.011	1		< 0.001
Positive	0.34	0.24 - 0.47		0.58	0.39 - 0.87		0.51	0.36 - 0.72	
HER-2									
Negative	1		0.415	1		0.296	1		0.40
Indetermined	1.59	0.63 - 4.00		2.00	0.78 - 5.10		1.64	0.70 - 3.84	
Positive	1.37	0.78 - 2.40		1.41	0.73 - 2.70		1.32	0.75 - 2.31	
Chemotherapy									
No	1		0.002	1		< 0.001	1		< 0.001
Non-anthracycline	1.61	1.22 - 2.12		2.01	1.47 - 2.75		1.80	1.37 - 2.37	
Anthracycline	1.72	1.06 - 2.79		1.71	0.97 - 2.99		1.78	1.10 - 2.85	
Other	1.86	0.98 - 3.54		2.21	1.07 - 4.52		1.98	1.04 - 3.77	
Adjuvant Hormones									
No	1		< 0.001	1		< 0.001	1		< 0.001
Tamoxifen	0.29	0.20 - 0.42		0.36	0.23 - 0.56		0.39	0.27 - 0.60	
AI	0.11	0.03 - 0.46		0.24	0.07 - 0.81		0.27	0.94 - 0.77	
Other (Tam+Zoladex)	0.15	0.02 - 1.10		0.21	0.03 - 1.57		0.18	0.24 - 1.33	

* TNM Tumour staging by Singletary et al al 2003.
** TNM Nodal Staging by Singletary et al 2003.
*** NPI Notthingham Prognostic Index. Galea et al 1992.

Positive ER status reduced the risk of dying in about 70% (HR: 0.34; 95%CI: 0.24-0.47; p<0.001). Similarly, both tamoxifen (HR:0.29; 95%CI:0.20-0,42; p<0.001) and AI (HR:0.11; 95%CI:0.03-0.46) were associated with better DDFS. Age was not associated with DDFS and DFS in the UVA.

Patients who received adjuvant non-anthracycline based regimen had twice the risk for developing DDFS than had patients who did not received chemotherapy (HR: 2.01; 95%CI:1.47-2.75; p<0.001). Although at a lower magnitude, the same is observed for OS and DFS. No association was found in this cohort between OS, DDFS and DFS and HER-2 status possibly because of insufficient number of patients having HER-2 testing.

2.2.2- Identification of Clinico-pathological variables associated with OS, DDFS and DFS in BC. Multivariate analysis.

Variables significantly associated with the endpoints were included in the MVA as shown in table 2.8. Age remained significantly associated with OS with patients > 70 years old having an increased risk of dying from any cause when compared to patients younger 50 years (which is expected in view of higher incidence of co-morbidities in elderly population). In the MVA analysis, HG stood as one of the main variables associated with OS, DDFS and DFS with only N3 disease conferring a higher risk of developing the endpoints. By contrast, HG2 lost its significance when adjusted for other variables in the MVA for OS, retaining its significance for DDFS and DFS.

Tumour stage remained significantly associated with all the endpoints, except for T3 disease in the OS analysis, probably due to the low number of observations in the T3 category (only 7 cases). For nodal staging, the risk of relapse/death from BC was higher as the number the positive LN increased. However, the magnitude of increased risk was lower between N2 to N3 disease than for patients with <3 positive nodes.

Adjuvant tamoxifen was a protective factor for all the endpoints. Use of chemotherapy, when adjusted for other variables, was associated with better outcome in DDFS and DFS particularly for anthracycline and “other” category mainly represented by taxane-containing regimen. As NPI is a complex variable, it was decided not to include it in the MVA, avoiding over-weighting some of the variables included in the NPI that were also individually included in the Cox regression models.

Table 2.8- Cox regression MVA OS, DDFS and DFS

Variable	OS		DDFS		DFS	
	HR	95% CI	HR	95% CI	HR	95% CI
Age						
≤50	1					
50-70	1.17	0.73 - 1.85				
≥70	2.16	1.25 - 3.75				
Histological Grade						
1	1		1		1	
2	1.45	0.67 - 3.16	2.72	1.34 - 5.49	2.55	1.37 - 4.7
3	3.26	1.48 - 7.15	6	2.98 - 12.08	6.06	3.30 - 11.11
TNM-T Staging *						
T1	1		1		1	
T2	1.49	1.02 - 2.19	2.41	1.70 - 3.43	2.37	1.74-3.23
T3	0.77	0.10 - 5.70	3.73	1.54 - 9.05	3.55	1.58-7.95
T4	3.4	1.80 - 6.44	4.00	2.34 - 6.86	3.64	2.25-5.86
TNM-N Staging **						
Negative	1		1		1	
N1	1.43	0.92 - 2.24	1.78	1.19 - 2.66	1.68	1.192-2.39
N2	3.27	2.00 - 5.34	5.37	3.45 - 8.35	4.60	3.14-6.75
N3	3.79	2.09 - 6.87	7.13	4.39 - 11.57	5.7	3.72-8.76
ER						
Negative	1					
Positive	0.54	0.36 - 0.83				
Chemotherapy						
No	1		1		1	
Non-anthracycline	0.8	0.51-1.27	0.74	0.53 - 1.06	0.69	0.51-0.94
Anthracycline	0.71	0.36-1.40	0.38	0.21 - 0.70	0.45	0.27-0.76
Other	0.34	0.11-1.04	0.33	0.15 - 0.71	0.31	0.16-0.65
Adjuvant Hormones						
No	1		1		1	
Tamoxifen	0.46	0.27 - 0.79	0.35	0.21 - 0.57	0.4	0.26-0.63
AI	0.84	0.01 - 0.64	0.17	0.05 - 0.60	0.21	0.72-0.62
Other (Tam+Zoladex)	0.87	0.10 - 7.03	0.26	0.30 - 1.95	0.24	0.33-1.82

* TNM Tumour staging by Singletary et al al 2003.
** TNM Nodal Staging by Singletary et al 2003.

2.3- Relevance of ER Status in HG3 tumours

ER lost its association with DDFS and DFS when adjusted for other variables, although remained significant in the OS analysis (ER+ HR: 0.54; 95%CI: 0.36-0.83). As the main prognostic factor evaluated in the present study was HG, UVA was repeated separately among HG1 and HG3 patients in order to assess the effect of HG on the prognostic information provided by ER status.

ER status was not associated with DDFS and DFS in HG3 cases but it was significantly associated with DDFS and of borderline significance for DFS in HG1 disease (HG3: DDFS HR:1.16; 95%CI:0.72-1.87, p:0.54 and DFS HR:1.07; 95%CI:0.72-1.60; p:0.73 and HG1: DDFS HR:0.09; 95%CI:0.01-0.87, p<0.037; DFS HR:0.58; 95%CI: 0.01-1.11; p:0.06).

For HG1, due to the good prognosis in this population of ER+/HG1 cases, the number of deaths or relapses was too low to provide significant results (number of ER+/HG1 dead cases contributing to the analysis were 9 as opposed to 105 in ER+/HG3). By contrast, any protective effect of ER was not apparent in patients with HG3 disease in respect of DDFS and DFS.

2.4- Tumour classification using IHC as a surrogate for GEP molecular sub-types.

Table 2.9 shows results of Cox regression MVA among different tumour groups with the survival endpoints adjusting for the effect of adjuvant hormones and chemotherapy. In the OS analysis, ER+/HG3 had 6.27 fold higher risk of dying (95%CI 2.74-14.38) in comparison to ER+/HG1 patients despite of adjusting for adjuvant hormones and chemotherapy. The risk was even higher when assessed for DDFS and DFS.

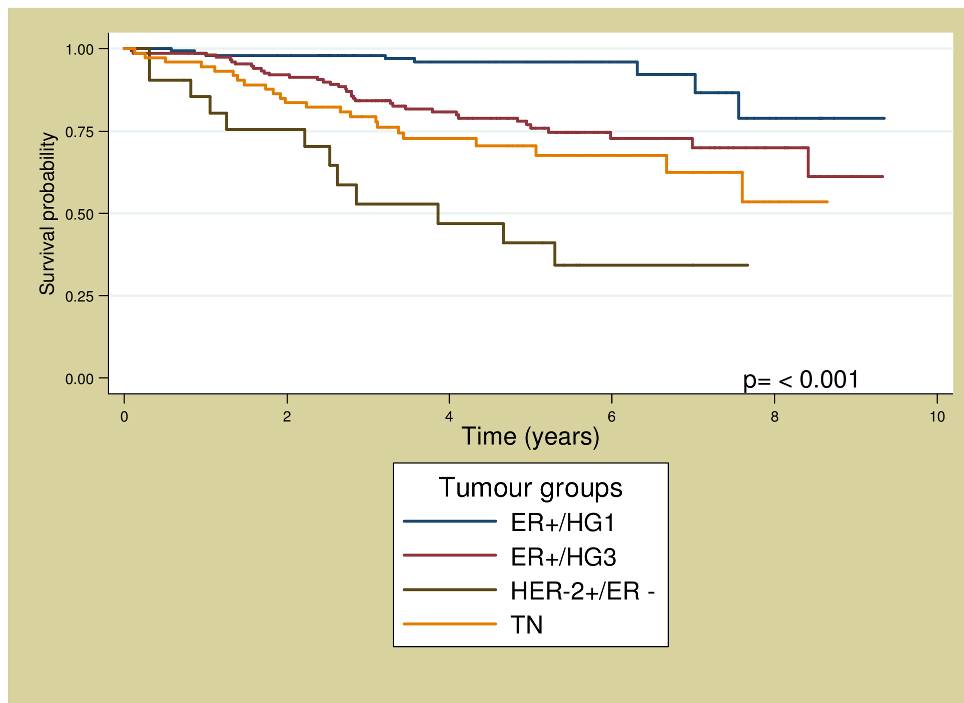
Table 2.9- Cox regression MVA according to tumour groups adjusting for adjuvant therapy for the survival endpoints OS, DDFS and DFS

Variable	OS		DDFS		DFS	
	HR	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Tumour Group						
ER+/HG1	1		1		1	
ER+/HG3	6.27	2.74-14.38	13.41	4.00-44.92	8.78	3.36-22.93
HER-2 +/ER -	19.52	6.86-55.58	26.88	6.20-116.56	21.79	6.72-70.63
TN	8.37	3.42-20.51	8.81	2.28-33.99	8.13	2.81-23.49
Adjuvant Hormones	0.49	0.25-0.98	0.36	0.15-0.87	0.46	0.22-0.98
Adjuvant Chemotherapy	0.45	0.27-0.74	0.56	0.31-1.00	0.57	0.34-0.96

TN patients had worse OS than ER+/HG3 tumours. Interestingly, the risk of relapse was higher in patients with ER+/HG3 disease compared to TN patients. 13/66 (19.69%) patients within the ER+/HG3 group were HER-2 positive. A sensitivity analysis excluding these cases did not alter the results. Cox regression analysis was repeated for N- patients with the same results. For OS in N- patients, ER+/HG3 imposed a risk of 3.4 fold (95%CI:1.01-10.84, p:0.03) with respect to the risk of death compared with ER+/HG1 patients.

Kaplan Meier survival estimates according to tumour groups are shown in Figure 2.5. This graph confirmed the excellent prognosis for patients in the ER+/HG1 group, with a survival at 5 years over 95%; relapses in this group occurred after a follow-up of more than 6 years. For ER+/HG3 patients, the curve overlapped that from ER+/HG1 during the first 18 months with increasing divergence from that point. The survival for TN patients was poorer than for ER+ patients, with deaths occurring from diagnosis. However, the curve for TN tumours closely followed that for ER+/HG3 disease.

Figure 2.5 - KM for OS estimates according to tumour groups



2.5- Discussion

The demographic data are consistent with our belief that this was a representative cohort of BC routinely presenting to hospital practice in the UK. In our cohort, young women (≤50 years) represented 19.4% of the population. This is entirely consistent with age specific incidence of BC in the UK where 81% of BC cases occurred in women aged 50 years and over (Office-for-National-Statistics, 2008). The peak of BC in both our cohort and data published in the UK showed that over half of the cases are diagnosed in the 50-70 age group. Similar statistics are described in the Danish Breast Cancer Cooperative Group (DBCG 04) with approximately 70% of women being diagnosed with BC having over 55 years of age (Moller et al., 2008).

Histological types are also consistent with the literature with around 85% of adenocarcinoma (ADC) of the breast having ductal, as opposed to lobular, histology (Moller et al., 2008, 2009). In some series, however, the proportion of special histological types among all the invasive carcinomas accounts for up to 25% (Reis-Filho and Lakhani, 2008). Histological tumour type has also prognostic significance as ductal carcinoma has been reported to have worse survival than other histological types (Li et al., 2003). However, histological type is not useful as a prognostic discriminator as the vast majority of cases, between 70-90%, are classified as ductal ADC depending on the series or even geographical regions (Enjoji et al., 1992).

If compared to recent data (Morrow et al., 2009), it seems that the proportion of BCS in our cohort is low, only 52.27%. However, data from 2004 showed comparable figures for BCS to ours (Moller et al., 2008). Undoubtedly, early tumour detection, since introduction of screening program, has contributed to the declining of mastectomy rates during the last decade. It is estimated that 1 in 8 women would be spared mastectomy as a results of earlier mammographic detection (Advisory-Committee-on-Breast-Cancer-Screening, 2006). The current reported mastectomy rates in most the BC departments in the UK are around 20-35% depending on the literature (Tataru et al., 2006).

The proportion of axillary clearance in this cohort is higher than the current rate which again contrasts with what is currently recommended (Harnett et al., 2009). However, by the time of completion of data collection sentinel lymph node biopsy (SLNB) was still considered experimental as the randomised trial favouring the use of SLNB in early-stage breast cancer with clinically negative nodes were published in later (Mansel et al., 2006).

We found that HG, tumour size and nodal staging were variables that associated with each other and with OS and CSS. Just over 55% of our patients had breast tumours measuring ≤2 cm at diagnosis, similar to the figures previously reported in the DBCG (Moller et al., 2008). Stage I disease -T1N0- according to the AJCC (Singletary and Greene, 2003) was diagnosed in over 40% of patients. Our data is comparable to a recent published review from the Surveillance and Epidemiology and End Results (SEER) where 60% of cases had disease localized in the breast at diagnosis (Horner 2008). Although it is undeniable that variables like nodal status and tumour size still remain as dominant pathological prognostic variables in BC, it is also becoming evident that the relative weight of these variables in the BC prognosis will diminish as more tumours are diagnosed earlier. In addition, it is also recognised that tumour and nodal staging do not define biologically distinct entities (Sotiriou et al., 2003).

HG, however, is probably the best surrogate for tumour biology compared with any other variables analysed in this cohort and it is a very useful tool for assessing the patient's risk in an affordable and accessible manner when compared with new prognostic genetic tools such as Oncotype DX©. This is particularly true as some of the recently described gene expression signatures have a high content of proliferation-related genes (Paik et al., 2004, Sorlie et al., 2001, van 't Veer et al., 2002). Cell proliferation, however, can also be assessed in a way by the mitotic activity, a component of the HG, or assessed directly in a more quantifiable way by Ki-67. There is also evidence that it is the mitotic activity within the HG assessment, as opposed to pleomorphism or tubule formation, that is the strongest predictor of survival (Parl and Dupont, 1982, Frkovic-Grazio and Bracko, 2002).

With very few exceptions, HG is commonly mentioned as one of the main prognostic factors associated with survival in BC (Soerjomataram et al., 2008). Some publications confirm the importance of HG as part of the Nottingham grading system (Frkovic-Grazio and Bracko, 2002, D'Eredita et al., 2001), NPI (D'Eredita et al., 2001) or its association with chemotherapy response (Pinder et al., 1998). In our and other cohorts, HG had a stronger association with the set endpoints than tumours size and nodal status. In addition, patients with HG3 tumours tended to be younger, had larger tumours, N+ disease, relapse quicker and have a higher overall and cancer-specific mortality than patients with HG1.

Gene-expression profile (GEP) remains as the gold standard for the classification of the intrinsic sub-type since the original publication by Perou et al (Perou et al., 2000). However, its practical application to both clinical and research settings remains limited. “Luminal A” tumours, as defined by GEP, include tumours with high expression of ER-related genes and low expression of proliferative genes such as CCNB1, MKI67, MYBL2 and HER-2 associated genes (Cheang et al., 2009). Although no accurate surrogate markers are available for the definition of BC sub-types, attempts for the clinical characterisation of cases and IHC combinations in every sub-groups are worthwhile. As only approximately 18-30% of “Luminal B” tumours are identified by HER-2 positivity (Prat and Perou, 2010a, Cheang et al., 2009), the major distinction with “Luminal A” tumours is a high proliferation rate. Cell proliferation, which seems the most determinant factor to affect prognosis among ER+ tumours, is what is also measured by the recurrent score (Paik et al., 2004) or Ki-67 (Cheang et al., 2009). Others have sub-classified Luminal tumours according to GGI (genomic grade Index) using high GGI levels to identify “Luminal B” tumours (Durbecq et al., 2008). It also is becoming more apparent that HG “mirrors” the proliferation differences between Luminal A and B tumours (Prat and Perou, 2010a) and that HG significantly correlates with gene-expression signatures (Fan et al., 2006). In the present study, we have demonstrated that significant differences in prognosis are found among ER+ tumours depending on HG and perhaps HG3 might be used as a surrogate for the definition of “Luminal B” tumours, similar to previous publications using GGI or Ki-67. Approximately 25% of both false negative and positive cases are found when Ki-67 is used as surrogate for the identification of

“Luminal B” tumours (Cheang et al., 2009) and it is likely that this figure increases when HG is used for the same purpose. Despite this limitation, it seems that defining “Luminal B” tumours as ER+/HG3 will identify a higher proportion of “Luminal B” cases than ER+/HER-2+ which has been used in recent publications (Blows et al., Wiechmann et al., 2009). Furthermore, the main consequence of the use of a definition with poor sensitivity would be having ER+/HG3 cases wrongly added to the comparison group (ER+/HG1). This would underestimate the magnitude of any association found between case status (i.e. ER+/HG3) and mortality. However, even under these circumstances, we are still showing a clear association between ER+/HG3 status and OS, with a significant HR even after controlling for other variables also associated independently to death. If our definition was more sensitive, the only possible effect on the analysis done would have been an even stronger association on the Cox regression model as it might be the case when Ki-67 or GGI are used.

Although they express ER, “Luminal B” tumours tend to have poor response to therapy with either tamoxifen (Cheang et al., 2009), aromatase inhibitors (Dowsett et al.) or even chemotherapy (Brenton et al., 2005, Cheang et al., 2009) in comparison with “Luminal A” tumours. In this study, the risk of dying from BC among ER+/HG3 tumours was 6.27 times higher than ER+/HG1 for OS even after adjusting for adjuvant treatment. Although, we cannot quantify the effect of adjuvant therapies on OS, it is obvious that any benefits from therapy do not compensate for the worse prognosis conferring by the higher HG.

However, although clear prognostic differences can be established between HG1 and HG3, there is still considerable heterogeneity within the group of tumours classified as HG2. There have recently been a number of attempts to separate this intermediate group into “good” and “poor” prognosis subgroups which behave clinically more like HG1 and HG3 respectively. These approaches have included the GGI based on RNA expression analysed by microarray (Sotiriou et al., 2006, Loi et al., 2007b, Ivshina et al., 2006), analysis of 8 genes by RT-PCR (Toussaint et al., 2009), by immunocytochemisty or Ki-67 (Klintman et al., 2009). In our study, clinical

correlation of HG2 for some variables was more similar to that obtained for HG3 than for HG1.

We have also shown that in HG3 disease, ER lost its correlation with DDFS and DFS. The fact that ER is no longer a protective factor in HG3 disease for some of the endpoints is an important issue to consider when assessing adjuvant CT for early stage BC. It is widely accepted that ER+ tumours have a better prognosis and predicts response for adjuvant hormonal treatment. However, our study suggests that ER+/HG3 disease should not necessarily be considered as being associated with a good prognosis, and that the treatment offered should be considered in the light of both the ER status and HG as ER+/HG3 tumours arguably identify the majority of "Luminal B" tumours (Prat and Perou, 2010a).

There are some limitations in the present study. Firstly, the retrospective design increased the proportion of missing data in the cohort, particularly for receptor status. The lack of HER-2 data is partially explained by the fact that at the time of data collection IHC for HER-2 was not routinely performed. This is the most likely explanation for the lack of association of HER-2 status with clinical endpoints in the MVA for the whole cohort. However, as seen figure 2.5, HER-2+ patients appear to have the worst survival among the four studied subgroups. It is important to emphasise that patients in this cohort did not receive adjuvant trastuzumab as treatment of the study population preceded the adoption of trastuzumab as standard adjuvant therapy for early stage BC (Piccart-Gebhart et al., 2005, Romond et al., 2005). Fortunately, trastuzumab therapy has currently changed the overall prognosis of this group of patients for the better (Jackisch, 2006).

Secondly, administered treatment in this cohort may differ from current practice. As an example, the majority of N- patients (81.5%) did not receive adjuvant chemotherapy. However, as many as 45% of N+ patients did not receive adjuvant CT either. Age certainly contributed to this as just over half of those were aged ≥70 years and even today there are no data to support routine use of chemotherapy in this age group. In contrast, adjuvant tamoxifen was prescribed to 1141 patients (85.21%) in our cohort; the majority of tamoxifen treated patients were ER+

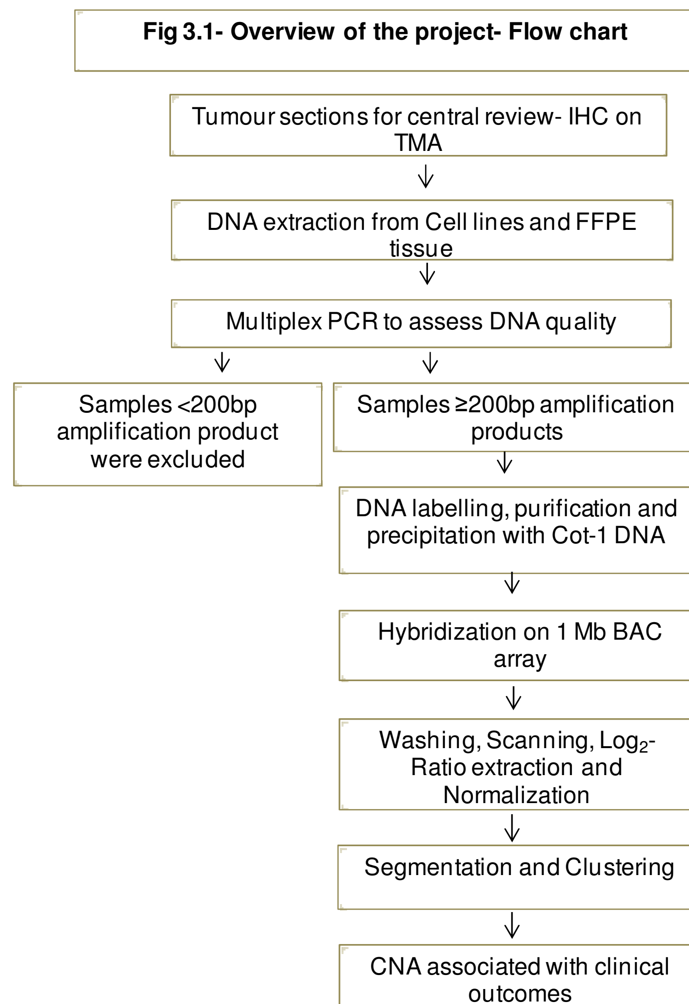
(81.29%). However, this was consistent with standard UK practice at that time where ER-unknown disease was assumed to be hormone-responsive.

In summary, our findings support the importance of HG in BC prognosis. The relative weight of HG in any prognostic equation will only increase as the numbers of Stage 1 BC increases as results of mammographic screening. It seems clearer that HG1 and HG3 tumours are different biological entities and that the majority of ER+/HG3 cases are a group with poorer prognosis compared with ER+/HG1 disease despite the increased use of adjuvant therapy in HG3 disease even at early stages. This is also supported by the finding that ER positivity loses its prognostic association in HG3 tumours suggesting that HG3 tumours may already be driven by other biological pathways. The challenge is to identify what biological properties confer the different prognoses to each of the sub-groups within HG3 tumours. As for HG2 tumours, this is a heterogeneous group with extremely variable prognosis that arguably is rendering the current pathological grading system as clinically inadequate. Additional markers such as Ki-67 (Cheang et al., 2009) or genomic grade (Ivshina et al., 2006, Loi et al., 2007a) should be explored further as a tool in helping to select the appropriate therapeutic options for these patients.

CHAPTER 3 – MATERIALS AND METHODS.

3.1 – Project overview.

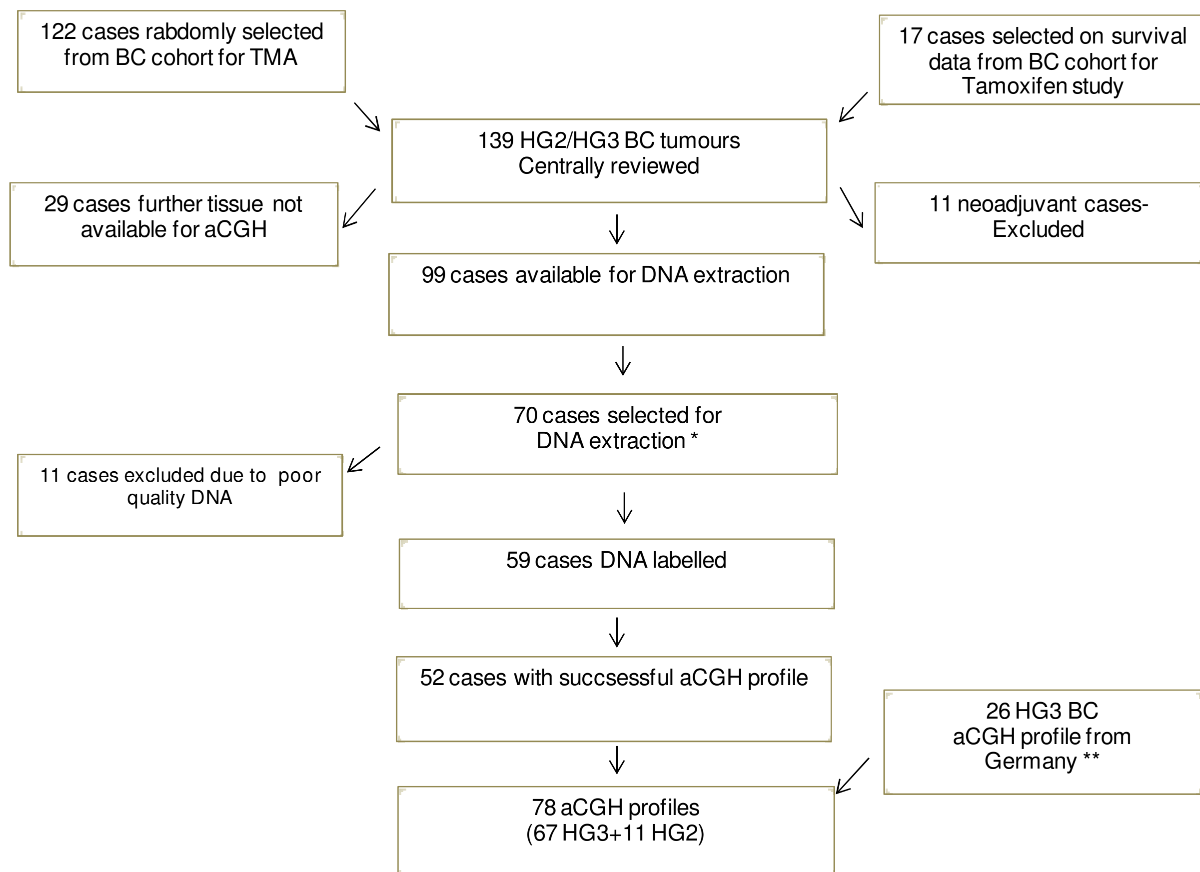
Patients were selected from a retrospective BC database containing clinical and histo-pathological data from 1339 non-metastatic patients presenting from 1996-2002 at the Oncology Department at Singleton Hospital, Swansea, South Wales, UK (chapter 2). Ethics approval for the study was obtained from the Local Research Ethics committee for Bridgend, Neath, Port Talbot and Swansea (Approval number 05/WMW02/129). An extension of this application was granted from February 2009 until 2012. An overview of the project is presented in Figure 3.1.



3.1- Patients selection for Tissue Microarray (TMA).

A total of 122 HG3 cases were randomly selected from the database and TMAs were constructed in the Histopathology Department, Singleton Hospital. Additional 17 cases were selected for the Tamoxifen study based on survival data (chapter 5). Patients who received neo-adjuvant treatment were excluded from the study. DNA was extracted from 70 consecutive cases out of 99 samples available due to limited resources (Figure 3.2).

Figure 3.2.- Patients selection for aCGH



* Limiting factor for the number of samples processed was availability of resources. Samples were analysed on consecutive order.
** Patients selected from same BC cohort as described in chapter 2. Samples arrayed abroad as part of previous collaborative study.

Only those blocks with sufficient tissue in order to take 4 x 4 um slices (2 of which were used for DNA extraction in the present study) and leave remaining tissue in the block were used. This limitation was set on ethical grounds as further studies might

be required in the patient's original tumours. TMA were constructed using (TMA-1, Manual tissue arrayer 1, Beecher Instruments Ltd; manufactured by K7 BioSystems Inc, Illinois, USA) where tissue cylinders of 0.6 mm^2 from BC FFPE blocks were transferred to the TMA paraffin-block. This work was performed in the Pathology Department of Singleton Hospital under the supervision of Professor Thomas.

Haematoxylin & Eosin (H&E) sections from the TMA were obtained in order to assess the cores and additional H&E full sections from the FFPE blocks were also obtained when possible for central review of HG and evaluation of percentage of invasive tumour present in relation to the whole section.

TMA map for these cases is available in appendix 2. Serial sections of the TMA were stained with antibodies to ER, PR, HER-2, Cyt 5/6 and EGFR at the diagnostic Laboratory of Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK.

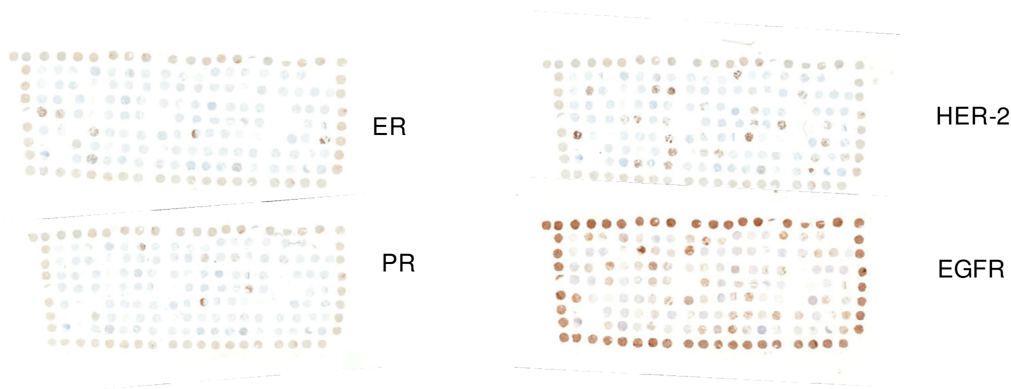
For ER, PR and HER-2 sections were stained using a BenchMark XT automated machine (Ventana Medical Systems, Tucson, AZ, USA). Bound antibodies were detected using an ultraview™ Universal DAB detection kit (Ventana Medical systems). Clone names for the antibodies used were SP1, DI2 and 4BS for the ER, PR and HER-2 respectively. They were provided in a ready to use solution. EGFR and Cyt 5/6 immunostaining were performed using a Bond-X automated immunostaining machine (Vision Biosystems, Newcastle, UK). Zymed EGFR antibody clone 31G7 (dilution of 1: 50) and Dako clone D5/16B4 (dilution 1:200) for Cyt 5/6 detection were used.

This IHC panel was chosen in order to classify HG3 tumours into the groups of different prognosis including sub-classification of TN tumours into basal-like and non-basal like subtypes (Nielsen et al., 2004). 17 additional samples were later provided for the Tamoxifen resistance project, including 4 additional cases of HG3 and 13 cases of HG2. In total 139 cases were centrally reviewed by Dr Mihir Gudi, a breast cancer pathologist who assessed all the cases following National guidelines for Pathology reporting of Breast tumours (NHSBSP, 2005).

H&E sections were obtained for each tumour sample and HG assessed. Published recommendations have tried to standardise the assessment of the HG (Page et al., 1995). Three components are evaluated namely tubule/glandular formation (1= >75%, 2= 10-75% and 3= <10%), nuclear atypia/pleomorphism (1 score for cells with small nuclei, uniform nuclear chromatin and regular outlines; 2= larger cells than normal, visible nucleoli and moderate variability in shape/size and 3= vesicular nuclei with marked variation in size/shape and vesicular nuclei) and the mitotic count per 10 high power filed (HPF) measured at the periphery of the tumour where cells are actively replicating. Field diameter should be measured and scoring category from 1-3 will depend on the number of mitotic figures seen (NHSBSP, 2005). HG is finally obtained by the sum of the independent scores as follow: grade1 for those tumours with scores 3-5; grade 2 for scores 6 or 7 and Grade 3 for 8 or 9.

TMA score sheets were generated by the pathologist, results marked in the appropriate box, crossing the box if core was missing, and marking as “IE” those samples with insufficient epithelium for assessment. 3 TMAs were constructed and results per core averaged to give a final result per tumour sample. Figure 3.3- shows a microscopic image of the TMA-1, as an example. Orientation of the TMA starts from the left-inferior corner with numbers along and letter by the side as shown in the TMA map (appendix 2).

Figure 3.3- TMA 1 with IHC staining for ER, PR, HER-2 and EGFR.

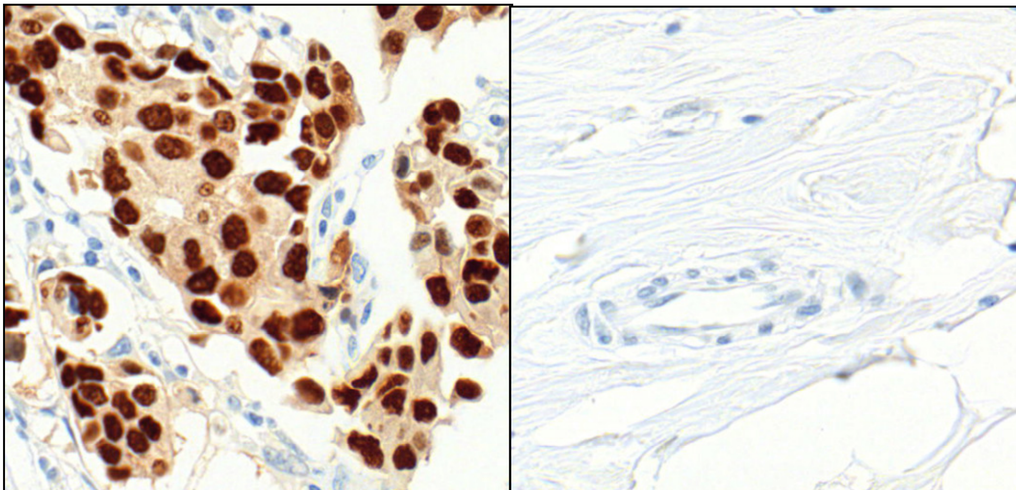


When there were discrepancies among results between the two histopathology laboratories, the most recent, centrally repeated and reviewed results were accepted Only one case had discrepancy in ER result (Case ID 415 previously ER+, changed to ER-). However, when no diagnosis could be made as IE or missing core, the

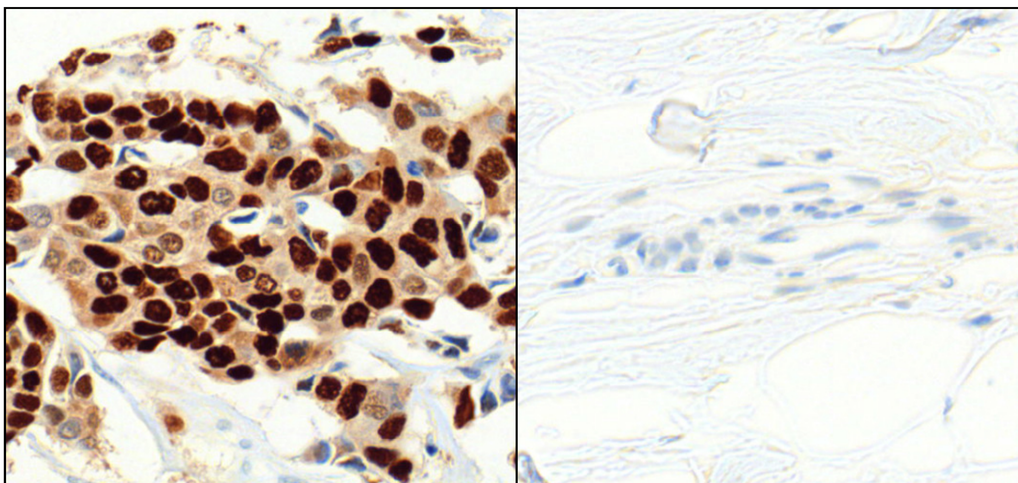
previous result was accepted when available. Results were added in the database and participated in the analysis of results in chapter 2.

