

Breast cancer prognostic classifiers: combining clinical with molecular profiles

Abir A. Abdelhadi Muftah

MBChB, MSc

This thesis has been submitted to fulfil the requirement of Doctorate of Philosophy from the University of Nottingham

September 2017

Department of Pathology, School of Medicine, University of Nottingham

Nottingham, UK

Abstract

Background and aims

While current prognostic tools for breast cancer (BC) provide valuable information on behaviour and outcome, there are growing concerns that these parameters are not sufficient to reflect the degree of heterogeneity or to guide management decisions at individual patient level. Therefore, further refinement of the existing prognostic tools is needed. The advent of multi-parameter gene signatures has increased our hope of refining BC prognostic classification; however, their cost and restricted application to certain subgroups of BC limit their clinical usefulness. Molecular taxonomy of BC using intrinsic gene sets has not only improved our understanding of BC biology, but has also provided important prognostic information. Yet, integration of this molecular classification with the clinical parameters remains a challenging task.

Our group has recently developed a prognostic tool that incorporates the molecular features of BC with the well-established prognostic morphological variables; the Nottingham Prognostic Index Plus (NPI+) that aimed at overcoming the limitations of using different molecular and clinicopathological prognostic parameters separately. However, NPI+ currently has limitations and needs further refinement to be applicable to BC management in routine practice. Therefore, this study aimed to investigate some relevant potential prognostic markers that can improve BC prognostic classification, in the context of combined molecular and morphological prognostic BC taxonomy, and to further refine the NPI+ to improve its prognostic value, while addressing some issues related to its components and performance. The study included four main objectives. The first was the integration of the proliferation biomarker Ki67 and addressing some technical issues related to its prognostic value and application in routine practice. The second was the evaluation of a high-throughput proteomic technique, reverse phase protein array (RPPA), for its potential use as a single-step quantification method of multiple proteins. The third aim proposed to determine the biological relevance of BC expressing low levels of oestrogen receptor (ER) using techniques at both transcriptome and protein levels. The

fourth aim was to investigate the incorporation of a novel cancer stem cell (CSC) biomarker as a potential prognostic variable that can refine BC classification.

Methodology

In this study, five large primary invasive BC cohorts were investigated at both transcriptome and protein levels. Tissue microarray (TMA) and whole tissue section (WTS) were used. Molecular techniques used in this study included immunohistochemistry (IHC), western blot (WB), different protein and RNA extraction techniques, laser capture microdissection (LCM), RPPA, real time-polymerase chain reaction (RT-PCR) and RNAscope.

Results

Although Ki67 expression can be examined using both WTS and TMA, the assessment of Ki67 in whole sections is preferred and using multiple or larger TMA cores has to be explored as an alternative to WTS. When assessing Ki67 in TMAs in BC, a cut-point of 20% appears to be optimum in concordance with WTS and patients outcome. However, our results support the use of Ki67 as a continuous variable, particularly in the stratification of patients into prognostic groups, either using TMA or WTS assessment (Muftah et al., 2017). Using the MIB-1 clone with different optimisation conditions is associated with cytoplasmic/membranous reactivity. In this regard, it is recommended that different anti-Ki67 clones could be used for clearer staining. The results show that Ki67 can successfully replace mitotic frequency in the updated prognostic index, NPI+.

Unlocking FFPE tissue lysates utilising RPPA is a reliable method for protein quantification. Data produced by this high-throughput technique could be used in concurrent analyses of protein profiles in a large number of clinical cases. Accordingly, RPPA could be consistently used in molecular classification of BC, such as the NPI+ (Negm et al., 2016). Adding Ki67 to the cluster using the RPPA technique improved molecular classification with reassignment of 16% of the unclassified patients.

Investigating ER in BC at both levels (transcriptome and protein) shows that its expression is essentially bimodal (Muftah et al., 2016). Additionally, our results question the advantage of hormonal therapy in the low ER (<10%) subgroup;

where, in the study cohort, nearly half (42.2%) of the Core Needle Biopsy (CNB) cases showed WTS negative. There is strong agreement between the IHC and *in situ* RNAscope results, particularly in focal heterogeneous ER staining areas. This study provides scientific evidence that the actual ER cut-off seems to be 10%. Furthermore, this research supports the clinical importance of B-cell specific Moloney leukaemia virus insertion site-1 (Bmi-1) as a favourable prognostic biomarker in BC and its ability to refine the CD44/CD24 phenotypes as well as slow proliferating tumours into prognostically relevant subgroups.

Conclusion

This study presented data enabling the updating and evaluation of existing prognostic parameters and indices using promising biomarkers and high-throughput techniques, while combining molecular and clinical variables to stratify BC patients into relevant prognostic subgroups. Further investigation of the potential and refinement of the existing BC prognostic parameters is needed in order to allow more precise BC classifications that can predict patient outcomes and potential response to therapy.

Publications

A: Book Chapter

1. **Molecular-Based Diagnostic, Prognostic and Predictive Tests in Breast Cancer**
   Abir A. Muftah, Mohammed A. Aleskandarany, Ian O. Ellis and Emad A Rakha. *Molecular Pathology Library*. 2015

B: Articles in Peer-reviewed journals

1. **Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections.**
   Muftah, A. A., Aleskandarany, M. A., Al-Kaabi, M. M., Sonbul, S. N., Diez-Rodriguez, M., Nolan, C. C., Caldas, C., Ellis, I. O., Rakha, E. A. & Green, A. R. *Breast Cancer Res Treat.*2017

2. **Further Evidence to Support Bimodality of Oestrogen Receptor Expression in Breast Cancer.**
   Muftah, A. A., Aleskandarany, M., Sonbul, S. N., Nolan, C. C., Diez Rodriguez, M., Caldas, C., Ellis, I. O., Green, A. R. & Rakha, E. A. *Histopathology.* 2016

3. **Clinical utility of reverse phase protein array for molecular classification of breast cancer**
   Ola Negm, **Abir Muftah** (**First joint author**), Mohammed Aleskandarany, Mohamed Hamed, Dena Jerjees, Christopher Nolan, Maria Diez-Rodriguez, Patrick Tighe, Ian Ellis, Emad Rakha and Andrew Green. *Breast Cancer Research and Treatment.*2016

4. **Bimodality of intratumor Ki67 expression is an independent prognostic factor of overall survival in patients with invasive breast carcinoma.**
   Arvydas Laurinavicius, Benoit Plancoulaine, Allan Rasmusson, Justinas Besusparis, Renaldas Augulis, Raimundas Meskauskas, Paulette Herlin, Aida Laurinaviciene, **Abir A. Abdelhadi Muftah**, Islam Miligy, Mohammed Aleskandarany, Emad A. Rakha, Andrew R. Green, Ian O. Ellis, *Virchows Arch*. 2016

5. **Prognostic and biological significance of peroxisome proliferator-activated receptor-gamma in luminal breast cancer**
   Rezvan Abduljabbar, Methaq Mueen Al-Kaabi, Ola H. Negm, Dena Jerjees, **Abir A. Muftah,** Abhik Mukherjee, Chun F. Lai, Laki Buluwela, Simak Ali, Patrick J. Tighe, Andrew Green, Ian Ellis, Emad Rakha. *Breast Cancer Research and Treatment*. 2015

6. **Checkpoint kinase1 (CHK1) is an important biomarker in breast cancer having a role in chemotherapy response**
   Meethaq Al-kaabi, Alaa Alshareeda, Dena Jerjees, **Abir Muftah,** Andy Green, Nouf Alsubhi, Chris Nolan, Stephen Chan, E Cornford, Srinivasan Madhusudan, Ian Ellis and Emad Rakha. *British Journal of Cancer*. 2015

7. **The mammalian target of rapamycin complex 1 (mTORC1) in breast cancer: the impact of oestrogen receptor and HER2 pathways.**
   Dena Jerjees, Ola Negm, M. Layth Alabdullah, Sameer Mirza, Methaq Alkaabi, Mohamed Hamed, Rezvan Abduljabbar, **Abir Muftah,** Chris Nolan, Andrew Green, Patrick Tighe, Vimla Band, Ian Ellis and Emad Rakha. *Breast Cancer Res Treat*. 2015

8. **ERK1/2 is related to oestrogen receptor and predicts outcome in hormone-treated breast cancer**
   Dena Jerjees, M. Alabdullah, Methaq Alkaabi, Rezvan Abduljabbar, **Abir Muftah,** Chris Nolan, Andrew Green, Ian Ellis, Emad Rakha. *Breast Cancer Res Treat*. 2014

Conference Communications

A: Poster presentations

1. **The proliferation marker Ki67 in breast cancer: a reliability study to assess its readiness for routine clinical management,** 26th European Congress of Pathology, London, (Aug 2014)
   Abir A. Muftah, Methaq M. Al-Kaabi, Mohammed A. Aleskandarany, Dena A. Jerjees, Nouf H. Alsubhi, , Rezvan F. Abduljabbar, Christopher C. Nolan, Andrew R. Green, Ian O. Ellis, Emad A.Rakha

2. **Successful Alignment of dissected FFPE Tissue to Nottingham Prognostic Index Plus (NPI+) Classes Using RPPA Technology,** Liverpool (NCRI November, 2015)
   Abir Muftah, Ola Nagem, Mohammed Aleskandarany, Christopher Nolan, Patrick Tighe, Andrew Green, Emad Rakha, Ian Ellis

3. **Successful Alignment of Macrodissected FFPE Tissue to Nottingham Prognostic Index Plus (NPI plus) Classes Using RPPA Technology**, 8th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, Dublin (June, 2015)
   Abir A Muftah, Ola Nagem, Mohammed Aleskandarany, Methaq Al-Kaabi, Dena Jerjees, Ian Ellis, Emad Rakha, Andy Green.

4. **Reverse Phase Protein Array is a Useful High Throughput Technique for Assessment of Multiple Proteins in Breast Cancer** 8th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, Dublin (June, 2015)
   Dena Jerjees, Ola Negm, M. Alabdullah, Methaq Alkaabi, Sameer Mirza, Rezvan Abduljabbar, **Abir Muftah** , Andrew Green, Patrick Tighe, Vimla Band, Ian Ellis and Emad Rakha

5. **Improvement of Ki67 assessment in breast cancer,** 8th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, Dublin (June, 2015)
   Abir A Muftah, Mohammed Aleskandarany, Methaq Al-Kaabi, Dena Jerjees, Rezvan Abduljabbar, Nouf Alsubhi, Ian Ellis, Emad Rakha, Andy Green.

6. **High-throughput proteomic study for FFPE tissue using RPPA**, 4th Global Reverse Phase Protein Array Workshop, Paris (October, 2014)
   Ola H Nagem, **Abir A. Muftah,** Andrew R. Green, Chris Nolan, Patrick J. Tighe, Ian O. Ellis, Emad A. Rakha

7. **Activated c-Jun is a luminal-related marker associated with better outcome in breast cancer,** 26th European Congress of Pathology, London, (Aug 2014).
   Dena Jerjees, Rezvan Abduljabbar, **Abir Muftah**, Christopher Nolan, Andrew Green, Ian Ellis, Emad Rakha

8. **Checkpoint Kinase 1 (CHK1) is an important biomarker in breast cancer related to DNAdamage repair mechanism and response to chemotherapy,** 26th European Congress of Pathology, London, (Aug 2014).
   Methaq Al-Kaabi, A. Alashareeda, Dena Jerjees, **Abir Muftah,** Nouf Alsubhi, Andrew Green, S. Chan, E. Cornford, S. Madhusudan, Ian Ellis, Emad Rakha

9. **Deciphering the role of PTEN and its role in the DNA damage response in breast cancer,** 26th European Congress of Pathology, London, (Aug 2014).
   Nouf Alsubhi, W. Flesher, A. Alshareeda, **Abir Muftah,** Dena Jerjees, Rezvan Abduljabbar, Methaq Al-Kaabi, Andrew Green, S. Madhusudan, Ian Ellis, Emad Rakha

B: Oral presentations

2. **Unlocking the pathological achieves: High-throughput proteomic study for FFPE tissue using RPPA, 26th European Congress of Pathology, London, (Aug 2014)**
   Ola H Negm, **Abir A Muftah**, Andrew R Green, Chris Nolan2, Patrick J. Tighe, Ian O. Ellis, Emad A. Rakha

3. **Further Evidence to Support That Oestrogen Receptor Expression in Breast Cancer Is Bimodal. 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, Nottingham (Jun 2016)**
   Abir A Muftah, Mohammed Aleskandarany, Sultan Sonbul, Christopher Nolan, Maria Diez Rodriguez, Carlos Caldas, Ian Ellis, Andy Green, Emad Rakha.