ER IHC reporting only considers nuclear staining. The quick (Allred) score is recommended (NHSBSP, 2005). This is based on the score of the proportion of nuclei staining and score for intensity of staining. The score for the proportion is as follows: 0= no staining, 1= < 1% staining, 2= 1-10 % staining, 3=11-33% staining, 4= 34-66 % staining and 5= 67-100% staining. Score for intensity is measured according to the following scale 0= no staining, 1=weak, 2=moderate and 3=strong staining. The scores are summed up to a maximum of 8. PR is measured following the same guidelines as ER receptor. Figure 3.4 show representative examples of positive and negative IHC staining for ER (Panel A) and PR (Panel B) at an objective magnification of 40X.

Figure 3.4- A IHC stating for ER positive (left- ID 1195) and negative (right ID 1653)

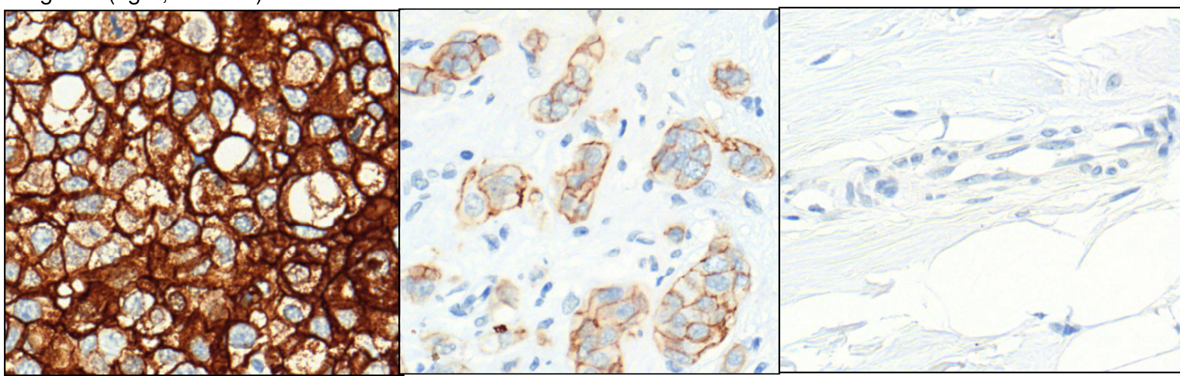


B- IHC staining for PR positive (left , ID 716) and negative (right, ID 1653)



By contrast, HER-2 receptor is a membrane structure as could be seen in a 3+ IHC results in figure 3.5, left panel. Well standardised interpretation of IHC for HER-2 is available with borderline 2+ cases by IHC (weak to moderate staining in >10 % of tumour cells) been characterised further by Fluorescent In-situ Hybridization (FISH). Those cases with a FISH ratio>2 have amplification of the HER-2 receptor and are classified as positive. These are the cases, as well as those with IHC 3+, that will benefit from trastuzumab therapy. Negative HER-2 IHC showed no staining as in Figure 3.5, right.

Figure 3.5- IHC for HER-2 showing a Positive results (Left, ID 906), Indetermined (2+, middle, ID 1543) and Negative (right, ID 1653)



In the case of IHC for EGFR and Cyt 5/6, they were considered positive in the presence of any (weak or strong) membrane or cytoplasmic staining, as it has been previously defined (Nielsen et al., 2004).

A summary of the IHC in all centrally repeated cases is shown in table 3.1. In comparison with results from the database presented in chapter 2, the positivity rate for ER decreases from 74.7% to 39.56% as expected for a group of mainly HG3 cases. All the ER negative patients were exclusively HG3. PR results behave similarly to ER. 92.30% of the HG2 patients were ER and PR positive. By contrast, HER-2 positivity was similar to that obtained when the full database analysed (18.9 and 17.26 % respectively). It was not possible to performed FISH on HER-2: 2+ cases by IHC due to the lack of recent tissue sections. Only 126 patients had reported ER, PR and HER-2 status (excluding HER-2 2+ and cases with unknown results) and were sub-classified further using the IHC definition of basal cases (TN

plus EGFR and/or Cyt 5/6 positive) and non-basal cases (5 negative staining). 57 out of 126 patients were TN (45.23%) with 31 patients with EGFR and/or Cyt5/6 expression positive and 26 patients showed no positivity for any of the 5 tested IHC staining. When looking at the immunostaining separately, all the patients that express EGFR were ER negative (28/28) and mainly HER-2 negative (21/27). This is also observed for Cyt 5/6 (26/33 for ER negative and 25/29 HER-2 negative).

Table 3.1 Results of TMA and centrally reviewed IHC cases according to HG.

Variable	Total	HG2		HG3	
	No	No	%	No	%
ER	**139**				
ER positive	55	13	23.6	42	76.36
ER negative	83	0	0	83	100
ER unknown	1	0	0	1	100
PR	**139**				
PR Positive	43	12	27.9	31	72.09
PR Negative	94	1	1.06	93	66.9
PR unknown	2	0	0	2	100
ER positive and PR	**53**				
ER / PR positive	41	12	29.27	29	70.73
ER positive/PR negative	12	1	8.33	11	91.67
ER positive/PR unknown	2	0	0	2	100
Her-2	**139**				
Positive (3+)	24	1	4.17	23	95.83
Indetermined (2+)	9	0	0	9	100
Negative (1+)	103	12	11.65	91	88.35
Unknown	3	0	0	3	100
Her-2 positive and ER	**24**				
Her-2/ ER positive	6	1	16.67	5	83.33
Her-2 positive /ER negative	18	0	0	18	100
Triple Negative	**57**				
Basal	31	0	0	31	100
Non-basal	26	0	0	26	100

3.2 DNA Isolation

3.2.1- DNA from Cell lines

DNA extracted from MCF-7 and Tamoxifen resistant (TAMR) cell line was provided by Prof Nicholson's team from Cardiff University. Wild type MCF-7 cells were grown at Cardiff University under the following conditions: in 500 ml RPMI 1640 without L-glutamine (Gibco 31870), 5% Fetal Calf Serum (FCS; Gibco 10108-157), 5 ml 1% Pen/Strep/Glutamine (Gibco 10378016) and 5 ml Fungizone (Gibco 15290018). Resistance to tamoxifen was induced by exposing MCF-7 wild type cell lines to 0.1 mM tamoxifen (Sigma H7904) for at least 6 months. TMAR cells were grown on 500 ml Phenol red free RPMI (32404), charcoal stripped FCS (Gibco 12676-011), 5 ml 1% Pen/Strep/Glutamine (Gibco 10378016), 5 ml Fungizone (Gibco 15290018) and 4 mM of Glutamine (Gibco 25030024).

3.2.2- DNA isolation from FFPE tissues

DNA was extracted from FFPE samples over 3 days using QIAmp DNA mini-Kit, (Qiagen,Germany) as per protocol adopted from VU Medical Centre in Amsterdam, Holland (See Appendix 3). Deparaffinization on Xylene was performed at 50°C for 3 minutes in a heat block or water bath to facilitate paraffin melting process as suggested by RecoverAll™ Ambion Protocol (Ambion, Cat #1975). Deparaffinised sections were centrifuged at 16.100 g to pellet the tissue and the supernatant was discarded. Sections were washed with 100% Ethanol followed by an overnight incubation in 1 M Sodium thiocyanate (NaSCN) at 38 °C in order to remove formalin-induced protein-protein cross-links. Samples were centrifuged, supernatant discarded and washed in PBS prior to the addition of ATL Buffer and Proteinase K for overnight digestion at 55°C. DNA was denatured by heating at 98°C for 10 minutes. Buffer ATL, AL and ethanol were added and lysate was transferred to the QIAmp MinElute Collumns. Columns were washed with AW1 and AW2 buffers. Samples were eluted with AE buffer (QIAmp micro-kit) and DNA stored at 4 °C. DNA quantity and 230/260 and 260/280 indexes were obtained using NanoDrop Nd-1000 UV/Vis 1µl Spectrophotometer (NanoDrop Technologies).

3.2.3 – FFPE cell pellets as positive control

Good quality DNA extracted from FFPE tissue was necessary to be used as positive controls of FFPE tissue extractions. Good quality DNA for this purpose was defined as the presence of 3 or 4 bands in 2% agarose gel after multiplex PCR as will be described in section 3.3. DNA was obtained from cell pellets made from MCF-7 cell line following a protocol attached to appendix 4. Briefly, MCF-7 cells were grown on serum calf media (SCM) until 80% confluence was reached. Cell were then dissociated with trypsine and transferred to 50 ml tube and centrifuge to form a pellet. Cell pellet was re-suspended, washed with PBS and transferred to a 2 ml eppendorf tube for further centrifugation until the pellet was again formed. Supernatant was discarded and cell pellet was re-suspended on 200 µl of Human plasma. Thrombin was added to form a soft clot that was transferred into the biopsy capsule, fixed for 12-24 hours on buffered formalin and embedded in a paraffin block. 2 sections of 4 µm thickness, the same as the FFPE cases, were obtained and placed in a 1.5 ml eppendorf tube.

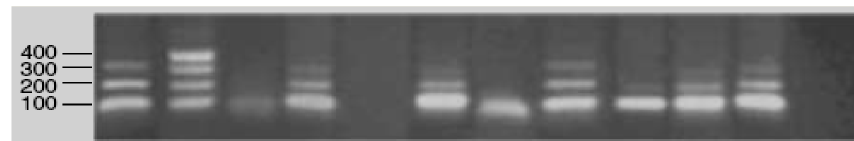
DNA extracted from the FFPE cell pellet was of excellent quality (4 bands on Multiplex PCR). They were used as a positive control for all the DNA extractions from FFPE tissues in this study.

3.3- Multiplex PCR

As discussed in Chapter 1, aCGH success greatly depends on the quality of the input DNA. It is well known that DNA from FFPE tissues, although extremely valuable, might be fragmented and of poor quality to serve as a template for hybridization in aCGH. Multiplex PCR has been previously described as a technique used to assess the DNA quality and to classify samples according to their probability of success for aCGH avoiding waste of time and resources. In principle, PCR is performed using a set of 4 primers pairs resulting in four non-overlapping amplimers (100bp, 200bp, 300bp and 400bp) of the housekeeping GAPDH gene. The poorer the DNA integrity and the higher the cross-links content in the sample DNA the less

specific the hybridization would be causing the disappearance of the DNA products in descending order. In other words, as the DNA quality decreases the 400 bp amplification products will disappear first followed by the 300, 200 and 100 bp fragment in that order (van Beers et al., 2006). The probability of aCGH success in samples ≥2-4 bands inclusive approaches 100%. However, for samples with ≤1 band the probability drops to approximately 30% (van Beers et al., 2006). Figure 3.6- shows that none of the cases with less than two bands resulted in a successful aCGH profile.

Figure 3.6- Agarose gel image after multiplex PCR showing PCR products of different lengths, the number of bands and correlation with aCGH results.



Number of Bands	3	3	1	3	0	2	1	3	1	2	3	0
Sucessful aCGH profile	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No

Modified from van Beers et al, 2006

Multiplex PCR was performed according to a protocol modified from van Beers (van Beers et al., 2006) shown in appendix 5. Briefly, the PCR primers (Metabion International GmbH) were mixed with 10x reaction buffer, MgCl2, dNTP and Taq DNA-Polymerase (FastStart High Fidelity PCR system, Roche). 100 ng of genomic DNA were added to 8.8 ul of the master-mix solution (table 3.2) to a total reaction volume of 30 µl. Pooled genomic DNA (Promega) was used as a positive control. Samples were placed in a thermocycler (GeneAmp 9700, Applied Biosystems) for initial denaturation at 96°C for 9 minutes, followed by 35 cycles of 1 minute denaturation at 94°C, annealing at 56°C for 1 minute and elongation at 72°C for 3 minutes, followed by 7 minutes final elongation at 72°C. Samples were stored at 4°C until analysed by gel electrophoresis.

Table 3.2- Pippeting scheme multiplex PCR

Multiplex PCR	Concentration per reaction 1X	Master Mix Solution(µl)
10X Buffer (pH 8.8)	10 mM	3
Primer-Mix	1.25µM each	3.2
MgCl2 (25 mM)	1.5 mM	1.8
dNTPs (10mM)	0.2 mM	0.6
Tag (5 UI/µl)	1 U	0.2
Total		8.8

Gel electrophoresis was performed in order to separate and visualise the PCR products. A 2% agarose gel (Sigma-Aldrich) was stained with SYBR® Safe DNA 10000x concentrate (Invitrogen; MP 33100) and a 100 bp DNA ladder (Invitrogen) was used as a length standard in order to estimate fragment sizes. The gel image was taken using the SynGene™ gel documentation system (SynGene-Synoptics,Cambridge, UK).

A summary of the DNA concentration of the analysed 70 cases along with 260/280 and 260/230 ratios and the number of DNA bands after multiplex PCR are shown in Appendix 6. Of note, DNA extraction was repeated in 3 cases (J10, K4 and K5) in view of very low initial DNA concentration. Only one of these samples, J10-2, showed good quality DNA in order to be included for aCGH. The drop-out rate following DNA extraction was 15.7%. Gel image for each case is shown in appendix 7.

3.4 - BAC Array

3.4.1- Random Prime labelling (RPL)

Array CGH uses differentially labelled test DNA (tumour sample) and reference DNA (genomic DNA from healthy individuals) that are co-hybridised onto a DNA array. Many approaches are available for labelling of DNA for aCGH. In RPL, the genomic DNA is denatured and random primers (8-10 mers) are subsequently annealed to the single stranded DNA. The primers binding sites are randomly distributed along

the genomic DNA and subsequently elongated (5’ to 3’) by the Klenow enzyme, a polymerase that retains its 5’-3’ polymerase activity and the 3’-5’ exo-nuclease activity (useful for proof-reading). The reaction mix containing unlabelled nucleotides and fluorescence labelled nucleotides are subsequently incorporated into the newly synthesised strands. RPL typically results in 10-40 fold amplification of the input DNA.

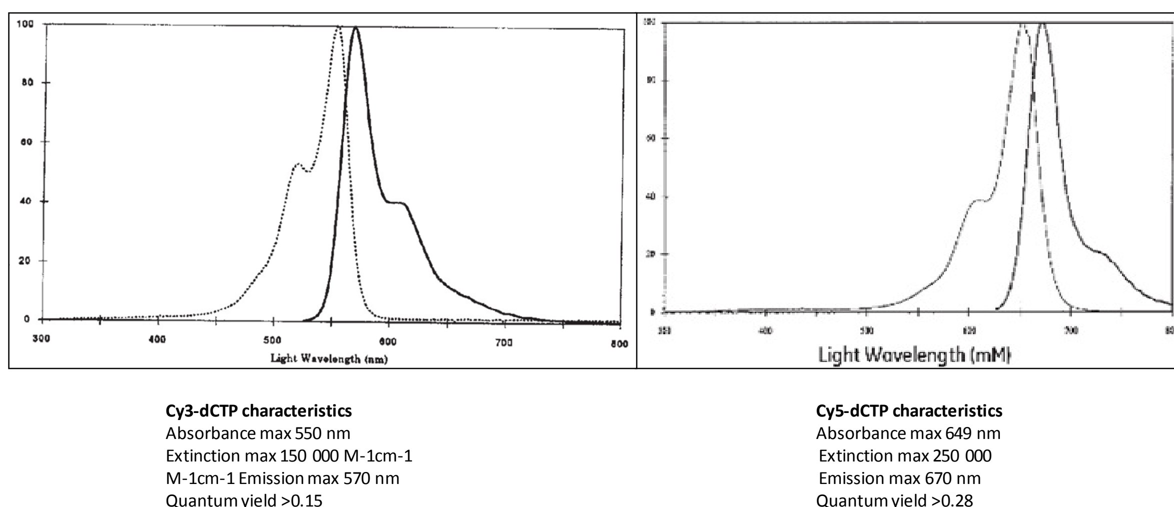
DNA labelling is, as many other steps in aCGH, a critical step as only high quality differentially labelled DNA will be able to produce good signal-to-noise ratio. In order to use the protocol giving the best possible fluorescence labelling in our laboratory, we compared three different commercially available labelling kits (BioPrime® DNA Labelling System (Invitroge), Bioprime® Total Genomic labelling System (Invitrogene) and Amersham CyScribe Array CGH genomic DNA labelling kit (GE Healthcare). Labelling efficiency was assessed by spectro-photometric measurement of the incorporation rate of labelled nucleotides using the NanoDrop Spectrophotometer (NanoDrop Technologies) in the microarray modus which allows measuring the absorbance of Cy3 at 550 nm and Cy5 at 649 nm, respectively (Figure 3.7).

3.4.1.1- BioPrime® DNA Labelling System - Invitrogen

The first labelling experiments in the present study were performed using Bioprime labelling (Invitrogen 18094-011). Briefly, 150 ng of genomic DNA were mixed with 20 µl of 2.5X Random primer solution (RPS) in the presence of 1 µl of 1 M NaCl and distilled water to a total reaction volume of 49 µl. Mixture was incubated for 5 min at 100°C in order to denature the DNA followed by immediately cooling on ice. 5 µl of 10X dNTP mixture, either 0.7 µl of 1 mM Cy3 dCTP (Amersham) or 1 mM Cy5 dCTP (Amersham) and 1µl of Klenow (40U/µl) were added and sample was incubated overnight at 37°C. Unincorporated nucleotides and primers were removed using a Illustra™MicroSpin G-50 columns (GE Healthcare, catalog 27-5330-02) according to the supplier’s protocol.

Samples were stored in the dark until determination of DNA yield (measuring absorbance at 260-320 nm) and labelling efficiency using NanoDrop Nd-1000 UV/Vis 1µl Spectrophotometer. It was decided to start with a 3X reaction using a modify version of the Manufacturer's protocol (http://tools.invitrogen.com/content/sfs/manuals/18094011.pdf). Female tumour samples were sex-mismatched using male reference DNA (Male Human genomic DNA Promega Catalog G1471; 238ng/ml) as an internal control. Manufacturer specifications web-link for Cy3 (Amersham PA53021) and Cy5 (Amersham; PA 55021) are shown appendix 8. Fluorescent dye excitation and spectral emission characteristics for Cy3 and Cy5 are shown in Figure 3.7.

Figure 3.7- Cy3 (left) and Cy5 (right) Amersham fluorescent dye excitation and emission spectral characteristics.



3.4.1.2- Bioprime ® Total Genomic labelling System- Invitrogen

In comparison with the previous kit, this labelling system (Cat. No. 18097-011) offered a higher signal-to-noise ratio and simplified the workflow by reducing the pipetting steps and efficiently label a wider range of input DNA (50 ng to 3 µg). It used Alexa Fluor® 3 and 5 dyes with already labelled dNTPs and random primers in a 2X reaction mix. It also used a mutant form of the DNA Polimerase, Exo-Klenow fragment, which retain 5' → 3' polymerase activity, but lack any exonuclease activity.

Briefly, 25 µl of Alexa Fluor® 3 2X Reaction Mix was added to the 22 µl of test DNA (4.4 µl of EDTA -5mM and distilled water) and 25 µl of Alexa Fluor® 5 2X reaction mix was added to the 22 µl of reference DNA. Mixtures were incubated at 95 ºC for 5 minutes followed by ice-cooling. 3 µl of Exo-Klenow was added and incubated overnight at 37ºC in protected from light. Labelled DNA was centrifuged and stored at 4ºC.

3.4.1.3 Amersham CyScribe Array CGH genomic DNA labelling kit (GE Healthcare 28909726)

This Kit allowed genomic DNA to be labelled to high specific activity with Cy^{TM}3 and Cy^{TM}5 labelled dCTP using a mutant fragment of DNA polymerase I (Klenow–Exo free), random nanomers (primers) and dCTP labelling mix. Labelling was performed according to manufacturer protocol attached to appendix 9.

In summary, 150 ng of genomic DNA was mixed with 10 µl of random Nonamers, 10 µl of reaction buffer and nuclease-free-water (all components supply with the kit) in a final reaction volume of 38 µl. Samples were incubated at 95 ºC for 5 minutes to denature DNA and allowed annealing to the primers and placed on ice for 5 min. 8 µl dCTP-labelling mix and either 3 µl of Cy3 (test DNA reaction) or Cy5 (reference DNA reaction) and 1 µl of Klenow Exo-free were added to the reaction mix and incubated at 37ºC overnight. Labelled DNA was purified using filter columns (Microcon YM-30, Milipore). Samples were stored at -20ºC protected from light. Probe quantification was performed by measuring the absorbance using the Nanodrop device.

3.4.1.4- Comparison of different genomic DNA labelling systems for aCGH

In order to establish reproducibility of the aCGH profiles using different kits, a cell line with a known CNV was arrayed. For this purpose we used RPEX, a retinoblastoma cell line with high level of amplification involving a specific region on

Chromosome 10. DNA extracted from this cell line was provided by the Institute of Molecular Radiobiology Molecular Cytogenetics, GSF-National Research Centre, Neuherberg, Germany.

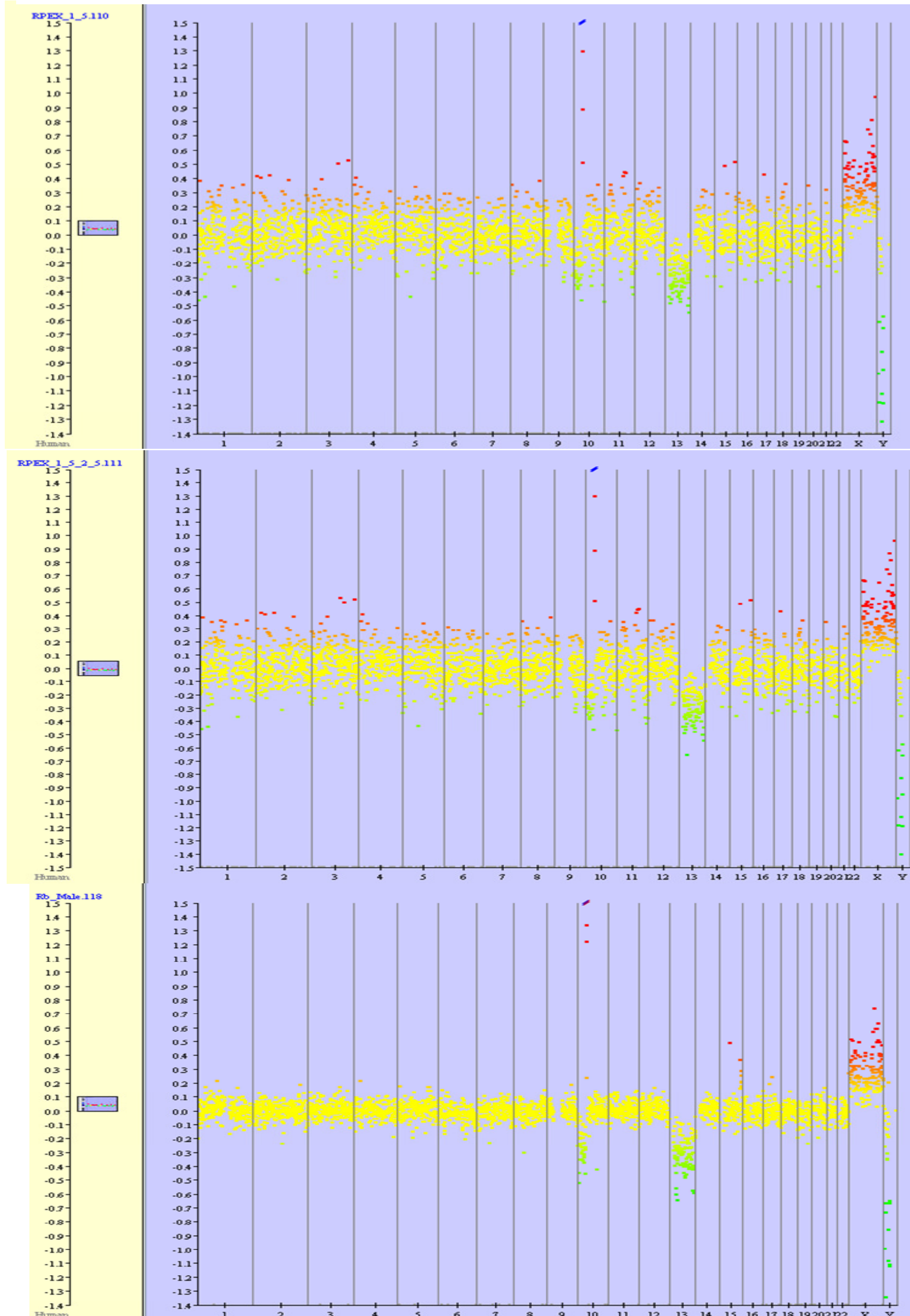
Table 3.3 shows DNA yield and incorporation of labelled nucleotide using the three different kits for RPEX cell line. The rationale behind testing different labelling kits was to assess its success and reproducibility in our laboratory rather than formally compare their labelling efficiency. Therefore, no conclusions can be drawn from the results apart that the three labelling system worked satisfactory both in reference, cell line and tumour DNA and gave reproducible results as shown in Figure 3.8. For financial reasons, it was decided to proceed with the GE kit. Initial experiments were performed with 5X reactions. However, as purification of labelled samples and washing protocol were also optimised it was possible to reduce the number of reaction to 3.5X in a final protocol already described in section 3.4.1.3. This protocol gave us more than 95% of surviving clones using a foreground/background ratio of 3 for all cell line cases and the majority of FFPE samples.

Table 3.3 Incorporation of labelled nucleotide in RPEX cell line using different labelling kits.

Sample ID	Dye 1 name	Dye1 pmol/ul	Dye 2 name	Dye2 pmol/ul	ng/ul	260/280
BioPrime Invitrogen (3X)						
TE	Cy3	0	Cy5	0	0	NaN
RPEX_Cy3	Cy3	50.38	Cy5	-0.17	349.29	1.73
Male Cy5	Cy3	10.25	Cy5	52.72	471.34	1.76
Bioprime Total Invitrogen (3X)						
TE	Cy3	0	Cy5	0	0	NaN
RPEX_Alexa3	Cy3	34.65	Cy5	-0.06	665.07	1.6
Male_Alexa5	Cy3	4.03	Cy5	33.37	718.55	1.69
Amersham CyScribe GE (3X)						
TE	Cy3	0	Cy5	0	0	NaN
RPEX_Cy3	Cy3	31.93	Cy5	-0.06	364.87	1.67
Male_Cy5	Cy3	3.91	Cy5	30.09	321.68	1.8

Similarly, aCGH profile showed high level of amplification of a region on Chromosome 10 and deletion of a region on Chromosome 13. Results were consistent and reproducible using the three labelling kits. Sex mismatch was detected in these experiments (as RPEX was derived from a female patients, male reference DNA was used).

Figure 3-8. Comparative aCGH profile in RPEX cell line using different labelling kits. Y axis shows Log-2 ratio along chromosome number (x axis). Gained regions (red dots) or lost region (green dots) are represented along the reference line 0. There is evidence of high level of amplification of Chromosome 10 (blue tick on the top). Sex mismatch is shown as amplification of Chromosome X and deletion of chromosome Y. Upper image using BioPrime® DNA Labeling System ,Invitrogen, 5X reaction; middle image Bioprime® Total Genomic System ,Invitrogen, 5X reaction and Amersham CyScribe Array CGH,GE, 5 X reaction at the bottom.



3.4.2- Summary of labelling results in BC tumour samples.

BC tumour samples were labelled using GE Healthcare (3.5X reaction). On average, the incorporation rate of labelled nucleotide for tumour samples (Cy3) was 41.48 pmol/µl, DNA yield 439.9 ng/µl and 260/280 index 1.75. A summary of the labelling reaction for all the tumour samples is presented in appendix 10.

3.4.3 Initial aCGH experiments in FFPE tissues.

Initial aCGH arrays were performed following a protocol described by Fiegler et al (Fiegler et al., 2007). From this protocol, the automatic array hybridization buffer was used but combined with manual washing as an automatic washing station was not available. Manual hybridization buffer was initially used but found to be more viscous, as contained double the concentration of Dextran Sulphate, causing uneven spatial array hybridization. Therefore, the automatic hybridization buffer was used as published by Fiegler et al (Fiegler et al., 2007). Samples were stored at –20℃.

Samples were centrifuged at maximum speed for 30 minutes at 4℃. Pellet was washed with 500 µl of 80% EtOH, and dried by centrifuging, re-suspended in 490 µl (1 array/slide) of automatic hybridization buffer and denatured by heating at 70℃ for 10 minutes. Sample was incubated at 37℃ for 60 minutes in the presence of Cot-1 DNA. Samples were then placed into the gasket slide (G2539-60006; 2 microarrays) loaded into the hybridization chamber and incubated at 37℃ for 40 h.

The slides were subsequently washed in three steps: Step 1: 500 ml of PBS/0.05% Tween 20 for 10 minutes at room temperature with shaking at maximum speed. Step 2: 50% formamide/2X SCC solution for 30 minutes at 42 ℃ with shaking and Step 3: 500 ml PBS/0.05% Tween 20 for 10 minutes at room temperature with shaking. Dry by spinning (150g for 1 min) and stored in a light-proof box until scanned.

Several reactions were performed using the protocol described above. Unfortunately, arrays had a very low signal-to-noise ratio due to a high background signal with a very low proportion of surviving clones. Optimisation of the protocol was attempted

for several months until successful array obtained. Several modifications were applied to the protocol including initial increment in the oven temperature from 42℃ to 65 ℃ in order to increase the stringency of the hybridization (reduce non-specific hybridization). This was later reduced to 60℃. Other changes included increments in the temperature of the washing step 2 between 40 to 45℃ (to increase stringency and reduce background signal), adding a final washing step with distilled water with a soaking time ranging from 30 seconds to 2 minutes, controlling pH of the formamide washing solution at 7 by adding HCL, repeating all washing steps twice, introducing a first step washing with SCC only in order to prevent phosphate precipitation, repeating the final centrifugation step twice to ensure complete dryness of the array and using new (not re-used) gasket slides. Finally, it was decided to change protocol to use the hybridization buffer and washing procedures to the Agilent oligonucleotide protocol and move to another laboratory with higher standard conditions for molecular biology work.

3.4.4- Agilent Oligonucleotide Array-based CGH for Genomic DNA protocol.

3.4.4.1- Preparation of Labelled Genomic DNA prior hybridization.

Experiments were started using 2-active array per slide using the Agilent Oligonucleotide protocol Array-based CGH for Genomic DNA analysis as per 105K array with a final hybridisation volume of 260 µl.
(version 4 June 2006; http://www.opengenomics.com/pdf/protocol_cgh_genomic_DNA.pdf) .
In our adapted protocol (table 3.4) labelled DNA was co-precipitated with 50 µg of Human Cot-1 DNA (Roche – cat.11581074001), mixed with 26 µl of 10X Blocking agent (Agilent; cat. 5188-5281) and 130 µl of 2X hybridization buffer (Agilent;cat.5190-0403). Co-precipitation with Cot-1 DNA was important to block the repetitive sequences and prevent non-specific hybridization so we decided to increase the amount of Cot-1 DNA reducing the labelled DNA volume so the same final reaction volume was achieved. Detailed protocol is shown in appendix 11.

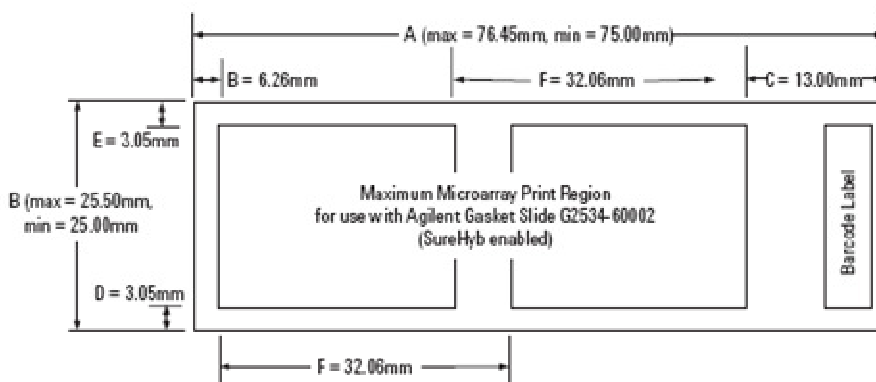
Table 3.4. Preparation of Labelled gDNA prior hybridisation

Component	Agilent Oligonucleotide Protocol	Our adapted protocol*
	Volume (µl) per hybridisation	
Cy3 and Cy5 –labeled gDNA mixture	79	54
Cot1 DNA (1.0 mg/µl)	25	50
Agilent 10X Blocking agent	26	26
Agilent 2X Hybridisation buffer	130	130
Final hybridisation volume	260	260

*Adapted from Agilent Oligonucleotide Array-based CGH for Genomic DNA analysis. Version 4.0, June 2006.

Samples were mixed by pipetting and incubated at 95 °C for 3 minutes followed by incubation at 37°C for 30 minutes. Hybridization gasket slides (G2539-60006; 2 microarray) were loaded onto the hybridization chamber (SureHyb enabled G2534A; see figure 3.9) and sample applied to the gasket slide dwell. Active array slide was placed on the top of the gasket slide being careful to prevent bubble formation. Hybridization chamber was assembled and vertically rotated to ensure no immobile air bubbles were present. Samples were hybridised in an Agilent Hybridization oven (G2545A) at 60°C for 67 hours at 22 rpm. Temperature was reduced from 65°C in the original protocol to 60°C in order to reduce the hybridization stringency. For aCGH 1 Mb BAC arrays (Fiegler et al., 2003) printed by the Centre for Microarray Resources in Cambridge were used (two arrays per slide, approx. 3400 clones printed in triplicate).

Figure 3.9- Hybridisation chamber dimensions (SureHyb enabled).



3.4.4.2- Washing

The washing steps were performed according to the Wash Procedure A – Agilent Oligonucleotide protocol Array-based CGH. Briefly, disassembly of the array slide was performed at room temperature while completely submerged in the Oligo aCGH Wash Buffer 1 (Agilent 5188 5221). Slides were placed in a slide rack and washed for 5 minutes using the Oligo aCGH Wash Buffer 1 on a magnetic stir plate at room temperature. Final washing step was performed in a pre-heated glass dish containing Oligo aCGH wash Buffer 2 (Agilent 5188-5222) at 37℃ for 1 minute on a magnetic stir plate. Arrays were let air dry and kept in the dark until scanned.

3.4.4.3- Scanning

Full slide area (61x21.6 mm) was scanned using an Agilent Microarray Scanner (Agilent; p/n G2565BA) using 5 µm scan resolution in a single pass. Photomultiplier tubes (PMT) was set to 100%. Image was cropped and rotated (upper-left to lower-right). Image splitting for green and red channels were performed using Tiff-page-Splitter Deluxe programme prior downloading images into an image processor programme.

3.6- Bioinformatic Analysis of CGH

3.6.1 Extraction of Fluorescence intensities.

Spot fluorescence intensities were extracted and written into text files using the software MAIA which is distributed under free license by the developer (http://bioinfo-out.curie.fr/projects/maia/ - MAIA (Microarray Image Analysis; Version 2.75 version; Institute Curie, Paris, France).