Acknowledgements

First and foremost, I would like to thank Allah for giving me the opportunity, strength and ability to finish this project and gain this achievement.

I would like to express my sincere gratitude to my supervisors, Prof Ian Ellis, Prof Emad Rakha and Dr Andrew Green for their valuable support and continuous encouragement. Their guidance and advice helped me during the time of research, writing of this thesis, publishing scientific papers and sharing at international conferences. Besides my advisor, I have great pleasure in acknowledging Dr Mohammed Aleskandarany and my colleagues Dr Ahmed Benhasouna, Dr Dena Ahmad, Dr Rezvan Abduljabbar Dr Meethaq Alkaabi and Sultan Sonbul for their insightful comments, encouragement and continuous support. I am grateful to all my collaborators at the Queen's Medical Centre, Nottingham City Hospital, the Computer Science department at the University of Nottingham, Nottingham Trent University and ACD Bio Group in the US, who are mentioned in this thesis and helped in this project or who gave access to their laboratories and research facilities. I express my apology that I could not mention them personally one by one. A very special gratitude goes out to Chris Nolan, Maria Diez-Rodriguez, Dr Chitra Joseph and Dr Madeleine Craze for their valuable technical support.

Words will not be enough to thank the biggest source of my power: my parents, my husband, my daughters; Rudeina and Nsiba and my little son Mansour, my in-laws, my sister and brothers, for their love, understanding and support given to me at all times.

I would like to dedicate this work with my warmest and deepest appreciation to my mother Mrs. Himeedah Albaseer and my father Mr. Abdulla Miftah whose dream for me brought about this achievement; without their loving, listening, sending me their blessings I would not have been where I am today. I also dedicate this work to my husband Mr. Tarek Abuhagar for helping me in difficult moments, standing by me and pushing me on to follow my dream of getting this degree. I thank them with all my heart.

Finally, special thanks go to the Ministry of Higher Education and Scientific Research, and the University of Benghazi, Libya for funding my PhD research.

Abbreviation

ACD	Advanced Cell Diagnostics
AR	Antigen retrieval
ASCO	American society of clinical oncology
BC	Breast cancer
BCI	Breast cancer index
BCIRG	Breast cancer international research group
BCSS	BC-specific survival
BM	Basement membrane
BSA	Bovine serum albumin
CAP	College of American pathologists
CC1	Conditioning solution 1
CI	Confidence interval
CNB	Core needle biopsy
CSC	Cancer stem cell
DFC	Dense fibrillar components
DNA	Deoxyribonucleic acid
ECRIC	Eastern cancer registration and information centre
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
ESA	Epithelial surface antigen
FAC	Fluorouracil, doxorubicin and cyclophosphamide
FC	Fibrillar centre
FDA	Food and Drug Administration
FFPE	Formalin-fixed paraffin-embedded
FHA	Forkhead-associated domain
GC	Granular component
GEP	Gene expression profiling
GGI	Genomic Grade Index
GHI-RS	Gene genomic health recurrence score
GPG	Good prognostic group
H&E	Haematoxylin and eosin stain

HDPP	HER2-derived prognostic predictor
HER2	Human epidermal growth factor receptors 2
HP1	Heterochromatin protein 1
HPRT	Hypoxanthine ribosyltransferase
HR	Hazard ratio
ICC	Intraclass correlation coefficient
IGS	Invasiveness gene signature
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
ISH	In situ hybridisation
LBA	ligand binding assays
LIMMA	Linear Models for Microarray Data
LN	Lymph node
MAPK	Mitogen-activated protein kinase
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MGI	Molecular Grade Index
MPG	Moderate prognostic group
NCDB	National Cancer Database
NGS	Nottingham grading system
NOS	Invasive ductal carcinomas
NPI	Nottingham prognostic index
PAM50	Prediction analysis of microarray 50
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PEPI	Preoperative endocrine prognostic index
PI3K	Phosphoinositol 3-kinase
PP1	Protein phosphatase 1
PPG	Poor prognostic group
PR	Progesterone receptor
RCB	Residual cancer burden
RNA	Ribonucleic acid

ROR	Risk-of-relapse
RPCB	Residual proliferative cancer burden
RPPA	Reverse phase protein array
RS	Recurrence score
RT-PCR	Real time- polymerase chain reaction
SSC	Saline-sodium citrate
TAC	Docetaxel, doxorubicin and cyclophosphamide
TBS	Tris buffered saline
TDLUs	Terminal ductal lobular units
WMCIU	West Midlands cancer intelligence unit
WTS	Whole tissue sections

Table of Contents

1. Introduction 1

1.1. Breast cancer (BC) 2
1.1.1. Basic anatomy and histology of the breast 2
1.1.2. Molecular pathogenesis of BC 3
1.1.3. Breast cancer incidence 8
1.2. Clinical prognostic/predictive factors, indices and classifiers 9
1.2.1. Histological types 10
1.2.2. Tumour size and lymph node stage 10
1.2.3. Histological tumour grade 11
1.2.4. Tumour-node-metastasis (TNM) staging system 14
1.3. Molecular prognostic/predictive factors, indices and classifiers 15
1.3.1. Prognostic/predictive markers 15
1.3.2. Online prognostic algorithms 26
1.3.3. Gene expression profiling and molecular classification 27
1.3.4. Immunohistochemistry-based indices 36
1.4. Hypothesis 40
1.5. Aims of this study 41

2. Material and Methods 42

2.1. Patient series 43
2.1.1. The first cohort (Primary Series, version one) 43
2.1.2. The second cohort (Core Biopsy ER series) 43
2.1.3. The third cohort (Primary Series, version two) 45
2.1.4. The fourth cohort (METABRIC data set) 45
2.1.5. The fifth cohort (Nottingham prospective series) 47
2.2. TMA construction 48
2.2.1. Steps of TMA construction 48
2.3. Immunohistochemistry 50
2.3.1. Antibody selection 50
2.3.2. Antibody optimisation 50
2.3.3. Controls 50

2.3.4. Steps of IHC procedure 51
2.3.5. Manual staining 51
2.3.6. Ventana Autostainer 52
2.3.7. Assessment of IHC staining 53
2.4. Western blot 54
2.4.1. Cell lysis preparation 54
2.4.2. Sample preparation for electrophoresis 55
2.4.3. Preparation of gel and running buffer 55
2.4.4. Electro-blotting transfer step 55
2.4.5. Primary and secondary antibodies incubations 56
2.5. Protein extraction from FFPE sections 56
2.5.1. Deparaffinisation and sampling of FFPE tissue sections 57
2.5.2. Extraction of protein from FFPE tissue section 57
2.6. Laser Capture Microdissection (LCM) 57
2.6.1. Optimisation of LCM 58
2.6.2. Deparaffinisation and sampling of FFPE tissue sections 58
2.7. Reverse phase protein array 59
2.7.1. Validation of primary antibodies for microarray experiments 59
2.7.2. Lysates array 60
2.8. Real time- polymerase chain reaction (RT-PCR) 62
2.8.1. RNA extraction from FFPE tissue sections 62
2.8.2. RNA extraction using RNeasy FFPE Kit (Qiagen) 62
2.8.3. Testing quality and quantity of RNA 63
2.8.4. Complementary DNA (cDNA) Generation 64
2.8.5. Primers design 64
2.8.6. Optimisation of primers 65
2.9. RNAscope technique 65
2.9.1. Preparation and pre-treatment of FFPE for the RNAscope assay 65
2.10. Ethical Approval 67

3. Evaluation of integration of Ki67 into BC classification .. 69

3.1. Introduction 70
3.1.1. Clinical importance in breast cancer 70
3.1.2. Ki67 and mitotic count 73

3.2. The hypothesis 76
3.3. Aim of this chapter 76
3.4. Methodology 77
3.4.1. Study group 77
3.4.2. Comparison of Ki67 expression between TMA and WTS 77
3.4.3. Assessment of cytoplasmic/membranous staining ... 80
3.4.4. Integration of Ki67 into NPI 84
3.4.5. Integration of Ki67 into NPI+ 85
3.5. Statistical analysis 87
3.6. Results 88
3.6.1. Comparison of Ki67 expression between TMA and WTS 88
3.6.2. The reproducibility of Ki67 scoring 97
3.6.3. The association with the important clinicopathological variables 99
3.6.4. The association with survival using different Cut-off values 106
3.6.5. Assessment of different immunostaining conditions 107
3.6.6. Evaluation of integration of Ki67 to NPI and NPI+ ..113
3.7. Discussion 114
3.8. Conclusion 122

4. Clinical utility of RPPA for molecular classification of BC 123

4.1. Introduction 124
4.2. Hypothesis.....128
4.3. Aim of this chapter 128
4.4. Methodology 130
4.4.1. Patient cohorts 130
4.4.2. Optimisation of protein extraction from FFPE tissues 130
4.4.3. Comparison between three sampling methods 132
4.4.4. Correlation between IHC and RPPA.....136
4.4.5. Analysis of RPPA data using consensus clustering and classification techniques including the role of Ki67 136
4.5. Results.....138
4.5.1. Protein extraction from FFPE tissues.....138
4.5.2. Comparison between three sampling methods 141

4.5.3. Correlation between IHC and RPPA.....145
4.5.4. Analysis of RPPA data using consensus clustering and classification techniques including the role of Ki67 148
4.6. Discussion 156
4.7. Conclusion 159

5. Evaluation of tumours expressing low ER in the molecular classification of BC 160

5.1. Introduction 161
5.1.1. The pattern of ER distribution 162
5.1.2. Luminal/ER-positive class and low ER subgroup 164
5.2. Hypothesis.....165
5.3. Aim of this chapter 165
5.4. Methodology 166
5.4.1. Assessment of ER expression pattern 166
5.4.2. Evaluation of clinical and molecular features of low ER group 169
5.5. Statistical analysis 177
5.6. Results.....178
5.6.1. Assessment of the ER expression pattern.....178
5.6.2. Evaluation of clinical and molecular features of low ER group 184
5.6.3. Comparison between the three different techniques 197
5.7. Discussion 198
5.8. Conclusion 205

6. Bmi-1 and a new prognostic subgroups 206

6.1. Introduction 207
6.1.1. Bmi-1 structure and function 208
6.1.2. Bmi-1 and BC 211
6.1.3. Bmi-1 and Ki67.....213
6.2. Hypothesis.....213
6.3. Aims 214
6.4. Methodology 214
6.4.1. Patient cohorts 214
6.4.2. *BMI1* gene expression 216
6.4.3. At protein level 216
6.5. Statistical analysis 217

6.6. Results 217
6.6.1. Transcriptomic level 217
6.6.2. Protein level 223
6.7. Discussion 232
6.8. Conclusion 233

7. General Discussion 235

7.1. Background and cardinal findings of study 236
7.2. Conclusion and recommendations 242
7.3. Strengths of the study 243
7.4. Limitations of the study 243
7.5. Future work 243

8. References 245

List of Tables

Table-1: Summary of method for examining histological grade in BC 13
Table-2: TNM breast cancer staging 15
Table-3: Clinicopathological characteristics of the first cohort 44
Table 4: Clinicopathological characteristics of the third cohort 46
Table-5: Clinicopathological characteristics of the four cohort 47
Table-6: Summary of the characteristics of the study cohort..... 78
Table-7: Different antibodies and different staining conditions for Ki67 staining 83
Table-8: NPI+ formulae for the Biological classes..... 86
Table-9: Demonstrates the results of the different cut-off points (TMA) with cut-off 10% (WTS) 90
Table-10: The correlation between the different cut-off values with the important clinicopathological parameters (TMA)..... 100
Table-11: The correlation between the results of the different cut-off points with the important clinicopathological parameters (WTS)..... 103
Table-12: Different cut-off values and the relations with BCSS (TMA) 106
Table-13: Cox proportional hazards analysis for predictors of BCSS (TMA) .. 106
Table-14: Different cut-off values and the relations with BCSS (WTS)..... 107
Table-15: Cox proportional hazards analysis for predictors of BCSS (WTS)... 107
Table-16: Clinicopathological criteria of the 25 BC cases used in this step of the study (Negm et al., 2016) 133
Table-17: Antibodies used in RPPA and the original NPI+ using IHC 135
Table-18: Correlation between RPPA and IHC results using NPI+ biomarkers (Negm et al., 2016) 147
Table-19: RPPA classification of BC according to NPI+ 148
Table-20: A total of 99 BC cases are consensus among K-means and PAM 150
Table-21: A total of 92 BC cases are consensus among K-means and PAM 150
Table-22: A total of 122 BC cases are consensus among K-means and PAM 151
Table-23: A total of 99 BC cases are consensus among K-means and PAM 151
Table-24: Primers sequences for genes used in RT-PCR..... 173
Table-25: Semi-quantitative scoring criteria for RNAscope staining..... 176
Table-26: Cross-tabulation of immunostaining ER expression groups and ESR1 transcriptomic groups (Muftah et al., 2016)..... 183
Table-27: ER scoring of CNB cases that remain positive on ER immunostaining of WTS 189
Table-28: Comparison between ER expression on CNB and WTS using IHC.. 189
Table-29: The details of clinicopathological parameters of the low/intermediate ER expression, WTS cases..... 190
Table-30: The details of clinicopathological parameters of the low ER expression, WTS cases (<10%)..... 191
Table-31: The results between the three different samples using different kits 192
Table-32: The details of clinicopathological parameters of the low/intermediate ER expression cases using RT-PCR 196
Table-33: The details of clinicopathological parameters of the low/intermediate ER expression cases using RNAscope..... 199
Table-34: Cox proportional hazards analysis for predictors of BC specific survival. Effect of *BMI1* mRNA expression, tumour grade, tumour size and lymph node stage..... 219
Table-35: Correlation of *BMI1* with *CD44, CD24* and *MKI67* genes expression220
Table-36: Correlation between Bmi-1 expression and the clinicopathological parameters in the studied cohort..... 226

Table-37: Correlation between Bmi-1 expression and the important biomarkers in the studied cohort 227
Table-38: Cox proportional hazards analysis for predictors of BCSS. Effect of Bmi-1 IHC expression, tumour grade, tumour size and lymph node stage..... 228

List of Figures

Figure-1: The anatomy of female breast..... 3
Figure-2: The bar chart demonstrating breast cancer incidence (blue)..... 9
Figure-3: Functional domain organization of nuclear receptors 18
Figure-4: Differences in PR functional domains..... 18
Figure-5: Ki67 structural elements 24
Figure-6: The diagram representing cell cycle phases 25
Figure-7: NPI+ biological classes 39
Figure-8: Representative diagram for preparation of membrane sandwich 56
Figure-9: Laser capture microdissection..... 59
Figure-10: Representative diagram showing steps of reverse phase protein array. 61
Figure-11: Representative diagram showing steps of RNAscope..... 68
Figure-12: A flowchart representing the cohorts and steps were used in the comparison between TMA and WTS of BC tissue 79
Figure-13: Correlation between Ki67 expressions..... 90
Figure-14: The bar chart demonstrates the percentage of matched cases 91
Figure-15: The bar chart demonstrates the percentage of matched cases 92
Figure-16: BCSS for Ki67 high/low groups..... 93
Figure-17: The bar chart demonstrates the percentage of matched cases 94
Figure-18: The bar chart demonstrates the percentage of matched cases 95
Figure-19: The bar chart demonstrates the percentage of matched cases 96
Figure-20: Correlation between *MKI67* mRNA and immunostained Ki67..... 98
Figure-21: Optimisation using different MIB-1 antibody concentrations using breast cancer tissues 109
Figure-22: Optimisation of MIB-1 using different antigen retrieval time..... 110
Figure-23: Optimisation of MIB-1 using different antigen retrieval methods using breast cancer tissues 111
Figure 24: Using different anti Ki67 clones (MIB-1, SP6, MKI67 and 30-9), each column representing the same breast cancer case stained with different clones 112
Figure-25: Correlation between Ki67 and mitosis..... 114
Figure-26: Survival curve of the three NPI groups 115
Figure-27: Survival curve of the six NPI groups 116
Figure-28: Survival curve of the three luminal A prognostic groups of original117
Figure-29: Survival curve of the two luminal B prognostic groups of original.. 118
Figure-30: A flowchart representing the steps of study in this Chapter 129
Figure-31: Quantification and reproducibility of protein extraction from FFPE samples..... 139
Figure-32: Representative microarray image 140
Figure-33: Comparison between the three sampling methods 142
Figure-34: The boxplot shows that full-face and macrodissection..... 143
Figure-35: The bar graph shows the stained (S) and unstained (US) 143
Figure-36: Western blotting validation of the primary antibodies 144
Figure-37: Inter-slide reproducibility..... 144
Figure-38: Immunostaining expression of NPI+ markers in breast cancer 146
Figure-39: Hierarchical cluster analysis of NPI+ biomarkers..... 147
Figure-40: Nottingham prognostic index plus (NPI+) classes..... 148
Figure-41: Correlation between Ki67 expressions using immunostaining..... 149
Figure-42: Boxplots of RPPA- breast cancer classification 152
Figure-43: Boxplots of RPPA- breast cancer classification 153
Figure-44: Boxplots of RPPA- breast cancer classification 154
Figure-45: Boxplots of RPPA- breast cancer classification 155

Figure-46: A flowchart representing the cohorts were used in this Chapter 168
Figure-47: Western blot demonstrates a specific band 170
Figure-48: Fold changes of the negative cases relative to the pooled calibrator which used in RT-PCR. The cut-off point was 1.00 where the standard deviation was 1.006± 0.100 174
Figure-49: Semi-quantitative scoring criteria for RNAscope 176
Figure-50: Bimodal distribution of ER 179
Figure-51: The distribution of PR 179
Figure-52: Kaplan Meier plots illustrating BCSS and ER immunostaining subgroups 180
Figure-53: The distribution of ER 181
Figure-54: Illustrated different levels of ER expression 181
Figure-55: The frequency distribution of ER 182
Figure-56: Bimodal distribution of *ESR1* gene (Muftah et al., 2016) 183
Figure-57: Kaplan Meier plots demonstrating of BC specific survival 184
Figure-58: Scatterplot diagrams demonstrate the relation between CNB 186
Figure-59: Scatterplot diagrams to demonstrate the relation between CNB 187
Figure-60: Shows scatterplot diagrams to demonstrate the relation between CNB 188
Figure-61: The quantity of RNA yield using different kits 193
Figure-62: Evaluation of RNA purity using 260/280 ratio 193
Figure-63: ERα, PR-A, PR-B and HPRT primers 194
Figure-64: Melting curves for each primer contain a single peak 195
Figure-65:The image shows focal weak IHC staining of breast cancer cases 200
Figure-66: Low ER-positive cases have low staining of breast cancer tissue 201
Figure-67: Bmi-1 protein structure. 208
Figure-68: The diagram demonstrating a role of Bmi-1 in Ink4a/Arf 211
Figure-69: A flowchart representing the cohorts and steps were used in this Chapter 215
Figure-70: Normally distributed *BMI1* gene expression for the METABRIC cohort 218
Figure-71: Kaplan–Meier survival plots for *BMI1* expression. 219
Figure-72: Distribution of *CD24* and *CD44* gene expression 221
Figure-73: Kaplan–Meier survival plot 222
Figure-74: Western blot confirming Bmi-1 antibody specificity 223
Figure-75: Immunohistochemical expression of Bmi-1 in invasive beast cancer 224
Figure-76: Distribution of Bmi-1 immunostaining expression in the studied cohort 225
Figure-77: Kaplan–Meier survival plots for Bmi-1 expression. 228
Figure-78: Kaplan–Meier survival plot 230
Figure-79: Kaplan–Meier survival plot 230
Figure-80: Kaplan–Meier survival plot 231

1. Introduction

1.1. Breast cancer (BC)

1.1.1. Basic anatomy and histology of the breast

The mammary gland is an organ consisting of skin and underlying connective tissue comprising a group of 15–20 large secretory lobes (modified sweat glands), embedded in adipose tissue (**Figure-1**). Morphologically, breast tissue heterogeneity is remarkable within a single breast and between breasts (Howard and Gusterson, 2000). A lobe comprises of lobules and excretory ducts that drain through a single lactiferous duct that opens onto the surface of the nipple. Terminal ductal lobular units (TDLUs) are composed terminal ductules and alveoli of the lobules. They represent the functional units of the human mammary gland and histologically resemble a bunch of grapes (Cardiff, 1998). The tubuloalveolar system is surrounded by a basement membrane (BM). The large ducts are lined with a stratified columnar epithelium while the small ducts are lined with a simple cuboidal epithelium. Luminal epithelial cells have little cytoplasm and an oval central nuclei, with one or more nucleoli, and marginal chromatin. Between the luminal cells and the BM is an incomplete layer of myoepithelial cells. In the large ducts, myoepithelial cells are numerous and they form a relatively complete layer (Woodward et al., 2005) (Johnson, 2010). In alveoli, the myoepithelial cells form a network of slight processes that is more attenuated. Myoepithelial cells contain smooth muscle-specific proteins, form gap junctions with each other and are separated from connective tissue by a BM. Moreover, macrophages and lymphocytes have been found travelling through the epithelium toward the lumen (Ferguson, 1985). Stem cells are highest in ducts (Villadsen et al., 2007) and they have a tendency to be quiescent and surrounded by patches of proliferating and differentiated cells (Clarke, 2005).

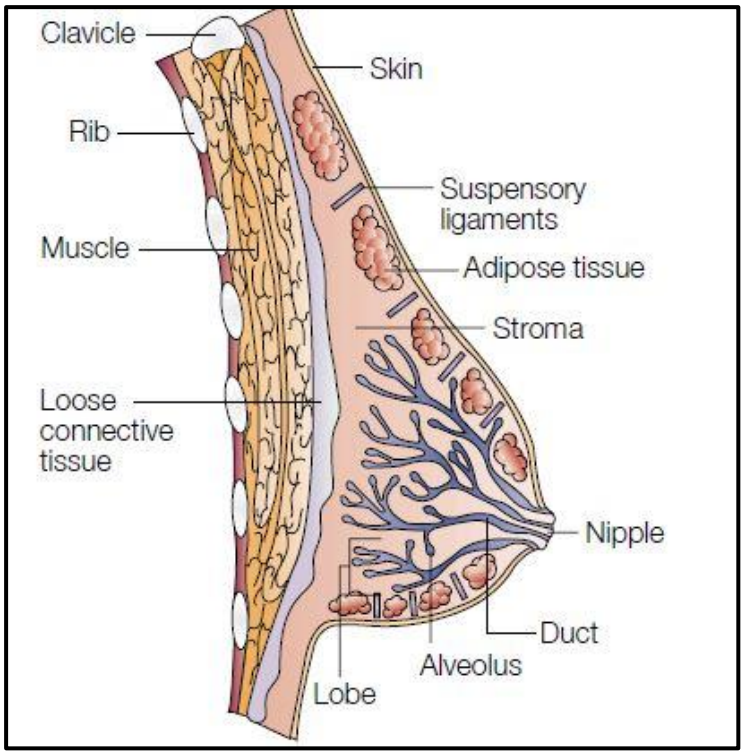


Figure-1: The anatomy of female breast
(Ali and Coombes, 2002)

1.1.2. Molecular pathogenesis of BC

No single alteration is involved for sporadic BC. The cause and reasons for its progression are unknown. Assumed oncogenic processes comprise mutated oncogenes and tumour suppressor genes, dysfunction of cell-number regulation or epigenetic controls and disturbed intercellular interaction (Prince et al., 2007). Two principal models of mammary carcinogenesis are suggested: the stochastic clonal evolution model and the cancer stem cell (CSC) model (Nguyen et al., 2012). The former suggests that random mutations can occur in any breast epithelial cell; when a cell accumulates enough numbers and types of mutations to contribute to tumour growth, uncontrolled cell division results. The CSC hypothesis postulates that a carcinogenic event is initiated in stem cells or progenitors instead of differentiated luminal or myoepithelial cells (Bombonati and Sgroi, 2011).

Stem cells are primitive cells, which are regarded to be progenitor of hundreds of cell types present in human. They are undifferentiated cells that are characteristically of the same lineage. They maintain the capability to divide throughout life and produce cells that can become highly specialised replacing cells that undergo apoptosis. There is two broad categories of stem cells: embryonic stem cells resulting from blastocyst and adult stem cells, which are present in adult tissue. In a developing embryo, stem cells have the ability to differentiate into all the specialised embryonic tissue while in adults, they perform as a repair system for replacing specialised damaged cells (Avasthi, 2008).