3.6.2- Primary data analysis

3.6.2.1- Primary data analysis in cell lines

This process includes normalisation, segmentation and copy number calling. It was performed using CAPweb-CGH array Analysis Platform on the Web (Hupe et al., 2004). CAPWeb converts the fluorescent intensities in the array to a green-to-red log_2 ratio profiles. CAPWeb uses the R package MANOR for data “normalization”, an important step in which biases introduced by experimental artefacts are discarded with preservation of the true biological signal in the array (Neuvial et al., 2006). Normalization also makes profiles comparable. In addition, bias correction for local spatial effect is performed by estimating the spatial trend on the array, segmenting the array into those regions with similar trend and identification and exclusion of those areas affected by local spatial bias; eg: those spots with very low signal-to-noise ratio, unique spots or clones with poor replication consistency (Liva et al., 2006, Neuvial et al., 2006).

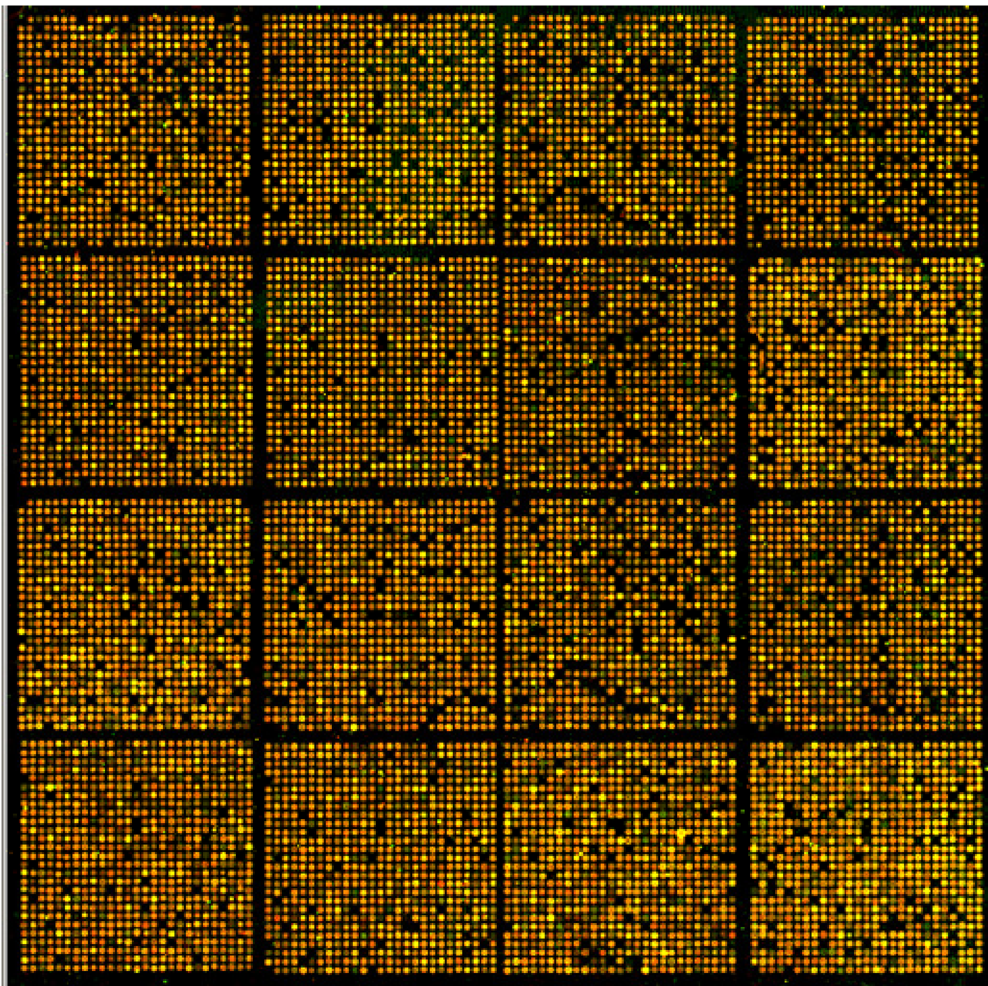
Data is then segmented into chromosomal regions with identical copy number using the GLAD algorithm. This process called “segmentation” will identify the point (breakpoint) at which a variation in the copy number is identified. The breakpoint identification process was performed in this study using default GLAD parameters recommended by the providers. Finally, GLAD assigned a status to the regions between the breakpoints (normal, gain, loss or amplified) and provides a summary report of the results.

VAMP software (Visualization and Analysis of array-CGH, transcriptome and other Molecular Profiles; http://bioinfo.curie.fr/vamp) was used for visualization of the aCGH data in a simplified and meaningful way (La Rosa et al., 2006) (http://bioinfo.curie.fr/vamp). Using this graphic interface connected to CAPWeb, the results can be retrieved by the whole genome, by chromosome, allow clone identification and comparison between arrays. Analysis for the cell lines was repeated using a CGHcall for calling of the log_2 ratios (van de Wiel et al., 2007).

As will be presented in chapter 5, MCF-7 and TAMR cell lines aCGH profiles were compared using both the CAPWeb and CGHcall within the R-platform. Same results were obtained using CAPWeb and CGHcall regarding regions involved in the CNV as it will be explained in section 5.4.1.1

A representative example of an aCGH image using the MAIA programme is shown in Figure 3.10.

Figure 3.10- Array image taken using MAIA program



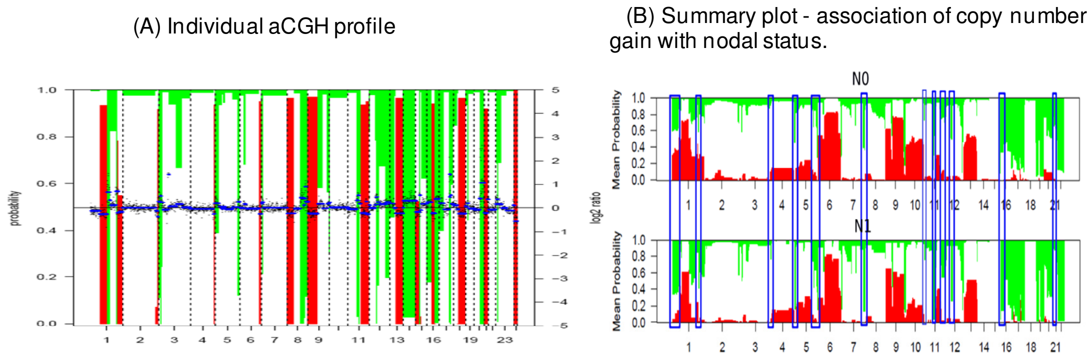
3.6.2.2- Primary data analysis in BC tissue

The fluorescence intensities were imported into the R statistical platform (R Development Core Team, 2009) transformed into log2 ratios and normalised by spatial normalisation using an algorithm implemented in the R package MANOR (Neuvial et al., 2006). Only data points were included that showed a foreground to background ratio of greater than 2. Datasets were excluded from the study if more than 20% of clones did not pass this quality assurance step or the standard error of replicates was greater than 10%. The normalised log2-ratios were segmented using the circular binary algorithm from the DNA copy R package (Olshen et al., 2004) whereas copy number calling was performed using the CGHcall package (van de Wiel et al., 2007). Moreover, the CGHcall method permitted correction for contamination with non-tumour cells or cellularity correction (van de Wiel et al., 2007).

CGHcall makes the assumption that breakpoints divide segments into different states of CNA. In this study, 4 copy number states are defined (-1=loss, 0= normal, +1= gain and +2= amplification). The assumptions is made that clones in the same segment belong to the same state and are classified together rather than having different states per clone. Segments are in the normal state (null hypothesis) and CNV differing from the normal states is measured as a p-value or False discovery rate (FDR).

Representative examples from our study of plots for a individual aCGH profile (A) and a summary plot (B) according to nodal status are shown in table 3.11. In individual profiles plots, the normalized $\log_2$-ratio is shown on the right Y axis, probability scales on the left, segments are plotted in the X inferior where the chromosome number is shown and vertical bars indicate gain or loss probability. This probability is a “call” when is >0.5 (van de Wiel et al., 2007). In the summary plot, the left Y axis shows the mean probability for a CNV of node-negative cases (top) and positive cases (bottom). The blue rectangles point out copy number regions which are significantly associated with copy number gains (see section 4.3).

Figure 3.11- Exemplary plots for aCGH results in our study



3.6.3 Secondary Analysis

3.6.3.1- Clustering of aCGH data

Hierarchical clustering analysis is a method to identify natural similarities between a number of observations and grouping them together according to common characteristics. Several methods have been used for clustering analysis. In this study unsupervised hierarchical clustering of array CGH profiles was performed using the WECCA (Weighted clustering of called aCGH data) method (Van Wieringen et al., 2008). Kaplan Meier plots were used to analyse survival. For every BAC clone KM curves (BCSS) were generated for the groups no gain/gain and no loss/loss and were tested for differences by log-rank Chi-square statistics (Harrington 1982). P-values were corrected for multiple testing error using false discovery rate (FDR) (Benjamini 1995) whereas FDR values < 0.2 were accepted as significant. Results were analysed only when at least 10 patients were in each of the groups. Multivariate testing was performed using Cox-proportional hazard regression analysis. P-values were considered significant when <0.05. Supervised univariate testing was performed by testing the counts of gain or loss for each BAC clone using Fisher's exact test.

In addition, analysis was performed to identify CNA in tamoxifen-treated patients with poorly differentiated tumours who died/relapse of BC within 5 years of diagnosis and compared to those patients alive with no distant relapse after 5 years. This was

performed using Permutation t-testing method. Common minimal region of alteration were also defined.

Candidate genes affected by CNA were then identified using the ***e!Ensembl*** GRCh37, in the section *Homosapiens* which is publicly available (http://www.ensembl.org/index.html).

3.6.4- In silico validation of candidates genes

In order to validate the candidate genes of the gained region on chromosome 5 which correlated with nodal status, at the mRNA expression level, a publicly available data set was used (Loi et al., 2007). The dataset GSE6532 (Loi et al., 2007b) was downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) and imported into R. The Affymetrix expression dataset (Human U133A/B U133 Plus 2.0) was normalised using the GeneChip robust multi-array average (GCRMA) method (Bioconductor). Only profiles were selected from patients who received adjuvant tamoxifen treatment and had tumours with an histopathology grade 2 or 3. For each gene from the region on chromosome 5 the corresponding Affymetrix probe was extracted querying the Ensembl database using the biomaRt package in R (Durinck, 2003, Durinck et al., 2005). Gene expressions were tested for differential expression in the node-positive and the node-negative group using Mann-Whitney test. Due to the low number of tests (n=49) no correction of p-values for MTE was performed.

CHAPTER 4 – MOLECULAR DIVERSITY IN HIGH GRADE BREAST CANCER. ASSOCIATION OF 5q GAIN WITH SURVIVAL

HG is one the main prognostic factors used in clinical practice, along with nodal status and tumour size, for the classification of patients into different prognostic groups that support the clinical decision making process and to tailor adjuvant treatment according to specific risk of relapse or death. HG is consistently included in the most clinically relevant BC prognostic tools such as Nottingham prognostic Index (NPI), St Gallen and Adjuvant Online (Galea et al., 1992, Goldhirsch et al., 2001, Ravdin et al., 2001, Ravdin et al., 1998). With very few exceptions, HG is commonly mentioned as one of the main prognostic factors associated with survival in BC (Soerjomataram et al., 2008). An increase in HG is also accompanied by an increase in the frequency of DNA CNA , with HG3 tumours showing a higher number of alterations relative to HG1 (Roylance et al., 1999). In addition, CNA profiles generated from aCGH studies have shown association with specific BC subtypes (Natrajan, 2009) and have been used to classify new groups within ER- BC (Chin et al., 2007).

Clinically, BC is a heterogeneous disease and there is considerable variability in the response to treatment even when patients are stratified on grade. A number of initiatives have been instigated to improve treatment tailoring with respect to selecting breast cancer patients for adjuvant chemotherapy on prognostic factors such as molecular profile (Paik et al., 2006). In this study we set out to identify heterogeneity at the DNA copy number level specifically in high grade BC using BAC aCGH, and to identify specific CNAs that associate with poorer prognosis in this high risk group of patients.

4.1- Patients cohort

Patients were selected from a retrospective BC database containing clinical and histopathological data from 1339 non-metastatic patients as described in chapter 2. A total of 78 cases (67 HG3 and 11 HG2) of high grade primary invasive ductal breast carcinoma of no special type (IDC NST) were selected from the database

(chapter 3). The median age of the patients was 67.78 (range 30.03 to 82.98). 53 cases were node negative. ER, PR and HER-2 status were assessed for the 67 HG3 tumours by immunocytochemistry on tissue microarrays (TMAs) produced by a manual tissue arrayer (Beecher Instruments), and on full face sections for the 11 HG2 cases, according to standard procedures. (chapter 3). Additional 26 HG3 IDC-NST node negative BC cases obtained from the same clinical cohort but previously arrayed in Institute of Molecular Radiobiology Molecular Cytogenetics, GSF-National Research Centre, Neuherberg, Germany were included for combined analysis of the aCGH data.

4.2 – Results

4.2.1 – Patients characteristics

In comparison with results from the database presented in chapter 2, the positivity rate for ER decreases from 74.7% to 44.9% (35/78) as expected for a group of mainly HG3 cases. Patients characteristics are shown in table 4.1. 53 cases were node negative. 36/78 cases were negative for ER, PgR and HER-2, 6 were ER-ve/HER-2 3+ positive, 6 were positive for both ER and HER-2. All ER negative patients were exclusively HG3. PR results behaved similarly to ER with the majority of PR negative cases being HG3. By contrast, HER-2 positivity was similar to that obtained when the full database was analysed (18.9 and 17.9 % respectively). It was not possible to performed FISH on those HER-2 2+ cases by IHC due to the lack of recent tissue sections.

Table 4.1- Demographics of BC patients with aCGH profile

Variable	HG2	HG3
	No (%)	No (%)
Total Number of patients	11 (14.1%)	67 (85.9%)
Age - Median (range)	71.17 (53.12-81.04)	67.51 (30.02-82.97)
≤50	0	16 (23.9%)
50-70	5 (45.5%)	20 (29.9%)
≥70	6 (54.5%)	31 (46.2%)
TNM T staging*		
T1 (≤ 2 cm)	7 (63.6%)	32 (47.7)%)
T2 (>2 - ≤5 cm)	4 (36.4%)	33 (49.3%)
T3 (>5 cm)	0	1 (1.5%)
T4	0	0
Unknown	0	1 (1.5%)
TNM N staging**		
No	7 (63.6%)	46 (68.7%)
N1 (1-3 nodes)	2 (18.2%)	11 (16.4%)
N2 (4-9 nodes)	0	5 (7.5%)
N3 (≥10 nodes)	2 (18.2%)	4 (6.0%)
Unknown - No	0	1 (1.5%)
ER		
Negative - No (%)	0 (0%)	43 (64.2%)
Positive - No (%)	11 (100%)	24 (35.8%)
PR		
Negative - No (%)	1 (9.1%)	50 (74.6%)
Positive - No (%)	10 (90.9%)	17 (25.4%)
HER-2		
Negative - No (%)	10 (90.1%)	49 (73.1%)
Indetermined - No (%)	0	5 (7.5%)
Positive - No (%)	1 (9.09%)	13 (19.4%)

* TNM Tumour staging by Singletary et al al 2003.

** TNM Nodal Staging by Singletary et al 2003.

4.2.2 – Association of Copy Number Changes with survival

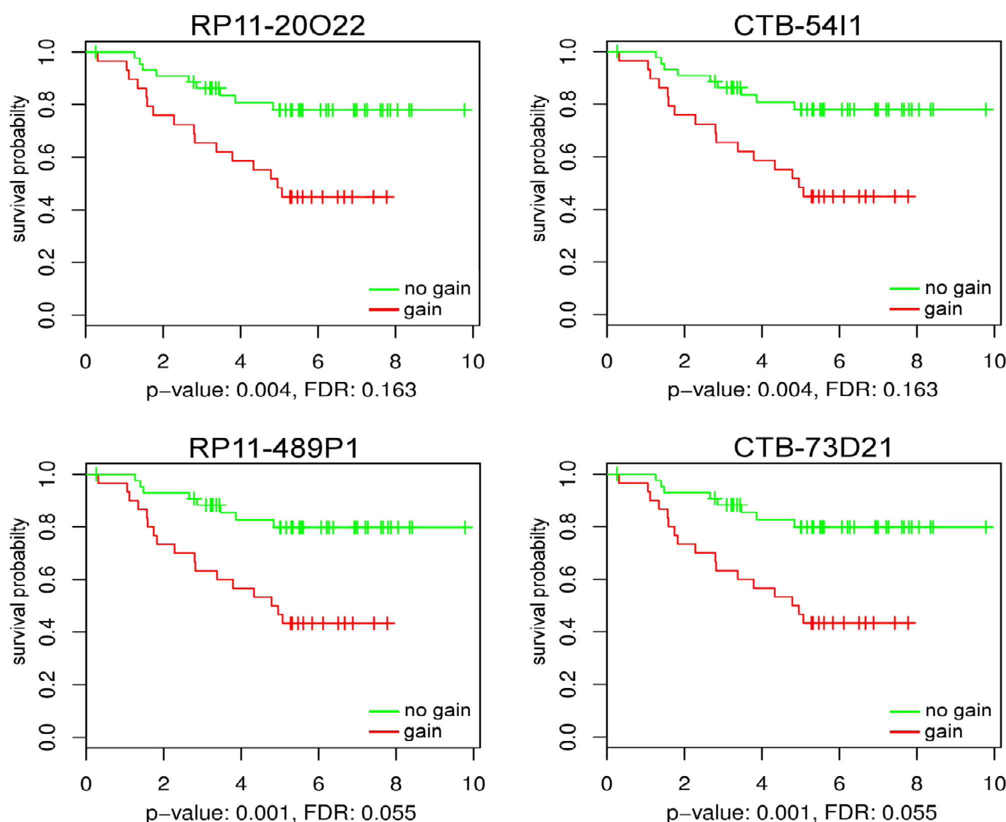
Testing of the association of copy number changes with cancer-specific survival (CSS) revealed association of a gained region on chromosome 5q35.1-2 which is defined by four BAC clones (RP11-20O22, CTB-54I21, RP11-489P1 and CTB-73D21) with CSS (p-value < 0.005, FDR < 0.2) as shown in table 4.2. Patients with tumours harbouring the gain on chromosome 5q35.1-2 showed poorer CSS (median survival time: 5 years) compared to patients with tumours which did not harbour the gain (Fig 4.1).

Table 4.2- Chromosomal region alteration associated with worse OS and CSS. P-values (uncorrected and corrected for Multiple Testing error are presented).

Clone No	Cr	Gene	CAN	OS		CSS	
	No	HGNC	Call	p (uncorrected)	FDR	p (uncorrected)	FDR
1121	5	RP11-20O22	Gain	0.033	0.908	0.004	0.16
1122	5	CTB-54I1	Gain	0.033	0.908	0.004	0.16
1123	5	RP11-489P1	Gain	0.0147	0.436	0.001	0.055
1124	5	CTB-73D21	Gain	0.0147	0.436	0.001	0.055

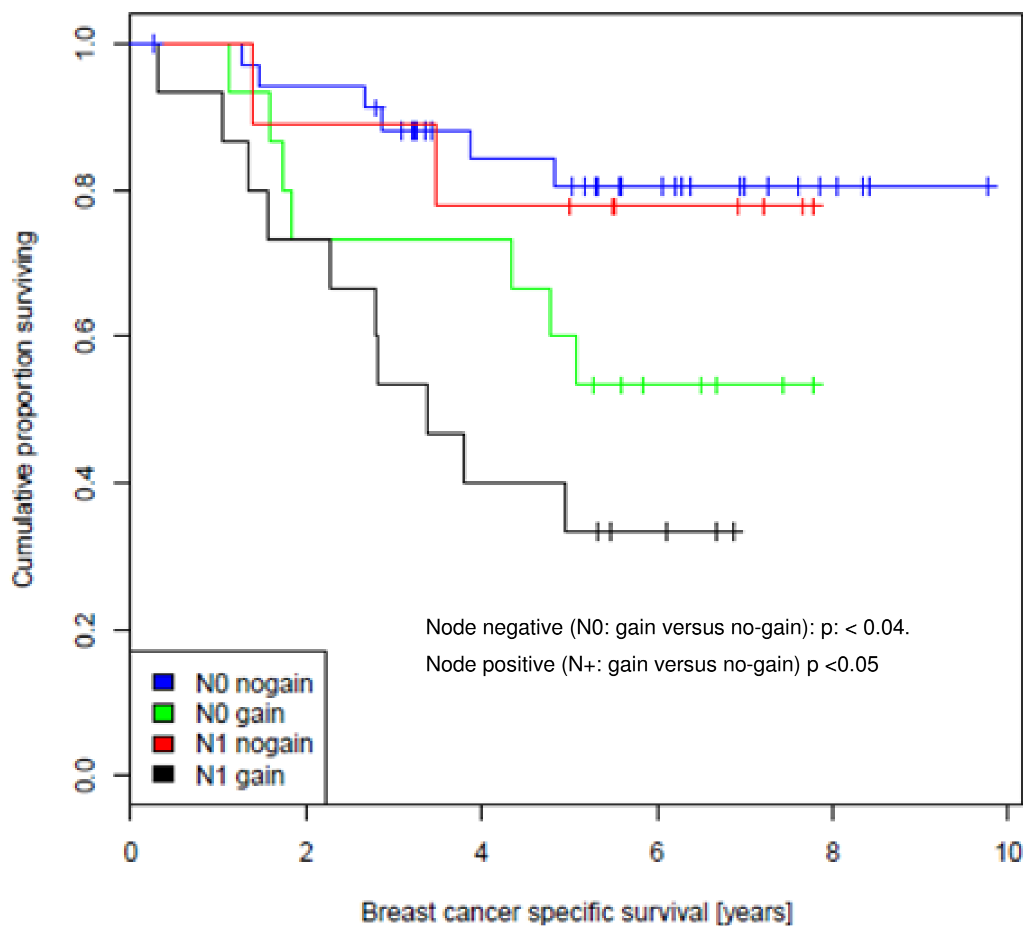
In UVA for CSS the p-value for gain 5q35.1-2 status was 0.001 and that of nodal status was 0.049 whereas in the MVA (nodal status, gain status 5q35.1-2) the hazard ratios were 1.3 (0.73-3.6 95% CI; p:0.23) for nodal status and 3.2 (1.4-7.7 95%CI; p:0.0076) for gain 5q35.1-2 status.

Figure 4. 1: Kaplan-Meier survival plots of patients with tumours harbouring copy number gain of the regions defined by BAC clones RP11-20O22, CTB-54I1, RP11-489P1 and CTB-73D21 (5q35.1-2, red) and those not harbouring the gain (green). Patients with gain of 5q35.1-2 show significantly reduced cancer specific survival (p-value < 0.004, FDR < 0.2) compared to patients not showing the gain.



Moreover in UVA analysis (log-rank testing of KM curves) gain of a portion of this region defined by the BAC clones RP11-489P1 and CTB-73D21 also separated node-negative (n=53) and node-positive (n=24) patients into two groups with a significantly different prognosis (p-values < 0.04 and < 0.05 in the node-negative and node-positive group, respectively, Fig. 4.2).

Figure 4.2: Kaplan Meier plots of survival in node-negative and node-positive patients with 5q gain-negative and 5q gain-positive tumours



Sub-analysis studies performed among different molecular tumour sub-groups showed significant association of 5q gain with worse OS (p:0.012) and CSS (p:0.005) in ER+ patients when compared with no-gain in the UVA in the same

population. MVA confirmed worse outcome for cases harbouring the gain in terms of OS (HR:2.2; C: 1.1-4.6, p:0.029) and CSS (HR:3.2; CI: 1.4-7.2, p:0.005). Significant association was also observed among HER-2 negative patients both for OS and CSS as shown in Figure 4.3. In the TN sub-group, no association was observed between 5q gain and OS or CSS (p:0.484 and p:0.052 respectively). Further analysis into basal and non-basal phenotype was not possible in view of low number of observations.

Figure 4.3- Kapplan Meier survival curves for ER+ subgroup for OS (A) and CSS (B) and for Her-2 negative patients for OS (C) and CSS (D).

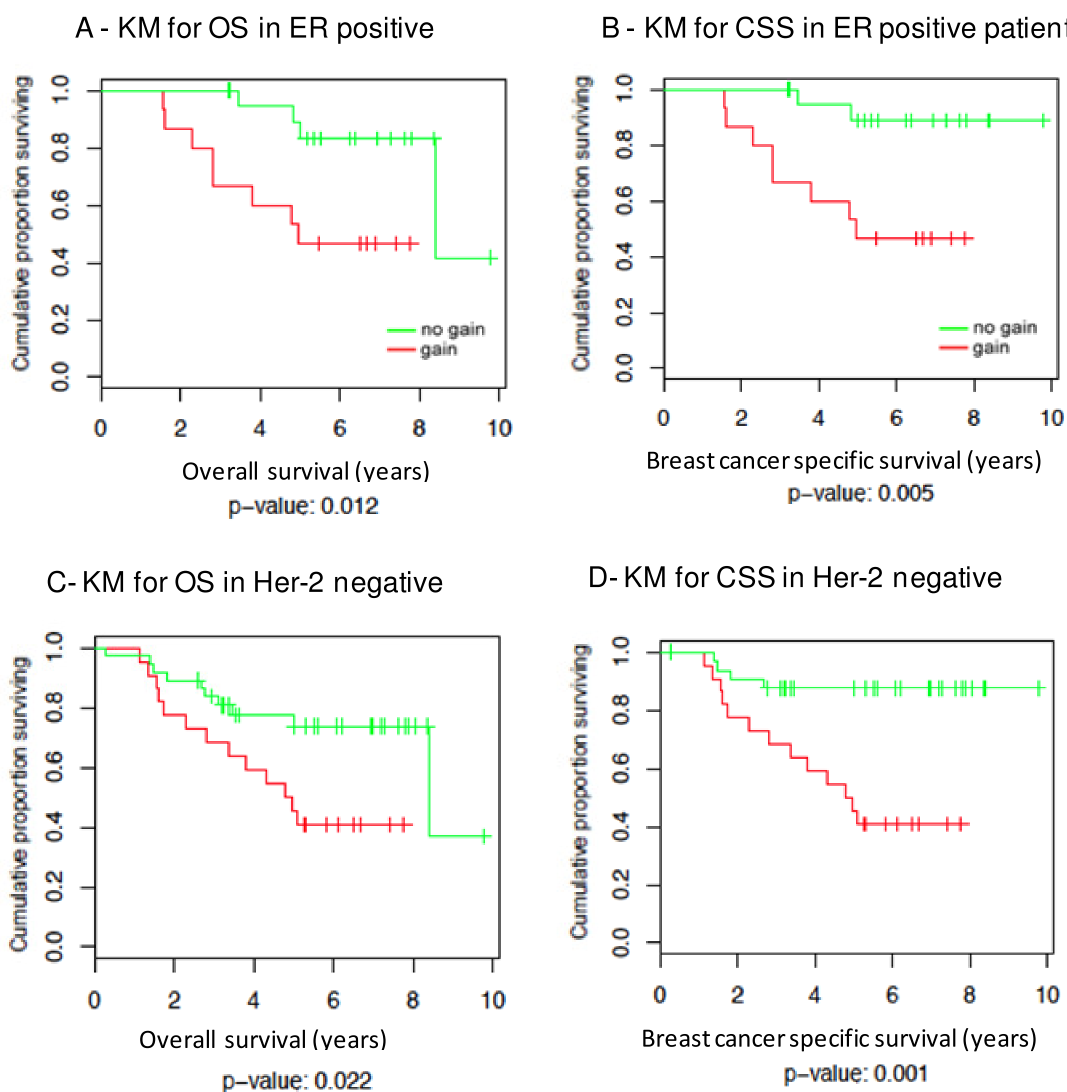


Table 4.3- Genes encoded by gained region associated with CSS (chromosome 5_171064773_174710119)

HGNC Symbol	Start position	End position	Gene description
MSX2	174151536	174157902	Homeobox protein MSX-2 (Hox-8)
C5orf47	173416162	173433142	Uncharacterized protein C5orf47
CPEB4	173315331	173388979	Cytoplasmic polyadenylation element-binding protein 4
BOD1	173034517	173043663	Protein FAM44B
STC2	172741716	172756506	Stanniocalcin-2 Precursor (STC-2)
NKX2-5	172659112	172662262	Homeobox protein Nkx-2.5.Cardiac-specific homeobox
BNIP1	172571445	172591390	BCL2/adenovirus E1B 19 kDa protein-interacting protein 1
C5orf41	172483370	172563961	UPF0474 protein C5orf41
ATP6V0E1	172410760	172461900	V-type proton ATPase subunit e 1 (V-ATPase subunit e 1)(V0 subunit e1
RPL26L1	172386427	172396774	60S ribosomal protein L26-like 1
ERGIC1	172261278	172379688	Endoplasmic reticulum-Golgi intermediate compartment protein 1
DUSP1	172195093	172198198	Dual specificity protein phosphatase 1. (MAP kinase phosphatase 1)(MKP-1)
NEURL1B	172068276	172118531	E3 ubiquitin-protein ligase NEURL3 (EC 6.3.2.-)(Neuralized-like protein 3)
SH3PXD2B	171760505	171881527	SH3 and PX domain-containing protein 2B (Adapter protein HOFI)
UBTD2	171636644	171711075	Ubiquitin domain-containing protein 2 (Dendritic cell-derived ubiquitin-like protein)
EFCAB9	171621177	171630349	EF-hand calcium-binding domain-containing protein 9
STK10	171469074	171615346	Serine/threonine-protein kinase 10 (EC 2.7.11.1)(Lymphocyte-oriented kinase)
FBXW11	171288556	171433877	F-box/WD repeat-containing protein 11 (F-box/WD repeat-containing protein 1B)
C5orf50	171212876	171221602	Uncharacterized protein C5orf50
SNORA74B	172447731	172447931	Small nucleolar RNA SNORA74

4.2.3 Association of copy number changes with Nodal Status

In order to identify copy number changes associated with nodal status of patients supervised association analysis was performed. A summary of the copy number changes associated with nodal status are given in table 4.4. Interestingly, the gain of the region on chromosome 5q35.1-2 that had already been found to be associated with patient survival overlaps the region on chromosome 5 that is associated with node-positive tumours. Consequently gain of 5q35.1-2 and nodal status are not independent and do not appear as independent markers in MVA.

Candidate genes

The consensus of the region on chromosome 5 (chr5:171064773-174541830) that was associated with positive nodal status and with unfavourable breast cancer specific survival was subsequently used for the definition of candidate genes (Table 4.3) by querying the Biomart Ensembl database (GrCH37).

Table 4.4 - Copy number alterations significantly associated with nodal status.

Chromosome number	**Copy number** alteration	**Region (bp)** Start	End	**p-value** Corrected	FDR*
Node positive					
1	Gain	928301	11843141	0.021	0.064
1	Gain	12428632	28962129	0.021	0.064
4	Gain	1625433	8102624	0.0064	0.039
5	Gain	2517762	10872654	0.0002	0.005
5	Gain	148662044	149998339	0.022	0.064
5	Gain	151339486	166617450	0.008	0.039
5	Gain	167659919	174541830	0.0038	0.036
7	Gain	148594274	158803231	0.0298	0.082
11	Gain	67661700	70629079	0.0116	0.048
11	Gain	70918961	76396807	0.0062	0.039
11	Gain	116088651	134478818	0.008	0.039
12	Gain	48284343	57831732	0.0286	0.082
16	Gain	3531439	31047088	0.0198	0.0645
Node negative					
1	Loss	928301	11843141	0.0042	0.0196
1	Loss	12428632	28962129	0.0186	0.057
1	Loss	109735370	119699888	0.0254	0.057
1	Loss	144007818	145057046	0.007	0.024
1	Loss	145438616	155948960	0.0232	0.0571
1	Loss	157789502	182629019	0.0232	0.0571
1	Loss	202152138	206584659	0.0052	0.0215
1	Loss	207021349	223286340	0.0008	0.0176
1	Loss	223823291	249063112	0.0024	0.0188
6	Loss	145997	29213068	0.0134	0.044
6	Loss	31923988	43857965	0.021	0.057
10	Loss	269607	5637110	0.0022	0.0189
10	Loss	69771091	88064908	0.0022	0.057
10	Loss	89607026	131107320	0.0334	0.066
10	Loss	131926164	133779590	0.0032	0.0189
10	Gain	135221600	135221940	0.019	0.064
21	Gain	14540608	31752113	0.0036	0.0363

* FDR: false discovery rate

4.2.4 – In-silico validation of candidate genes.

In order to ascertain whether the gain in copy number of candidate genes obtained from the chromosomal region associated with a positive nodal status in this study is

also associated with an increase in mRNA gene expression, a publicly available BC gene expression set (Loi et al., 2007b) was analysed. In total, 49 probes of the microarray matching the 20 candidate genes from the region on 5q35.1-35.2, the region specifically gained in node-positive carcinomas in our study were tested for differential expression in lymph node-positive and node-negative IDC. A total of 36 probes matching 14 of the 20 different genes showed significantly increased expression in the node-positive cases relative to node-negative cases. Table 4.5 summarises test results of significantly differentially expressed genes in node-negative and node-positive patients from the study by Loi et al. The five genes with the lowest p-values were ATP6VOE1, NEURL1B, DUSP1, MSX2 and C5ORF41. It is known that whilst there is, on the whole, good association with datasets that analyse BAC array data and mRNA expression data derived from the same cases, absolute association would not be expected (Jarvinen et al., 2008). This may be as a result of gene silencing by hypermethylation or other changes in gene regulation, for example at the miRNA level.

Table 4.5- In-silico validation of genes encoded by gained region Cr 5 according to its association with CSS and ascending p-value

	Affy Probe	p - value
ATP6V0E1	201172_x_at	0.0000
DUSP1	201044_x_at	0.0001
NEURL1B	225355_at	0.0001
MSX2	205555_s_at	0.0003
C5orf41	238476_at	0.0006
CPEB4	224829_at	0.0006
UBTD2	224834_at	0.0009
STK10	237875_at	0.0012
RPL26L1	218830_at	0.0019
BNIP1	37226_at	0.0074
SH3PXD2B	231823_s_at	0.0201
BOD1	225030_at	0.0355
FBXW11	209455_at	0.0410
STC2	203439_s_at	0.0413

4.2.5- Hierarchical Unsupervised Clustering Analysis

Unsupervised hierarchical clustering of cases based on CNA revealed no association of clusters with outcome and nodal status nor any of the classical BC markers such as HER-2, ER, and PR. Figure 4.3 shows the clusterogram.