1.1.2.1. Mammary stem cells (MaSCs)

The hallmark characteristics of any stem cell population are the ability to differentiate into several cell lineages and to sustain proliferation and to self-renew to yield more stem cells (Shackleton et al., 2006). The discovery of mammary stem cells (MaSCs) occurred in the late 1950s, when DeOme et al. (1959) published an article describing how the transplanted portions of the normal mammary epithelia from donor mice into recipient fat pads cleared of endogenous epithelium could expand and reproduce the full functions of the mammary epithelial gland (Deome et al., 1959). Successful repopulation of the mammary gland could be obtained with any segment of the mammary epithelial tree (Smith and Medina, 1988).

As stem cells exist in the mammary gland. One of the morphologically distinct cells called Cap cells were hypothesised to represent MaSCs. The cap cells line the outside of the terminal end buds located at the tips of the ducts during puberty (Williams and Daniel, 1983). The other cells are the pale or light-staining cells; these have low cellular complexity and are visualised by electron microscopy and known as undifferentiated large light cells and small light cells (Smith and Medina, 1988) (Chepko and Smith, 1997).

BC stem cells (BCSCs)

The origin

The origin of BCSCs remains debatable. Depending on the original cells which led to CSCs two hypotheses seem to be the more accepted. The first assumes that BCSCs are derived from epithelial– mesenchymal transition (EMT). The cells that undergo EMT have similar behaviour and characteristics to both normal stem cells and CSCs. These cells also have an increased ability to form mammospheres capable of surviving and proliferating in a non-adherent non-differentiated culture thus proposing a highly tumourigenic phenotype (Mani et al., 2008). The second hypothesis is that the BCSCs derive from the deregulation of the normal, self-renewal and differentiation pathways of MaSCs. The reasons which support this hypothesis are the high similarity between BCSCs and MaSCs and the long lifespans of MaSCs, which makes them highly susceptible to mutations and, subsequently, oncogenic transformations (Al-Hajj et al., 2003, Ponti et al., 2005). While this has been revealed for leukaemia, (Bonnet and Dick, 1997) evidence in solid tumours is not yet as clear cut. Moreover, how stem cells transform into cancer is also controversial. Several hypotheses have arisen to explain this transformation:

1) **Defeat of the asymmetric division**: Asymmetric division offers stem cells the capacity for self-renewal. The stem cell divides into daughter cells: one is similar to the mother and the other generates more differentiated progenitor cells with a proliferative potential toward terminally differentiated cells. If the CSCs miss the apical basal-polarity axis, both daughter cells are then identical to the mother, which will cause an accumulation of stem cells that result in malignant behaviour (Clevers, 2005) (Wodarz and Gonzalez, 2006). This hypothesis is supported by studies in neuronal stem cells in *Drosophila melanogaster* (Caussinus and Gonzalez, 2005).

2) **Non-regulation by the microenvironment:** the regulatory effect of the microenvironment plays an essential role in the stem cell cellular cycle.

Stem cells which are recruited to other sites can cause neoplastic processes in the context of chronic inflammation. This could be because of loss of control making inhibitory signals disappear from the extracellular matrix when stem cells are outside their usual niche (Houghton et al., 2004).

3) **Cell fusion:** this is necessary for certain physiological processes and is essential for developmental and behavioural patterning in organisms; for instance, in fertilisation (fusion of sperm and egg) and in the immune response (fusion of macrophages to form giant cells) (Shemer and Podbilewicz, 2003) (Ogle et al., 2005). Increased incidence of cell fusion might be linked to cancer initiation. Many tumour cells are fusogenic and fusion between tumour and normal cells produces hybrid cells that are often more malignant than the original ones (Pawelek, 2000). Human stem cells belonging to a grafted kidney, for example, might leave the grafted organ and reach the skin, differentiate or fuse to it as keratinocytes and thence undergo transformation (Aractingi et al., 2005).

4) **Horizontal gene-transfer:** this is the same mechanism used in the establishment of antibiotic resistance in bacteria. Apoptotic bodies released from tumour cells can induce p53-deficient fibroblasts to form tumours *in vivo* and colonies *in vitro*, and chromosomal fragments can be transferred through phagocytosis into recipient cells (Bergsmedh et al., 2001). Stem cells, through their phagocytic capacity, could introduce apoptotic bodies, and the rescheduled genes might have an important function in tumour initiation and progression (Bjerkvig et al., 2005).

Pathways

In MaSCs, Hedgehog (Hh), Wnt and Notch pathways are the main responsible signalling pathways for self-renewal. BCSCs and MaSCs share cellular markers and pathways, but with an atypical activation that may initiate cancer.

Notch pathway

This pathway comprises four Notch transmembrane receptors (Notch1, Notch2, Notch3 and Notch4) and five ligands: jagged proteins (JAG1 and JAG2) and delta-like proteins (DLL1, DLL3 and DLL4). In the context of MaSCs, the Notch pathway is essential for proliferation, self-renewal and the regulation of cell outcome (Dontu et al., 2004). The activation of this pathway is said to exist in 50% of BCs (Karamboulas and Ailles, 2013). The stimulation of this pathway activates mammosphere development, and it has been proposed that it has a role in lineage-specific differentiation. However, the down-regulation of the Notch pathway decreases the mammosphere formation, the frequency of $CD44^{+}/CD24^{-/low}$ cells and the possibility of producing tumours *in vivo* (Harrison et al., 2010).

Wnt pathway

Wnt is another important pathway in BCSCs; it consists of 19 ligands and a number of receptors. Two pathways have been recognised for Wnt signalling: the canonical pathway, where Wnt activity occurs via β-catenin and is employed in cell outcome determination; and the non-canonical pathway, where Wnt signalling, through a β-catenin independent way, is utilised in cell mobility and polarity. Wnt1 was initially recognized as a mammary oncogene in infected mice with mouse mammary tumour virus (MMTV), which was incorporated into the Wnt1 gene leading to its activation. In this model, hyperplasia was initially observed and, later, carcinogenesis occurred, proposing that Wnt has a role in tumour initiation. Further, activated β-catenin was also adequate to stimulate mammary tumour formation (Nusse and Varmus, 1982) (Karamboulas and Ailles, 2013).

Hedgehog pathway

Hh pathway signalling is necessary for normal development of breast tissue (Charafe-Jauffret et al., 2008). In BCSCs, the Hh pathway is activated while its inhibition leading to a reduction of mammosphere development. It has a pivotal

role in the EMT process, which is required for malignant tumour invasion and metastasis (Katoh and Katoh, 2008). The Hh ligands bind to a transmembrane protein patched (PTCH) receptor site and, as a result, an effect on smoothened (SMO) occurs (Pepinsky et al., 2000). SMO then translocate to the primary cilium where transcription factors GLI-1 and GLI-2 are activated, leading to transcription of GLI target genes (Pepinsky et al., 2000). Data from various tumours, comprising BC, propose that Hh signalling has an essential role in CSC regulation (Liu et al., 2006). Furthermore, previous studies provide evidence that Hh signalling regulates self-renewal, which is a crucial feature for the preservation of the malignant clone (Merchant and Matsui, 2010). Liu et al. revealed that stimulation of hedgehog signalling enhances mammosphere-initiating cell number and mammosphere size, while suppression of the pathway leads to a reduction of these impacts. These influences are mediated by Bmi-1. Moreover, the pathway activation using the Hh ligand, GLI-1 and GLI-2 changes the expression of Bmi-1, a central regulator of self-renewal in normal stem cells. $CD44^{+}/CD24^{-/low}$ cells exhibit high expression of Hh pathway which constitutes PTCH1, Gli1 and Gli2 by about 1.7, 30 and 6-fold respectively, and for BMI-1 it is increased 5-fold compared with the cells separated from the same tumour, which are deficient in cancer stem cell biomarkers (Liu et al., 2006).

1.1.3. Breast cancer incidence

In 2016, in the United States, 1,685,210 new cancer cases and 595,690 cancer deaths were expected to occur corresponding to about 1,600 deaths per day. About 61,000 cases of carcinoma *in situ* of the female breast and 246,660 of BC were diagnosed in 2016, with about 40,450 estimated deaths. BC is the second most diagnosed cancer in the US and is forecast to account for 29% of all new cancer diagnoses in women. According to the recorded number of deaths in 2012, BC is the leading cause of cancer death in women aged 20 to 59 years. However, in women aged 60 years or older, this is replaced by lung cancer (Siegel et al., 2016). In the UK, when incidences and mortality rates pertaining to the period of 2011-2014 are calculated, for every five females newly

diagnosed with BC, one female dies of the disease (**Figure-2**) (Dubey et al., 2015).

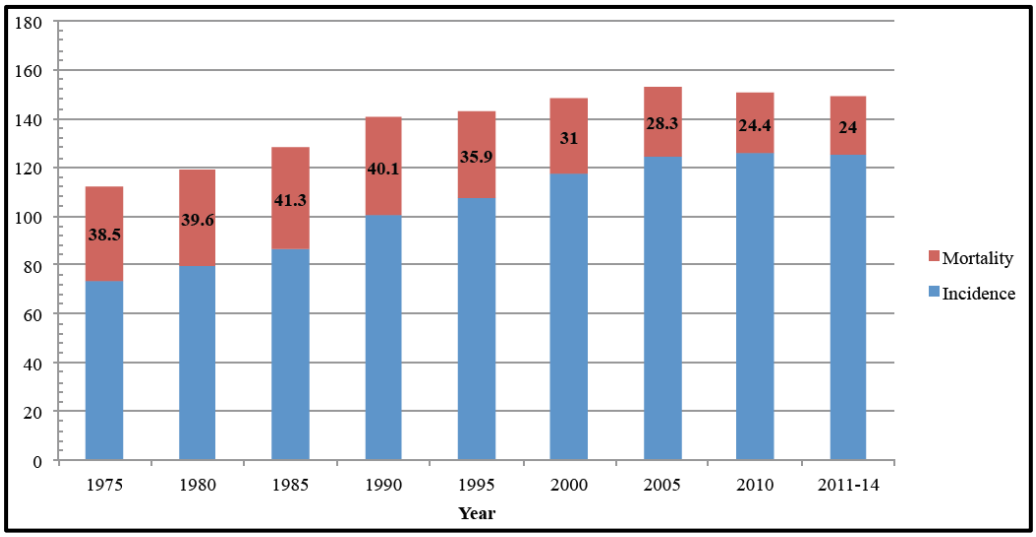


Figure-2: The bar chart demonstrating breast cancer incidence (blue) and mortality rates (red), UK 1975-2014 (Dubey et al., 2015)

1.2. Clinical prognostic/predictive factors, indices and classifiers

In general, a prognostic factor is any quantifiable parameter able to provide information on patient clinical outcome, i.e. evaluating the risk of disease recurrence at the time of primary diagnosis, unrelated to therapy. Prognostic factors are usually reflected by tumour growth, invasiveness and metastatic probability. A predictive factor is any measurable parameter that can give an information on the likelihood of response to a certain therapeutic modality (Gasparini et al., 1993) (Hayes et al., 1998). Prognostic or predictive factors could be either a single trait or signature of traits that can classify patients into different groups. Although prognostic and predictive parameters could be separately classified, several factors in BC can be both prognostic and predictive parameters. Oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptors 2 (HER2), tumour size, histological grade,

type and lymph node stage are the routine practical prognostic and predictive parameters (Ellis et al., 2016).

1.2.1. Histological types

Histological types represent specific morphological and cytological patterns of tumour. The World Health Organization (WHO) classification recognises the existence of at least 17 distinct histological special types (Weigelt et al., 2010b), some of which have prognostic significance including mucinous, medullary and tubular BC (Diab et al., 1999) (Rakha et al., 2010a). Approximately 70–80 % of breast carcinomas are defined as invasive ductal carcinoma not otherwise specified (NOS) or of no special type (NST), while about 25 % are defined as histological special types. The common histological types of BC are; invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). For both invasive ductal carcinoma and invasive lobular carcinoma, the overall 5-years survival rate is almost 85 % (Pourteimoor et al., 2016). Although the recent edition of WHO's classification offers a more broad interpretation of epidemiology, clinical features, genetics, macroscopic appearance and prognostic and predictive characteristics (Sinn and Kreipe, 2013), a comprehensive classification of BC based on the distinct histological patterns and biological features has not been recommended by the scientific association (Weigelt and Reis-Filho, 2009). The main drawback of this classification is that most of the BCs ultimately join either one of the two major histopathological types; NOS or ILC. This reveals that the classification may be unable to reflect the actual heterogeneity of BC because tumours that have a very different biological and clinical profile cluster together within the same group (Viale, 2012).