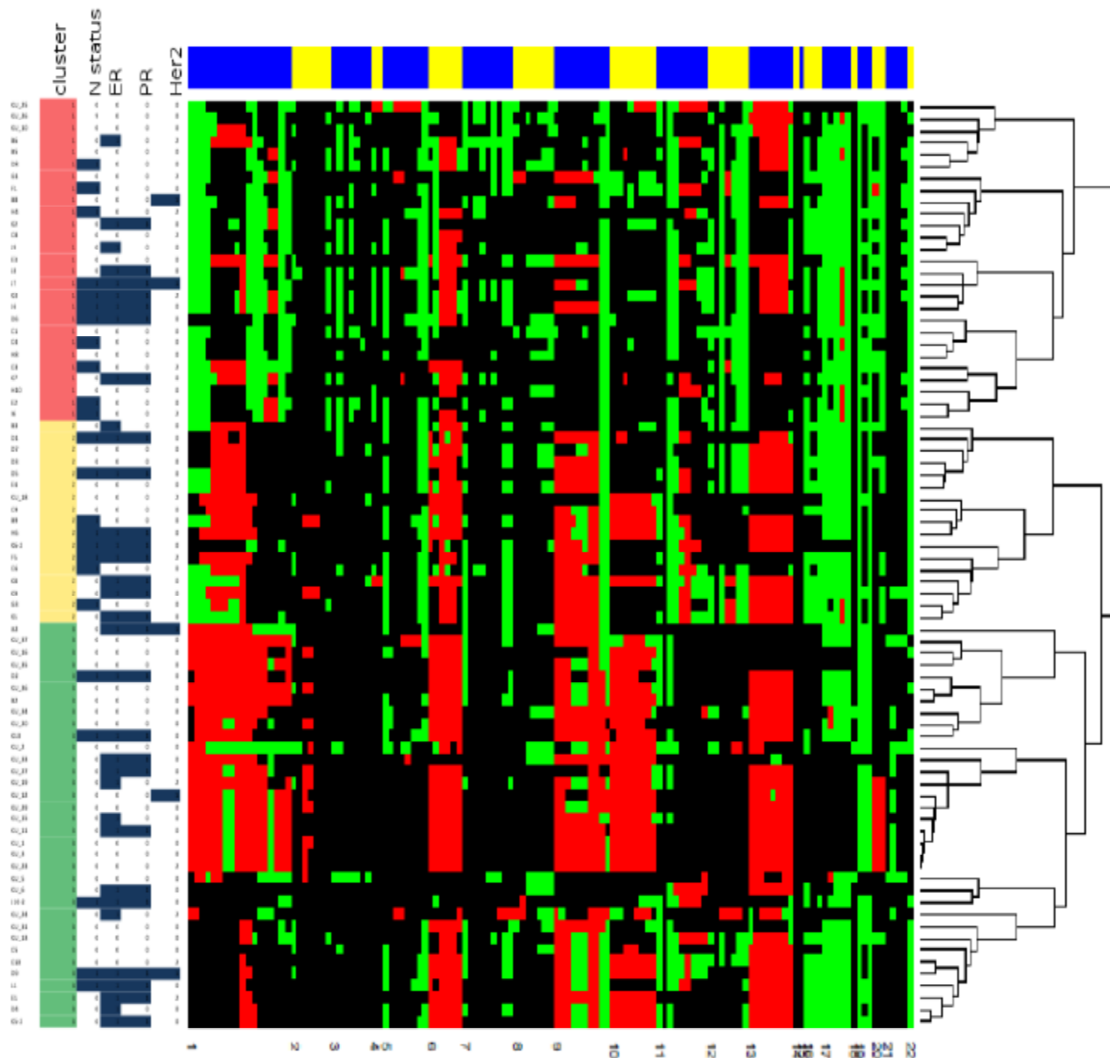


Figure 4.4: Unsupervised and supervised analysis of copy number alterations in relation to BC clinical parameters. Heatmap was generated with options total, agree, all.equal and False. Red indicates a loss,black a normal, and green a gain in DNA copy number. The regions are ordered according to their location on the genome, starting from the top with the first region on chromosome 1. The numbers at the bottom side represent chromosome numbers. The blue/yellow bar on the top indicates the p and q arms of the chromosomes. The labels on the left-hand side are the names of the samples.

4.3- Discussion

There have been a number of papers that have reported an association between CNA and survival in BC (Hicks et al., 2006). However, these papers have included a spectrum of histological grades. The aim of our study was to investigate whether we could identify a CNA that could be used as an adjunct to conventional prognostic markers in high grade BC as an aid to treatment tailoring in this group of patients. This study of 78 high grade (HG2 and HG3) breast cancers shows significant statistical association between a region of copy number gain on Chromosome 5 (171064773-174710119) and BCSS and of a portion of this region (Chromosome 5:173869298-174710119) with BCSS in the node-negative and the node-positive groups of the same cohort. Patients with tumours harbouring the gain on chromosome 5q35.1-2 showed poorer CSS compared to patients with tumours which did not harbour the gain (Fig 4.1). A comprehensive search of the literature in all articles published in PubMed was performed using “5q”, “5q gain” and “Breast cancer”. Thirteen articles were identified some of which contained relevant information to our work. Loveday et al demonstrated that the occurrence of CVA varied across the genome with some chromosomes such as 1p/q, 5q, 8p/q frequently showing alterations (Loveday et al., 2000). They also found that CNA in 5q occurred at comparable frequency both in Invasive ductal and invasive lobular carcinoma (Loveday et al., 2000).

In a recent publication Johnson et al identified four regions of gain (5q, 16p, 19q and 20) that were commonly affected in Invasive ductal carcinoma but not in DCIS. Specifically, 5q gain included the region between 170.12- 170.21 and 170.69 172.75, similar to the region gained in our study (Johnson et al., 2011). Although DCIS cases shared the same molecular subtypes as invasive ductal carcinomas these differentially expressed genes might identify genetic markers of progression. A region in 5q was also found to be gained in invasive ductal carcinoma when compared with DCIS in another publication (Yao et al., 2006).

In our study, distribution of the 5q gain according to tumour subgroups showed that 5q gained was present in 15 patients among those ER+ (15/35, 42,9%), 6 cases among those HER-2+ (6/12, 50%) and 12 cases among triple negative patients (12/36, 33,3%). Survival analysis showed significant association between OS and CSS in ER + and Her-2 negative cases for those with 5q gain versus no-gain (Figure 4.3).

Several authors have described a relative deletion of 5q in ER- cases (Fang et al., 2011, Wang et al., 2004, Loo et al., 2004, Pierga et al., 2007) in comparison to ER+. Loo et al found a >40% more frequent losses in 5q31-q35 among other regions when compared with ER+ tumours. Other authors also found that 5q deletion in ER- tumours was among the significant CNAs found relative to ER+ tumours (Fang et al., 2010), also supporting our findings.

A previous publication found three regions (-3p.13,+5q35.2 and +8p12) were associated with OS in the combined HER-2 negative reference data set (log rank P:0.004, 0.002 and 0.003, respectively) adding either independent or near independent prognostic information in the MVA (P:0.08, 0.007 and 0.06, respectively). The association with outcome for the 5q35.2 region was stronger for HER-2-/ER- tumours compared to HER2-/ ER+ tumours (log-rank P:0.006 and 0.07 respectively (Staaf et al., 2010) Another author also found losses of 5q33.3–q4 among others was associated with HER-2 amplification (Pierga et al., 2007).

Although previously publications have found 5q gain associated with invasive breast cancer and worse prognosis as discussed above the body of literature is very limited to a small number of publications. In our study 5q gain was the only gene significantly associated with OS and CSS particularly in ER+ and HER-2 negative subgroup and node positive disease. Therefore, It was decided to concentrate in the analysis of those candidate genes that are encoded by this altered region, including genes previously reported to show an association with carcinogenesis.

Association of candidate genes with *in silico* validation in the Loi dataset (Loi et al., 2007b) confirmed significant gene expression of some genes. The most interesting

of these are DUSP1 which is encoded by both gained regions, a dual specificity protein phosphatase and MSX-2, a homeobox gene which is in the centre of the smaller region that correlates with BCSS in the node-negative and node-positive groups.

4.3.1- DUSP-1 gene

Mitogen-activated protein kinases (MAPKs) constitute a family of related kinases that include c-Jun NH2-terminal kinase (JNK 1,2 and 3), p38 protein kinase (α, β, δ and γ) and the extracellular signal-regulated kinases (ERK1 and 2). These kinases play an important role transferring information from the extracellular signals to intracellular pathways that control diverse cellular processes such as proliferation, differentiation, migration, apoptosis and stress response (Keyse, 2008). The balance between activated (phosphorilated) and inactivated (de-phosporilated) state of MAPKs depends on the activity of the dual specificity MAP kinases (MEK) and dual specificity MAPK phosphatases (MKPs). The MKPs are the main endogenous mechanism to negatively regulate MAPKs.

An emerging family of MKPs have now being described with at least 10 different phosphatases grouped into 3 categories (Table 4.6) according to their sequences similarities, gene structure, cellular location and substrate specificity (Theodosiou and Ashworth, 2002). These phosphatases share a common anatomical structure with a C-terminal (catalytic domain) and an N-terminal which contains the interaction motif (Keyse, 2008).

Table 4.6- Classification of DUSP family according to Keyse et al*

Group	DUSP member	Substrate	Cellular location
1	DUSP1/MKP1	ERK, p38 and JNK	Nuclear
	DUSP2/PAC1	ERK, p38	Nuclear
	DUSP4/MKP-2	ERK, p38 and JNK	Nuclear
	DUSP5/hvH-3	ERK	Nuclear
2	DUSP 6/MKP-3	ERK	Cytoplasm
	DUSP7/MKP-X	ERK	Cytoplasm
	DUSP9/MKP-4	ERK>p38	Cytoplasm
3	DUSP8//hVH-5	p38 and JNK	Cytoplasmic and nuclear
	DUSP10/MKP-5	p38 and JNK	Cytoplasmic and nuclear
	DUSP16/MKP-7	p38 and JNK	Cytoplasmic and nuclear

**Modified from Keyse et al. Cancer Metastasis Rev, 27, 253-61. 2008*

DUSP-1 is the prototype of the MKPs with higher activity for p38 than for JNKs and ERK1 and 2 (Boutros et al., 2008) although the substrate specificity is variable and depends on cell type and biological conditions (Wu et al., 2005). DUSP-1 gene is considered an immediate early response gene under the influence of many stimuli such as cellular stress and growth factors (Keyse, 1995).

Although cancer mutation usually involves Ras/Raf pathway within the MAPK pathway, stress-activated proteins such as JNKs and p38 are playing a more protagonist role in cancer development (Dhillon et al., 2007). Interestingly, JNKs biological function widely vary between different cell type with an anti-proliferative and pro-apoptotic activity in some tissues and proliferation and anti-apoptotic function in others (Dhillon et al., 2007). DUSP1 has been found to be over-expressed in a variety of solid tumours including lung, ovarian, bladder, osteosarcoma and breast cancer (Loda et al., 1996, Rojo et al., 2009). Many human cancers exhibit increased expression of DUSP1, but the level of expression varies with the type of cancer, and in some tissues, the stage of tumour progression. For example, in the prostate carcinoma, over-expression of DUSP1 is found only in the early stages of disease and is lower in hormone refractory prostate carcinoma (Rauhala et al., 2005). Higher levels of DUSP1 are observed in gastric adenocarcinomas (Bang et al., 1998) and, interestingly, are associated with a shorter progression-free survival in ovarian cancer (Denkert et al., 2002).

DUSP1/MKP1 has been implicated in the development of anti-cancer drug resistance (Montagut et al., 2010). High levels of DUSP1 expression are also associated with chemoresistance in a number of cancer cell lines, including non small cell lung cancer [NSCLC] (Chattopadhyay et al., 2006) and BC (Small et al., 2007). A further study (Moncho-Amor et al.) suggests that DUSP1 over-expression may also be involved in angiogenesis and metastasis. Other members of the DUSP family, including DUSP4 have also been shown to be up-regulated in BC (Wang et al., 2003). DUSP1 is also reported as one of the genes in the bone metastatic signature both in BC cell lines (Kang et al., 2003) and human BC tumours (Lu et al., 2010). DUSP1 is commonly repressed in BC patients after exposure to doxorubicin, a finding that co-exists with induction of the apoptotic protein JNKs enhancing its anti-tumour effect. However, a sub-group of BC patients seems to have high-expression of DUSP1 that is not repressed after anthracycline therapy and is associated with worse DFS in univariate and multivariate analysis (Rojo et al., 2009). MKP-1 has also been associated with cisplatin resistance in other solid tumours (Wang et al., 2006) and in a publication reduction in levels of DUSP1 expression was accompanied by reduction of cell growth and cisplatin-induced cell death (Chattopadhyay et al., 2006). More importantly, there is evidence that loss of MKP-1 restored or improved cisplatin sensitivity in ovarian cancer patients (Wang et al., 2007). Similarly, DUSP1 over-expression has also been associated with anti-EGFR drugs resistance in K-ras wild type colon cancer patients (Montagut et al., 2010).

Other authors have also found DUSP1 over-expression in BC (Wang et al., 2003) even in poorly differentiated and advanced tumours which contrast with data reported in other tumour such us colon, prostate and bladder where MKP-1 expression was inversely related to histological grade (Loda et al., 1996). Hypothetically, advanced or poorly differentiated BC tumours remain dependent on MAPK pathway until late stages in the natural history of the disease (Loda et al., 1996) playing an important anti-apoptotic role (Small et al., 2004). A strong association was also found between MKP1 over-expression and elevated levels of HER-2 and EGFR (Loda et al., 1996). Recently DUSP1 over-expression in non-small cell lung cancer was found associated with tumour growth, invasion and

angiogenesis. Angiogenesis is believed to be mediated by an increased expression of vascular endothelial growth factor C [VEGFC] (Moncho-Amor et al.).

Our study therefore supports the evidence in the literature that up-regulation of DUSP1, in this case via a mechanism of gain of DNA copy number, is significant in both invasion and worse prognosis in high grade BC. Whilst gain of DUSP1 is more frequently found in node-positive cases, it is also present in some node-negative cases. Further studies will be needed to ascertain whether gain of copy number or expression of this gene in node negative high grade BC could be used to stratify patients for adjuvant treatment.

4.3.2- MSX2

MSX2 is a member of the superfamily of Homeobox genes (Hox genes). Homeobox protein MSX2 is a transcription factor involves in embryogenesis of different organs, including mammary tissues (Satoh et al., 2004). Members of this superfamily play roles in both cell fate and differentiation during development but have also been found to be commonly deregulated in cancer. Deregulation of the Homebox-gene family might lead to enhanced proliferation and inhibits cell differentiation facilitating tumour formation (Satoh et al., 2004). One of the mechanisms through which MSX2 induces breast cell proliferation is Cyclin D1 up-regulation, a finding present in approximately 40% of BC. In BC, over-expression HoxB13 has been associated with tamoxifen resistance in ER positive breast cancer, especially when associated with down-regulation of expression of choline dehydrogenase (CHDH) and Il17β receptor, Il17RB (Ma et al., 2004).

Over-expression of HoxB7 has also been implicated in epithelial/mesencymal transition (EMT) and associated with bone metastasis in vivo (Wu et al., 2006). Interestingly, MSX2 is also associated with EMT and also significantly induces Twist-1 expression. Recent reports show that in BC this is strongly correlated with high tumour grade, chromosome instability, vascularisation in addition to EMT (Mironchik et al., 2005). Evidence is also emerging to support the role of Twist in EMT and its

relationship with Basal-like BC phenotype, which is known to indicate a worse clinical prognosis (DiMeo et al., 2009) and with CD24(-/lo)CD44(+) cells (Tomaskovic-Crook et al., 2009). MSX2 is also a relevant target for RAS/Raf pathway which is involved in many different epithelial solid tumours. In particular, around 90% of pancreatic cancers contain K-ras mutation. Satoh et al showed that cell proliferation, migration and vascular invasion are increased in MSX2 expressing pancreatic cancer cell lines and that MSX2 expression was significantly correlated with high histological tumour grade (Satoh et al., 2008). Recently, it has been suggested that MSX2 expression could be used as a diagnostic tool to differentiate chronic pancreatitis from pancreatic carcinoma, another tumour type that displays marked EMT (Satoh et al., 2010).

However, a recent publication showed that MSX2 protein expression, mainly its cytoplasmic component, was associated with prolonged OS in MVA in BC patient. Further assessment of the role of MSX2 in BC cell lines showed that its effect is likely to be related to an increment of apoptotic tumour cell death (Lanigan et al., 2010). Another study showed no induction of apoptosis in MSX2 over-expression (di Bari et al., 2009) in mouse mammary epithelial cells.

It is clear that MSX2 expression varies among BC tumour sub-types as they also found that MSX2 mRNA expression was high in the Luminal B and HER-2 but low in the basal subtype (Lanigan et al., 2010). Furthermore, MSX2 function depends on its intra-cellular location and it is also possible that the downstream intracellular pathways depends on the genetic background of the cell line used to assess its functionality (Lanigan et al., 2010).

In conclusion, this study, using BAC array CGH on a series of high grade breast cancers, has demonstrated that gain of 5q35.1-35.2 is associated with decreased BCSS and node positivity. We have validated that there is increased expression of genes encoded in this region using a publicly available dataset (Loi et al., 2007b). Two of the genes coded for in this region (DUSP1 and MSX2) have documented roles in angiogenesis, invasion and EMT. Further validation studies are needed to ascertain whether either evaluation of gain of this region by FISH, or gene

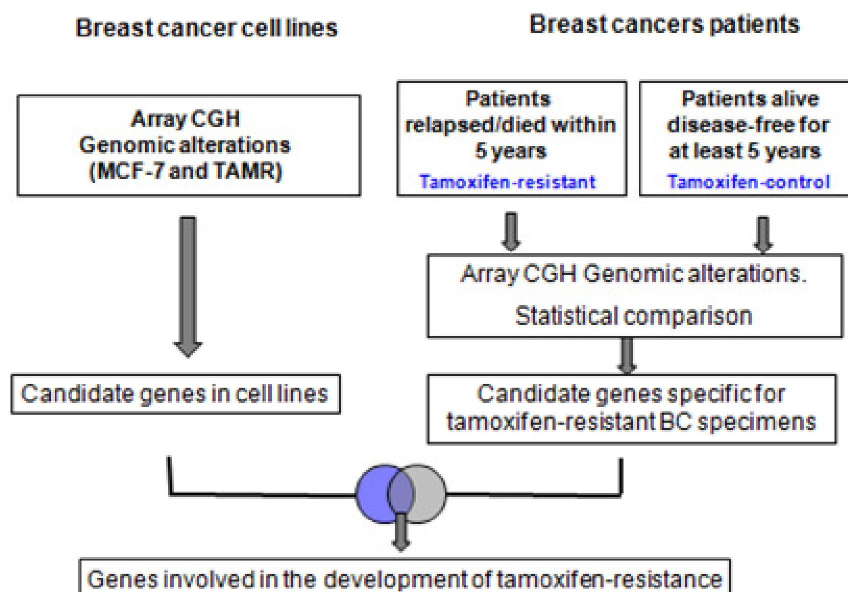
expression studies using qRT-PCR would prove to be a useful tool in treatment tailoring for high grade BC.

CHAPTER 5 – BAC ARRAY ANALYSIS AND RESISTANCE TO TAMOXIFEN

5.1- Aims of the study

The present study intended to identify differences in aCGH profile between breast cancer cell lines sensitive (MCF-7) and resistant to Tamoxifen (TAMR) in an attempt to define genomic alterations involved in the development of hormone resistance in vitro. However, it is well recognised that genetic or expression profile in cell lines might change over time as is highly dependent of cell passage (Wenger et al., 2004) and do not always represent changes in tumours (Olsavsky et al., 2007). Therefore, aCGH profile from BC tumours were obtained from ER+ patients who relapse and died within 5 years from diagnosis despite of tamoxifen therapy and compared with those obtained from recurrence-free patients for at least 5 years from diagnosis. Finally, a comparison is made among the genomic alterations obtained from the aCGH profiles in cell lines and tissue specimens that might help to elucidate specific pathway involved in the development of a resistant and more aggressive phenotype. Figure 5.1 shows a general overview of the study.

Figure 5.1- General design for Tamoxifen resistant study



5.2- Sample selection

DNA extracted from MCF-7 and TAMR cell lines was provided by Prof Nicholson's team from Cardiff University (section 3.3.1). Tissue specimens for this study were selected from demographic and survival data from the BC cohort described in Chapter 2. Patients included ER+ status who received tamoxifen only as adjuvant therapy and had good quality DNA/aCGH profile (patients who received neo/ adjuvant chemotherapy were excluded). Patients were classified as Tamoxifen-resistant-group (TAMRG) if had died from BC within 5 years of diagnosis as opposed as Tamoxifen-controls-group (TAMCG) who were recurrence-free for at least 5 years from diagnosis. Patients characteristics are summarised in table 5.1.

Table 5.1- Patients characteristics of Tamoxifen-resistant cases and controls.

Characteristics	**TAMRG** No (%)	**TAMCG** No (%)	**p-value**
Total Number of patient	11	13	
Age - Median (range)	76.25 (59.19-81.04)	67.57 (60.24-73.66)	
≤50	0	0	0.02
50-70	1 (12.5)	7 (87.5)	
≥70	10 (62.5)	6 (37.5)	
Histological grade			
1	0 (0)	0 (0)	0.093
2	3 (27.3)	8 (72.7)	
3	8 (61.5)	5 (38.5)	
TNM-T staging*			
T1 (≤ 2 cm)	1 (10.0)	9 (90.0)	0.003
T2 (>2 - ≤5 cm)	10 (71.4)	4 (28.6)	
TNM-N staging**			
No	3 (27.3)	8 (72.7)	0.132
N1 (1-3 nodes)	3 (42.9)	4 (57.1)	
N2 (4-9 nodes)	3 (100.0)	0 (0)	
N3 (≥10 nodes)	2 (66.7)	1 (33.3)	
PR			
Negative - No (%)	2 (66.7)	1 (33.3)	0.439
Positive - No (%)	9 (42.9)	12 (57.1)	
HER-2			
Negative - No (%)	8 (44.4)	10 (55.6)	0.695
Indetermined No (%)	2 (66.7)	1 (33.3)	
Positive - No (%)	1 (33.3)	2 (66.7)	

* TNM Tumour staging by Singletary et al al 2003.

** TNM Nodal Staging by Singletary et al 2003.

5.3- Materials and method

DNA extracted from MCF-7 and Tamoxifen resistant (TAMR) cell line was provided by Prof Nicholson's team from Cardiff University, Wales. Cell line DNA and male reference DNA was labelled using Amersham CyScribe Array CGH genomic DNA labelling kit as described in section 3.4.1.3. The BAC DNA used 1.0 Mb platform provided by the Sanger Institute, Cambridge, UK. This contains 3038 BAC clones in triplicate spaced 1 Mb interval across the whole genome. Sample preparation, washing and scanning was performed as described in section 3.4.4. Primary data analysis for cell lines was obtained using CAPWeb (MANOR/GLAD) and with CGHcall (MANOR/GLAD) within the R-platform section (3.6.2.1). GLAD is a pure single-case approach that works with single fixed thresholds. GLAD performs very sensitive and the output seems to be more sensible for cell lines compared with CGHcall.

The CGHcall is a multi-array approach which makes use of all CGH profiles included in a study for calculating the calling probabilities of a particular CGH profile. Analysis was performed using both platforms as CGHcall performs sub-optimally with regard to sensitivity of calling CNA when very low case number is analysed (<20). The basis for calculating calling probabilities is very narrow when the number of cases is very low hence the calling probabilities of very small copy number alterations are often underestimated. Therefore for the biostatistical analysis of the cell lines CAPWeb was used. However, CGHcall was highly suitable for multi-case studies with a greater number of cases and therefore it was used for the analysis of tumour samples. Comparative analysis of CNA between both platforms is performed.

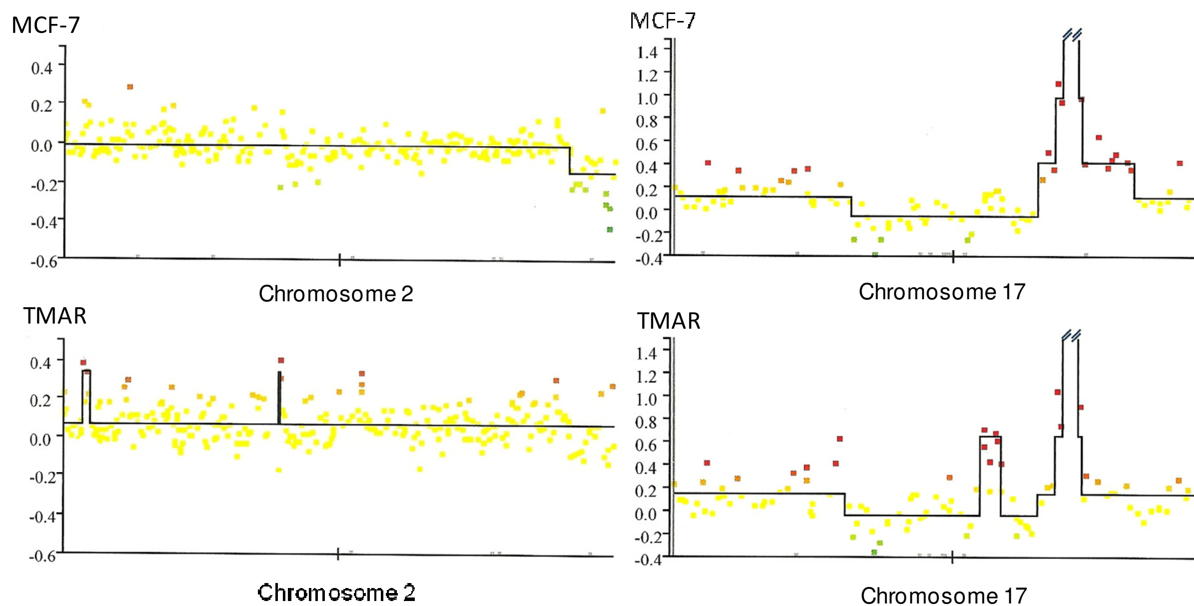
Primary endpoint was BC relapse using a threshold for significant p-value of <0.05. Biostatistical analysis of tumour specimens was performed as described in section 3.6.2.2.

5.4- Results

5.4.1- Comparative aCGH profile between MCF-7 and TAMR cell lines.

Three regions were found to be differentially amplified between MCF-7 and TAMR profiles as shown in Figure 5.2. Two of those regions were found amplified in Chromosome 2 namely 2p25.1: 8595937-104702995 and 2q11.2: 95674761-96161127. A third region of amplification in the TAMR cell lines was found involving 17q21.32-q21.33: 46603673-48556752 (clones RP11361K8 and RP1194C24).

Figure 5.2 Summary plot of CNA in Chromosome 2 (left) and 17 (right) for MCF-7 (top) and TAMR (bottom). Lo2ratios of BAC clones are ordered according to their genomic position from left to right. Yellow dots represent normal regions, red dots copy number gain, green dots copy number losses and blue dots represents amplifications. The black line gives the median Log2ratio of copy number segments.



A summary of genes encoded by amplified regions is shown in table 5.2.

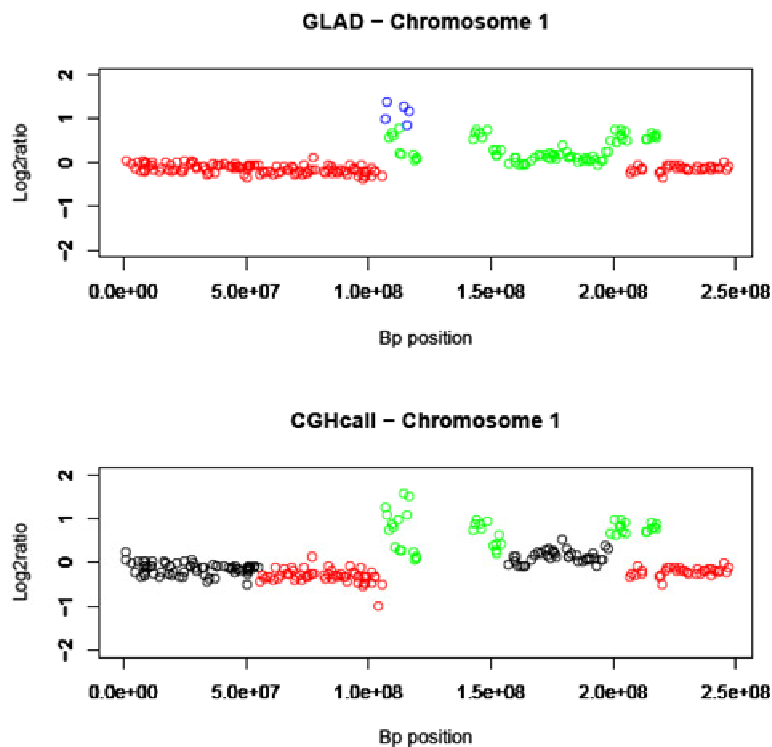
Table 5.2- Regions differentiatlly amplified in TAMR in comparison with MCF-7 cell lines showing regions of alterations and HGNC symbols.

Region	HGNC symbol
Chr 2- p25.1	ID2, MBOAT2, DDEF2, ITGB1BP1, CPSF3,IAH1,ADAM17,YWHAQ,TAF1B,GRHL1,KLF11, CYS1,RRM2, C2orf48, HPCAL1
Chr 2- q11.2	ASTL
Chr 17- q21.32-q21.33	HOXB1, HOXB2, HOXB3, HOXB4, HOXB5, HOXB6, HOXB7, HOXB8, HOXB9, **HOXB13**, TTLL6,CALCOCO2, ATP5G1, UBE2Z, SNF8, GIP, **IGF2BP1**, **B4GALNT2**, GNGT2, **ABI3**, PHOSPHO1, ZNF652, PHB, NGFR, NXPH3, SPOP, SLC35B1, FAM117A , MYST2, TAC4, DLX4, DLX3, ITGA3, PDK2, SAMD14, PPP1R9B, SGCA, HILS1, COL1A1, TMEM92, XYLT2, MRPL27, EME1, LRRC59,CHAD, RSAD1, MYCBPAP, EPN3, , SPATA20, CACNA1G, **ABCC3**, ANKRD40,C17orf73,WFIKKN2,**TOB1**

5.4.1.1 Comparison of the performance of GLAD and CGHcall and cell lines

In the following a "head-to-head" comparison of the performance of CGHcall and GLAD, another segmentation and calling algorithm was performed. Comparison of genomic alterations found using both platforms was relevant as the R package CGHcall (van de Wiel et al.) was used for analysing the array CGH data generated from the tissue samples. In the figure 5.3 shows that the log2-ratio of alterations detected by the GLAD and aCGH algorithm is comparable using the chromosome 1 profile as example.

Figure 5.3- Exemplary comparisonof aCGH profile in MCF-7 cell lines analysed using different "calling" platforms: CAPweb (top) and CGHcall (bottom).



5.3.2 – Array CGH profile in Tamoxifen resistant tumours

A total of 11 and 13 aCGH profiles were included in the TAMRG and TAMCG respectively. Summary plot is shown in Figure 5.4.

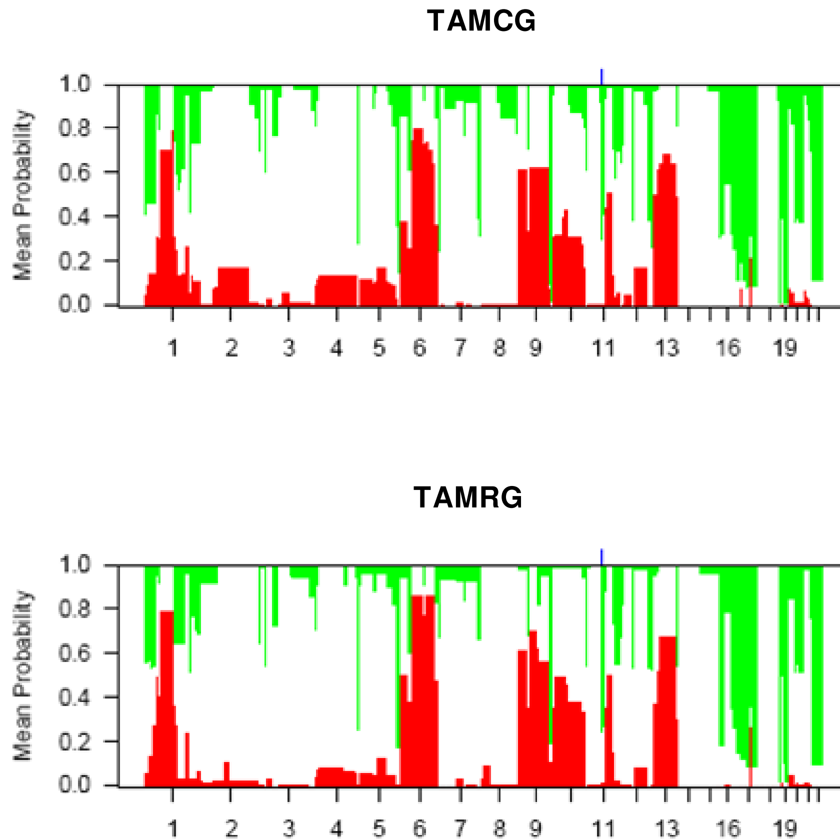


Figure 5.4: Summary Plots of TAMCG patients who did not relapse/die within 5 years (top) and those of TMARG who died/relapsed within 5 years (bottom). The summary plot gives the mean probability of a copy number region being gained (green bars coming from the top) or lost (red bars coming from the bottom) with respect to the whole group. E.g. the first copy number region (gain/green) on chromosome 1 shows a mean probability of about 0.5 in the TAMCG and about 0.4 in the TAMRG. Mean probabilities indirectly reflect the proportion of cases showing an alteration. Amplications are indicated by blue ticks on top of the box.

Permutation t-testing of copy number regions for association with BC relapse showed a region in Chromosome 7 significantly gained (p: 0.046) in the TAMCG (not corrected for MTE in view of low number of observations). This region was gained in 10 cases in the TMACG as opposed to only 3 patients in the TAMRG. Extract genes from this region showed a start BAC clone CTB-164D18 (94136- 161450) and end BAC RP11-425P5 (6267527- 6268062). Manual retrieval of up-to-date (GRCh37) annotation information using MySQL Query Browser showed the region of interest

located 94136- 6268062 bp (Chromosome 7p22.3-p22.1). Genelist extraction from Ensembl Biomart database is shown in table 5.3.

Table 5.3- Genelist extracted from Ensembl Biomart database from region Chr7: 94136- 6268062 showing HGNC and gene description.

	Description
ACTB	Actin, cytoplasmic 1
ADAP1	Arf-GAP with dual PH domain-containing protein 1
AIMP2	Aminoacyl tRNA synthetase complex-interacting multifunctional protein 2
AMZ1	Archaemetzincin-1
ANKRD61	Ankyrin repeat domain-containing protein
C7orf20	UPF0363 protein C7orf20
C7orf27	HEAT repeat-containing protein C7orf27 Precursor
C7orf28A	UPF0550 protein C7orf28
C7orf50	Uncharacterized protein C7orf50
CARD11	Caspase recruitment domain-containing protein 11
CHST12	Carbohydrate sulfotransferase 12
COX19	Cytochrome c oxidase assembly protein COX19
CYP2W1	Cytochrome P450 2W1
CYTH3	Cytohesin-3
EIF2AK1	Eukaryotic translation initiation factor 2-alpha kinase 1
EIF3B	Eukaryotic translation initiation factor 3 subunit B
FAM20C	Dentin matrix protein 4 Precursor
FBXL18	F-box/LRR-repeat protein 18
FOXK1	Forkhead box protein K1
FSCN1	Fascin
FTSJ2	Putative ribosomal RNA methyltransferase 2
GNA12	Guanine nucleotide-binding protein subunit alpha-12
GPER	G-protein coupled estrogen receptor 1
GPR146	Probable G-protein coupled receptor 146
HEATR2	HEAT repeat-containing protein 2
INTS1	Integrator complex subunit 1 (Int1)
IQCE	IQ domain-containing protein E
KIAA0415	Uncharacterized protein KIAA0415
LFNG	Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe
MAD1L1	Mitotic spindle assembly checkpoint protein MAD1
MAFK	Transcription factor MafK
MICALL2	MICAL-like protein 2
MMD2	Monocyte to macrophage differentiation factor 2
NUDT1	7,8-dihydro-8-oxoguanine triphosphatase
OCM	oncomodulin
PAPOLB	Poly(A) polymerase beta
PDGFA	Platelet-derived growth factor subunit A Precursor
PMS2	Mismatch repair endonuclease PMS2
PRKAR1B	cAMP-dependent protein kinase type I-beta regulatory subunit
PSMG3	Proteasome assembly chaperone 3
RADIL	Ras-associating and dilute domain-containing protein
RBAK	RB-associated KRAB zinc finger protein
RNF216	E3 ubiquitin-protein ligase RNF216
RNF216L	Putative protein RNF216-like
RSPH10B	Radial spoke head 10 homolog B
SDK1	Protein sidekick-1 Precursor
SLC29A4	Equilibrative nucleoside transporter 4
SNX8	Sorting nexin-8
TFAMP1	transcription factor A, mitochondrial pseudogene 1
TMEM184A	Transmembrane protein 184A
TNRC18	Trinucleotide repeat-containing gene 18 protein
TTYH3	Protein tweety homolog 3 (hTTY3)
UNCX	Homeobox protein unc-4 homolog
USP42	Ubiquitin carboxyl-terminal hydrolase 42
WIPI2	WD repeat domain phosphoinositide-interacting protein 2
ZFAND2A	AN1-type zinc finger protein 2A
ZNF815	zinc finger protein 815 (ZNF815),

5.3.3 – Association of gene expression located in Chromosome 7 region with early and late relapse in Tamoxifen-treated patients

Association of aCGH finding was performed using a publically available dataset at gene Expression Omnibus (GEO) GSE6532 at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi (Loi et al., 2007a). Clinical criteria used for data selection included tamoxifen-treated patients with HG2 and HG3 who relapse within 5 years of diagnosis (TAMRG) compared with those relapse-free for at least 5 years (TMACG). In this cohort, 131 cases were identified in the TAMCG as opposed to 62 cases in the TAMRG. The list of Affymetrix probes matching region on Chromosome 7 was extracted and log2-ratios collected for each Affymetrix-probe within the region (Table 5.4). Mann-Withney test was performed to assess differential expression between the TMACG and TAMRG. Each probe gives one box-plot along with the according test p-value. The box-plots are ordered according to the position of the corresponding Affymetrix probe from p to q.

Two Affymetrix probes giving a significant p-values <0.05 were found for gene Wipi2 and SNX8. The SNX8 boxplot is shown in Figure 5.5 as an example.

Figure 5.5 shows the boxplot for the expression of SNX-8 gene.

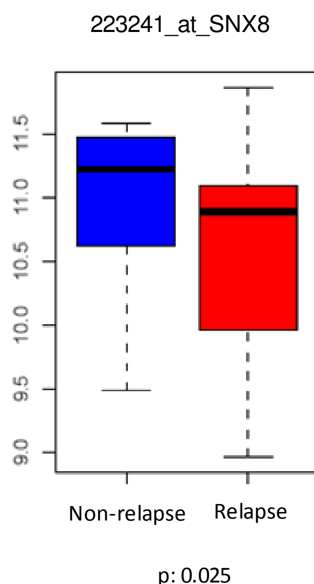


Table 5.4. Probelist extracted GEO for region Chr7: 94136- 6268062

HGNC symbol	Affymetrix probe
ACTB	200801_x_at,213867_x_at , 224594_x_at
AIMP2	202138_x_at, 209971_x_at
C7orf27	225437_s_at
C7orf28A	201974_s_at, 201973_s_at, 215024_at, 208310_s_at
C7orf50	240900_at, 224478_s_at, 240147_at
CARD11	223514_at
CHST12	218927_s_at, 222786_at
COX19	219150_s_at, 90265_at, 235533_at, 231831_at
CYP2W1	220562_at
CYTH3	206523_at, 240400_at, 236136_at, 243752_s_at, 225147_at
EIF2AK1	217736_s_at, 217735_s_at
EIF3B	211501_s_at,208688_x_at,203462_x_at,236274_at,242550_at
FAM20C	226722_at
FBXL18	215068_s_at,220896_at,227500_at
FOXK1	230342_at
FSCN1	210933_s_at,201564_s_at
FTSJ2	218356_at,222130_s_at
GNA12	221737_at,224681_at,242514_at
GPER	210640_s_at,211829_s_at
GPR146	228770_at
HEATR2	218460_at
INTS1	212210_at,212212_s_at
IQCE	217124_at,204202_at
KIAA0415	209912_s_at,209913_x_at
LFNG	215270_at,228762_at
MAD1L1	204857_at,239756_at,233921_s_at,240092_at
MAFK	206750_at,226206_at
MICALL2	219332_at,241478_at
MMD2	230826_at
NUDT1	204766_s_at
OCM	207944_at
PAPOLB	208271_at,242158_at
PDGFA	216867_s_at,205463_s_at
PMS2	209805_at
PRKAR1B	212555_at,212559_at,238183_at
PSMG3	223363_at
RADIL	223693_s_at
RBAK	230439_at,228571_at,241366_at
RNF216	218425_at,218426_s_at,227065_at,244810_at
SDK1	229407_at,229912_at
SLC29A4	227281_at
SNX8	223241_at
TMEM184A	235817_at
TNRC18	214964_at,242187_s_at,229257_at,232658_at,226691_at
TTYH3	224674_at
UNC84A	212074_at,206487_at,214169_at,230210_at,223811_s_at,224808_s_at
USP42	226176_s_at, 226174_at,226669_at
WIPI2	204710_s_at, 202031_s_at, 214699_x_at, 226986_at
ZFAND2A	226650_at
ZNF815	232863_at

Two Affymetrix probes were significantly expressed in the region of interest between the TAMRG and TAMCG namely probe 204710_s_at (0.0098) for gene Wipi2 and probe 223241_at for gene SNX8 (p: 0.0246). The fact that the Wipi2 gene has a probe set consisting in four different probes being statistically significant only one of them make the probe set less specific and unlikely to be biologically relevant. By contrast, the probe selection for the SNX8 gene is unique to a single transcript with regard to binding to a particular transcript of genes and therefore more likely to be specific for the studied event. Over-expression of Snx8 in the TAMCG group is in line with the fact that the region is gained in the non-relapsed group.

5.5- Discussion

5.5.1- Cell lines results

Elucidating the mechanisms involved in the development of hormone resistance in BC continue to be relevant as the majority of BC tumours express ER receptors and therefore benefit from adjuvant tamoxifen therapy. Research on tamoxifen resistance mechanisms has been facilitated by the development of TAMR cell lines. Using a wide-genome aCGH analysis we have found that several genes were differentially amplified in the TAMR cell lines previously found to be related to hormone resistance in BC including HOXB13 (Homeobox protein B13) and TOB1 (Protein Tob1; traducer of ErB-2).

5.5.1.1- Homeobox protein HOX-B13

Genes of this family has been found to be highly conserved among species (Zeltser et al., 1996). It belongs to a group of genes, Homeobox gene family, which encodes for transcriptional regulators of cell growth, differentiation and embryogenesis. It has previously been described to be up-regulated in BC cell lines when compared with normal epithelium (Cantile et al., 2003, Ma et al., 2004) and might directly contribute to tumour progression (Ma et al., 2004). Emerging literature has also linked

expression of HOX-B13 with hormone-related malignancies such as BC (Jerevall et al., 2008, Ma et al., 2004), prostate cancer (Miller et al., 2003, Kim et al.) and ovarian cancer (Miao et al., 2007). In cell lines, HOX-B13 potentiates other biological pathways that stimulate cancer cell progression such us EGFR (Nicholson et al., 1994). Furthermore, HOX-B13 has proven to inhibit the histone acetyltransferase (HAT) activity of CBP/p300, a co-activator mediating ER-dependent nuclear-receptor-activated gene transcription (Hanstein et al., 1996, Chakravarti et al., 1996). In another words, HOX-B13 blocks the CBP/p300 effect of augment the oestrogen-dependent ER activation and therefore modulates the ER signalling pathway.

Several authors have demonstrated that HOX-B13 expression predicts for clinical response to tamoxifen treatment suggesting that it might be involved in tamoxifen resistance. In a series of studies published by Ma et al, it was demonstrated that HOX-B13 was over-expressed in tamoxifen-recurrence cases as opposed to interleukin-17 receptor B (IL17BR), which was over-expressed in non-recurrence cases. The two gene expression ratio HOX-B13/IL17BR was found to be a highly significant (p: 0.0003) predictor of outcome in ER+ patients treated with adjuvant tamoxifen monotherapy (Ma et al., 2004). This results has been clinically validated in a larger independent patient cohort not only at gene expression levels using RT-PCR (Ma et al., 2004, Ma et al., 2006) but also at a protein level analysis using immunohistochenmistry (Jerevall et al., 2010). It has also proven predictive of tamoxifen response in the metastatic settings (Jansen et al., 2007). Taken together, there is evidence that HOX-B13 acts as both prognostic for BC recurrence and predictive for tamoxifen response in ER+ patient with node negative disease treated with adjuvant tamoxifen. This is of clinical relevance as it is estimated that 25% of ER+/PR+ patients and 66% of ER+/PR- patients will fail to respond to tamoxifen even in this good prognostic sub-group of BC. Interestingly, the prognostic significance of the HOX-B13/IL17BR was enhanced when combined with Molecular grade Index (MGI), a five-gene tumour grade signature (Ma et al., 2008).

5.5.1.2- TOB-1: Protein TOB-1 (Transducer of erbB-2 1)

TOB1 is thought to be a tumour suppression gene as TOB expression is lost in lung (Iwanaga et al., 2003) and thyroid cancer (Ito et al., 2005). However, not only little is known about the role of TOB1 in BC but also the very limited body of evidence is contradictory. A study reported TOB1 expression levels in MCF-7 BC cell lines to be inversely correlated with tumour formation and metastatic potential and therefore a tumour suppression role for TOB1 was proposed (O'Malley et al., 2009). TOB1 has an anti-proliferative activity mediated by suppression of Cyclin D, a protein that drives the G1/S phase transition, inducing G1-S arrest.

By contrast, a study performed in a cohort of 160 node negative Stage I-II BC patients, high mRNA TOB1 expression levels was significantly associated with shorter 5-year metastasis-free survival (P = 0.0143) (Helms et al., 2009). The same result was found in the UVA and in silico validation analysis by the same authors using a publically available data using the NCBI dataset GSE2034 (Wang et al., 2005). However, TOB1 phosphorilation status is positively correlated with Ki-67 suggesting TOB1 phosphorilation eliminates its anti-proliferative effect in BC (Helms et al., 2009). Helms et al also found high HER-2 and EGF protein levels were significantly correlated with TOB1 expression which may support a new regulatory role of TOB1 in BC.

Matsuda et al reported that the proliferative function of TOB1 is inhibited by p185ERBB2 as it negatively regulates TOB1 anti-proliferative function resulting in cell growth (Matsuda et al., 1996). Interestingly, authors found no association between TOB mRNA and c-erbB2 expression postulating that TOB1 is also under the regulation of other receptor-type-protein tyrosine kinase.

Our data in cell lines supports more the proposal of TOB1 might play a role in tamoxifen resistance. However, there have been no reports on the role of TOB1 in tamoxifen resistant BC and further research is needed to elucidate this possible mechanism.

5.5.1.3- 17q21.3~q23 gain/amplification is a common event in Tamoxifen resistant development in BC cell lines.

Previous publications have studied the genomic alterations between MCF-7, a tamoxifen sensitive BC cell lines and tamoxifen resistant cell lines such as CL-9. Similar to our finding, both cell lines also demonstrated very similar aCGH profiles but consistent differences were developed during the acquisition of a resistant clone (Achuthan et al., 2001, Forozan et al., 2000). Among other regions, 17q21.3~q23 gain/amplification has been found consistently gained in previous publications (Kytola et al., 2000, Achuthan et al., 2001, Davidson et al., 2000). Interestingly, the genomic regions encoding for ER receptors, namely ERα (chromosome 6q25.1) and ERβ (chromosome 14q) were not found differentially altered among these cell lines suggesting ER-independent mechanism of tamoxifen resistance, supporting the fact that only around 10% of BC harbour ER mutations.

5.5.2- Sorting-Nexin 8 (SNX-8) over-expression might have a protective effect for ER+ patients treated with tamoxifen

Sorting-Nexin 8 gene was found to be amplified in TMACG implying this gene might have a protective role among the ER+ patients treated with tamoxifen. SNX8 belongs to a sorting nexins (SNXs) protein family which function is related to endosomal sorting and transporting of endosomal vesicles to their correct final destination (Dyve et al., 2009). At least 25 human SNXs have been described (Worby and Dixon, 2002) although precise biological function is established for only some of them. The SNXs are formed by three classes of proteins including SNX-BAR, SNX-PX and SNX-other (Attar and Cullen, 2010) depending on the presence of a common domain.

The SNX-PX domain family proteins are both located in the cytoplasm and membrane and share a common sequence of 100-130 aminoacids called PX domain that binds Phosphatidylinositol-3- phosphates (PI3P). SNX1 has been the most widely studied SNX-PX protein. SNX1 has been found to have homology with Mvp1,

involved in the regulation of membrane trafficking of some important membrane receptors such as EGFR (Worby and Dixon, 2002). Along other proteins, SXN1 made up the retromer complex, a pentamer unit that has been involved in cancer biology. Previous reports have found that SNX1 silencing results in increase proliferation and reduced apoptosis (Attar and Cullen, 2010), an effect likely to be mediated by its relation with EGFR. Under this proposal, an inverse functional connection has been established between the retromer and the EGFR so over-expression of SNX1 results in increase degradation of EGFR signal pathway and viceversa (Chin et al., 2001, Cozier et al., 2002). SNX2 has also been found to be involved regulating endosomal sorting of EGFR and is also a member of the retromer complex (Gullapalli et al., 2004). Clinically, SNX1 has been found to be down-regulated in a high number of colon cancers (Nguyen et al., 2006). However, even though Gullapani et al demonstrated that both SNX1 and SNX2 are involved in EGFR regulation, neither of these proteins are essential for this process. Conversely to the SNX1 and SNX2 function, over-expression of SNX5 has been reported to inhibit EGFR degradation and therefore has an antagonistic role with SNX1 function (Liu et al., 2006).

Interestingly, mammary gland gene expression of SNX5 has been described to be negatively regulated by alpha-oestrogen receptor (Aboghe et al., 2009). This finding might lead to an interesting hypothesis that downregulated of SNX5 might be related to BC (Attar and Cullen, 2010), a theory that remains unproven.

SNX8 contains both SNX-PX and BAR domain. It is localised in early endosomes and retromer complex (Dyve et al., 2009). Very limited body of literature is available regarding the SNX8 function. It was not until relatively recently that it was proven that knockdown of SNX8 results in an increment in the transport to Golgi, an opposite effect to that described for SNX1 and 2. Therefore, SNX8 seems to act as a “brake” downregulating the endosomal transport to Golgi with special affinity for highly tubular membranes containing PI(3)P (Dyve et al., 2009).

There has been no previous report of SNX8 in association with cancer to date. However, we have found evidence of SNX8 amplification in the TAMCG and

therefore in patients with hormone-sensitive disease. To the best of our knowledge, it is the first report of SNX8 having a "protective effect" in BC recurrence. Similarly, there has been no previous reports of SNX8 being involved in the ER pathway or even tamoxifen response and therefore validation of these results and further research is needed to confirm the role of SNX8 in BC.

5.5.3- Findings differ between CNA between Tamoxifen resistant cell lines and tumours.

Previous studies have found similar patterns of CNV in cell lines and BC tumours increasing the relevance of results obtained from cell line research (Naylor et al., 2005). This is particularly the case when aCGH profiles are performed in cell lines derived from their parental BC tumours. However, we have found different set of genes differentially expressed in TAMR cell lines and tumours when each of them was hybridised against normal reference DNA. Cell lines growth conditions, mechanism to induce tamoxifen resistance, parental tumour origin and genomic stability are among those factors responsible for discrepancies of experimental results when using cell lines. Similarly, BC tumours are subjected to several in-vivo conditions difficult to reproduce in the laboratory. One of the main difficulties in the study of tamoxifen resistance is the lack of standardisation of the definition of cases and tissue availability. The higher the numbers of criteria applied for case selection, the lower the number of cases available for analysis. In our study, after selecting for all the clinical criteria and tissue availability, aCGH profile was obtained from only 11 BC tamoxifen-resistant tumours, although small it does seems comparable to previous published studies (Han et al., 2006, Vendrell et al., 2008). As only HG2 and HG3 cases were included, the case number was equally low for the controls, expected to be alive and well after 5 years, a criteria usually met by HG1 cases. Despite this limitation, we have found that SNX8 gene was amplified in the TAMCG and over-expressed in similar population using gene expression publically available dataset.

Previous studies have also looked at prognostic genomic alterations involved in tamoxifen resistance. A study of Tamoxifen-treated ER+ BC was also performed by

Han et al. They performed aCGH 9 patients in the recurrence group and 19 patients in the control (non-relapse) group. Different from our study, they included patients with all histological grades and most patients received adjuvant chemotherapy. They found that loss of 11p15.5-p15.4, 11q13.1, 11p11.2 and 1p36.33 and gain 8q21 were significantly associated with BC distant recurrence within 5 years (Han et al., 2006). Several studies have also addressed the question of tamoxifen resistant using GEP (Vendrell et al., 2008, Jansen et al., 2005, Chanrion et al., 2008). In a publication from Vendrell et al GEP was performed in 10 cases considered unfit for primary surgery and had progressed on neo-adjuvant tamoxifen and age-matched to 8 cases free from BC-recurrence for 5 years after primary surgical resection treated with adjuvant tamoxifen alone. Authors identified a 47 gene signature of tamoxifen resistance mainly low expression mRNA levels of immune response genes and high proliferation genes. Further analysis demonstrated a strong prognostic value of the BCL2/FOS signature among the initial 47 genes. When compared the chromosomal distribution of all the genes in the 47 gene-signature there was over-representation of genes in chromosomes 6p21, 11q13, 11q23, 17q11-q21,19p13, and 19q13 (Vendrell et al., 2008).

In another publication, BC patients with advanced disease treated with primary tamoxifen were divided between responders and resistant tumours clinically and using GEP 81 genes were found to be differentially expressed. From those genes a predictive signature of 44 genes was validated to predict for tamoxifen response (Jansen et al., 2005). In another series of 132 primary tumours from patients who received adjuvant tamoxifen, a 36-gene signature was developed with 79% accuracy to correctly classified patients with and without relapse BC (Chanrion et al., 2008).

Interestingly, although comparisons among different signatures have shown high levels of concordance in outcome prediction, there is very little overlap in the involved genes. As a representative example there was only one gene (*FLT1*) identified in the 47-gene signature from Vendrell et al and the 70-gene signature by van't Veer and colleagues (van 't Veer et al., 2002). Despite this low gene-concordance, signatures were likely to be tracking similar biological pathways (Fan et al., 2006). SNX8 gene was not found over-expressed in the above cited studies

suggesting it is a novel candidate gene involved in the development of tamoxifen resistance and therefore further research is need to confirmed this results and elucidate the possible mechanism through which SNX8 seems to reduce the appearance of tamoxifen resistant clones.

CHAPTER 6 – CONCLUSION

Improving BC prognostic classification continues to be a challenge and a key factor for treatment tailoring according to specific clinical and biological factors. The present research work was based on a retrospective analysis of a BC cohort of 1339 patients that provided not only the clinico-pathological parameters for sample selection but also survival data for subsequent molecular biology studies including CNA analysis to enrich our understanding of BC biological pathways and prognosis.

Genome wide analysis has become a very powerful technique for the identification of potential genes involved in cancer development and metastasis. As presented in chapter 2, histological grading was the most important prognostic factor that defined tumour biology among all the routinely assessed clinico-pathological parameters. Therefore, we decided to concentrate on molecular studies within a defined grade. We therefore further characterised a group of HG3 tumours by performing wide-genome analysis aiming to identify CNAs associated with poorer prognosis that might be novel and potentially targetable. Similarly, aCGH was used to identify CNA both in cell lines and tumours considered to be sensitive and resistant to tamoxifen.

6.1- Research contributions

6.1.1- Histological grade appears to be the most important biological prognostic parameter routinely assessed in breast cancer tumours in clinical practice.

Statistical analysis of the BC cohort has revealed very interesting findings. Stage I disease was diagnosed in over 40% of patients and this number is expected to increased as more BC patients are diagnosed earlier due to national screening programme and improvements in adjuvant therapy. HG stood as one of the main variables associated with OS, DDFS and DFS with only N3 disease conferring a higher risk of developing the endpoints. In addition, patients with HG3 tumours tended to be younger, had larger tumours, N+ disease, relapse quicker and have a higher overall and cancer-specific mortality than patients with HG1. We have also

shown that in HG3 disease ER lost its association with DDFS and DFS. The fact that ER was no longer a protective factor in HG3 disease for some of the endpoints is an important issue to consider when assessing adjuvant treatment for early stage BC.

In addition, our study suggests that ER+/HG3 disease should not necessarily be considered as being associated with a good prognosis, and that the treatment offered should be considered in the light of both the ER status and HG as ER+/HG3 tumours that arguably identify the majority of "Luminal B" tumours, which according to our findings and supported by previous publications, have a significantly worse prognosis that ER+/HG1 tumours (Prat and Perou, 2010b, Sotiriou et al., 2003) and be less responsive to therapy with either tamoxifen (Cheang et al., 2009), aromatase inhibitors (Dowsett et al.) or even chemotherapy (Brenton et al., 2005, Cheang et al., 2009). In this study, the risk of dying from BC among ER+/HG3 tumours was about 6 times higher than ER+/HG1 for OS even after adjusting for adjuvant treatment.

Taken together, and even acknowledging the limitations of this study (see section 6.2), we believe it provides additional supporting evidence for the prognostic role of histological grade combined with receptors status in identifying markers of poor prognosis in BC. It also yields further insight into the sub-classification of BC sub-types using pathological parameters used in clinical practice. Furthermore, it provided the study with the demographic and survival data needed for sample selection of molecular biology studies.

6.1.2- Molecular heterogeneity of high grade breast cancer – association of gain of a chromosomal region on 5q with survival.

Beyond clinico-pathological parameters that do not define different biological entities (Sotiriou et al., 2003), genomic analysis has started to establish tumour classifications that correlate with clinical outcome, mainly using GEP (van 't Veer et al., 2002, Paik et al., 2004). Some of these signatures are approved for their application in clinical setting in those cases where a clinical decision based on standard parameters is unclear. Although Oncotype DX® (Paik et al., 2004) is a useful tool in the decision making process, it has significant cost implications, which

is also the case for Mammaprint® (van 't Veer et al., 2002). There is evidence that although the concordance rate among genes in these signatures is low, they have very good outcome prediction and probably track common biological pathways (Fan et al., 2006). Proliferation related genes are recognised as a common factor among these signatures with histological grade reflecting this proliferation differences (Prat and Perou, 2010b).

Having previously established the importance of HG in BC, the present study aimed to use aCGH as a technique to define genomic CNA among poorly differentiated BC patients. aCGH was not, however, established in our laboratory prior to this research and significant time and efforts were dedicated for the successful application of this technique. Array CGH optimisation resulted in successful application of aCGH on FFPE tissue and the development of an original and simplified aCGH washing protocol (section 3.4.4.2). A combined biostatistical analysis of 78 aCGH profiles was carried out. Results showed a gained region on chromosome 5q35.1-2 defined by four BAC clones (RP11-20O22, CTB-54I21, RP11-489P1 and CTB-73D21) which was associated with poorer CSS (p-value < 0.005, FDR < 0.2) even after adjustment by other factors significantly associated with the outcome (HR 3.2; 1.4-7.7 95%CI; p:0.0076). Sub-analysis studies performed among different molecular tumour sub-groups showed significant association of 5q gain with worse OS (p:0.012) and CSS (p:0.005) in ER+ patients when compared with no-gain in the UVA in the same population. Significant association was also observed among HER-2 negative patients both for OS and CSS as shown section 4.2.2.

6.1.3- DUSP-1 and MSX2 were up-regulated in poorly differentiated breast cancers using In-silico validation.

DUSP-1 and MSX2 were among those genes encoded by 5q35.1-2 and validated using a public database. DUSP-1 is the prototype of the MAPK phosphatases (MKPs) involved in diverse cellular processes such as proliferation, differentiation, migration, apoptosis and stress response that has previously been reported associated with BC. Our study therefore supports the evidence in the literature that

up-regulation of DUSP1, in this case via a mechanism of gain of DNA copy number, is significant in both invasion and prognosis in high grade breast cancer.

MSX2 is a transcription factor involves in embryogenesis of different organs, including mammary tissues (Satoh et al., 2004). One of the mechanisms through which MSX2 induces breast cell proliferation is Cyclin D1 up-regulation, a finding present in approximately 40% of BC. Interestingly, MSX2 is also associated with EMT and also significantly induces Twist-1 expression. Recent reports show that in BC this is strongly correlated with high tumour grade, chromosome instability and vascularisation in addition to EMT (Mironchik et al., 2005).

As discussed in section 6.2, our results need to be validated and the real implications of this gain should be explored. Other alterations may be identified in larger and therefore better powered studies. However, this study added to the scientific evidence of the role of DUSP-1 and MSX2 genes in an aggressive phenotype in poorly differentiated BC.

6.1.4- HOXB13 and TOB1 genes are gained in TAMR cell lines.

aCGH was also used to identify differences in aCGH profile between breast cancer cell lines sensitive (MCF-7) and resistant to Tamoxifen (TAMR) in an attempt to define genomic alterations involved in the development of hormone resistance in vitro. A region of amplification in the TAMR cell lines was found involving 17q21.32-q21.33: 46603673-48556752 (clones RP11361K8 and RP1194C24). Several genes were differentially amplified in the TAMR cell lines in this region previously found to be related to hormone resistance in BC including HOXB13 (Homeobox protein B13) and TOB1 (Protein Tob1; traducer of ErB-2) as described in chapter 5.

In particular, HOXB13 has been found to predict for clinical response to tamoxifen treatment suggesting that it might be involved in tamoxifen resistance and it is part of the two gene expression ratio HOX-B13/IL17BR which was found to be a highly significant (p: 0.0003) predictor of outcome in ER+ patients treated with adjuvant

tamoxifen monotherapy (Ma et al., 2004), results that has been validated in a larger independent patient cohort not only at gene expression levels using RT-PCR (Ma et al., 2004, Ma et al., 2006) but also at a protein level analysis using Immunohistochemistry (Jerevall et al., 2010). Taken together, there is evidence that HOX-B13 acts as both prognostic for BC recurrence and predictive for tamoxifen response in ER+ patient with node negative disease treated with adjuvant tamoxifen.

The role of TOB1 in BC is however not well established and the very limited body of evidence is contradictory. A study reported TOB1 expression levels in MCF-7 BC cell lines to be inversely correlated with tumour formation and metastatic potential and therefore a tumour suppression role for TOB1 was proposed (O'Malley et al., 2009). By contrast, a study performed in a cohort of 160 node negative Stage I-II BC patients, high mRNA TOB1 expression levels was significantly associated with shorter 5-year metastasis-free survival (Helms et al., 2009). Helms et al also found high HER-2 and EGF protein levels were significantly correlated with TOB1 expression which may support a new regulatory role of TOB1 in BC. Our data in cell lines supports the proposal that TOB1 might play a role in tamoxifen resistance. However, there have been no reports on the role of TOB1 in tamoxifen resistant BC and further research is needed to elucidate this possible mechanism.

Comparable to our findings, other authors have demonstrated very similar aCGH profiles but consistent differences were developed during with the acquisition of a resistant clone (Achuthan et al., 2001, Forozan et al., 2000). Our findings are in the line of previous published literature where 17q21.3~q23 gain/amplification has been found consistently gained in BC cell lines (Kytola et al., 2000, Achuthan et al., 2001, Davidson et al., 2000).

6.1.5- SNX8 gained – allowing Tamoxifen sensitivity?

The main difficulty encountered by our and other research groups (Han et al., 2006) for the study of Tamoxifen-resistant tumours was sample availability particularly when a stringent case definition was applied. Small sample size will inevitably limit the statistical power and the biological significance of the results. In this study a total

of 11 and 13 aCGH profiles were included in the TAMRG and TAMCG respectively. We found no concordance among TAMR cell lines and Tamoxifen resistant tumours aCGH profile, which in this study mainly consisted in HG2/HG3 cases. A possible explanation is that there are different mechanisms involved in the development of hormone resistant phenotype among poorly differentiated tumours that have certainly added to the complexity of the endocrine resistant issue. Despite of starting sample selection from a relatively large cohort we found a relatively small number of samples from patients considered tamoxifen resistant by our definition. Histological grade again had a significant impact on sample selection, eg: it was difficult to find HG3 samples alive and well after 5 years of diagnosis or HG1 samples from patient who relapsed and died within 5 years who also met other criteria. Therefore, we decided to include HG2 and HG3 patients and both groups were found balanced in all clinico-pathological characteristics other than age and TNM-T staging.

Despite these limitations, a region in Chromosome-7 p22.3-p22.1 was found significantly gained in the TAMCG (p: 0.046). Correlation with gene expression found SNX8 gene amplified in the Tamoxifen sensitive group implying this gene might have a protective role among the ER+ patients treated with Tamoxifen. SNX8 seems to act as a “brake” downregulating the endosomal transport to Golgi with special affinity for highly tubular membranes containing PI(3)P (Dyve et al., 2009). There has been no previous report of SNX8 in association with cancer to date and to the best of our knowledge it is the first report of SNX8 having a “protective effect” in BC recurrence. Similarly, there have been no previous reports of SNX8 being involved in the ER pathway or Tamoxifen response and therefore validation of these results and further research is needed to confirm the role of SNX8 in BC.

6.2- Research limitations.

There are some limitations in the present study. Firstly, the retrospective design of the BC cohort increased the proportion of missing data, particularly for receptor status. The lack of HER-2 data is partially explained by the fact that at the time of data collection IHC for HER-2 was not routinely performed at that time and although

we did repeat HER-2 status centrally, it was only performed for cases within our study in a mainly grade 3 population.

Secondly, administered treatment in this cohort may differ from current practice. As an example as many as 45% of N+ patients did not receive adjuvant CT. Age certainly contributed to this as just over half of those patients were aged ≥70 years and even today there are no data to support routine use of chemotherapy in this age group. In contrast, adjuvant tamoxifen was prescribed to 1141 patients (85.21%) in our cohort; the majority of tamoxifen treated patients were ER+ (81.29%). However, this was consistent with standard UK practice at that time where ER-unknown disease was assumed to be hormone-responsive.

Several technical difficulties had to be overcome for the successful culmination of this research project. Suboptimal laboratory conditions for the development of molecular biology techniques caused significant delays in the establishment of aCGH. It was not until the appropriate laboratory conditions were available that we started obtaining good results.

Several publications have demonstrated the role of laser microdissection prior aCGH experiments in order to minimise normal cell contamination and increase detection sensitivity (Garnis et al., 2005, van Beers and Nederlof, 2006). Following several discussion within the group this step was not considered essential as the Biostatistical package used would apply corrections to minimise the impact of tissue heterogeneity on the genomic profile being able to detect CNV even with only 30% of tumour cell content (Garnis et al., 2005).

Finally, we were only able to performed In-silico validation of our aCGH findings due to the lack of funding and time. In-silico validation is however becoming more widely accepted as a validation state for aCGH experiments. This possibly reflects the successfully evolution of microarray technology to a more advanced state of reproducible and reliable results. It has the main advantage of being cost-effective, quick and reliable particularly when the dataset used for validation is from a recognised research group. We have also demonstrated reproducible aCGH results

using different commercially available kits as described in chapter 3. Despite this, it is recognised that biological validation by FISH, or gene expression studies using qRT-PCR are needed to confirm our finding and be ascertain whether either evaluation of gain of these regions would prove to be a useful tool in treatment tailoring for high grade breast cancer.

6.3- Recommendations for future research

One of the future challenges is to identify what biological pathways confer the different prognoses to each of the sub-groups within HG3 tumours as it is a heterogeneous entity in terms of survival, receptors status and genomic profiles. Further research finding adequate surrogates for the intrinsic sub-type classification of Luminal A and B tumours is also of paramount importance that will determine their applicability into clinical practice and future incorporation into prognostic algorithms. Further characterisation of HG2 cases is also required to ultimately separate them into high and low grade tumours that will probably dictate selection of appropriate therapy. Few techniques are already available such as genomic signatures (Sotiriou et al., 2006, Loi et al., 2007b, Ivshina et al., 2006), analysis of 8 genes by RT-PCR (Toussaint et al., 2009) or Ki-67 (Klintman et al., 2009) but their clinical prospective validation in large RCT is essential.

Research in microarray technology is rapidly evolving with platforms of high resolution oligonucleotide arrays being available along with well established tissue banks of fresh frozen tissues linked with prospective collection of clinical data. Under these circumstances, continuous development of further BC classification using high technology promise unstoppable or perhaps just limited by affordability. Established research groups with the appropriate expertise are more likely to be successful in applying aCGH, expression arrays and proteomics in the same clinical samples that will more rapidly identify targetable genes.

Further biological validation of this research finding is currently being considered within this research team. Functional studies to assess the biological relevance of DUSP-1 and MSX2 genes driving an aggressive phenotype and the role SNX8 gene in tamoxifen sensitivity would be the next scientific step.

REFERENCES

(2009) Office for National Statistics.

. *Cancer Statistics registrations: Registrations of cancer diagnosed in 2006, England. <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8843> Series MB1 no.37. 2009, National Statistics: London.*

, http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8843.

ABD EL-REHIM, D. M., BALL, G., PINDER, S. E., RAKHA, E., PAISH, C., ROBERTSON, J. F., MACMILLAN, D., BLAMEY, R. W. & ELLIS, I. O. (2005) High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int J Cancer,* 116**,** 340-50.

ABOGHE, D. H., YOSHIOKA, M., PHANEUF, D. & ST-AMAND, J. (2009) Regulation of gene expression by estrogen in mammary gland of wild type and estrogen receptor alpha knockout mice. *J Steroid Biochem Mol Biol,* 113**,** 116-26.

ACHUTHAN, R., BELL, S. M., ROBERTS, P., LEEK, J. P., HORGAN, K., MARKHAM, A. F., MACLENNAN, K. A. & SPEIRS, V. (2001) Genetic events during the transformation of a tamoxifen-sensitive human breast cancer cell line into a drug-resistant clone. *Cancer Genet Cytogenet,* 130**,** 166-72.

ADVISORY-COMMITTEE-ON-BREAST-CANCER-SCREENING (2006) Screening for breast cancer in England: past and future. *J Med Screen,* 13**,** 59-61.

ANDERSON, T. J., ALEXANDER, F. E., LAMB, J., SMITH, A. & FORREST, A. P. (2000) Pathology characteristics that optimize outcome prediction of a breast screening trial. *Br J Cancer,* 83**,** 487-92.

ARRIOLA, E., LAMBROS, M. B., JONES, C., DEXTER, T., MACKAY, A., TAN, D. S., TAMBER, N., FENWICK, K., ASHWORTH, A., DOWSETT, M. & REIS-FILHO, J. S. (2007) Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation. *Lab Invest,* 87**,** 75-83.

ATTAR, N. & CULLEN, P. J. (2010) The retromer complex. *Adv Enzyme Regul,* 50**,** 216-36.

BACLESSE, F. & DENOIX, P. F. (1957) [Presentation of a project of clinical classification of malignant tumors of the larynx adjusted by the International Union against Cancer.]. *J Radiol Electrol Arch Electr Medicale,* 38**,** 187-93.

BANG, Y. J., KWON, J. H., KANG, S. H., KIM, J. W. & YANG, Y. C. (1998) Increased MAPK activity and MKP-1 overexpression in human gastric adenocarcinoma. *Biochem Biophys Res Commun,* 250**,** 43-7.

BEJJANI, B. A. & SHAFFER, L. G. (2006) Application of array-based comparative genomic hybridization to clinical diagnostics. *J Mol Diagn,* 8**,** 528-33.

BEMBENEK, A. & SCHLAG, P. M. (2000) Lymph-node dissection in breast cancer. *Langenbecks Arch Surg,* 385**,** 236-45.

BENDER, L. M. & NAHTA, R. (2008) Her2 cross talk and therapeutic resistance in breast cancer. *Front Biosci,* 13**,** 3906-12.

BENJAMINI , H. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B.57(1) 189.

BERGAMASCHI, A., KIM, Y. H., WANG, P., SORLIE, T., HERNANDEZ-BOUSSARD, T., LONNING, P. E., TIBSHIRANI, R., BORRESEN-DALE, A. L. & POLLACK, J. R. (2006) Distinct patterns of DNA copy number alteration are

associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer,* 45**,** 1033-40.

BLOOM, H. J. & RICHARDSON, W. W. (1957) Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer,* 11**,** 359-77.

BLOWS, F. M., DRIVER, K. E., SCHMIDT, M. K., BROEKS, A., VAN LEEUWEN, F. E., WESSELING, J., CHEANG, M. C., GELMON, K., NIELSEN, T. O., BLOMQVIST, C., HEIKKILA, P., HEIKKINEN, T., NEVANLINNA, H., AKSLEN, L. A., BEGIN, L. R., FOULKES, W. D., COUCH, F. J., WANG, X., CAFOUREK, V., OLSON, J. E., BAGLIETTO, L., GILES, G. G., SEVERI, G., MCLEAN, C. A., SOUTHEY, M. C., RAKHA, E., GREEN, A. R., ELLIS, I. O., SHERMAN, M. E., LISSOWSKA, J., ANDERSON, W. F., COX, A., CROSS, S. S., REED, M. W., PROVENZANO, E., DAWSON, S. J., DUNNING, A. M., HUMPHREYS, M., EASTON, D. F., GARCIA-CLOSAS, M., CALDAS, C., PHAROAH, P. D. & HUNTSMAN, D. Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. *PLoS Med,* 7**,** e1000279.