1.2.2. Tumour size and lymph node stage

Tumour size is a time-dependent prognostic parameter and patients with small tumour sizes have longer survival time than those who harbour larger ones

(Elston et al., 1999). Tumour sizes show a positive association with lymph node status (Weiss et al., 2003). However, they both are independent measurements of outcome. Despite large tumours, the number of lymph nodes significantly involved does impact on the outcome. Therefore, if there is increased lymph node involvement, the survival status will decrease. Simultaneously, if the tumour size is increased, survival is also reduced regardless of lymph node involvement. However, in general, the larger the diameter of the tumour, the more the number of lymph nodes with metastatic cancer increases (Carter et al., 1989) (Foulkes, 2012). Conversely, this relation between tumour size, lymph node and survival is diminished in triple-negative BC (tumours that are ER, PR, and HER2-negative). Dent et al. reported that with a tumour <5 cm in the triple-negative group, there is no relation between the size of the tumour and lymph node status, as 55% of patients with a tumour equal to or less than 1cm has one positive lymph node as a minimum (Dent et al., 2007).

In the early stage of BC, lymph node grade represents one of the most important prognostic tools. Patients who are lymph node-positive have a higher mortality than node-negative ones (Arriagada et al., 2006) and there is a direct correlation between the number of metastatic lymph nodes and the risk of death and recurrence (Weiss et al., 2003). In addition, depending on international guidelines, patients exhibiting BC with positive lymph nodes will generally be exposed to treatment with adjuvant chemotherapy, without considering of their hormonal status, in view of the fact that this class of BC patients demonstrates the highest benefit from this management approach. For the other group of BC patients (lymph node-negative), evaluation for further predictive and prognostic markers has to be taken before considering adjuvant treatment (Zhang et al., 2012).

1.2.3. Histological tumour grade

An additional strong and known clinical prognostic factor in early BC is the histological grade (Arriagada et al., 2006) (Soerjomataram et al., 2008). Although the validity of the histological grade has been confirmed by different

studies (Dalton et al., 1994) (Pereira et al., 1995), initially there was disagreement about using grading in management decisions with BC patients, for the reason that there is a lack of reducibility. This concern was reduced after the modification of the Bloom-Richardson grading system to the modern Nottingham (Elston- Ellis) Grading System (NGS) and the demonstration of its prognostic relevance in 1991 (Elston and Ellis, 1991). Subsequently, it has been validated in various discrete studies (Sundquist et al., 1999) (Lundin et al., 2001, Blamey et al., 2010) as well as being approved by different international professional organisations: the Royal College of Pathologists (UK RCPath), the WHO and the European Union (EU) (Rakha et al., 2010b) (Ellis et al., 2016).

Because of its strong prognostic value, the NGS was joined with lymph node stage and tumour size to form the Nottingham Prognostic Index (NPI) (Galea et al., 1992). Additionally, it has been incorporated the St Gallen guidelines (Goldhirsch et al., 2009) and algorithms such as Adjuvant! Online (Mook et al., 2009). The NGS depends on the assessment of three components: (1) the tubular formation; (2) the degree of nuclear pleomorphism; and (3) mitotic count in a specific field area. Each of these features is scored from 1 to 3 (**Table-1**) (Elston and Ellis, 1991). The sum of the scores of these criteria provides a total ultimate score. The matching grades are as follows: (1) scores of 3–5 represent grade I tumours, which are well-differentiated and have the best prognosis; (2) scores of 6–7 are associated with grade II tumours, which are moderately differentiated with intermediate prognosis; and finally, (3) scores of 8–9 express grade III tumours, which are poorly differentiated and present the worst prognosis (Pourteimoor et al., 2016). Evaluating the survival rates of 22,616 BC cases, Henson and colleagues showed that patients who had grade I tumours of <2 cm in diameter had 5-years survival rate of 99% (Henson et al., 1991). This result is emphasised by the study of a long-term follow-up of 2,219 operable BC cases from the Nottingham group (Rakha et al., 2008). The result was that the grade and lymph node used together would have the ability to affect the outcomes of the BC patients, as the grade is the important determinant of outcome. However, 30-60% of tumours are categorised as grade II which is less informative than grade I or III in terms of clinical decision (Ignatiadis and Sotiriou, 2008).

Table-1: Summary of method for examining histological grade in BC

Features	Score
Tubular formation • Majority of tumour (>75%) • Moderate degree (10-75%) • Little or none (<10 %)	1 2 3
Nuclear pleomorphism • Small regular uniform cells • Moderate increase in size and variability • Marked variation	1 2 3
Mitotic count • Dependent on microscopic field area	1-3

The NPI is an approved and widely accepted method for prognosis as well as survival prediction in operable cases of primary BC (Blamey et al., 2007b). It is one of the earliest indices to be developed. In 1982, it was calculated throughout a retrospective study of 387 women with primary, operable BC by multivariate regression analysis (Haybittle et al., 1982). In 1991, the prognostic importance of NPI in BC was initially expressed (Elston and Ellis, 1991). Then, after the long-term follow-up (Galea et al., 1992) and independent validation by different centres, it was confirmed and applied widely (Brown et al., 1993) (Balslev et al., 1994) (D'Eredita et al., 2001). It is widely used in the UK for its simplicity, clinical reliability and because it mainly achieves well in different populations (Elston et al., 1999). It is structured from the three previous mentioned prognostic determinants (lymph node stage, tumour size and histological grade) to classify patients into good, moderate and poor prognostic groups. In the NPI, these are < 3.4, 3.4–5.4 and > 5.4, respectively. Based on data from the 1980s, the estimation of the original 15-years overall survival rate for each prognostic group is revealed as 80%, 42% and 13%, correspondingly (Galea et al., 1992). Updated estimation of survival from the 1990s has been published with further stratification of the prognostic groups (Blamey et al., 2007a).

Calculating NPI

The formula for NPI is:

lymph node (LN) stage (1–3) + histological grade (1–3) + (0.2) x tumour size (cm), where stage 1 means lymph node-negative; stage 2 means three or fewer nodes involved by metastatic carcinoma; stage 3 is four or more nodes, or apical node, or any axillary plus internal mammary node invaded by carcinoma (Ellis et al., 2016). In the UK, about 20%, 30% and 50% of symptomatic BCs are grade I, II and III respectively (Ellis et al., 2016). In addition to the ability to classify the patients into clinically meaningful subgroups, histological grade, interestingly, revealed to be associated with the genetic and transcriptomic features of BCs (Sotiriou et al., 2006) (Weigelt et al., 2010a).

1.2.4. Tumour-node-metastasis (TNM) staging system

The TNM staging system consists of the size of the primary tumour (T), lymph node involvement (N) and the distant metastases (M) (**Table-2**). According to Singletary and Connolly (1942), TNM was developed by Pierre Denoix based on these three main prognostic parameters in BC. In 1958, a BC clinical classification dependent on the TNM system was presented by the International Union against Cancer (UICC). Later, in 1977 the American Joint Committee on Cancer (AJCC) released a staging system for BC based on TNM system in their first manual. After that, revisions have been updated regularly with the latest progress in prognosis as well as in treatment. The variations between AJCC and UICC versions of the TNM were removed in 1987 (Singletary and Connolly, 2006). BC staging is used by clinicians to estimate prognosis. Bland and colleagues supported the prognostic value of the staging system by using data from 1.3 million cases from the National Cancer Database (NCDB). They reported a noticeable difference between the stages: during the first 10 years after diagnosis, in stages I/II, from 12% to 34% of patients died, while > 60% of stage III and > 90% of stage IV patients died in the same period (Bland et al., 1998).

Table-2: TNM breast cancer staging

ANATOMIC STAGE/PROGNOSTIC GROUPS			
Stage 0	Tis	N0	M0
Stage IA	T1*	N0	M0
Stage IB	T0	N1mi	M0
	T1*	N1mi	M0
Stage IIA	T0	N1**	M0
	T1*	N1**	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1*	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

https://cancerstaging.org/references tools/quickreferences/Pages/default.aspx

1.3. Molecular prognostic/predictive factors, indices and classifiers

1.3.1. Prognostic/predictive markers

The difference in specific characteristics between the tumour and corresponding normal tissue can be measured by tumour markers (Levenson, 2007).

Identification of tumour markers with prognostic or predictive power for different BC treatment approaches is an area of intense research. Although some tumour markers are already accepted as important prognostic markers for BC, there are still those that need further standardisation and updating to meet the target of personalised medicine.

1.3.1.1. Hormonal receptors

The expression of ER and PR act as prognostic and predictive parameters and the evaluation of their expression is an established standard process for breast pathology. The Early Breast Cancer Trialists' Collaborative Group demonstrated that how treatment of patients with tamoxifen (anti-oestrogen therapy), which can block cellular proliferation in BC, had an effect on BC mortality and recurrence in 5 years (EBCTCG, 2005). PR is an ER-regulated gene, therefore the expression of this receptor is highly related to the presence of the ER. An ER-negative/PR-positive result is uncommon accounting for less than 1% of BCs (Hefti et al., 2013). Risk factors and incidence rates for BC differ according to ER status. Therefore, dividing BCs depending on the ER status of the tumour can accurately estimate the BC risk (Colditz et al., 2004). Moreover, clear determination of ER status is fundamental for optimally choosing patients for hormonal treatment. In accordance with a recommendation by the American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP), the detection of ER is deemed positive from 1% of tumour nuclei immunostaining (Hammond et al., 2010b). However, identification of the percentage of positively staining cells that should be used to categorise tumours as ER-negative or ER-positive remains problematic, particularly with those tumour with low ER staining <10% (Deyarmin et al., 2013).

Structure of ER and PR

ER belongs to the nuclear receptor superfamily which contains six subgroups; a subfamily for Retinoic Acid Receptors, Thyroid Hormone Receptors, Vitamin D Receptor, Peroxisome Proliferator Receptor, Ecdysone Receptor and numerous orphan receptors. The second subfamily contains Hepatocyte Nuclear

Factor 4, Retinoic X Receptors and Chicken Ovalbumin Upstream Promotor-transcription Factor. The third category for steroid receptor with ER, PR, Androgen Receptors, Glucocorticoid Receptors and Mineralocorticoid Receptors. Another subfamily consists of the nerve growth factor inducible factor I-B group of orphan receptors. Further, a small subgroup exists for Steroidogenic Factor-1 and Drosophila Factor-1 Receptor. Finally, there is a group containing the Germ Cell Nuclear factor-1 Receptor.

In 1979, Wrange et al. showed three distinct domains of Glucocorticoid Receptors; the DNA-binding domain, the hormone (ligand) binding domain and N-terminal domain (Wrange et al., 1979). The three domains representative in all nuclear receptors. Two isoforms of genomic nuclear ER exist: ERα (Jensen and Jacobson, 1962) (Walter et al., 1985) and ERβ (Kuiper et al., 1996, Mosselman et al., 1996). Moreover, a non-genomic membrane bound receptor like GPER or GPR30 has been identified (Carmeci et al., 1997). Both ERα and ERβ share the same general structure, they are composed of multiple domains starting from the N-terminal to C- terminal; N-terminal A/B domain contains the autonomous transactivation function AF-1, C region containing DNA binding domain (DBD), the hinge region or D domain and E region having ligand-binding domain (LBD) and AF-2 (**Figure-3**) (Ruff et al., 2000). The two receptors share a high degree of homology except in the N-terminal/AF-2 domains (Eyster, 2016). Moreover, the molecular weight of ERα is 67kDa and contains 595 amino acids while ERβ has 59kDa and 530 amino acids (Green et al., 1986) (Ogawa et al., 1998). A clinical role has only been established for ERα (Duffy, 2006). The expression of ERα is a recognised predictor of a positive response to Tamoxifen (which binds ERα) therapy and patients with higher ERα expression express better response to Tamoxifen than cases with lower receptor levels (Davies et al., 2011). Furthermore, the clinical responsiveness to the different types of hormonal treatments, which suppress the production of oestrogen such as the aromatase inhibitors, is also reliant on ERα expression, and only positive tumours have a response (Howell et al., 2005) (Dowsett et al., 2005). The role of ERβ in BC remains unclear and it is currently not considered as a target in diagnostics or treatment (Rondon-Lagos et al., 2016).

Both PR isoforms (PR-A and PR-B) are produced by a single gene located on chromosome 11 at q22-23. PR-B is a longer protein isoform (116kDa) than PR-A form (94kDa). PR-A and PR-B are co-expressed in the breast as well as in several other progesterone target organs. PR-B isoform is required for normal development of the mammary gland (Mulac-Jericevic et al., 2003). Although they are similar hormones with DNA-binding activities, the two isoforms differ with regard to their transcriptional activities (Giangrande and McDonnell, 1999) (Richer et al., 2002). PR-B is considered a stronger activator than PR-A, probably due to its third transactivation domain (AF- 3) that present in the N-terminal region (Sartorius et al., 1994). Both PR-B and PR-A isoforms contain the basic structure that found in other nuclear receptors: a C-terminal ligand-binding domain (LBD); a hinge region (H); a DNA-binding domain (DBD) as shown in **Figure-4** (Leonhardt et al., 2003).

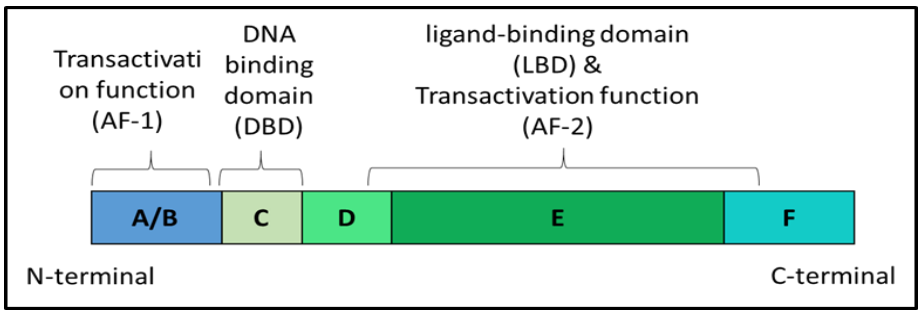


Figure-3: Functional domain organization of nuclear receptors including ER*α*. N-terminal A/B domain contains the autonomous transactivation function AF-1, C region containing DNA binding domain (DBD), the hinge region or D domain and E region having ligand-binding domain (LBD) and AF-2

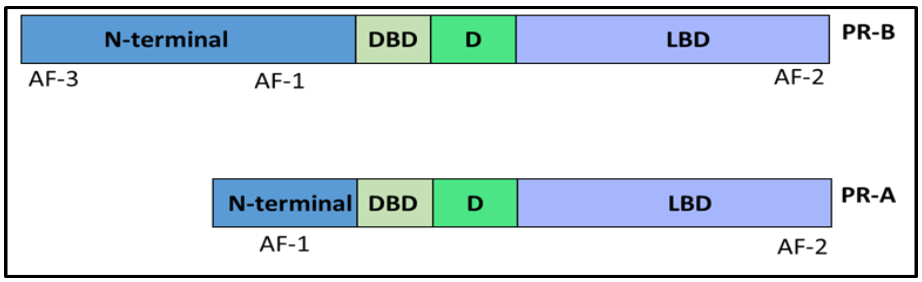


Figure-4: Differences in PR functional domains
Mainly due to the third transactivation domain (AF- 3) that present in the N-terminal region

Genomic and non-genomic function of ER

Oestrogen dependent function can be divided into genomic and non-genomic signalling. Additionally, the genomic signalling is divided into direct or indirect depending on binding of the oestrogen-ER complex to DNA.

Direct genomic signalling

Direct genomic signalling is deemed as the classical pathway of oestrogen signalling in which it acts as a ligand-dependent transcription factor and controls the expression of a variety of genes. In the cellular cytoplasm of target tissue, the binding of 17β-estradiol to either ERα or ERβ produces configurational transformation that permits receptor dimerisation and translocation to the nucleus to bind the oestrogen response elements (EREs). These specific sites are located in or close to the promoters of target genes. Additionally, the ligand-receptor complex stimulates recruitment of different co-regulators in a complex that changes chromatin structure and enables attracting of the RNA polymerase II transcriptional machinery. Using this mechanism, oestrogen-ER complex works as a transcriptional activator enhancing gene expression (Vrtacnik et al., 2014).

Indirect genomic signalling

In indirect genomic signalling, 17β-estradiol can regulate expression of genes that lack EREs in their promoter regions. Ligand- activated ERs occurs through protein-protein interactions with other groups of transcription factors at their response elements instead of direct binding to DNA. This enables activation or suppression of gene expression and considerably increases the regulatory influence of oestrogen. For instance, interaction of oestrogen-ER complex with Jun proto-oncogene (JUN) and FBJ murine osteosarcoma viral oncogene homolog (FOS) proteins at binding sites of the activator protein 1 (AP-1) in genes encoding insulin-like growth factor 1 (IGF1), cyclin D1 (CCND1) collagenase, choline acetyltransferase and ovalbumin. The activated signal rely on the type of the ligand and ER subtype. Furthermore, Sp1 transcription factor,

GATA binding protein 1 (GATA1), nuclear factor κB (NFκB), CCAAT/enhancer binding protein β (C/EBPβ) and signal transducer and activator of transcription 5 (STAT5) are transcription factors that promote oestrogen signalling (Bjornstrom and Sjoberg, 2005).

Non-genomic signalling

The action of steroid hormones via non-genomic signalling is common and usually related to activation of different protein-kinase cascades that ultimately produces indirect variations in gene expression owing to phosphorylation of transcription factors. This signalling is mostly linked to membrane bound ER and certain variants of ERα and ERβ. The mobilisation of intracellular calcium due to binding of oestrogens to ERs located at the cell surface can cause activation of the phosphoinositol 3-kinase (PI3K) signalling pathway, stimulation of adenylate cyclase activity and cyclic adenosine monophosphate, activation of membrane tyrosine kinase receptors and activation of the mitogen-activated protein kinase (MAPK) pathway (Bjornstrom and Sjoberg, 2005) (Ajj et al., 2013).

1.3.1.2. HER2

HER2 is a member of the transmembrane epidermal growth receptor family: HER1, HER2, HER3 and HER4 (Prenzel et al., 2001). HER2 overexpression occurs in 15-20% of BC cases (Slamon et al., 1987) (Slamon et al., 1989). Its overexpression is an established poor prognostic indicator of invasive BC (Cui et al., 2005, Rakha et al., 2007, Dobrescu et al., 2011). HER2 is the principal predictive tool for anti-HER2 targeted treatment such as Herceptin, Lapatinib (HER2 dual inhibitor tyrosine kinase) and pertuzumab (HER dimerization inhibitor) (Cortes et al., 2012) (Baselga et al., 2012) (Blackwell et al., 2012). In addition to its prognostic value as a single marker, HER2, when combined with the hormone receptor, can provide addition prognostic information similar to prognostic gene signatures with multi-parameters (Dowsett et al., 2013). Furthermore, HER2 amplification in breast tumours can be utilised to indicate malignancy, as benign tumours do not exhibit HER2 amplification; however,

some benign tumours present weak to moderate non-amplified HER2 membrane staining (Lewis et al., 2004).

1.3.1.3. Ki67

Cell proliferation represents one of the recognised prognostic factors in BC (Clahsen et al., 1999). Nuclear Ki67 protein is considered as an indicator of tumour proliferation (Lehr et al., 1999) (Thor et al., 1999). Ki67 has been extensively examined and reported as a poor prognostic biomarker in invasive BC with its higher expression level (Domagala et al., 1996) (Trihia et al., 2003) (Viale et al., 2008) (Aleskandarany et al., 2010) (Brown et al., 2014) (Ingolf et al., 2014). Furthermore, various meta-analyses and systematic reviews have shown the independent prognostic importance of Ki67; for instance, de Azambuja et al. conducted a meta-analysis of 46 studies including 12,155 patients that assessed the effect of Ki67 on survival in early BC. They suggested that Ki67 positivity gives a worse survival and higher risk of relapse in early BC patients (de Azambuja et al., 2007). Another systematic review and meta-analysis of 85 studies in 32,825 patients concluded that it is associated with worse survival outcome in early BC (Stuart-Harris et al., 2008). Luporsi et al. evaluated studies of Ki67 extracted from papers published between January 1990 and July 2010 and included 71 references in their review. The results support the role of Ki67 as a prognostic marker (Luporsi et al., 2012).

As a predictive biomarker, the researchers in the Breast Cancer International Research Group (BCIRG) 001 trial classified the population into four BC subtypes: luminal A, luminal B, HER2 and triple-negative. Further, patients were assigned into two different regimens: FAC (fluorouracil, doxorubicin and cyclophosphamide) versus TAC (docetaxel, doxorubicin and cyclophosphamide). In this trial, Taxane therapy revealed a significant benefit for cases assigned to the luminal B subtype, with a 3-year disease-free survival rate of 85.2% versus 79.0% for FAC. Thus, high Ki67 expression could be a predictive biomarker for Taxane therapy (Hugh et al., 2009) (Dumontet et al., 2010).

In the PACS01 study, Ki67 was assessed in 798 patients with ER-positive BC for relapse linked to docetaxel. The results were 0.51 (95% CI 0.26–1.01) in both ER and Ki67 expressed tumours and 1.03 (0.69–1.55) in the ER-positive/Ki67-negative tumours. The researchers concluded that Ki67 expression recognises a subgroup: ER-positive BC patients who perhaps sensitive to docetaxel therapy (Penault-Llorca et al., 2009). As well, a study from the European institute of oncology utilising 1,241 patients categorised as luminal BC with one to three axillary lymph nodes showed that Ki67 predicted the benefit adjuvant chemotherapy additional to hormonotherapy in luminal B cases (Criscitiello et al. 2014).

In the neoadjuvant setting, Ki67 has been significantly linked to clinical or pathological responses. Measurement of Ki67 expression pre- and post-therapy offers a reflection of the responsiveness of BC to treatment. Ki67 index was predictive for pathological responses in various studies (de Azambuja et al., 2007) (Luporsi et al., 2012). Furthermore, Ki67 evaluation showed its clinical utility as a predictor of outcome for patients not achieving a pathologic complete response (Jones et al. 2009). However, Yerushalmi et al. concluded that, after reviewing a large number of studies, various methods for assessing Ki67 make comparisons difficult. They suggested developing a standard methodology and an accepted cut-off point for proper assessment of this biomarker (Yerushalmi et al., 2010).

Ki67 has been endorsed by the St Gallen consensus panel as a marker for definition of intrinsic subtypes of BC to differentiate between luminal A and luminal B subgroups (Goldhirsch et al., 2011) (Goldhirsch et al., 2013), and it is included in IHC4 (Cuzick et al., 2011). Ki67 was one of the variables, which assessed for relapse-free survival utilising a preoperative endocrine prognostic index (PEPI). This was developed using 228 postmenopausal BC cases and validated in an independent study of 203 postmenopausal women in the IMPACT trial. This study assessed five post-therapy variables: tumour size, node status, ER expression, histological grade and Ki67 index (Ellis et al., 2008). Ki67 was also used in Residual Proliferative Cancer Burden (RPCB) (Sheri et

al., 2015), which applied a formula that generated a sum of relative event rates for the post-therapy Ki67 index and the Residual Cancer Burden (RCB) (Symmans et al., 2007).

On account of the clinical relevance of Ki67, it is important to include it in the BC classification. However, the Ki67 biomarker is not readily available in many cancer centres and is not routinely used in a widespread manner in clinical practice, owing to numerous issues, including tissue sampling type, reproducibility and optimal cut-offs; many different cut-points for Ki67 have been reported, however, it has not been possible to determine or agree about an optimal one and cut-offs for Ki67 are still under debate. These reasons have led many investigators to try to optimise and standardise this crucial biomarker using different methods and cut-offs or to search for other techniques to determine Ki67, other than IHC (Kamel et al., 2012).

Structure, function and cellular localisation of Ki67

In 1983, Ki67 was identified by Gerdes et al. as a monoclonal antibody reacting with a nuclear antigen present in proliferating cells and absent in quiescent cells, the name was originated from the city of its discovery (Kiel, Germany) (Gerdes et al., 1983). Subsequently, the same group published that immunostaining of this antigen with a monoclonal antibody Ki67 presented a reliable means of evaluating the growth fraction of neoplastic cell group (Gerdes et al., 1984). As the antigen was not initially characterised, it was referred to as the Ki67 antigen. When the antigen was identified as a protein, it showed no homology to any known protein, therefore the initial name Ki67 was adopted (Scholzen and Gerdes, 2000). Ki67 is encoded by the *MKI67* gene. Schluter et al. cloned and sequenced the full length cDNA, identified two differentially spliced isoforms of mRNA. Double bands with apparent molecular weights of 345 and 395 kDa were detected in immunoblots of proteins from proliferating cells (Schluter et al., 1993).