BOIESEN, P., BENDAHL, P. O., ANAGNOSTAKI, L., DOMANSKI, H., HOLM, E., IDVALL, I., JOHANSSON, S., LJUNGBERG, O., RINGBERG, A., OSTBERG, G. & FERNO, M. (2000) Histologic grading in breast cancer--reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. *Acta Oncol,* 39**,** 41-5.

BONNEFOI (2007) To bury TNM classification in breast cancer or to praise it? *European Journal of Cancer Supplements,* 5**,** 23-29.

BOUTROS, T., CHEVET, E. & METRAKOS, P. (2008) Mitogen-activated protein (MAP) kinase/MAP kinase phosphatase regulation: roles in cell growth, death, and cancer. *Pharmacol Rev,* 60**,** 261-310.

BRENTON, J. D., CAREY, L. A., AHMED, A. A. & CALDAS, C. (2005) Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol,* 23**,** 7350-60.

CADY, B. (1997) New era in breast cancer. Impact of screening on disease presentation. *Surg Oncol Clin N Am,* 6**,** 195-202.

CAMPBELL, H. E., TAYLOR, M. A., HARRIS, A. L. & GRAY, A. M. (2009) An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. *Br J Cancer,* 101**,** 1074-84.

CANTILE, M., PETTINATO, G., PROCINO, A., FELICIELLO, I., CINDOLO, L. & CILLO, C. (2003) In vivo expression of the whole HOX gene network in human breast cancer. *Eur J Cancer,* 39**,** 257-64.

CHAKRAVARTI, D., LAMORTE, V. J., NELSON, M. C., NAKAJIMA, T., SCHULMAN, I. G., JUGUILON, H., MONTMINY, M. & EVANS, R. M. (1996) Role of CBP/P300 in nuclear receptor signalling. *Nature,* 383**,** 99-103.

CHANRION, M., NEGRE, V., FONTAINE, H., SALVETAT, N., BIBEAU, F., MAC GROGAN, G., MAURIAC, L., KATSAROS, D., MOLINA, F., THEILLET, C. & DARBON, J. M. (2008) A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res,* 14**,** 1744-52.

CHATTOPADHYAY, S., MACHADO-PINILLA, R., MANGUAN-GARCIA, C., BELDA-INIESTA, C., MORATILLA, C., CEJAS, P., FRESNO-VARA, J. A., DE

CASTRO-CARPENO, J., CASADO, E., NISTAL, M., GONZALEZ-BARON, M. & PERONA, R. (2006) MKP1/CL100 controls tumor growth and sensitivity to cisplatin in non-small-cell lung cancer. *Oncogene,* 25**,** 3335-45.

CHEANG, M. C., CHIA, S. K., VODUC, D., GAO, D., LEUNG, S., SNIDER, J., WATSON, M., DAVIES, S., BERNARD, P. S., PARKER, J. S., PEROU, C. M., ELLIS, M. J. & NIELSEN, T. O. (2009) Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst,* 101**,** 736-50.

CHIN, K., DEVRIES, S., FRIDLYAND, J., SPELLMAN, P. T., ROYDASGUPTA, R., KUO, W. L., LAPUK, A., NEVE, R. M., QIAN, Z., RYDER, T., CHEN, F., FEILER, H., TOKUYASU, T., KINGSLEY, C., DAIRKEE, S., MENG, Z., CHEW, K., PINKEL, D., JAIN, A., LJUNG, B. M., ESSERMAN, L., ALBERTSON, D. G., WALDMAN, F. M. & GRAY, J. W. (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell,* 10**,** 529-41.

CHIN, L. S., RAYNOR, M. C., WEI, X., CHEN, H. Q. & LI, L. (2001) Hrs interacts with sorting nexin 1 and regulates degradation of epidermal growth factor receptor. *J Biol Chem,* 276**,** 7069-78.

CHIN, S. F., TESCHENDORFF, A. E., MARIONI, J. C., WANG, Y., BARBOSA-MORAIS, N. L., THORNE, N. P., COSTA, J. L., PINDER, S. E., VAN DE WIEL, M. A., GREEN, A. R., ELLIS, I. O., PORTER, P. L., TAVARE, S., BRENTON, J. D., YLSTRA, B. & CALDAS, C. (2007) High-resolution array-CGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol,* 8**,** R215.

CINGOZ, S., ALTUNGOZ, O., CANDA, T., SAYDAM, S., AKSAKOGLU, G. & SAKIZLI, M. (2003) DNA copy number changes detected by comparative genomic hybridization and their association with clinicopathologic parameters in breast tumors. *Cancer Genet Cytogenet,* 145**,** 108-14.

COOMBES, R. C., HALL, E., GIBSON, L. J., PARIDAENS, R., JASSEM, J., DELOZIER, T., JONES, S. E., ALVAREZ, I., BERTELLI, G., ORTMANN, O., COATES, A. S., BAJETTA, E., DODWELL, D., COLEMAN, R. E., FALLOWFIELD, L. J., MICKIEWICZ, E., ANDERSEN, J., LONNING, P. E., COCCONI, G., STEWART, A., STUART, N., SNOWDON, C. F., CARPENTIERI, M., MASSIMINI, G., BLISS, J. M. & VAN DE VELDE, C. (2004) A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. *N Engl J Med,* 350**,** 1081-92.

COZIER, G. E., CARLTON, J., MCGREGOR, A. H., GLEESON, P. A., TEASDALE, R. D., MELLOR, H. & CULLEN, P. J. (2002) The phox homology (PX) domain-dependent, 3-phosphoinositide-mediated association of sorting nexin-1 with an early sorting endosomal compartment is required for its ability to regulate epidermal growth factor receptor degradation. *J Biol Chem,* 277**,** 48730-6.

D'EREDITA, G., GIARDINA, C., MARTELLOTTA, M., NATALE, T. & FERRARESE, F. (2001) Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution. *Eur J Cancer,* 37**,** 591-6.

DALTON, L. W., PAGE, D. L. & DUPONT, W. D. (1994) Histologic grading of breast carcinoma. A reproducibility study. *Cancer,* 73**,** 2765-70.

DAVIDSON, J. M., GORRINGE, K. L., CHIN, S. F., ORSETTI, B., BESRET, C., COURTAY-CAHEN, C., ROBERTS, I., THEILLET, C., CALDAS, C. & EDWARDS, P. A. (2000) Molecular cytogenetic analysis of breast cancer cell lines. *Br J Cancer,* 83**,** 1309-17.

DAVIES, J. J., WILSON, I. M. & LAM, W. L. (2005) Array CGH technologies and their applications to cancer genomes. *Chromosome Res,* 13**,** 237-48.

DENKERT, C., SCHMITT, W. D., BERGER, S., RELES, A., PEST, S., SIEGERT, A., LICHTENEGGER, W., DIETEL, M. & HAUPTMANN, S. (2002) Expression of mitogen-activated protein kinase phosphatase-1 (MKP-1) in primary human ovarian carcinoma. *Int J Cancer,* 102**,** 507-13.

DENOIX, P. F. & BACLESSE, F. (1955) [Proposed clinical classification of breast cancers; statement of the International Cancer Union.]. *Bull Assoc Fr Etud Cancer,* 42**,** 433-40.

DEVRIES, S., NYANTE, S., KORKOLA, J., SEGRAVES, R., NAKAO, K., MOORE, D., BAE, H., WILHELM, M., HWANG, S. & WALDMAN, F. (2005) Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J Mol Diagn,* 7**,** 65-71.

DHILLON, A. S., HAGAN, S., RATH, O. & KOLCH, W. (2007) MAP kinase signalling pathways in cancer. *Oncogene,* 26**,** 3279-90.

DI BARI, M. G., GINSBURG, E., PLANT, J., STRIZZI, L., SALOMON, D. S. & VONDERHAAR, B. K. (2009) Msx2 induces epithelial-mesenchymal transition in mouse mammary epithelial cells through upregulation of Cripto-1. *J Cell Physiol,* 219**,** 659-66.

DIMEO, T. A., ANDERSON, K., PHADKE, P., FAN, C., PEROU, C. M., NABER, S. & KUPERWASSER, C. (2009) A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res,* 69**,** 5364-73.

DOWSETT, M., CUZICK, J., WALE, C., FORBES, J., MALLON, E. A., SALTER, J., QUINN, E., DUNBIER, A., BAUM, M., BUZDAR, A., HOWELL, A., BUGARINI, R., BAEHNER, F. L. & SHAK, S. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol,* 28**,** 1829-34.

DURBECQ, V., AMEYE, L., VEYS, I., PAESMANS, M., DESMEDT, C., SIRTAINE, N., SOTIRIOU, C., BERNARD-MARTY, C., NOGARET, J. M., PICCART, M. & LARSIMONT, D. (2008) A significant proportion of elderly patients develop hormone-dependant "luminal-B" tumours associated with aggressive characteristics. *Crit Rev Oncol Hematol,* 67**,** 80-92.

DURINCK, H. (2003) Interface to BioMart databases (e.g. Ensembl, COSMIC ,Wormbase and Gramene).Bioconductor version: Release (2.7).http://www.bioconductor.org/help/bioc-views/2.7/bioc/html/biomaRt.html.

DURINCK, S., MOREAU, Y., KASPRZYK, A., DAVIS, S., DE MOOR, B., BRAZMA, A. & HUBER, W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics,* 21**,** 3439-40.

DYVE, A. B., BERGAN, J., UTSKARPEN, A. & SANDVIG, K. (2009) Sorting nexin 8 regulates endosome-to-Golgi transport. *Biochem Biophys Res Commun,* 390**,** 109-14.

EBCTCG (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet,* 365**,** 1687-717.

ELSTON, C. W. & ELLIS, I. O. (1991) 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,* 19**,** 403-10.

ENJOJI, M., MATSUKUMA, A., SAKAMOTO, G., TOYOSHIMA, S., HIROTA, T., WADA, A., ISHIDA, T., ENOMOTO, K., KOYAMA, H., YAMAMOTO, H. & ET AL. (1992) Invasive ductal carcinoma of the breast with a predominant intraductal component: cooperative clinicopathologic study in seven Japanese centers. *Jpn J Clin Oncol,* 22**,** 84-91.

FAN, C., OH, D. S., WESSELS, L., WEIGELT, B., NUYTEN, D. S., NOBEL, A. B., VAN'T VEER, L. J. & PEROU, C. M. (2006) Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med,* 355**,** 560-9.

FANG, M., TOHER, J., MORGAN, M., DAVISON, J., TANNENBAUM, S. & CLAFFEY, K. (2010) Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. *Cancer*.

FANG, M., TOHER, J., MORGAN, M., DAVISON, J., TANNENBAUM, S. & CLAFFEY, K. (2011) Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. *Cancer,* 117**,** 2024-34.

FIEGLER, H., CARR, P., DOUGLAS, E. J., BURFORD, D. C., HUNT, S., SCOTT, C. E., SMITH, J., VETRIE, D., GORMAN, P., TOMLINSON, I. P. & CARTER, N. P. (2003) DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer,* 36**,** 361-74.

FIEGLER, H., REDON, R. & CARTER, N. P. (2007) Construction and use of spotted large-insert clone DNA microarrays for the detection of genomic copy number changes. *Nat Protoc,* 2**,** 577-87.

FISHER, B., BAUER, M., WICKERHAM, D. L., REDMOND, C. K., FISHER, E. R., CRUZ, A. B., FOSTER, R., GARDNER, B., LERNER, H., MARGOLESE, R. & ET AL. (1983) Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. *Cancer,* 52**,** 1551-7.

FISHER, B., DIGNAM, J., TAN-CHIU, E., ANDERSON, S., FISHER, E. R., WITTLIFF, J. L. & WOLMARK, N. (2001) Prognosis and treatment of patients with breast tumors of one centimeter or less and negative axillary lymph nodes. *J Natl Cancer Inst,* 93**,** 112-20.

FISHER, B., REDMOND, C., FISHER, E. R., BAUER, M., WOLMARK, N., WICKERHAM, D. L., DEUTSCH, M., MONTAGUE, E., MARGOLESE, R. & FOSTER, R. (1985) Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *N Engl J Med,* 312**,** 674-81.

FOROZAN, F., MAHLAMAKI, E. H., MONNI, O., CHEN, Y., VELDMAN, R., JIANG, Y., GOODEN, G. C., ETHIER, S. P., KALLIONIEMI, A. & KALLIONIEMI, O. P. (2000) Comparative genomic hybridization analysis of 38 breast cancer cell

lines: a basis for interpreting complementary DNA microarray data. *Cancer Res,* 60**,** 4519-25.
FOULKES, W. D., GRAINGE, M. J., RAKHA, E. A., GREEN, A. R. & ELLIS, I. O. (2009) Tumor size is an unreliable predictor of prognosis in basal-like breast cancers and does not correlate closely with lymph node status. *Breast Cancer Res Treat,* 117**,** 199-204.
FRIERSON, H. F., JR., WOLBER, R. A., BEREAN, K. W., FRANQUEMONT, D. W., GAFFEY, M. J., BOYD, J. C. & WILBUR, D. C. (1995) Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am J Clin Pathol,* 103**,** 195-8.
FRKOVIC-GRAZIO, S. & BRACKO, M. (2002) Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma. *J Clin Pathol,* 55**,** 88-92.
FULFORD, L. G., EASTON, D. F., REIS-FILHO, J. S., SOFRONIS, A., GILLETT, C. E., LAKHANI, S. R. & HANBY, A. (2006) Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. *Histopathology,* 49**,** 22-34.
GALEA, M. H., BLAMEY, R. W., ELSTON, C. E. & ELLIS, I. O. (1992) The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat,* 22**,** 207-19.
GARCIA, M. J., POLE, J. C., CHIN, S. F., TESCHENDORFF, A., NADERI, A., OZDAG, H., VIAS, M., KRANJAC, T., SUBKHANKULOVA, T., PAISH, C., ELLIS, I., BRENTON, J. D., EDWARDS, P. A. & CALDAS, C. (2005) A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. *Oncogene,* 24**,** 5235-45.
GARNIS, C., COE, B. P., LAM, S. L., MACAULAY, C. & LAM, W. L. (2005) High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics,* 85**,** 790-3.
GOLDHIRSCH, A., GLICK, J. H., GELBER, R. D., COATES, A. S. & SENN, H. J. (2001) Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol,* 19**,** 3817-27.
GOLDHIRSCH, A., INGLE, J. N., GELBER, R. D., COATES, A. S., THURLIMANN, B. & SENN, H. J. (2009) Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol,* 20**,** 1319-29.
GOSPODAROWICZ, M., BENEDET, L., HUTTER, R. V., FLEMING, I., HENSON, D. E. & SOBIN, L. H. (1998) History and international developments in cancer staging. *Cancer Prev Control,* 2**,** 262-8.
GOULD, V. E., KOUKOULIS, G. K., JANSSON, D. S., NAGLE, R. B., FRANKE, W. W. & MOLL, R. (1990) Coexpression patterns of vimentin and glial filament protein with cytokeratins in the normal, hyperplastic, and neoplastic breast. *Am J Pathol,* 137**,** 1143-55.
GULLAPALLI, A., GARRETT, T. A., PAING, M. M., GRIFFIN, C. T., YANG, Y. & TREJO, J. (2004) A role for sorting nexin 2 in epidermal growth factor receptor down-regulation: evidence for distinct functions of sorting nexin 1 and 2 in protein trafficking. *Mol Biol Cell,* 15**,** 2143-55.
GUSTERSON, B. A., WARBURTON, M. J., MITCHELL, D., ELLISON, M., NEVILLE, A. M. & RUDLAND, P. S. (1982) Distribution of myoepithelial cells and

basement membrane proteins in the normal breast and in benign and malignant breast diseases. *Cancer Res,* 42**,** 4763-70.

GUTIERREZ, M. C., DETRE, S., JOHNSTON, S., MOHSIN, S. K., SHOU, J., ALLRED, D. C., SCHIFF, R., OSBORNE, C. K. & DOWSETT, M. (2005) Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. *J Clin Oncol,* 23**,** 2469-76.

GUVAKOVA, M. A. & SURMACZ, E. (1997) Overexpressed IGF-I receptors reduce estrogen growth requirements, enhance survival, and promote E-cadherin-mediated cell-cell adhesion in human breast cancer cells. *Exp Cell Res,* 231**,** 149-62.

HAN, W., HAN, M. R., KANG, J. J., BAE, J. Y., LEE, J. H., BAE, Y. J., LEE, J. E., SHIN, H. J., HWANG, K. T., HWANG, S. E., KIM, S. W. & NOH, D. Y. (2006) Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer,* 6**,** 92.

HANSTEIN, B., ECKNER, R., DIRENZO, J., HALACHMI, S., LIU, H., SEARCY, B., KUROKAWA, R. & BROWN, M. (1996) p300 is a component of an estrogen receptor coactivator complex. *Proc Natl Acad Sci U S A,* 93**,** 11540-5.

HARNETT, A., SMALLWOOD, J., TITSHALL, V. & CHAMPION, A. (2009) Diagnosis and treatment of early breast cancer, including locally advanced disease--summary of NICE guidance. *BMJ,* 338**,** b438.

HARRINGTON , F. (1982) A class of rank test procedures for censored survival data. *Biomelrika* 69**,** 553-66.

HAVERTY, P. M., FRIDLYAND, J., LI, L., GETZ, G., BEROUKHIM, R., LOHR, S., WU, T. D., CAVET, G., ZHANG, Z. & CHANT, J. (2008) High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer,* 47**,** 530-42.

HELMS, M. W., KEMMING, D., CONTAG, C. H., POSPISIL, H., BARTKOWIAK, K., WANG, A., CHANG, S. Y., BUERGER, H. & BRANDT, B. H. (2009) TOB1 is regulated by EGF-dependent HER2 and EGFR signaling, is highly phosphorylated, and indicates poor prognosis in node-negative breast cancer. *Cancer Res,* 69**,** 5049-56.

HICKS, J., KRASNITZ, A., LAKSHMI, B., NAVIN, N. E., RIGGS, M., LEIBU, E., ESPOSITO, D., ALEXANDER, J., TROGE, J., GRUBOR, V., YOON, S., WIGLER, M., YE, K., BORRESEN-DALE, A. L., NAUME, B., SCHLICTING, E., NORTON, L., HAGERSTROM, T., SKOOG, L., AUER, G., MANER, S., LUNDIN, P. & ZETTERBERG, A. (2006) Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res,* 16**,** 1465-79.

HORNER , R. L., KRAPCHO M, NEYMAN N, AMINOU R, HOWLADER N, ET AL (2008) SEER Cancer Statistics Review, 1975-2006, National Cancer Institute. Bethesda, MD. Online. Available at http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site, 2009.

HOWELL, A., ROBERTSON, J. F., QUARESMA ALBANO, J., ASCHERMANNOVA, A., MAURIAC, L., KLEEBERG, U. R., VERGOTE, I., ERIKSTEIN, B., WEBSTER, A. & MORRIS, C. (2002) Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast

cancer progressing after prior endocrine treatment. *J Clin Oncol,* 20**,** 3396-403.

HU, Z., FAN, C., OH, D. S., MARRON, J. S., HE, X., QAQISH, B. F., LIVASY, C., CAREY, L. A., REYNOLDS, E., DRESSLER, L., NOBEL, A., PARKER, J., EWEND, M. G., SAWYER, L. R., WU, J., LIU, Y., NANDA, R., TRETIAKOVA, M., RUIZ ORRICO, A., DREHER, D., PALAZZO, J. P., PERREARD, L., NELSON, E., MONE, M., HANSEN, H., MULLINS, M., QUACKENBUSH, J. F., ELLIS, M. J., OLOPADE, O. I., BERNARD, P. S. & PEROU, C. M. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics,* 7**,** 96.

HUPE, P., STRANSKY, N., THIERY, J. P., RADVANYI, F. & BARILLOT, E. (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics,* 20**,** 3413-22.

HYMAN, E., KAURANIEMI, P., HAUTANIEMI, S., WOLF, M., MOUSSES, S., ROZENBLUM, E., RINGNER, M., SAUTER, G., MONNI, O., ELKAHLOUN, A., KALLIONIEMI, O. P. & KALLIONIEMI, A. (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res,* 62**,** 6240-5.

ICHIMURA, K., MUNGALL, A. J., FIEGLER, H., PEARSON, D. M., DUNHAM, I., CARTER, N. P. & COLLINS, V. P. (2006) Small regions of overlapping deletions on 6q26 in human astrocytic tumours identified using chromosome 6 tile path array-CGH. *Oncogene,* 25**,** 1261-71.

ISHKANIAN, A. S., MALLOFF, C. A., WATSON, S. K., DELEEUW, R. J., CHI, B., COE, B. P., SNIJDERS, A., ALBERTSON, D. G., PINKEL, D., MARRA, M. A., LING, V., MACAULAY, C. & LAM, W. L. (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet,* 36**,** 299-303.

ITO, Y., SUZUKI, T., YOSHIDA, H., TOMODA, C., URUNO, T., TAKAMURA, Y., MIYA, A., KOBAYASHI, K., MATSUZUKA, F., KUMA, K., YAMAMOTO, T. & MIYAUCHI, A. (2005) Phosphorylation and inactivation of Tob contributes to the progression of papillary carcinoma of the thyroid. *Cancer Lett,* 220**,** 237-42.

IVSHINA, A. V., GEORGE, J., SENKO, O., MOW, B., PUTTI, T. C., SMEDS, J., LINDAHL, T., PAWITAN, Y., HALL, P., NORDGREN, H., WONG, J. E., LIU, E. T., BERGH, J., KUZNETSOV, V. A. & MILLER, L. D. (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res,* 66**,** 10292-301.

IWANAGA, K., SUEOKA, N., SATO, A., SAKURAGI, T., SAKAO, Y., TOMINAGA, M., SUZUKI, T., YOSHIDA, Y., J, K. T., YAMAMOTO, T., HAYASHI, S., NAGASAWA, K. & SUEOKA, E. (2003) Alteration of expression or phosphorylation status of tob, a novel tumor suppressor gene product, is an early event in lung cancer. *Cancer Lett,* 202**,** 71-9.

JACKISCH, C. (2006) HER-2-positive metastatic breast cancer: optimizing trastuzumab-based therapy. *Oncologist,* 11 Suppl 1**,** 34-41.

JACQUEMIER, J., GINESTIER, C., ROUGEMONT, J., BARDOU, V. J., CHARAFE-JAUFFRET, E., GENEIX, J., ADELAIDE, J., KOKI, A., HOUVENAEGHEL, G., HASSOUN, J., MARANINCHI, D., VIENS, P., BIRNBAUM, D. & BERTUCCI, F. (2005) Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. *Cancer Res,* 65**,** 767-79.

JANSEN, M. P., FOEKENS, J. A., VAN STAVEREN, I. L., DIRKZWAGER-KIEL, M. M., RITSTIER, K., LOOK, M. P., MEIJER-VAN GELDER, M. E., SIEUWERTS, A. M., PORTENGEN, H., DORSSERS, L. C., KLIJN, J. G. & BERNS, E. M. (2005) Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol,* 23**,** 732-40.

JANSEN, M. P., SIEUWERTS, A. M., LOOK, M. P., RITSTIER, K., MEIJER-VAN GELDER, M. E., VAN STAVEREN, I. L., KLIJN, J. G., FOEKENS, J. A. & BERNS, E. M. (2007) HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol,* 25**,** 662-8.

JARVINEN, A. K., AUTIO, R., KILPINEN, S., SAARELA, M., LEIVO, I., GRENMAN, R., MAKITIE, A. A. & MONNI, O. (2008) High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. *Genes Chromosomes Cancer,* 47**,** 500-9.

JEREVALL, P. L., BROMMESSON, S., STRAND, C., GRUVBERGER-SAAL, S., MALMSTROM, P., NORDENSKJOLD, B., WINGREN, S., SODERKVIST, P., FERNO, M. & STAL, O. (2008) Exploring the two-gene ratio in breast cancer-independent roles for HOXB13 and IL17BR in prediction of clinical outcome. *Breast Cancer Res Treat,* 107**,** 225-34.

JEREVALL, P. L., JANSSON, A., FORNANDER, T., SKOOG, L., NORDENSKJOLD, B. & STAL, O. (2010) Predictive relevance of HOXB13 protein expression for tamoxifen benefit in breast cancer. *Breast Cancer Res,* 12**,** R53.

JOHNSON, C. E., GORRINGE, K. L., THOMPSON, E. R., OPESKIN, K., BOYLE, S. E., WANG, Y., HILL, P., MANN, G. B. & CAMPBELL, I. G. (2011) Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. *Breast Cancer Res Treat.*

JOHNSTON, S. R., SACCANI-JOTTI, G., SMITH, I. E., SALTER, J., NEWBY, J., COPPEN, M., EBBS, S. R. & DOWSETT, M. (1995) Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer. *Cancer Res,* 55**,** 3331-8.

JONSSON, G., NAYLOR, T. L., VALLON-CHRISTERSSON, J., STAAF, J., HUANG, J., WARD, M. R., GRESHOCK, J. D., LUTS, L., OLSSON, H., RAHMAN, N., STRATTON, M., RINGNER, M., BORG, A. & WEBER, B. L. (2005) Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. *Cancer Res,* 65**,** 7612-21.

KALLIONIEMI, A., KALLIONIEMI, O. P., PIPER, J., TANNER, M., STOKKE, T., CHEN, L., SMITH, H. S., PINKEL, D., GRAY, J. W. & WALDMAN, F. M. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A,* 91**,** 2156-60.

KALLIONIEMI, A., KALLIONIEMI, O. P., SUDAR, D., RUTOVITZ, D., GRAY, J. W., WALDMAN, F. & PINKEL, D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science,* 258**,** 818-21.

KEYSE, S. M. (1995) An emerging family of dual specificity MAP kinase phosphatases. *Biochim Biophys Acta,* 1265**,** 152-60.

KEYSE, S. M. (2008) Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev,* 27**,** 253-61.

KIM, Y. R., OH, K. J., PARK, R. Y., XUAN, N. T., KANG, T. W., KWON, D. D., CHOI, C., KIM, M. S., NAM, K. I., AHN, K. Y. & JUNG, C. HOXB13 promotes

androgen independent growth of LNCaP prostate cancer cells by the activation of E2F signaling. *Mol Cancer,* 9**,** 124.
KING, C. R., KRAUS, M. H. & AARONSON, S. A. (1985) Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science,* 229**,** 974-6.
KIRCHHOFF, M., GERDES, T., MAAHR, J., ROSE, H., BENTZ, M., DOHNER, H. & LUNDSTEEN, C. (1999) Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. *Genes Chromosomes Cancer,* 25**,** 410-3.
KLINTMAN, M., BENDAHL, P. O., GRABAU, D., LOVGREN, K., MALMSTROM, P. & FERNO, M. (2009) The prognostic value of Ki67 is dependent on estrogen receptor status and histological grade in premenopausal patients with node-negative breast cancer. *Mod Pathol*.
KNOWLDEN, J. M., HUTCHESON, I. R., JONES, H. E., MADDEN, T., GEE, J. M., HARPER, M. E., BARROW, D., WAKELING, A. E. & NICHOLSON, R. I. (2003) Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells. *Endocrinology,* 144**,** 1032-44.
KROMAN, N., JENSEN, M. B., WOHLFAHRT, J., MOURIDSEN, H. T., ANDERSEN, P. K. & MELBYE, M. (2000) Factors influencing the effect of age on prognosis in breast cancer: population based study. *BMJ,* 320**,** 474-8.
KYTOLA, S., RUMMUKAINEN, J., NORDGREN, A., KARHU, R., FARNEBO, F., ISOLA, J. & LARSSON, C. (2000) Chromosomal alterations in 15 breast cancer cell lines by comparative genomic hybridization and spectral karyotyping. *Genes Chromosomes Cancer,* 28**,** 308-17.
LA ROSA, P., VIARA, E., HUPE, P., PIERRON, G., LIVA, S., NEUVIAL, P., BRITO, I., LAIR, S., SERVANT, N., ROBINE, N., MANIE, E., BRENNETOT, C., JANOUEIX-LEROSEY, I., RAYNAL, V., GRUEL, N., ROUVEIROL, C., STRANSKY, N., STERN, M. H., DELATTRE, O., AURIAS, A., RADVANYI, F. & BARILLOT, E. (2006) VAMP: visualization and analysis of array-CGH, transcriptome and other molecular profiles. *Bioinformatics,* 22**,** 2066-73.
LANIGAN, F., GREMEL, G., HUGHES, R., BRENNAN, D. J., MARTIN, F., JIRSTROM, K. & GALLAGHER, W. M. (2010) Homeobox transcription factor muscle segment homeobox 2 (Msx2) correlates with good prognosis in breast cancer patients and induces apoptosis in vitro. *Breast Cancer Res,* 12**,** R59.
LEGRAND, B., MAZANCOURT, P., DURIGON, M., KHALIFAT, V. & CRAINIC, K. (2002) DNA genotyping of unbuffered formalin fixed paraffin embedded tissues. *Forensic Sci Int,* 125**,** 205-11.
LI, C. I., MOE, R. E. & DALING, J. R. (2003) Risk of mortality by histologic type of breast cancer among women aged 50 to 79 years. *Arch Intern Med,* 163**,** 2149-53.
LIU, H., LIU, Z. Q., CHEN, C. X., MAGILL, S., JIANG, Y. & LIU, Y. J. (2006) Inhibitory regulation of EGF receptor degradation by sorting nexin 5. *Biochem Biophys Res Commun,* 342**,** 537-46.
LIVA, S., HUPE, P., NEUVIAL, P., BRITO, I., VIARA, E., LA ROSA, P. & BARILLOT, E. (2006) CAPweb: a bioinformatics CGH array Analysis Platform. *Nucleic Acids Res,* 34**,** W477-81.
LODA, M., CAPODIECI, P., MISHRA, R., YAO, H., CORLESS, C., GRIGIONI, W., WANG, Y., MAGI-GALLUZZI, C. & STORK, P. J. (1996) Expression of mitogen-activated protein kinase phosphatase-1 in the early phases of human epithelial carcinogenesis. *Am J Pathol,* 149**,** 1553-64.

LOI, S., HAIBE-KAINS, B., DESMEDT, C., LALLEMAND, F., TUTT, A. M., GILLET, C., ELLIS, P., HARRIS, A., BERGH, J., FOEKENS, J. A., KLIJN, J. G., LARSIMONT, D., BUYSE, M., BONTEMPI, G., DELORENZI, M., PICCART, M. J. & SOTIRIOU, C. (2007a) Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol,* 25**,** 1239-46.

LOI, S., PICCART, M. & SOTIRIOU, C. (2007b) The use of gene-expression profiling to better understand the clinical heterogeneity of estrogen receptor positive breast cancers and tamoxifen response. *Crit Rev Oncol Hematol,* 61**,** 187-94.

LOO, L. W., GROVE, D. I., WILLIAMS, E. M., NEAL, C. L., COUSENS, L. A., SCHUBERT, E. L., HOLCOMB, I. N., MASSA, H. F., GLOGOVAC, J., LI, C. I., MALONE, K. E., DALING, J. R., DELROW, J. J., TRASK, B. J., HSU, L. & PORTER, P. L. (2004) Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res,* 64**,** 8541-9.

LOVEDAY, R. L., GREENMAN, J., SIMCOX, D. L., SPEIRS, V., DREW, P. J., MONSON, J. R. & KERIN, M. J. (2000) Genetic changes in breast cancer detected by comparative genomic hybridisation. *Int J Cancer,* 86**,** 494-500.

MA, X. J., HILSENBECK, S. G., WANG, W., DING, L., SGROI, D. C., BENDER, R. A., OSBORNE, C. K., ALLRED, D. C. & ERLANDER, M. G. (2006) The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol,* 24**,** 4611-9.

MA, X. J., SALUNGA, R., DAHIYA, S., WANG, W., CARNEY, E., DURBECQ, V., HARRIS, A., GOSS, P., SOTIRIOU, C., ERLANDER, M. & SGROI, D. (2008) A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res,* 14**,** 2601-8.

MA, X. J., WANG, Z., RYAN, P. D., ISAKOFF, S. J., BARMETTLER, A., FULLER, A., MUIR, B., MOHAPATRA, G., SALUNGA, R., TUGGLE, J. T., TRAN, Y., TRAN, D., TASSIN, A., AMON, P., WANG, W., ENRIGHT, E., STECKER, K., ESTEPA-SABAL, E., SMITH, B., YOUNGER, J., BALIS, U., MICHAELSON, J., BHAN, A., HABIN, K., BAER, T. M., BRUGGE, J., HABER, D. A., ERLANDER, M. G. & SGROI, D. C. (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell,* 5**,** 607-16.

MAKRETSOV, N. A., HUNTSMAN, D. G., NIELSEN, T. O., YORIDA, E., PEACOCK, M., CHEANG, M. C., DUNN, S. E., HAYES, M., VAN DE RIJN, M., BAJDIK, C. & GILKS, C. B. (2004) Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. *Clin Cancer Res,* 10**,** 6143-51.

MANSEL, R. E., FALLOWFIELD, L., KISSIN, M., GOYAL, A., NEWCOMBE, R. G., DIXON, J. M., YIANGOU, C., HORGAN, K., BUNDRED, N., MONYPENNY, I., ENGLAND, D., SIBBERING, M., ABDULLAH, T. I., BARR, L., CHETTY, U., SINNETT, D. H., FLEISSIG, A., CLARKE, D. & ELL, P. J. (2006) Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. *J Natl Cancer Inst,* 98**,** 599-609.

MATSUDA, S., KAWAMURA-TSUZUKU, J., OHSUGI, M., YOSHIDA, M., EMI, M., NAKAMURA, Y., ONDA, M., YOSHIDA, Y., NISHIYAMA, A. & YAMAMOTO, T. (1996) Tob, a novel protein that interacts with p185erbB2, is associated with anti-proliferative activity. *Oncogene,* 12**,** 705-13.

MIAO, J., WANG, Z., PROVENCHER, H., MUIR, B., DAHIYA, S., CARNEY, E., LEONG, C. O., SGROI, D. C. & ORSULIC, S. (2007) HOXB13 promotes ovarian cancer progression. *Proc Natl Acad Sci U S A,* 104**,** 17093-8.
MILLER, G. J., MILLER, H. L., VAN BOKHOVEN, A., LAMBERT, J. R., WERAHERA, P. N., SCHIRRIPA, O., LUCIA, M. S. & NORDEEN, S. K. (2003) Aberrant HOXC expression accompanies the malignant phenotype in human prostate. *Cancer Res,* 63**,** 5879-88.
MIRONCHIK, Y., WINNARD, P. T., JR., VESUNA, F., KATO, Y., WILDES, F., PATHAK, A. P., KOMINSKY, S., ARTEMOV, D., BHUJWALLA, Z., VAN DIEST, P., BURGER, H., GLACKIN, C. & RAMAN, V. (2005) Twist overexpression induces in vivo angiogenesis and correlates with chromosomal instability in breast cancer. *Cancer Res,* 65**,** 10801-9.
MOLLER, S., JENSEN, M. B., EJLERTSEN, B., BJERRE, K. D., LARSEN, M., HANSEN, H. B., CHRISTIANSEN, P. & MOURIDSEN, H. T. (2008) The clinical database and the treatment guidelines of the Danish Breast Cancer Cooperative Group (DBCG); its 30-years experience and future promise. *Acta Oncol,* 47**,** 506-24.
MONCHO-AMOR, V., IBANEZ DE CACERES, I., BANDRES, E., MARTINEZ-POVEDA, B., ORGAZ, J. L., SANCHEZ-PEREZ, I., ZAZO, S., ROVIRA, A., ALBANELL, J., JIMENEZ, B., ROJO, F., BELDA-INIESTA, C., GARCIA-FONCILLAS, J. & PERONA, R. DUSP1/MKP1 promotes angiogenesis, invasion and metastasis in non-small-cell lung cancer. *Oncogene*.
MONTAGUT, C., IGLESIAS, M., ARUMI, M., BELLOSILLO, B., GALLEN, M., MARTINEZ-FERNANDEZ, A., MARTINEZ-AVILES, L., CANADAS, I., DALMASES, A., MORAGON, E., LEMA, L., SERRANO, S., ROVIRA, A., ROJO, F., BELLMUNT, J. & ALBANELL, J. (2010) Mitogen-activated protein kinase phosphatase-1 (MKP-1) impairs the response to anti-epidermal growth factor receptor (EGFR) antibody cetuximab in metastatic colorectal cancer patients. *Br J Cancer,* 102**,** 1137-44.
MOOK, S., SCHMIDT, M. K., RUTGERS, E. J., VAN DE VELDE, A. O., VISSER, O., RUTGERS, S. M., ARMSTRONG, N., VAN'T VEER, L. J. & RAVDIN, P. M. (2009) Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study. *Lancet Oncol*.
MORI, S., CHANG, J. T., ANDRECHEK, E. R., MATSUMURA, N., BABA, T., YAO, G., KIM, J. W., GATZA, M., MURPHY, S. & NEVINS, J. R. (2009) Anchorage-independent cell growth signature identifies tumors with metastatic potential. *Oncogene,* 28**,** 2796-805.
MORROW, M., JAGSI, R., ALDERMAN, A. K., GRIGGS, J. J., HAWLEY, S. T., HAMILTON, A. S., GRAFF, J. J. & KATZ, S. J. (2009) Surgeon recommendations and receipt of mastectomy for treatment of breast cancer. *JAMA,* 302**,** 1551-6.
NATRAJAN, R. (2009) Tiling Path Genomic Profiling of Grade 3 Invasive Ductal Breast Cancers. *Clinical cancer research,* 15**,** 2711-2722.
NAYLOR, T. L., GRESHOCK, J., WANG, Y., COLLIGON, T., YU, Q. C., CLEMMER, V., ZAKS, T. Z. & WEBER, B. L. (2005) High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res,* 7**,** R1186-98.