Ki67 protein comprises a forkhead-associated domain (FHA), which are group of unrelated proteins identified by Hofman and Bucher through advanced sequence profiling methods (Hofmann and Bucher, 1995). FHA and a protein phosphatase 1 (PP1) domain (Moorhead et al., 2008) (Booth et al., 2014) present at the N-terminus of the protein while C-terminal part is enriched in leucine and arginine residue pairs, known as the LR domain (Takagi et al., 1999). The LR domain is able to bind to heterochromatin protein 1 (HP1), a well characterised particle as a mediator of heterochromatin formation. Ki67 is partially co-localised with HP1 at foci on chromosomes in an early interphase of cell cycle (Kametaka et al., 2002) (Scholzen et al., 2002). The central region of the protein contains a conserved domain (CD) and Ki67 repeats, 16 tandem repeats which contains a highly conserved named Ki67 motif (Schluter et al., 1993). **Figure-5** shows the different Ki67 protein domains (Sobecki et al., 2016).

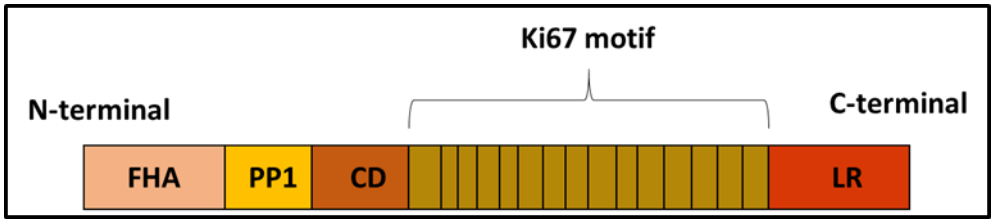


Figure-5: Ki67 structural elements
Forkhead-associated domain (FHA), protein phosphatase 1 (PP1) domain, conserved domain (CD, Ki67 motif, leucine and arginine residue (LR) domain

In 1987, a Ki67 score (positive cells/total tumour cells) was determined by Barnard et al. using 60 cases of primary BC (Barnard et al., 1987). They investigated the relationships between a number of clinical and pathological parameters where Ki67 was identified as a potential objective indicator of biological behaviour and of clinical significance. The expression of Ki67 increased with progression of cell cycle mainly in the second half of S-phase, compared with early S-phase, and achieved a maximum at G2 and M phases (Sasaki et al., 1987), **Figure-6**.

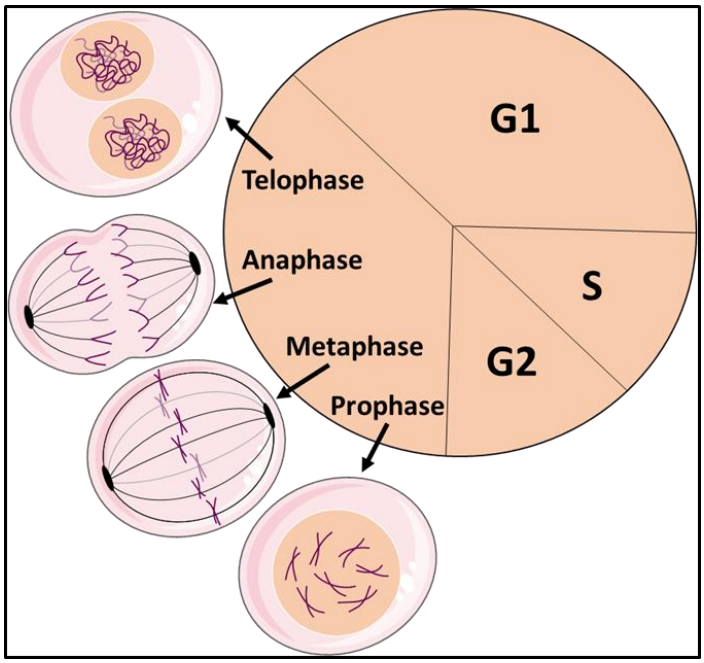


Figure-6: The diagram representing cell cycle phases

Despite intensive research about Ki67, little is known about the exact cellular function of this protein. However, Verheijen et al. studied the intracellular localisation of the Ki67 antigen in human cells showing that in interphase cells, the antigen is mainly present in the nucleoli predominantly localised in the nucleolar cortex in the dense fibrillar components (DFC) (Verheijen et al., 1989a). The DFC is one of three elements in the nucleolus beside the fibrillar centre (FC) and the granular component (GC). Subsequently, the same group suggested that the antigen may be associated with the non-histone proteinaceous structure known as the chromosome scaffold. In addition, they described different localisation of Ki67 antigen in relation to the cell cycle; at metaphase the antigen was distributed in a reticulate structure surrounding the condensed chromosomes while at late telophase it was observed as punctuated staining of the entire nucleoplasm (Verheijen et al., 1989b). During S phase, the expression of Ki67 expression increases and reaches the peak during metaphase and gradually decreasing thereafter (Verheijen et al., 1989a) (du Manoir et al., 1991).

The role of Ki67 in growth and cell division is supported by experiments revealing that the application of Ki67 antigen-specific antisense inhibits cell growth (Schluter et al., 1993). Furthermore, Bullwinkel et al. published that in

resting cells, Ki67 co-localises with components of the polymerase I transcription apparatus and also in proliferating cells, protein can be identified at sites linked to ribosomal RNA synthesis during interphase and mitosis (Bullwinkel et al., 2006). Moreover, studies demonstrate that the regulation of the Ki67 protein through the cell cycle is related to Ki67 phosphorylation through the activity of the cyclin B/cdc2 complex during mitosis. These post-translational modifications of the Ki67 protein are linked to the redistribution of the protein from the inside of the nucleus to the margin of the condensed chromosomes and vice versa. Further, it was revealed that the Ki67 protein is a member of the family of MPM-2 reactive phosphoproteins, which comprises proteins that are essential for the control and timing of mitosis (MacCallum and Hall, 1999) (Endl and Gerdes, 2000). Recently, additional mechanism was added to the role of Ki67 in cell growth and proliferation as a recent study shows that Ki67 is needed to maintain individual mitotic chromosomes dispersed in the cytoplasm after their release from the nuclear envelope. This function might be provided through a surfactant mechanism at the phase boundary between mitotic chromatin and the cytoplasm (Cuylen et al., 2016).

1.3.2. Online prognostic algorithms

Some prognostic algorithms have been established and published online. These algorithms utilise molecular biomarkers, predominantly ER, HER2 and Ki67, combined with other well-established prognostic parameters. They predict BC outcomes with regard to the probability of recurrence within a specific period of time. Their key aim is to predict who is likely to benefit from chemotherapy and who has to be saved from such toxic treatment. They are not precisely predictive and they do not have any diagnostic value.

1.3.2.1. Adjuvant! Online

Adjuvant! Online is a free and widely accepted prognostic and predictive online calculator (https://www.adjuvantonline.com/) for risk stratification of BC patients. It allows entry of data including a patient's age, tumour size, grade, co-

morbidity, menopausal status, stage, number of positive lymph nodes and ER status in order to predict mortality and disease recurrence at 10 years, and the potential benefit provided by adjuvant therapy. This is achieved by estimating a patient's risk of a negative outcome and then multiplying that by the proportion of negative measures that a given adjuvant therapy is recognized to prevent (Ravdin et al., 2001). The tool was established utilising a database that was recorded in the Surveillance Epidemiology and End Results (SEER) registry. This model has been validated in different studies (Campbell et al., 2009) (Hajage et al., 2011) (Bhoo-Pathy et al., 2012). However, it is drawn from published data and gives population-based rather than personalised information.

1.3.2.2. Predict

Predict (http://www.predict.nhs.uk/predict.html) is a mathematical online model, a prognostic tool to predict overall survival utilising cancer registration data recorded by the Eastern cancer registration and information centre (ECRIC). The study population was 5694 BC patients, validated with different set of 5468 patients documented by the West Midlands cancer intelligence unit (WMCIU) (Wishart et al., 2010). The prognostic effect of HER2 status was incorporated by Wishart et al. to produce the revised version, PREDICT+, using the British Columbia cohort which was applied to validate the original PREDICT (Olivotto et al., 2005) (Wishart et al., 2012). It forecasts the benefit of hormone treatment, trastuzumab and chemotherapy at 10-year time points. An improvement to the tool performance and clinical decision-making for ER-positive patients was achieved by including Ki67 in PREDICT (Wishart et al., 2014).

1.3.3. Gene expression profiling and molecular classification

Although the above-mentioned confirmed clinical and histological parameters reveal a potent relation with prognosis and outcome, there are increasing

concerns that these tools have a restricted role for assessing the diversity and heterogeneity of BC behaviour and could be insufficient to personalise management in BC patients. Therefore, the introduction of modern technologies that review huge numbers of genes in a particular assay, combined with complicated analytical programs, gives an opportunity for classifying BC into distinct groups relying on its gene expression profiles (GEP) (Sorlie et al., 2003, Sotiriou et al., 2003). Microarray-based gene expression profiling has facilitated simultaneous analyses of thousands of gene expression status in given tumour samples (Ioannidis et al., 2009). Through different methods, GEP has brought a new insight for a better understanding of the heterogeneity and behavioural differences in BC, the assessment of cellular heterogeneity in the form of molecular classification, and the development of some prognostic and predictive multigene signatures (Perou et al., 2000) (Sorlie et al., 2001) (Hu et al., 2006). GEP studies have revealed two key themes: the presence of the molecular subtypes of BC and the identification of ER-positive and ER-negative BC, as they have fundamental differences in the transcriptomic level (Perou et al., 2000) (Lusa et al., 2007) (Sotiriou et al., 2006). Initially, BC classified into four distinctive molecular groups utilising the cDNA microarray analyses of 38 invasive BCs, three normal breast samples, one ductal carcinoma in situ and one fibroadenoma (Perou et al., 2000). These groups are: luminal group, HER2-positive group, basal-like group and normal breast-like group. Further studies divided the luminal group into two or three subgroups, mostly luminal A and luminal B subtypes, which vary in clinical outcome and biological processes (Sorlie et al., 2001) (Sorlie et al., 2003) (Hu et al., 2006). Moreover, another type -claudin low type- was identified in 2007 (Herschkowitz et al., 2007).

Luminal A is the most common subtype, constituting 50–60% of the total. Its expression patterns resemble normal luminal epithelial cells, comprising ER, genes associated with an active ER pathway and a low demonstration of cell proliferation genes (Perou et al., 2000). The luminal A IHC profile is featured in the expression of ER and cytokeratin CK8/18, with an absence of HER2 expression and a low histological grade. Luminal A tumours have a higher level of PR expression than luminal B tumours (94% vs 74%) (Bastien et al., 2012).

Patients with this subgroup of cancer have a good prognosis with relapse rate of 27.8% and longer survival from the time of relapse than for other subtypes (Kennecke et al., 2010). The luminal B subtype, which represents 10- 20% of all BCs, is a more aggressive phenotype and a higher histological grade than luminal A. Moreover, it has a higher expression of proliferative genes, a worse prognosis and lower survival from time of relapse (Kennecke et al., 2010) (Sorlie et al., 2003) (Geyer et al., 2009). The essential difference between the two subtypes lies the changes in the expression of the proliferation genes mainly, *MKI67*, which can be translated at protein level using immunostaining of Ki67 as a possible marker (Cheang et al., 2009). However, immunohistochemical assessment of Ki67 in BC remains controversial and has yet to be standardised, thereby adding a variability factor in the evaluation of this biomarker. Considering this marker is the most widely used for measuring cell proliferation, more efforts are needed to reach a consensus on evaluating it and, subsequently, in using it in BC classification.

The HER2 group is distinguished by high expression of the HER2 gene and comprises 15-20% of all BCs. Although these tumours display an overexpression of proliferative genes, the group does not express genes of the basal-like group and, further, it can exhibit a low expression of characteristic luminal genes. Only 70% of HER2-positive tumours using the microarray have the protein over-expressed by IHC. Therefore, at the protein level, the ER-negative/HER-positive immunostaining group does not reproduce perfectly in the intrinsic subtype (Prat and Perou, 2011). Moreover, a fraction of the HER2 subgroup is not classified as HER2 subtype by microarray-related methods, but is actually categorised as a luminal B subtype (Hu et al., 2006) (Parker et al., 2009). Staaf et al. divided HER2 expression tumours into three subtypes with divergent biologic characteristics independent of ER status. One cluster had significantly worse survival than the other two, based on gene expression clustering. Further, they constructed a HER2-derived prognostic predictor (HDPP) that separated these HER2-positive tumours into two groups with significantly different prognoses using 158 genes. HDPP was not directly associated with the expression of genes from the HER2 pathway or proliferation

genes; rather, it was for genes associated with immune response, tumour invasion and metastasis (Staaf et al., 2010).