NEUVIAL, P., HUPE, P., BRITO, I., LIVA, S., MANIE, E., BRENNETOT, C., RADVANYI, F., AURIAS, A. & BARILLOT, E. (2006) Spatial normalization of array-CGH data. *BMC Bioinformatics,* 7**,** 264.

NGUYEN, L. N., HOLDREN, M. S., NGUYEN, A. P., FURUYA, M. H., BIANCHINI, M., LEVY, E., MORDOH, J., LIU, A., GUNCAY, G. D., CAMPBELL, J. S. & PARKS, W. T. (2006) Sorting nexin 1 down-regulation promotes colon tumorigenesis. *Clin Cancer Res,* 12**,** 6952-9.

NHSBSP (2005) Pathology reporting of breast disease. *NHSBSP Publication. No 58, January 2005. NHS Cancer Screening Programmes website (www.cancerscreening.org.uk) and The Royal College of Pathologists' website (www.rcpath.org)*.

NICHOLSON, R. I., HUTCHESON, I. R., HARPER, M. E., KNOWLDEN, J. M., BARROW, D., MCCLELLAND, R. A., JONES, H. E., WAKELING, A. E. & GEE, J. M. (2002) Modulation of epidermal growth factor receptor in endocrine-resistant, estrogen-receptor-positive breast cancer. *Ann N Y Acad Sci,* 963**,** 104-15.

NICHOLSON, R. I., HUTCHESON, I. R., KNOWLDEN, J. M., JONES, H. E., HARPER, M. E., JORDAN, N., HISCOX, S. E., BARROW, D. & GEE, J. M. (2004) Nonendocrine pathways and endocrine resistance: observations with antiestrogens and signal transduction inhibitors in combination. *Clin Cancer Res,* 10**,** 346S-54S.

NICHOLSON, R. I., MCCLELLAND, R. A., GEE, J. M., MANNING, D. L., CANNON, P., ROBERTSON, J. F., ELLIS, I. O. & BLAMEY, R. W. (1994) Epidermal growth factor receptor expression in breast cancer: association with response to endocrine therapy. *Breast Cancer Res Treat,* 29**,** 117-25.

NIELSEN, T. O., HSU, F. D., JENSEN, K., CHEANG, M., KARACA, G., HU, Z., HERNANDEZ-BOUSSARD, T., LIVASY, C., COWAN, D., DRESSLER, L., AKSLEN, L. A., RAGAZ, J., GOWN, A. M., GILKS, C. B., VAN DE RIJN, M. & PEROU, C. M. (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res,* 10**,** 5367-74.

NUGOLI, M., CHUCHANA, P., VENDRELL, J., ORSETTI, B., URSULE, L., NGUYEN, C., BIRNBAUM, D., DOUZERY, E. J., COHEN, P. & THEILLET, C. (2003) Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications. *BMC Cancer,* 3**,** 13.

NYANTE, S. J., DEVRIES, S., CHEN, Y. Y. & HWANG, E. S. (2004) Array-based comparative genomic hybridization of ductal carcinoma in situ and synchronous invasive lobular cancer. *Hum Pathol,* 35**,** 759-63.

O'MALLEY, S., SU, H., ZHANG, T., NG, C., GE, H. & TANG, C. K. (2009) TOB suppresses breast cancer tumorigenesis. *Int J Cancer,* 125**,** 1805-13.

OFFICE-FOR-NATIONAL-STATISTICS (2008) Registrations of cancer diagnosed in 2006, England_Series MB1 No 37. Online. Available at http://www.statistics.gov.uk/downloads/theme_health/MB1-37/MB1_37_2006.pdf.

OFFICE-FOR-NATIONAL-STATISTICS (2010) Breast Cancer Statistics. Online. Available at http://www.statistics.gov.uk/cci/nugget.asp?id=575. Accessed 7th May 2011.

OLIVOTTO, I. A., BAJDIK, C. D., RAVDIN, P. M., SPEERS, C. H., COLDMAN, A. J., NORRIS, B. D., DAVIS, G. J., CHIA, S. K. & GELMON, K. A. (2005) Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol,* 23**,** 2716-25.

OLSAVSKY, K. M., PAGE, J. L., JOHNSON, M. C., ZARBL, H., STROM, S. C. & OMIECINSKI, C. J. (2007) Gene expression profiling and differentiation assessment in primary human hepatocyte cultures, established hepatoma cell lines, and human liver tissues. *Toxicol Appl Pharmacol,* 222**,** 42-56.

OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. & WIGLER, M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics,* 5**,** 557-72.

OSBORNE, C. K., PIPPEN, J., JONES, S. E., PARKER, L. M., ELLIS, M., COME, S., GERTLER, S. Z., MAY, J. T., BURTON, G., DIMERY, I., WEBSTER, A., MORRIS, C., ELLEDGE, R. & BUZDAR, A. (2002) Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. *J Clin Oncol,* 20**,** 3386-95.

PAIK, S., SHAK, S., TANG, G., KIM, C., BAKER, J., CRONIN, M., BAEHNER, F. L., WALKER, M. G., WATSON, D., PARK, T., HILLER, W., FISHER, E. R., WICKERHAM, D. L., BRYANT, J. & WOLMARK, N. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med,* 351**,** 2817-26.

PAIK, S., TANG, G., SHAK, S., KIM, C., BAKER, J., KIM, W., CRONIN, M., BAEHNER, F. L., WATSON, D., BRYANT, J., COSTANTINO, J. P., GEYER, C. E., JR., WICKERHAM, D. L. & WOLMARK, N. (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol,* 24**,** 3726-34.

PARL, F. F. & DUPONT, W. D. (1982) A retrospective cohort study of Histologic risk factors in breast cancer patients. *Cancer,* 50**,** 2410-6.

PEROU, C. M., JEFFREY, S. S., VAN DE RIJN, M., REES, C. A., EISEN, M. B., ROSS, D. T., PERGAMENSCHIKOV, A., WILLIAMS, C. F., ZHU, S. X., LEE, J. C., LASHKARI, D., SHALON, D., BROWN, P. O. & BOTSTEIN, D. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A,* 96**,** 9212-7.

PEROU, C. M., SORLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H., AKSLEN, L. A., FLUGE, O., PERGAMENSCHIKOV, A., WILLIAMS, C., ZHU, S. X., LONNING, P. E., BORRESEN-DALE, A. L., BROWN, P. O. & BOTSTEIN, D. (2000) Molecular portraits of human breast tumours. *Nature,* 406**,** 747-52.

PICCART-GEBHART, M. J., PROCTER, M., LEYLAND-JONES, B., GOLDHIRSCH, A., UNTCH, M., SMITH, I., GIANNI, L., BASELGA, J., BELL, R., JACKISCH, C., CAMERON, D., DOWSETT, M., BARRIOS, C. H., STEGER, G., HUANG, C. S., ANDERSSON, M., INBAR, M., LICHINITSER, M., LANG, I., NITZ, U., IWATA, H., THOMSSEN, C., LOHRISCH, C., SUTER, T. M., RUSCHOFF, J., SUTO, T., GREATOREX, V., WARD, C., STRAEHLE, C., MCFADDEN, E., DOLCI, M. S. & GELBER, R. D. (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med,* 353**,** 1659-72.

PIERGA, J. Y., REIS-FILHO, J. S., CLEATOR, S. J., DEXTER, T., MACKAY, A., SIMPSON, P., FENWICK, K., IRAVANI, M., SALTER, J., HILLS, M., JONES, C., ASHWORTH, A., SMITH, I. E., POWLES, T. & DOWSETT, M. (2007) Microarray-based comparative genomic hybridisation of breast cancer patients receiving neoadjuvant chemotherapy. *Br J Cancer,* 96**,** 341-51.

PINDER, S. E., MURRAY, S., ELLIS, I. O., TRIHIA, H., ELSTON, C. W., GELBER, R. D., GOLDHIRSCH, A., LINDTNER, J., CORTES-FUNES, H., SIMONCINI, E., BYRNE, M. J., GOLOUH, R., RUDENSTAM, C. M., CASTIGLIONE-GERTSCH, M. & GUSTERSON, B. A. (1998) The importance of the histologic grade of invasive breast carcinoma and response to chemotherapy. *Cancer,* 83**,** 1529-39.

PINKEL, D., SEGRAVES, R., SUDAR, D., CLARK, S., POOLE, I., KOWBEL, D., COLLINS, C., KUO, W. L., CHEN, C., ZHAI, Y., DAIRKEE, S. H., LJUNG, B. M., GRAY, J. W. & ALBERTSON, D. G. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet,* 20**,** 207-11.

POLLACK, J. R., PEROU, C. M., ALIZADEH, A. A., EISEN, M. B., PERGAMENSCHIKOV, A., WILLIAMS, C. F., JEFFREY, S. S., BOTSTEIN, D. & BROWN, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet,* 23**,** 41-6.

POLLACK, J. R., SORLIE, T., PEROU, C. M., REES, C. A., JEFFREY, S. S., LONNING, P. E., TIBSHIRANI, R., BOTSTEIN, D., BORRESEN-DALE, A. L. & BROWN, P. O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A,* 99**,** 12963-8.

PRAT, A. & PEROU, C. M. (2010a) Deconstructing the molecular portraits of breast cancer. *Mol Oncol*.

PRAT, A. & PEROU, C. M. (2010b) Deconstructing the molecular portraits of breast cancer. *Mol Oncol,* 5**,** 5-23.

RAKHA, E. A., EL-SAYED, M. E., REIS-FILHO, J. S. & ELLIS, I. O. (2008) Expression profiling technology: its contribution to our understanding of breast cancer. *Histopathology,* 52**,** 67-81.

RASTELLI, F. & CRISPINO, S. (2008) Factors predictive of response to hormone therapy in breast cancer. *Tumori,* 94**,** 370-83.

RAUHALA, H. E., PORKKA, K. P., TOLONEN, T. T., MARTIKAINEN, P. M., TAMMELA, T. L. & VISAKORPI, T. (2005) Dual-specificity phosphatase 1 and serum/glucocorticoid-regulated kinase are downregulated in prostate cancer. *Int J Cancer,* 117**,** 738-45.

RAVDIN, P. M., SIMINOFF, I. A. & HARVEY, J. A. (1998) Survey of breast cancer patients concerning their knowledge and expectations of adjuvant therapy. *J Clin Oncol,* 16**,** 515-21.

RAVDIN, P. M., SIMINOFF, L. A., DAVIS, G. J., MERCER, M. B., HEWLETT, J., GERSON, N. & PARKER, H. L. (2001) Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol,* 19**,** 980-91.

REDON, R., RIO, M., GREGORY, S. G., COOPER, R. A., FIEGLER, H., SANLAVILLE, D., BANERJEE, R., SCOTT, C., CARR, P., LANGFORD, C., CORMIER-DAIRE, V., MUNNICH, A., CARTER, N. P. & COLLEAUX, L. (2005) Tiling path resolution mapping of constitutional 1p36 deletions by array-CGH: contiguous gene deletion or "deletion with positional effect" syndrome? *J Med Genet,* 42**,** 166-71.

REIS-FILHO, J. S. & LAKHANI, S. R. (2008) Breast cancer special types: why bother? *J Pathol,* 216**,** 394-8.

REIS-FILHO, J. S., MILANEZI, F., CARVALHO, S., SIMPSON, P. T., STEELE, D., SAVAGE, K., LAMBROS, M. B., PEREIRA, E. M., NESLAND, J. M.,

LAKHANI, S. R. & SCHMITT, F. C. (2005a) Metaplastic breast carcinomas exhibit EGFR, but not HER2, gene amplification and overexpression: immunohistochemical and chromogenic in situ hybridization analysis. *Breast Cancer Res,* 7**,** R1028-35.
REIS-FILHO, J. S., SIMPSON, P. T., JONES, C., STEELE, D., MACKAY, A., IRAVANI, M., FENWICK, K., VALGEIRSSON, H., LAMBROS, M., ASHWORTH, A., PALACIOS, J., SCHMITT, F. & LAKHANI, S. R. (2005b) Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity. *J Pathol,* 207**,** 1-13.
RENNSTAM, K., AHLSTEDT-SOINI, M., BALDETORP, B., BENDAHL, P. O., BORG, A., KARHU, R., TANNER, M., TIRKKONEN, M. & ISOLA, J. (2003) Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res,* 63**,** 8861-8.
RIGGINS, R. B., SCHRECENGOST, R. S., GUERRERO, M. S. & BOUTON, A. H. (2007) Pathways to tamoxifen resistance. *Cancer Lett,* 256**,** 1-24.
ROBBINS, P., PINDER, S., DE KLERK, N., DAWKINS, H., HARVEY, J., STERRETT, G., ELLIS, I. & ELSTON, C. (1995) Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol,* 26**,** 873-9.
ROJO, F., GONZALEZ-NAVARRETE, I., BRAGADO, R., DALMASES, A., MENENDEZ, S., CORTES-SEMPERE, M., SUAREZ, C., OLIVA, C., SERVITJA, S., RODRIGUEZ-FANJUL, V., SANCHEZ-PEREZ, I., CAMPAS, C., COROMINAS, J. M., TUSQUETS, I., BELLOSILLO, B., SERRANO, S., PERONA, R., ROVIRA, A. & ALBANELL, J. (2009) Mitogen-activated protein kinase phosphatase-1 in human breast cancer independently predicts prognosis and is repressed by doxorubicin. *Clin Cancer Res,* 15**,** 3530-9.
ROMOND, E. H., PEREZ, E. A., BRYANT, J., SUMAN, V. J., GEYER, C. E., JR., DAVIDSON, N. E., TAN-CHIU, E., MARTINO, S., PAIK, S., KAUFMAN, P. A., SWAIN, S. M., PISANSKY, T. M., FEHRENBACHER, L., KUTTEH, L. A., VOGEL, V. G., VISSCHER, D. W., YOTHERS, G., JENKINS, R. B., BROWN, A. M., DAKHIL, S. R., MAMOUNAS, E. P., LINGLE, W. L., KLEIN, P. M., INGLE, J. N. & WOLMARK, N. (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med,* 353**,** 1673-84.
ROUZIER, R., PEROU, C. M., SYMMANS, W. F., IBRAHIM, N., CRISTOFANILLI, M., ANDERSON, K., HESS, K. R., STEC, J., AYERS, M., WAGNER, P., MORANDI, P., FAN, C., RABIUL, I., ROSS, J. S., HORTOBAGYI, G. N. & PUSZTAI, L. (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res,* 11**,** 5678-85.
ROYLANCE, R., GORMAN, P., HARRIS, W., LIEBMANN, R., BARNES, D., HANBY, A. & SHEER, D. (1999) Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. *Cancer Res,* 59**,** 1433-6.
SATOH, K., GINSBURG, E. & VONDERHAAR, B. K. (2004) Msx-1 and Msx-2 in mammary gland development. *J Mammary Gland Biol Neoplasia,* 9**,** 195-205.
SATOH, K., HAMADA, S., KANNO, A., ISHIDA, K., ITO, H., HIROTA, M., MASAMUNE, A., EGAWA, S., UNNO, M. & SHIMOSEGAWA, T. (2010) Evaluation of MSX2 mRNA in brush cytology specimens distinguished pancreatic carcinoma from chronic pancreatitis. *Cancer Sci,* 102**,** 157-61.

SATOH, K., HAMADA, S., KIMURA, K., KANNO, A., HIROTA, M., UMINO, J., FUJIBUCHI, W., MASAMUNE, A., TANAKA, N., MIURA, K., EGAWA, S., MOTOI, F., UNNO, M., VONDERHAAR, B. K. & SHIMOSEGAWA, T. (2008) Up-regulation of MSX2 enhances the malignant phenotype and is associated with twist 1 expression in human pancreatic cancer cells. *Am J Pathol,* 172**,** 926-39.

SIMS, A. H., ONG, K. R., CLARKE, R. B. & HOWELL, A. (2006) 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,* 8**,** 214.

SINGLETARY, S. E. & GREENE, F. L. (2003) Revision of breast cancer staging: the 6th edition of the TNM Classification. *Semin Surg Oncol,* 21**,** 53-9.

SLAMON, D. J., CLARK, G. M., WONG, S. G., LEVIN, W. J., ULLRICH, A. & MCGUIRE, W. L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science,* 235**,** 177-82.

SMALL, G. W., SHI, Y. Y., EDMUND, N. A., SOMASUNDARAM, S., MOORE, D. T. & ORLOWSKI, R. Z. (2004) Evidence that mitogen-activated protein kinase phosphatase-1 induction by proteasome inhibitors plays an antiapoptotic role. *Mol Pharmacol,* 66**,** 1478-90.

SMALL, G. W., SHI, Y. Y., HIGGINS, L. S. & ORLOWSKI, R. Z. (2007) Mitogen-activated protein kinase phosphatase-1 is a mediator of breast cancer chemoresistance. *Cancer Res,* 67**,** 4459-66.

SOBIN, L. H. (2001) TNM: principles, history, and relation to other prognostic factors. *Cancer,* 91**,** 1589-92.

SOERJOMATARAM, I., LOUWMAN, M. W., RIBOT, J. G., ROUKEMA, J. A. & COEBERGH, J. W. (2008) An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat,* 107**,** 309-30.

SOLINAS-TOLDO, S., LAMPEL, S., STILGENBAUER, S., NICKOLENKO, J., BENNER, A., DOHNER, H., CREMER, T. & LICHTER, P. (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer,* 20**,** 399-407.

SORLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., THORSEN, T., QUIST, H., MATESE, J. C., BROWN, P. O., BOTSTEIN, D., EYSTEIN LONNING, P. & BORRESEN-DALE, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A,* 98**,** 10869-74.

SORLIE, T., TIBSHIRANI, R., PARKER, J., HASTIE, T., MARRON, J. S., NOBEL, A., DENG, S., JOHNSEN, H., PESICH, R., GEISLER, S., DEMETER, J., PEROU, C. M., LONNING, P. E., BROWN, P. O., BORRESEN-DALE, A. L. & BOTSTEIN, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A,* 100**,** 8418-23.

SOTIRIOU, C., NEO, S. Y., MCSHANE, L. M., KORN, E. L., LONG, P. M., JAZAERI, A., MARTIAT, P., FOX, S. B., HARRIS, A. L. & LIU, E. T. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A,* 100**,** 10393-8.

SOTIRIOU, C., WIRAPATI, P., LOI, S., HARRIS, A., FOX, S., SMEDS, J., NORDGREN, H., FARMER, P., PRAZ, V., HAIBE-KAINS, B., DESMEDT, C., LARSIMONT, D., CARDOSO, F., PETERSE, H., NUYTEN, D., BUYSE, M.,

VAN DE VIJVER, M. J., BERGH, J., PICCART, M. & DELORENZI, M. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst,* 98**,** 262-72.
STAAF, J., JONSSON, G., RINGNER, M., VALLON-CHRISTERSSON, J., GRABAU, D., ARASON, A., GUNNARSSON, H., AGNARSSON, B. A., MALMSTROM, P. O., JOHANNSSON, O. T., LOMAN, N., BARKARDOTTIR, R. B. & BORG, A. (2010) High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res,* 12**,** R25.
STEPHEN, R. L., SHAW, L. E., LARSEN, C., CORCORAN, D. & DARBRE, P. D. (2001) Insulin-like growth factor receptor levels are regulated by cell density and by long term estrogen deprivation in MCF7 human breast cancer cells. *J Biol Chem,* 276**,** 40080-6.
TAN, D. S. & REIS-FILHO, J. S. (2008) Comparative genomic hybridisation arrays: high-throughput tools to determine targeted therapy in breast cancer. *Pathobiology,* 75**,** 63-74.
TATARU, D., ROBINSON, D., MOLLER, H. & DAVIES, E. (2006) Trends in the treatment of breast cancer in Southeast England following the introduction of national guidelines. *J Public Health (Oxf),* 28**,** 215-7.
THEODOSIOU, A. & ASHWORTH, A. (2002) MAP kinase phosphatases. *Genome Biol,* 3**,** REVIEWS3009.
TINNEMANS, J. G., WOBBES, T., HOLLAND, R., HENDRIKS, J. H., VAN DER SLUIS, R. F. & DE BOER, H. H. (1989) Treatment and survival of female patients with nonpalpable breast carcinoma. *Ann Surg,* 209**,** 249-53.
TIRKKONEN, M., TANNER, M., KARHU, R., KALLIONIEMI, A., ISOLA, J. & KALLIONIEMI, O. P. (1998) Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer,* 21**,** 177-84.
TOMASKOVIC-CROOK, E., THOMPSON, E. W. & THIERY, J. P. (2009) Epithelial to mesenchymal transition and breast cancer. *Breast Cancer Res,* 11**,** 213.
TOUSSAINT, J., SIEUWERTS, A. M., HAIBE-KAINS, B., DESMEDT, C., ROUAS, G., HARRIS, A. L., LARSIMONT, D., PICCART, M., FOEKENS, J. A., DURBECQ, V. & SOTIRIOU, C. (2009) 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,* 10**,** 424.
UICC (1958) International Union Against Cancer , Committee on Clinical Stage Classification and Applied Statistics: Clinical stage classification and presentation of results, malignant tumours of the breast and larynx.
.
VAN 'T VEER, L. J., DAI, H., VAN DE VIJVER, M. J., HE, Y. D., HART, A. A., MAO, M., PETERSE, H. L., VAN DER KOOY, K., MARTON, M. J., WITTEVEEN, A. T., SCHREIBER, G. J., KERKHOVEN, R. M., ROBERTS, C., LINSLEY, P. S., BERNARDS, R. & FRIEND, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature,* 415**,** 530-6.
VAN BEERS, E. H., JOOSSE, S. A., LIGTENBERG, M. J., FLES, R., HOGERVORST, F. B., VERHOEF, S. & NEDERLOF, P. M. (2006) A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer,* 94**,** 333-7.
VAN BEERS, E. H. & NEDERLOF, P. M. (2006) Array-CGH and breast cancer. *Breast Cancer Res,* 8**,** 210.
VAN DE VIJVER, M. J., HE, Y. D., VAN'T VEER, L. J., DAI, H., HART, A. A., VOSKUIL, D. W., SCHREIBER, G. J., PETERSE, J. L., ROBERTS, C.,

MARTON, M. J., PARRISH, M., ATSMA, D., WITTEVEEN, A., GLAS, A., DELAHAYE, L., VAN DER VELDE, T., BARTELINK, H., RODENHUIS, S., RUTGERS, E. T., FRIEND, S. H. & BERNARDS, R. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med,* 347**,** 1999-2009.

VAN DE WIEL, M. A., KIM, K. I., VOSSE, S. J., VAN WIERINGEN, W. N., WILTING, S. M. & YLSTRA, B. (2007) CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics,* 23**,** 892-4.

VAN WIERINGEN, W. N., VAN DE WIEL, M. A. & YLSTRA, B. (2008) Weighted clustering of called array CGH data. *Biostatistics,* 9**,** 484-500.

VENDRELL, J. A., ROBERTSON, K. E., RAVEL, P., BRAY, S. E., BAJARD, A., PURDIE, C. A., NGUYEN, C., HADAD, S. M., BIECHE, I., CHABAUD, S., BACHELOT, T., THOMPSON, A. M. & COHEN, P. A. (2008) A candidate molecular signature associated with tamoxifen failure in primary breast cancer. *Breast Cancer Res,* 10**,** R88.

WA, C. V., DEVRIES, S., CHEN, Y. Y., WALDMAN, F. M. & HWANG, E. S. (2005) Clinical application of array-based comparative genomic hybridization to define the relationship between multiple synchronous tumors. *Mod Pathol,* 18**,** 591-7.

WALTERS S, R. B., WESTLAKE S, COOPER N, RASULO D, HORN O, COLEMEN M (2009) Cancer survival, England, patients diagnosed 2001-2006 and followed up to 2007:one-year and five-year survival for 21 common cancers, by sex and age. *Office for National Statistics website: http://www.statistics.gov.uk/downloads/theme_health/cancer-survival-Eng-2001-2006.pdf*.

WANG, H. Y., CHENG, Z. & MALBON, C. C. (2003) Overexpression of mitogen-activated protein kinase phosphatases MKP1, MKP2 in human breast cancer. *Cancer Lett,* 191**,** 229-37.

WANG, J., ZHOU, J. Y. & WU, G. S. (2007) ERK-dependent MKP-1-mediated cisplatin resistance in human ovarian cancer cells. *Cancer Res,* 67**,** 11933-41.

WANG, Y., KLIJN, J. G., ZHANG, Y., SIEUWERTS, A. M., LOOK, M. P., YANG, F., TALANTOV, D., TIMMERMANS, M., MEIJER-VAN GELDER, M. E., YU, J., JATKOE, T., BERNS, E. M., ATKINS, D. & FOEKENS, J. A. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet,* 365**,** 671-9.

WANG, Z., XU, J., ZHOU, J. Y., LIU, Y. & WU, G. S. (2006) Mitogen-activated protein kinase phosphatase-1 is required for cisplatin resistance. *Cancer Res,* 66**,** 8870-7.

WANG, Z. C., LIN, M., WEI, L. J., LI, C., MIRON, A., LODEIRO, G., HARRIS, L., RAMASWAMY, S., TANENBAUM, D. M., MEYERSON, M., IGLEHART, J. D. & RICHARDSON, A. (2004) Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res,* 64**,** 64-71.

WEISS, M. M., HERMSEN, M. A., MEIJER, G. A., VAN GRIEKEN, N. C., BAAK, J. P., KUIPERS, E. J. & VAN DIEST, P. J. (1999) Comparative genomic hybridisation. *Mol Pathol,* 52**,** 243-51.

WENGER, S. L., SENFT, J. R., SARGENT, L. M., BAMEZAI, R., BAIRWA, N. & GRANT, S. G. (2004) Comparison of established cell lines at different passages by karyotype and comparative genomic hybridization. *Biosci Rep,* 24**,** 631-9.

WESSELS, L. F., VAN WELSEM, T., HART, A. A., VAN'T VEER, L. J., REINDERS, M. J. & NEDERLOF, P. M. (2002) Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. *Cancer Res,* 62**,** 7110-7.

WETZELS, R. H., KUIJPERS, H. J., LANE, E. B., LEIGH, I. M., TROYANOVSKY, S. M., HOLLAND, R., VAN HAELST, U. J. & RAMAEKERS, F. C. (1991) Basal cell-specific and hyperproliferation-related keratins in human breast cancer. *Am J Pathol,* 138**,** 751-63.

WIECHMANN, L., SAMPSON, M., STEMPEL, M., JACKS, L. M., PATIL, S. M., KING, T. & MORROW, M. (2009) Presenting features of breast cancer differ by molecular subtype. *Ann Surg Oncol,* 16**,** 2705-10.

WORBY, C. A. & DIXON, J. E. (2002) Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol,* 3**,** 919-31.

WU, W., PEW, T., ZOU, M., PANG, D. & CONZEN, S. D. (2005) Glucocorticoid receptor-induced MAPK phosphatase-1 (MPK-1) expression inhibits paclitaxel-associated MAPK activation and contributes to breast cancer cell survival. *J Biol Chem,* 280**,** 4117-24.

WU, X., CHEN, H., PARKER, B., RUBIN, E., ZHU, T., LEE, J. S., ARGANI, P. & SUKUMAR, S. (2006) HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition. *Cancer Res,* 66**,** 9527-34.

YAO, J., WEREMOWICZ, S., FENG, B., GENTLEMAN, R. C., MARKS, J. R., GELMAN, R., BRENNAN, C. & POLYAK, K. (2006) Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res,* 66**,** 4065-78.

ZELTSER, L., DESPLAN, C. & HEINTZ, N. (1996) Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development,* 122**,** 2475-84.

ZHANG, D. H., SALTO-TELLEZ, M., CHIU, L. L., SHEN, L. & KOAY, E. S. (2003) Tissue microarray study for classification of breast tumors. *Life Sci,* 73**,** 3189-99.

APPENDICES

APPENDIX 1 Coding System used for the Statistical Analysis of the Breast cancer Database

Variable	**Code Number - meaning**
Patient Sex	0 : Female 1 : Male 9: Default
Menopausal Status	0: Post- menopausal 1: Pre-menopasual 2: Peri-menopausal 3: N/A (male) 9: Default
HRT	0: yes 1: No 7: N/A (male/premenopausal) 9: default
Surgery Side	0: Right 1: Left 7:N/A 9: Default
Surgery Method	0: Mastectomy 1: WLE 2: Bx only 9: Default
Axillary Dissection	0: Axillary Clearance 1: Axillary Sampling 2: N/A or not had 9: Default
Current Status	0: Dead 1: Alive 2: Alive with Recurrence 9: Missing value /Default
Cancer Related Death	0: No 1: Yes 2: Unknown 7: Not applicable (patient alive)
Pathology	0: IDC 1: ILC 2: Mixed 3: Other 9: Default
DCIS	0: Present 1: Absent
Tumor size	In mm
Nodal Status	0: Negative

	1: Positive 9: Unknown
No of positive Nodes	number
No of total nodes	number
ER Status	0: Negative 1: Positive 9: Unknown
PR	0: Negative 1: Positive 9: Unknown
HER-2	0: Negative (1+) 1: Indetermined (2+) 2: Positive (3+) 9: Unknown
Tumour Grade	0: grade 1 1: grade 2 2: grade 3 9: Unknown
T TNM	0: T1 (<2 cms) 1: T2 (2-5 cms) 2: T3 (5.1 cms) 3: T4 7: N/A 9: Unknown
N TNM	0: Negative nodes 1: N1 (1-3 nodes) 2: N2 (4-9 nodes) 3: N3(>10 nodes) 9: Unknown
M TNM	0 : No 1: Yes
Tissue Array	0 : No 1: Yes
RT	0 : Breast/ Chest wall 1: Breast/Chest wall +SCF 2: Breast/ Chest wall + SCF + axilla 3: Not specified 4: No RT 7: N/A 9: Missing
Chemotherapy	0 : No 1: CMF 2: Anthracycline 3: Other 7: N/A 9: Missing
Adjuvant Hormones	0 : No 1: Tamoxifen 2: AI

	3: Other (Tam+Zoladex) 7: N/A 9: Missing
First Relapse	0 : No 1: Loco-regional 2: Bone 3: Visceral/ Brain 4: Other (pelvic, pleural, skin, peritoneal, mediastinum, stomach, orbit) 9: Missing
NPI	0: ≤3.4 1: 3.5 - ≤5.4 2: > 5.4
IHC3	1: ER+ or PR + ; HER-2 negative 2: HER-2 + (regardless ER or PR) 3: Triple negative (ER - ; PR -; HER-2 -)
IHC5 (Triple negative)	0 : (five-negative): ER - ; PR -; HER-2 -; EGFR -; Cyt 5/6 - 1: (basal): triple negative plus either EGFR + or Cyt 5/6 +

APPENDIX 2 – Tissue Microarray Map

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	
A	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			ANIMAL TISSUE	A
B	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	B
C	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	C
D	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	D
E	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	E
F	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	F
G	ANIMAL TISSUE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ANIMAL TISSUE	G
	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE	ANIMAL TISSUE		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		

APPENDIX 3 Genomic DNA isolation from FFPE tissue blocks

HUMAN CANCER STUDIES GROUP

SOP reference MB06

Standard Operating Procedure for

Genomic DNA isolation from FFPE tissue blocks

Version number …………1……………………………………………………

Date created ………24/07/2008

Author:

Name Dr Kristian Unger / Dr Sileida Oliveros

Authorised by:

Name Professor GA Thomas

Summary
This protocol explains the procedure for the isolation of genomic DNA from FFPE tissue blocks either with or without prior microdissection

Safety Considerations

Protective clothing and gloves should be worn throughout.

Materials

- xylene
- methanol
- ethanol (absolute, 96%, 70%)
- NaSCN (Sigma, S-7757)
- Haematoxyline (only for DNA isolation from microdissected samples)
- QIAmp DNA mini kit
- Proteinase K (Qiagen) or from another supplier (20 mg/ml)
- Nuclease-free water
- 0.22 µm syringe filters

Preparation of reagents

NaSCN (Sodium thiocyanate, 1M)

Dissolve 8.1 g NaSCN in 100 ml nuclease-free water. Sterilize the solution by filtering through a 0.22 µm syringe filter.

Proteinase K (if used from other supplier than Quiagen)

Dissolve 100 mg proteinase K in 5 ml nuclease-free water and sterilize the solution by filtering trhough a 0.22 µm syringe filter.

DNA isolation (adopted from VUMC, Microarray Facility, Amsterdam)

Day 1 – Deparaffinization (RecoverAll Ambion protocol)

- Add Xylene 1 ml, vortex briefly to mix and incubate the tube for 10 min at 50℃ in heat-block or water bath to melt the paraffin.
- Centrifuge the sample for 2 min at room temp and maximum speed to pellet the tissue.
- (Optional) If the sample does not form a tight pellet, recentrifuge for an additional 2 min. If a tight pellet still does not form, then proceed with caution to the next step.
- Remove the xylene without disturbing the pellet. Discard the xylene. If the pellet is loose, you may need to leave some xylene in the tube to avoid removing any tissue pieces.
- Add 1 ml Ethanol 100% and vortex to mix. Centrifuge for 2 min at maximum speed to pellet the tissue and remove the supernatant.
- Add 1 ml Ethanol 100% and vortex to mix. Centrifuge for 2 min at maximum speed to pellet the tissue and remove the supernatant.
- Briefly centrifuge again to collect any remaining drops of ethanol in the bottom of the tube. Remove as much residual ethanol as possible without disturbing the pellet.
- Air dry the pellet for 10- 15 min.
- Add 1ml 1 M NaSCN to each micro centrifuge tube and shake by vortexing. Incubate the tube at 38℃ overnight. (MW: 81.06 gr/l; 4.9 gr in 60 mls dH20=60 samples)

Day 2- Protease digestion.

- Spin the tube for 5 min at maximum speed (the tissue will not form a pellet). Remove the supernatant.
- Wash 2x with 1 mL PBS and spin down for 5 min. at full speed, discard the PBS by pipetting
- Add 60 ul ATL buffer (QIAmp DNA micro-kit, Qiagen)
- Add 40 ul prot K (20 mg/mL) and vortex for 15 sec! Make sure all tissue is in the liquid
- Incubate o/n at 55℃ in heat-block or waterbath and vortex regularly!

When the isolation procedure is started in the morning, add another volume of 20 ul prot K. at the end of the day. When started in the afternoon, add another volume of

20 ul prot K in the morning of day 3 and continue in the afternoon. Check if all tissue is digested!

Day 3- DNA isolation

- Incubate at 98°C for 10 min
- Add 100 ul ATL buffer (QIAmp DNA micro-kit, Qiagen)
- Add 200 ul AL buffer (QIAmp DNA micro-kit, Qiagen) and mix very well by vortexing!
- Add 200 ul Ethanol 100% and mix very well by vortexing!
- Incubate 5 min. at RT
- Spin down the eppendorf cups (maximum speed for 3 min).
- Transfer the lysate to the QIAamp MinElute Column *
- The column can contain max. 600 ul of sample. Apply 300 ul of sample per time. Repeat step marked with * until all of the sample is loaded on the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube *
- Place the QI Column in a clean 2 ml collection tube, discard the flow-through *
- Add 500 ul AW1 buffer (QIAmp micro-kit) to the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube
- Add 500 ul AW2 buffer (QIAmp micro-kit) to the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube
- Spin down for 3 min at full speed to dry the membrane
- Place the QI column in a properly marked eppendorf cup and discard collection tube
- Add 20 ul (depending on sample size) of AE buffer (QIAmp micro-kit) to the column and incubate 5 min. at RT
- Spin down for 3 min at 8.000 rpm
- Dispose the column, close the eppendorf cup and add a parafilm and store the DNA at 4°C

APPENDIX 4- Protocol for the preparation of FFPE Cell Pellets.

Dr Sileida Oliveros/ Dr Bill Mathison.

Reagents

-Human plasma (Sigma P 9523) ; reconstitute by adding 1 ml dH20)

-Bovine Thrombin (Sigma T 4648-1) ; 500 U/ml- 1000 U per vial; contains 236 NIH unit/mg protein ; reconstitute by adding 2 mls dH20.

-Trypsin 0.05% in 0.53 mM EDTA (Gibco 25300)

-PBS (Dulbecco`s phosphate buffered saline w/o Sodium Bicarbonate)

-Biopsy capsule

-10% buffered formalin solution

Sterile area

- Proceed with the protocol when a 500 cm2 Triple Flask (Nunc) is 80% confluent.

-Remove the culture medium.

-Wash the cell in warm PBS.

- Add 10 ml Trypsin and place flask in the incubator for approximately 3- 5 minutes, until cells are detached from the flask (assessed under a light microscope).

- Transfer the cell suspension into a 50 ml centrifuge tube, fill with culture medium and centrifuge at 1000 rpm for 5 minutes.

- A visible cell pellet is formed at the bottom of the tube. Discard the supernatant and resuspend the cell pellet with PBS and centrifuge again.Discard the supernatant.

- Resuspend pellet in 2 ml of PBS and transfer to 2 ml Eppendorf tube and centrifuge at 2000 rpm for 15 seconds.

-Remove supernatant carefully.

Non-sterile area

-Resuspend each pellet in. 200 µl plasma

-Quickly add 350 µl of thrombin, mix and let it agglomerate for 5-10 min.

-Gently dislodge the cell pellet into the biopsy capsule, labelled it with pencil and close the capsule carefully.

-Soak the capsule in 10% buffered formalin 12-24 hours.

-Dehydrate the pellet (ascending Ethanol (70, 96, 100 %) series and Xylol and embed in Paraffin (ore use the embedding machine VIP, some machines do the ethanol step automatically).

APPENDIX 5.- Multiplex-PCR check of genomic DNA isolated from FFPE tissue for its usability in array CGH analysis

HUMAN CANCER STUDIES GROUP

SOP reference MB05

Standard Operating Procedure for

Multiplex-PCR check of genomic DNA isolated from FFPE tissue for its usability in array CGH analysis

Version number …………1……………………………………………………

Date created ………22/09/07

Author:

Name Kristian Unger

Authorised by:

Name Professor GA Thomas

Summary
This protocol explains the procedure for the isolation of genomic DNA from FFPE tissue blocks either with or without prior microdissection

Safety Considerations

Protective clothing and gloves should be worn throughout.

Materials

- xylene
- methanol
- ethanol (absolute, 96%, 70%)
- NaSCN (Sigma, S-7757)
- Haematoxyline (only for DNA isolation from microdissected samples)
- QIAmp DNA mini kit
- Proteinase K (Qiagen) or from another supplier (20 mg/ml)
- Nuclease-free water
- 0.22 µm syringe filters

Preparation of reagents

NaSCN (Sodium thiocyanate, 1M)

Dissolve 8.1 g NaSCN in 100 ml nuclease-free water. Sterilize the solution by filtering through a 0.22 µm syringe filter.

Proteinase K (if used from other supplier than Quiagen)

Dissolve 100 mg proteinase K in 5 ml nuclease-free water and sterilize the solution by filtering trhough a 0.22 µm syringe filter.

DNA isolation (adopted from VUMC, Microarray Facility, Amsterdam)

Day 1 – Deparaffinization (RecoverAll Ambion protocol)

- Add Xylene 1 ml, vortex briefly to mix and incubate the tube for 10 min at 50℃ in heat-block or water bath to melt the paraffin.
- Centrifuge the sample for 2 min at room temp and maximum speed to pellet the tissue.
- (Optional) If the sample does not form a tight pellet, recentrifuge for an additional 2 min. If a tight pellet still does not form, then proceed with caution to the next step.
- Remove the xylene without disturbing the pellet. Discard the xylene. If the pellet is loose, you may need to leave some xylene in the tube to avoid removing any tissue pieces.
- Add 1 ml Ethanol 100% and vortex to mix. Centrifuge for 2 min at maximum speed to pellet the tissue and remove the supernatant.
- Add 1 ml Ethanol 100% and vortex to mix. Centrifuge for 2 min at maximum speed to pellet the tissue and remove the supernatant.
- Briefly centrifuge again to collect any remaining drops of ethanol in the bottom of the tube. Remove as much residual ethanol as possible without disturbing the pellet.
- Air dry the pellet for 10- 15 min.
- Add 1ml 1 M NaSCN to each micro centrifuge tube and shake by vortexing. Incubate the tube at 38℃ overnight. (MW: 81.06 gr/l; 4.9 gr in 60 mls dH20=60 samples)

Day 2- Protease digestion.

- Spin the tube for 5 min at maximum speed (the tissue will not form a pellet). Remove the supernatant.
- Wash 2x with 1 mL PBS and spin down for 5 min. at full speed, discard the PBS by pipetting
- Add 60 ul ATL buffer (QIAmp DNA micro-kit, Qiagen)
- Add 40 ul prot K (20 mg/mL) and vortex for 15 sec! Make sure all tissue is in the liquid
- Incubate o/n at 55℃ in heat-block or waterbath and vortex regularly!

When the isolation procedure is started in the morning, add another volume of 20 ul prot K. at the end of the day. When started in the afternoon, add another volume of

20 ul prot K in the morning of day 3 and continue in the afternoon. Check if all tissue is digested!

Day 3- DNA isolation

- Incubate at 98℃ for 10 min
- Add 100 ul ATL buffer (QIAmp DNA micro-kit, Qiagen)
- Add 200 ul AL buffer (QIAmp DNA micro-kit, Qiagen) and mix very well by vortexing!
- Add 200 ul Ethanol 100% and mix very well by vortexing!
- Incubate 5 min. at RT
- Spin down the eppendorf cups (maximum speed for 3 min).
- Transfer the lysate to the QIAamp MinElute Column *
- The column can contain max. 600 ul of sample. Apply 300 ul of sample per time. Repeat step marked with * until all of the sample is loaded on the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube *
- Place the QI Column in a clean 2 ml collection tube, discard the flow-through *
- Add 500 ul AW1 buffer (QIAmp micro-kit) to the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube
- Add 500 ul AW2 buffer (QIAmp micro-kit) to the column
- Spin down for 1 min at 8.000 rpm and place the column in a new tube
- Spin down for 3 min at full speed to dry the membrane
- Place the QI column in a properly marked eppendorf cup and discard collection tube
- Add 20 ul (depending on sample size) of AE buffer (QIAmp micro-kit) to the column and incubate 5 min. at RT
- Spin down for 3 min at 8.000 rpm
- Dispose the column, close the eppendorf cup and add a parafilm and store the DNA at 4℃

APPENDIX 4- Protocol for the preparation of FFPE Cell Pellets.

Dr Sileida Oliveros/ Dr Bill Mathison.

Reagents

-Human plasma (Sigma P 9523) ; reconstitute by adding 1 ml dH20)

-Bovine Thrombin (Sigma T 4648-1) ; 500 U/ml- 1000 U per vial; contains 236 NIH unit/mg protein ; reconstitute by adding 2 mls dH20.

-Trypsin 0.05% in 0.53 mM EDTA (Gibco 25300)

-PBS (Dulbecco`s phosphate buffered saline w/o Sodium Bicarbonate)

-Biopsy capsule

-10% buffered formalin solution

Sterile area

- Proceed with the protocol when a 500 cm2 Triple Flask (Nunc) is 80% confluent.

-Remove the culture medium.

-Wash the cell in warm PBS.

- Add 10 ml Trypsin and place flask in the incubator for approximately 3- 5 minutes, until cells are detached from the flask (assessed under a light microscope).

- Transfer the cell suspension into a 50 ml centrifuge tube, fill with culture medium and centrifuge at 1000 rpm for 5 minutes.

- A visible cell pellet is formed at the bottom of the tube. Discard the supernatant and resuspend the cell pellet with PBS and centrifuge again.Discard the supernatant.

- Resuspend pellet in 2 ml of PBS and transfer to 2 ml Eppendorf tube and centrifuge at 2000 rpm for 15 seconds.

-Remove supernatant carefully.

Non-sterile area

-Resuspend each pellet in. 200 µl plasma

-Quickly add 350 µl of thrombin, mix and let it agglomerate for 5-10 min.

-Gently dislodge the cell pellet into the biopsy capsule, labelled it with pencil and close the capsule carefully.

-Soak the capsule in 10% buffered formalin 12-24 hours.

-Dehydrate the pellet (ascending Ethanol (70, 96, 100 %) series and Xylol and embed in Paraffin (ore use the embedding machine VIP, some machines do the ethanol step automatically).

APPENDIX 5.- Multiplex-PCR check of genomic DNA isolated from FFPE tissue for its usability in array CGH analysis

HUMAN CANCER STUDIES GROUP

SOP reference MB05

Standard Operating Procedure for

Multiplex-PCR check of genomic DNA isolated from FFPE tissue for its usability in array CGH analysis

Version number …………1……………………………………………………

Date created ………22/09/07

Author:

Name Kristian Unger

Authorised by:

Name Professor GA Thomas

Summary

This protocol explains the procedure for testing the usability of genomic DNA in array CGH experiments, following use of SOP number MB01.

Safety Considerations

Protective clothing and gloves should be worn throughout.

A Preparing of PCR Primers

PCR Primers (Table 1) are purchase ordered at Metabion International GmbH in a 0,02 µmol scale, HPLC purified and lyophilized. The oligos are dissolved in nuclease-free water (Sigma-Aldrich) to 100 µM (stock solution) according to suppliers datasheet.

Table 1.: PCR Primers used for Multiplex PCR

	Name	**Sequence**	**Number of bases**
100bp-fragment*			
	100F	5'- gttccaatatgattccaccc -3'	20
	100R	5'- ctcctggaagatggtgatgg -3'	20
200bp-fragment*			
	200F	5'- aggtggagcgaggctagc -3'	18
	200R	5'- ttttgcggtggaaatgtcct -3'	20
300bp-fragment*			
	300F	5'- aggtgagacattcttgctgg -3'	20
	300R	5'- tccactaaccagtcagcgtc -3'	20
400bp-fragment*			
	400F	5'- acagtccatgccatcactgc -3'	20
	400R	5'- gcttgacaaagtggtcgttg -3'	20

* amplification products of non-overlapping fragments within human GDP-gene (chromosome 12), van Beers et al, BrJCanc. 2006, 30;94(2):333-7

For the preparation of a ready-to-use equimolar primer-solution, mix 10 µl of each of the 100 µmol primer stock-solutions and add 720 µl nuclease-free water (800 µl in total; 1,25 µM).

B Preparation of the PCR-Reaction

Although a hotstart Taq-polymerase is used, please prepare the reaction on ice!

Used Materials:

FastStart High Fidelity Enzyme (5U/ µl)*
GeneAmp 10X PCR Buffer*
25 mM $MgCl_2$ Solution*
100 mM dNTP Mix (Applied Biosystems, Cat#: N8080007)
Nuclease-free water (Sigma, Cat#: W4502)
Human genomic DNA (Promega, male, Cat#: G1471 or female, Cat#: G1521)
* part of FastStart High Fidelity PCR System, Roche. Catalog No 03 553 426 001)
100 bp Ladder (Invitrogen)

Pipetting scheme: 1x Reaction

GeneAmp 10X PCR Buffer	3.0 µl
MgCl2 (25 mM)	1.8 µl
dNTP-mix (10 mM each)	0.6 µl
Primer-Mix (1.25 µM each)	3.2 µl
Taq (5 U/ul)	0.2 µl
total	8.8 µl

100ng genomic DNA to be tested is diluted with nuclease-free water to a volume of 21.2µl and mixed with 8.8µl master-mix.

For a positive control use male or female genomic DNA and for negative control use nuclease-free water.

Place PCR reaction tubes in the thermocycler and run the following PCR program:

Step 1:96℃ 9’
Step 2:94℃ 1’
Step 3:56℃ 1’

Step 4:72℃ 3'

Step 5: Goto step 2 for additional 34 cycles

Step 6:72℃ 7'

Step 7: 4℃ for ever

C Electrophoretic seperation of PCR amplification products

Used Materials:

10X Bluejuice gel loading buffer (Invitrogen, Cat#: 10816015)
GeneRuler 100 bp DNA length standard (Fermentas, Cat#: SM1143)
Agarose for molecular biology (Sigma-Aldrich, Cat#: A9539)
SYBR Safe DNA gel stain 10000x concentrate (Invitrogen, Cat#: S33102)
TAE buffer X50 (National Diagnostics, Cat#: EC-872)

Make X1 TAE buffer with SYBR Safe for gel and to fill tank:

X50 TAE buffer	20 ml
dH2O	1000 ml

Use 500 mls 1XTAE buffer and add 50 µl SYBR Safe DNA gel stain to fill the tank. Place 2 g of Agarose into a small conical flask and add 100ml of X1 TAE buffer containing 10 µl SYBR Safe. Heat in the microwave until the agarose has fully dissolved (take care not to let the agarose boil over). Run cold water over the conical flask whilst swirling the gel to cool the gel down. The agarose should be approximately 50° C before pouring (check this by placing the flask on the palm of your hand, if it is uncomfortable then it needs to be cooled more). Pour the cooled agarose into the gel tank apparatus and allow to set.

Once set, remove the combs and add enough TAE buffer to cover the gel.

Mix 9 ul of DNA with 1 ul of 10X loading buffer and load the samples into the gel. Load 10 ul of 100 bp ladder into the gel.

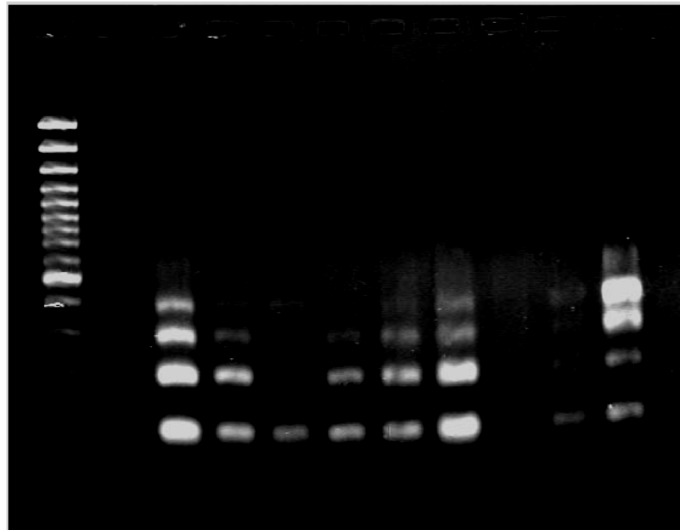
Run the gel at 150V for 30 minutes.

D Interpretation of Results

Place gel in a gel documentation system and make a picture. The number of visible bands represents integrity of the genomic DNA tested. The lane with the positive control should show 4 distinct bands (100 bp, 200 bp, 300 bp and 400 bp in length). No bands should be visible in the lane with the negative control.

DNA samples showing (Fig.1) at least 3 bands (100 bp, 200 bp and 300 bp) can be used for array CGH analysis. Samples showing bands smaller than 300 bp are not used for array CGH analysis.

Fig. 1: Example for gel electrophoresis of multiplex PCR products



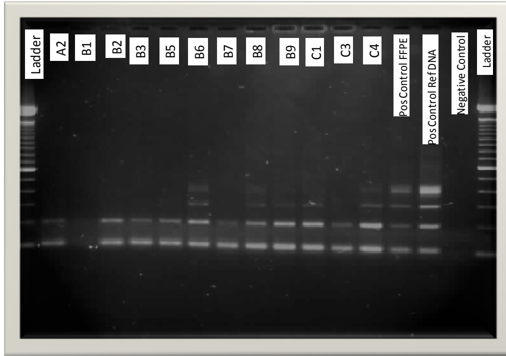
Lane1: 100 bp DNA standard; lane 2:negative control - no bands, negative ; lane 3: 4 bands, positive; lane 4: 3 bands, positive; lane 5: 1 band, negative; lane 6: 3 bands, positive; lane 7: 3 bands, positive; lane 8: 4 bands, positive; lane 9: no bands, negative; lane 10: 1 band, negative; lane 11:positive control - 4 bands, positive;

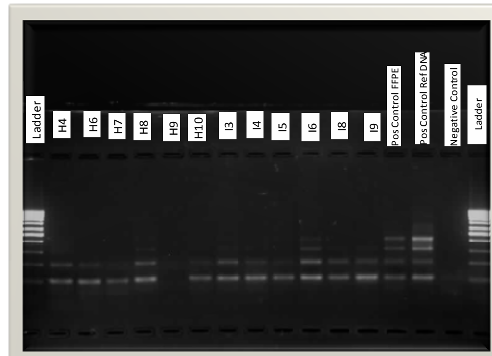
APPENDIX 6.- Summary of the DNA concentration

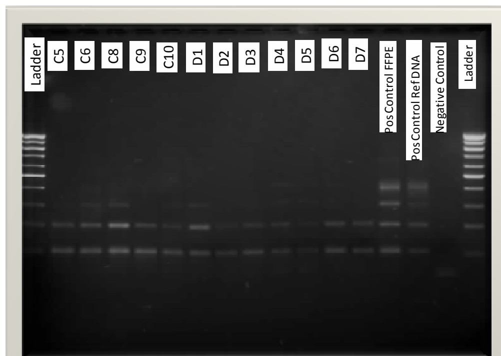
Appendix 6- Summary of DNA extraction from FFPE tissue

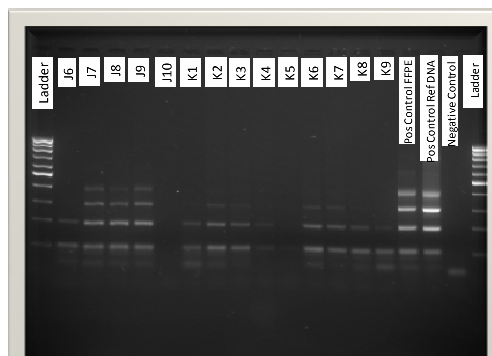
Sample	Concent (ng/ul)	260/280	260/230	Mult_ PCR Gel	Sample	Concent (ng/ul)	260/280	260/230	Mult_ PCR Gel
A2	191.27	1.94	1.31	2 bands	H6	69.53	1.69	1.61	2 bands
B1	30.01	1.13	1.26	No bands	H7	48.32	1.96	1.24	1 band
B2	195.98	1.93	1.36	2 bands	H8	88.37	2.01	0.55	3 bands
B3	183.88	1.97	1.73	2 bands	H9	4.81	2.05	0.21	No bands
B5	315.9	1.95	2.05	2 bands	H10	119.46	1.93	1.5	2 bands
B6	226.34	1.94	1.95	4 bands	I3	83.82	1.95	1.17	2 bands
B7	324.29	1.9	1.95	2 bands	I4	160.86	1.98	1.8	2 bands
B8	454.55	1.91	2.01	4 bands	I5	30.03	2.01	0.48	1 band
B9	433.2	1.93	2.1	4 bands	I6	324.19	1.96	2.1	4 bands
C1	213.34	1.96	2.04	4 bands	I8	42.7	2.03	1.29	3 bands
C3	557.3	1.93	2.17	2 bands	I9	84.25	1.95	1.61	3 bands
C4	659.28	1.92	0.35	4 bands	J1	14.46	1.94	0.46	1 band
C5	139.8	1.91	1.86	2 bands	J2	20.44	2.01	0.44	1 band
C6	443.72	1.89	1.94	3 bands	J4	32.46	1.94	1.39	2 bands
C8	1191.56	1.99	2.01	3 bands	J5	83.95	2.03	1.42	1 band
C9	250.49	1.96	1.98	2 bands	J6	26.53	1.79	1.5	2 bands
C10	173.76	1.96	1.27	2 bands	J7	93.35	1.96	1.07	4 bands
D1	465.13	1.9	1.64	3 bands	J8	32.24	1.96	0.75	4 bands
D2	170.69	1.91	1.47	2 bands	J9	148.02	1.99	1.88	4 bands
D3	101.34	1.93	1.51	2 bands	J10	7.17	1.47	0.42	No bands
D4	162.4	1.97	1.61	2 bands	K1	41.51	1.96	1.61	2 bands
D5	384.84	1.9	2.06	2 bands	K2	40.29	1.79	1.55	3 bands
D6	143.3	1.92	1.65	2 bands	K3	27.25	1.77	1.07	3 bands
D7	36.35	1.85	1.11	2 bands	K4	13.6	1.63	0.74	1 band
D8	490.73	1.93	2.11	4 bands	K5	13.01	1.71	0.67	No bands
D9	179.74	1.9	1.98	4 bands	K6	129.06	1.84	1.91	3 bands
E1	128.09	1.92	1.74	2 bands	K7	22.16	1.64	1.09	3 bands
E2	84.1	1.89	1.85	2 bands	K8	34.78	1.89	1.5	2 bands
E3	63.76	1.94	1.45	2 bands	K9	194.53	1.94	1.9	2 bands
E4	200.33	1.95	1.82	2 bands	J10-2	50.87	1.85	1.31	3 bands
E8	15.65	1.69	0.83	1 band	K4-2	30.76	1.94	1.09	No bands
F1	94.94	2	1.89	4 bands	K5-2	39.35	1.97	1.05	1 band
F5	68.55	1.93	1.35	3 bands	K10	121.56	1.98	1.44	3 bands
F6	161.02	1.96	1.68	4 bands	L1	62.53	1.88	1.44	2 bands
G1	91.08	1.96	1.6	4 bands	L2	2.81	3.55	0.22	No band
G3	42.86	2.1	1.56	4 bands	L3	13.4	2.07	0.53	No band
H4	310.57	1.96	2.05	2 bands					

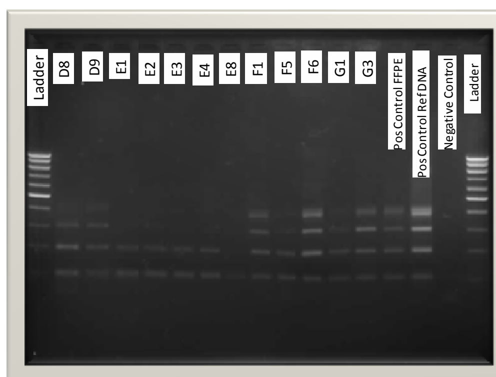
APPENDIX 7.- Agarose gel image after Multiplex PCR of tumour samples.

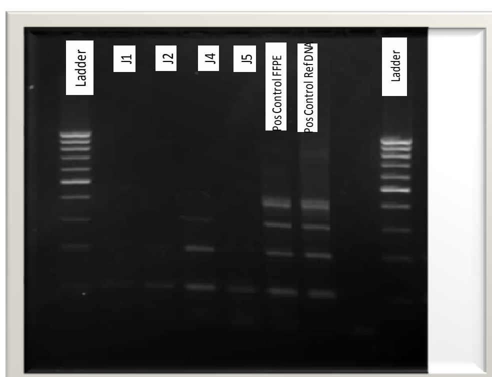












APPENDIX 8.- GE Healthcare. Amersham.

Cy^{TM}3PA 53021:

http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/24D727B03B42E720C125762 8001CDC58/$file/PA53021PL_Rev_B_2007_web.pdf.

Cy^{TM}5 PA 55021
http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/B3D24C67751DA666C125762 8001CDC65/$file/PA55021PL_Rev_B_2006_WEB.pdf

APPENDIX 9.- Random Primer Labelling

GE healthcare protocol

http://www5.gelifesciences.com/aptrix/upp00919.nsf/content/9FF6770A0EF6C9C0C1257628001D372A?OpenDocument&Path=Catalog&Hometitle=Catalog&entry=1&newrel&LinkParent=C1256FC4003AED40-BCB6D35E387DB352C1257322003C98A5_RelatedLinksNew-C821BEC677D8448BC1256EAE002E3030&newrel&hidesearchbox=yes&moduleid=167211

HUMAN CANCER STUDIES GROUP

SOP reference MB08

Standard Operating Procedure for

Labelling of genomic DNA using Random Priming with Cy3 and Cy5 labelled dCTP's

Version number …………1

Date created ………11.08.08

Date of review ……11.08.08

Author:

Name Dr Kristian Unger

Authorised by:

Name Professor GA Thomas

Summary

This protocol decribes labelling of genomic DNA in order to use it in BAC array CGH hybridization (1Mb BAC array/Agilent SureHyb 2 array chamber/240 ul reaction volume).

Safety Considerations

Protective clothing and gloves should be worn throughout.

A Equipment and reagents

Amersham CyScribe Array CGH Genomic DNA Labelling Kit (GE Healthcare 28909726)
Randon Nonamers (part of kit)
Reaction buffer (part of kit)
Nuclease Free water (part of kit)
Cy3/Cy5-dCTP (part of kit)
dCTP labelling mix (part of kit)
Klenow (part of kit)
Microcon Columns, YM-30 (Milipore)
Low-EDTA TE buffer (TEKnova)

B Method for Labelling gDNA

1.- Add the following reaction components sequentially to a 1.5 ml tube on ice:

Components	**reference**	**tumour**
Genomic DNA	525 ng	525 ng
Random Nonamers	35 ul	35 ul
Reaction Buffer	35 ul	35 ul
Nuclease free water	X ul	X ul
Total Volume	133	133

2.- Mix the reaction components by pipeting gently

3.- Incubate the reaction mixture at 95 ℃ for 5 minutes in a heat block followed by ice cooling for 5 minutes.

4.- Centrifuge the tubes briefly to collect all reactions components at the bottom of the tube and place on ice. Add the following components:

Components	reference	tumour
dCTP-labelling Mix	28 ul	28 ul
Labelled-dCTP	Cy5: 10.5 ul	Cy3: 10.5 ul
Klenow (ExoFree)	3.5 ul	3.5 ul
Total Volume	175 ul	175 ul

5.- Mix reactions by pipetting gently and spin briefly to collect the components at the bottom of the tube.

6.- Incubate the reaction at 37°C overnight in a water bath.

7.- For purification of the labelled genomic DNA add 175 ul TE buffer to each tube (total volume 480ul).

8.- Place the Microcon YM-30 filter into 1.5 ml microcentrifuge tube and load each labelled gDNA into the filter. Spin 10 minutes at 8000 x **g** in a microcentrifuge at room temperature. Discard the flow-through.

9.- Add 480 ul of TE to each filter. Spin for 10 min at 8000 x g in a microcentrifuge at room temperature. Discard the flow-through. Repeat this step twice.

10. Invert the filter into a fresh 1.5 ml microcentrifuge tube. Spin for 1 minute at 8000 x g in a microcentrifuge at room temperature to collect the purified sample.

11.- Measure and record volume (ul) for each eluate. If sample volume exceeds 27 ul, return sample to its filter and spin 1 minute 1 minute at 8000 x g in a microcentrifuge at room temperature. Discard the flow-through.

12.- Repeat steps 10 and 11 until sample volume ≤ 27 ul. Bring total volume up to 27 ul by adding 1 X TE if required.

13.- Determine the yield and specific activity by using Nanodrop Spectrophotometer (use TE for reference measurement).

14.- Combine the appropriate Cy3 and Cy5 samples for a total mixture volume of 54 ul in a new 1.5 ul microfuge tube. Labelled gDNA can be stored for weeks at -20 °C.

APPENDIX 10- Summary of labelling results in BC tumour samples

Summary of Labeling for tumour samples using GE (3.5X reaction) Cy3									
Sample	**ID**	**Dye1 pmol/ul**	**ng/ul**	**260/280**	**Sample**	**ID**	**Dye1 pmol/ul**	**ng/ul**	**260/280**
A2	716	32.7	376.42	1.75	G1	906	38.14	259.38	1.71
B2	131	30.39	375.65	1.74	G3	1202	38.9	336.23	1.76
B3	119	41.88	392.71	1.75	H4	1254	41.92	355.16	1.77
B5	1120	30.49	328.27	1.78	H6	408	31.26	365.73	1.8
B6	604	67.43	455.25	1.74	H8	415	42.87	328.77	1.76
B8	128	62.35	467.82	1.74	H10	323	20.17	186.68	1.77
B9	594	41.96	429.9	1.79	I3	266	71.89	608.15	1.7
C1	144	45.68	401.54	1.78	I4	577	35.17	343.55	1.77
C3	179	33.24	324.72	1.82	I5	552	39.52	347.72	1.79
C4	178	33.33	256.67	1.75	I6	550	45.06	558.8	1.78
C5	122	64.2	557.42	1.71	I8	105	22.98	328.54	1.82
C6	184	48.66	446.2	1.76	I9	219	25.66	432.11	1.82
C8	82	40.75	410.09	1.75	J4	255	23.96	319.63	1.77
C9	1522	20.79	277.69	1.8	J5	1638	31.94	544.3	1.8
C10	81	35.67	386.84	1.81	J6	524	45.28	625.75	1.72
D1	1280	41.27	347.87	1.75	J7	850	70.17	861.27	1.71
D2	1376	23.76	250.41	1.77	J8	625	50.65	653.32	1.75
D3	1416	31.35	358.58	1.8	J9	866	52.07	732.91	1.75
D4	90	29.59	323.97	1.65	J10-2	1721	55.1	557.27	1.73
D5	1512	52.99	483.26	1.77	K1	1267	12.91	377.89	1.82
D6	1553	65.83	472.66	1.65	K2	413	52.67	680.41	1.73
D7	1689	25.37	244.92	1.77	K3	419	43.43	562.63	1.75
D8	59	48.49	425.41	1.76	K5-2	1016	48.36	690.48	1.74
D9	1543	41.8	428.94	1.61	K6-2	1372	63.51	714.48	1.72
E1	1452	54.3	513.9	1.73	K7	1423	50.42	601.34	1.74
E2	1566	26.19	283.03	1.82	K8	1431	33.21	524.89	1.77
E3	254	31.56	280.35	1.76	K9	1479	11.91	331.48	1.81
E4	1653	44.43	381.22	1.7	K10	1177	63.84	752.34	1.73
F1	736	51.85	308.04	1.68	L1	439	58.46	730.7	1.72
F5	793	27.87	253.09	1.76					

APPENDIX 11.- Preparation of Labelled Genomic DNA for Array

(Adapted from Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Protocol Version 4.0, June 2006)

Preparation of 10X Blocking Agent:

-Add 1350 µL of nuclease-free water to the vial containing lyophilized 10X Blocking Agent (supplied with Agilent Oligo aCGH Hybridization Kit).

-Leave at room temperature for 60 minutes and mix on a vortex mixer to reconstitute sample before use or storage. The 10X Blocking Agent can be stored at -20°C.

Preparation of the samples prior hybridization:

-Equilibrate water baths or heat blocks to 95°C and 37°C.

-Add the components in Table 3.5 in the order indicated in a nuclease-free tube.

-Mix the sample by pipetting up and down, then quickly spin in a microcentrifuge to drive contents to the bottom of the reaction tube.

-Transfer sample tubes to a circulating water bath or heat block at 95°C. Incubate at 95°C for 3 minutes.

Preparation of Labelled gDNA prior hybridisation

Component	**Agilent Oligonucleotide Protocol**	**Our adapted protocol***
	Volume (µl) per hybridisation	
Cy3 and Cy5 –labeled gDNA mixture	79	54
Cot1 DNA (1.0 mg/µl)	25	50
Agilent 10X Blocking agent	26	26
Agilent 2X Hybridisation buffer	130	130
Final hybridisation volume	260	260

*Adapted from Agilent Oligonucleotide Array-based CGH for Genomic DNA analysis. Version 4.0, June 2006.

-Immediately transfer sample tubes to a circulating water bath or heat block at 37°C. Incubate at 37°C for 30 minutes.

-Remove sample tubes from the water bath or heat block. Spin 1 minute at 17,900 × g in a microcentrifuge to collect the sample at the bottom of the tube.

-The hybridization sample mixture is applied directly onto the gasket well in a “drag and dispense” manner.

-Then the active side of the microarray slide is placed on top of the gasket slide to form a "sandwich slide pair".

Washing Procedure

-Always use fresh Oligo aCGH Wash Buffer 1 and Oligo aCGH Wash Buffer 2 for each wash group (up to five slides).

Table 1- Wash conditions from the Wash Procedure A.

	Dish	Wash buffer	Temperature	Time
Disassembly	#1	Oligo aCGH wash. Buffer 1	Room temperature	
1st wash	#2	Oligo aCGH wash. Buffer 1	Room temperature	5 minutes
2nd wash	#3	Oligo aCGH wash. Buffer 2	37 °C	1 minute

-Completely fill slide-staining dish #1 with Oligo aCGH Wash Buffer 1 at room temperature.

-Place a slide rack into slide-staining dish #2. Add a magnetic stir bar. Fill slide-staining dish #2 with enough Oligo aCGH Wash Buffer 1 at room temperature to cover the slide rack. Place this dish on a magnetic stir plate.

-Put the prewarmed 500 ml glass dish filled with water and containing slide-staining dish #3 on a magnetic stir plate with heating element. Fill the slide-staining dish #3 approximately three-fourths full with Oligo aCGH Wash Buffer 2 (warmed to 37℃). Add a magnetic stir bar. Turn on the heating element and maintain temperature of Oligo aCGH Wash Buffer 2 at 37℃; monitor using a digital thermometer.

-Remove one hybridization chamber from incubator and record time. Record whether bubbles formed during hybridization and if all bubbles are rotating freely: place the hybridization chamber assembly on a flat surface and loosen the thumbscrew, turning counter-clockwise, slide off the clamp assembly and remove the chamber cover, with gloved fingers, remove the array-gasket sandwich from the chamber base by grabbing the slides from their ends and quickly transfer the sandwich slide to dish #1, Without letting go of the slides, submerge the array-gasket sandwich into slide-staining dish #1 containing Oligo aCGH Wash Buffer 1.

-With the sandwich completely submerged in Oligo aCGH Wash Buffer 1, pry

the sandwich open from the barcode end only. Do this by slipping one of the blunt ends of the forceps between the slides and then gently turn the forceps upwards or downwards to separate the slides. Let the gasket slide drop to the bottom of the staining dish. Remove the microarray slide and place into slide rack in the slide-staining dish #2 containing Oligo aCGH Wash Buffer 1 at room temperature. Minimize exposure of the slide to air. *Touch only the barcode portion of the microarray slide or its edges!*
-Repeat step 4 through step 6 for up to four additional slides in the group. A maximum of five disassembly procedures yielding five microarray slides is advised at one time in order to facilitate uniform washing.
-When all slides in the group are placed into the slide rack in slide-staining dish #2, stir using setting 4 for 5 minutes. Adjust the setting to get good but not vigorous mixing.
-Transfer slide rack to slide-staining dish #3 containing Oligo aCGH Wash Buffer 2 at 37℃, and stir using setting 4 for 1 minute.
-Slowly remove the slide rack trying to minimize droplets on the slides. It should take 5 to 10 seconds to remove the slide rack.
-Discard used Oligo aCGH Wash Buffer 1 and Oligo aCGH Wash Buffer 2.
-Repeat step 1 through step 11 for the next group of five slides using fresh Oligo aCGH Wash Buffer 1 and Oligo aCGH Wash Buffer 2 pre-warmed to 37℃.