The basal-like subtype represents 10–20% of all breast carcinomas and expresses genes generally present in basal/myoepithelial cells: cytokeratin as CK5 and CK17, P-cadherin, caveolin 1 and 2, nestin, CD44 and HER1. Additionally, the genes of luminal epithelium are expressed, but with lower levels than the luminal class such as CK8/18 (Choo and Nielsen, 2010). At the protein level, the basal-like group can be identified using IHC profiling by the selection of five markers: ER, PR, HER2, HER1 and CK5. These markers categorise this subtype with a sensitivity of 76% and specificity of 100% (Nielsen et al., 2004). Although in clinical practice, the terms triple-negative and basal-like are often substituted, the basal-like subtype shows distinct clinical and immunophenotypic differences (Rakha et al., 2009). Significantly, a study by Lehmann et al. described six different molecular triple-negative subtypes, including two basal-like, a mesenchymal, a mesenchymal stem-like, a luminal androgen receptor and an immune-modulatory subtype (Lehmann et al., 2011). The basal-like subtype has a high rate of P53 mutation; this could be the reason for its aggressiveness and worse prognosis (Sorlie et al., 2001), with a higher relapse rate in the first 3 years (Dent et al., 2007).

The normal breast-like group is still poorly characterised and can be accounted for in the classification with fibroadenoma and normal breast samples (Perou et al., 2000). These can be classified as triple negative because of the absence of ER, HER2 and PR expression, without being considered basal-like as they are negative for CK5 and HER1; further, they express genes that are related to adipose tissue (Peppercorn et al., 2008). It is presumed that this subgroup is a technical artefact from high contamination with normal tissue (Weigelt et al., 2010c). The claudin-low subtype has low levels of claudin and E-cadherin expression. This subtype is characterised by high activity of specific genes closely linked to epithelial mesenchymal transition (EMT), cancer stem cell-like features and immune response processes, and lower activity of MYC and PI3K pathways (Hennessy et al., 2009) (Prat et al., 2010) (Sabatier et al., 2014).

Gene expression signatures

Certainly, molecular subtype classification advances prognostic and predictive information for standard clinical variables in BC. Furthermore, it increasingly reflects how different gene expression signatures have to be applied to predict outcomes for various subgroups of BC. Several forms of multigene classifiers have been developed; including:

1.3.3.1. Oncotype DX

Oncotype DX, (21-gene Genomic Health Recurrence Score [GHI-RS]), a multistep approach was followed to produce a commercially obtainable assay of 21 prospectively selected genes for invasive BC, based on RT-PCR. The initial step utilised RT-PCR to examine the gene expression from FFPE tissue sections. The initial 250 genes were selected from genomic databases, literature review and microarray data (using fresh-frozen tissues). Using the results of these studies, a panel of 21 genes (16 cancer-related and 5 other reference genes) was selected for measuring the possibility of distant recurrence from FFPE tissue samples. An algorithm was designed to measure a recurrence score (RS) for every sample depending on the levels of expression of the genes. The cancer-related genes include genes from the ER group (ER, PR, BCL2 and SCUBE2), invasion group (CTSL2 and MMP11), HER2 group (HER2 and GRB7), cell proliferation group (Ki67, CCNB1, Survivin, STK15 and MYBL2) and GSTM1, BAG1 and CD68. The reference genes include ACTB (β-actin), GAPDH, RPLPO, GUS and TFRC. The RS of these 21 genes is expressed as a continuous variable ranging from 0 to 100, where higher scores reflect a greater possibility of recurrence. The RS is classified into three categories: low risk group (<18), intermediate risk group (≥18 <31) and high risk group (≥31) (Paik et al., 2004). The Oncotype DX BC assay predicts the 10-year risk of BC recurrence in newly diagnosed patients with ER-positive and node-negative early stage BC, which treated with tamoxifen. Oncotype DX can provide further information, including to what extent the patient would benefit from chemotherapy and tamoxifen therapy (Paik et al., 2006).

1.3.3.2. MammaPrint

The MammaPrint (70-gene signature) classifier was developed by Van't Veer and colleagues using the microarray analysis of 78 frozen samples from lymph node-negative patients aged under 55 years with tumour size smaller than 5 cm. This classifier divides BC patients into two groups in terms of distant metastasis-free survival. The good profile group comprises patients more likely to remain free of distant metastasis, while patients who are classified in the poor profile group have a high risk of developing distant metastasis (van de Vijver et al., 2002). The US Food and Drug Administration (FDA) approved MammaPrint for marketing in 2007 to be used with other clinicopathologic parameters as a prognostic test (Liu et al., 2008a). However, this test needs fresh tumour tissue, with at least 30% malignant cells, which then has to be sent to the company within five days after tumour resection in the kit's shipping medium (Rakha and Ellis, 2011), which is impractical and which makes it difficult for enrolment in daily laboratory assays. Furthermore, it has limited clinical use in ER-negative BC as only up to 4% are classified as having a good prognosis (Cardoso et al., 2008).

1.3.3.3. Genomic Grade Index (GGI) (MapQuant Dx)

Sotiriou et al. used five datasets of gene expression with a total of 661 BC patients. They examined the association between the histologic grade and gene expression profiles of BCs to improve histologic grading by measuring the expression of 97 genes. Therefore, instead of the classic grades I, II and III, GGI divides histologic grades into low and high risk. This signature supports the view that low-and high-grade disease can be sub-classified ; that is, grade II BC into grade I-like (GGI low grade) and grade III-like (GGI high grade). Only 9% histological grade I cancers are stratified as high GGI and 14 % of histological grade III cancers as a low GGI (Sotiriou et al., 2006).

1.3.3.4. Breast Cancer Index (BCI)

BCI is an index model of the ratio of expression of two genes, HOXB13 and IL17BR (H: I), joined with Molecular Grade Index (MGI) to test the risk of recurrence in early invasive, ER-positive and LN-negative BC patients. Initially, a genome-wide microarray analysis of frozen tumour samples of 60 ER-positive BC patients subjected to tamoxifen therapy only was applied (Ma et al., 2008). Then, a two-gene expression ratio was recognised, which is highly predictive of patient outcome. This expression ratio was revised to use analysis depending on the standard FFPE, which was established utilising an independent set of 20 FFPE tissue samples (Ma et al., 2004). The BCI risk score ranges from 0 to 10 and categories the patients into three groups: low-risk group when the score is <5 while between 5 and 6.3 represents the intermediate-risk group. High-risk group is ≥6.4 (Jerevall et al., 2011). Subsequent studies have revealed that both genes (HOXB13 and IL17BR) have prognostic and predictive values (Ma et al., 2006) (Goetz et al., 2006) (Jerevall et al., 2008).

1.3.3.5. CSC signature; Invasiveness gene signature (IGS)

The main evidence for the presence of CSCs in BC lies in the study conducted by Al-Hajj et al., which demonstrated that $CD44^{+}$/ $CD24^{-/low}$, where CD24 is a cell surface glycoprotein and CD44 is a cell surface receptor for hyaluronan, human BC cells-transplanted in immune-deficient mice were more tumourigenic compared to other tumour cells. Moreover, tumours induced by these $CD44^{+}/CD24^{-/low}$ cells display the heterogeneity of the genuine tumour. The $CD44^{+}/CD24^{-/low}$ cells have a capability for serial transplantation, given by evidence that they have characteristics of CSCs. Moreover, the tumour initiation potential of $CD44^{+}/CD24^{-/low}$ cells can be developed by splitting them further, depending on the epithelial surface antigen (ESA) expression: $ESA^{+}/CD44^{+}/CD24^{-/low}$ cells are more tumourigenic compared to $ESA^{-}/CD44^{+}/CD24^{-/low}$ cells (Al-Hajj et al., 2003).

Gene expression profiling of this small subpopulation ($CD44^{+}/CD24^{-/low}$) of tumourigenic BC cells was used to develop prognostic tools to evaluate survival in patients with BC. Liu et al. identified 186 genes based on significant differences of expression between tumourigenic BC and normal breast cells. They developed an invasiveness gene signature (IGS) (Liu et al., 2007) that varied substantially from previously described gene signatures in BC. The correlation between the IGS and both cancer recurrence and overall survival was significant, not only in BC cases, but also in patients with lung cancer, medulloblastoma and prostate cancer.

1.3.3.6. PAM50 (Prediction Analysis of Microarray 50)

PAM50 is a gene expression assay using 50 genes. Also called NanoString Prosigna assay, it represents a reduced gene set developed on a quantitative RT-PCR. It can classify BC samples into intrinsic subtypes, including luminal A and B, HER2-enriched, basal-like and normal breast-like from FFPE tissue. Moreover, PAM50 generates risk-of-relapse (ROR) for the patient's distant recurrence possibility by calculating the molecular subtype associations, a subset of genes reflecting proliferation and tumour size. The automatic software applies all of the quality thresholds to the given data. ROR and the intrinsic subtype are considered as prognostic indicators of the risk of disease recurrence for BC; both are calculated using a clinically validated algorithm. Either frozen or FFPE samples can be utilised in multiplexed gene-specific probes to measure gene status. PAM50 comprises proliferation genes and markers of basal phenotype and can also reflect the underlying biology associated with the ER and HER2 pathways. PAM50 ROR scores show prognostic value in both untreated and tamoxifen-treated patient populations (Parker et al., 2009) (Nielsen et al., 2010).

The main aim of GEP is to classify tumours into clinically meaningful categories, which can offer both prognostic and therapeutic information. However, making these options universally accessible and cost-effective

worldwide is also required. The cost of the previously mentioned signatures is high compared to IHC: about $4,000 for MammaPrint and $3,500 for Oncotype DX. These assays are not only expensive, but their benefits could be restricted in cases with intermediate features (Allison, 2012). A further limitation of first-established genomic signatures is that nearly all patients with ER-negative BC have been allocated to the high-risk group. Another important disadvantage of applying genomic research on a daily clinical basis is that these trials were based on small samples and, subsequently, have a lack of validation. For this reason, more recent research concentrates on using a larger number of samples and more genetic analysis than these initial trials, thereby generating a genomics scene of BC (Ellis and Perou, 2013).

With the current development of proteomic profiling, there is no reason to limit this technology to RNA. Proteomics can be regarded not only as a complement to the wealth of knowledge generated by genomics, but can also necessitate personalised management in BC for several reasons. RNA levels do not always necessarily correlate with counterpart protein levels (Gygi et al., 1999) (Ginestier et al., 2002) (Nishizuka et al., 2003). Post-translational modifications, including phosphorylations, glycosylations and acetylations, are an additional complexity and changes can alter the physiological activity of protein. These modifications are not detectable at the RNA level; however, they have important roles in protein localisation, interactions, stability and functions. Furthermore, proteins are more representative of the current state of the cellular microenvironment and are more accessible and applicable as therapeutic targets. Therefore, proteomic profiling can probably give clearer answers to functional and pharmacological questions.

IHC is an alternative robust, affordable and more accessible method and can easily be implemented into the standard pathology workflow. It is an important technique to determine surrogate protein biomarkers for the classification of BC. The next section will highlight the IHC-based indices.

1.3.4. Immunohistochemistry-based indices

Recently, as an alternative of using a single marker approach, panels of biomarkers have been examined to predict the prognosis and responses to specific treatment. This approach was introduced following the successful classification of BC using GEP, and the introduction of molecular classification based on the status of several genes with clinical importance. Multiple IHC biomarkers are utilised in combination to offer routine cost-effective replacements to GEP. Examples of IHC-based assays include Mammostrat, IHC4 and NPI+.

1.3.4.1. Mammostrat

Ring et al. designed a multiple marker test, utilising genes depending on IHC assessment, which examines the possibility of developing an IHC-based tool by using data from different gene expression researches. In their study, three retrospective BC cohorts were enrolled. The first was a discovery cohort (n=466), while the other two were independent validating cohorts (n=299 and 344 patients, respectively). The primary study utilising conventional FFPE samples led to an IHC assay which calculated a relative risk of recurrence; currently, this is commercially offered as the Mammostrat assay (Ring et al., 2006). Further validation of Mammostrat has been presented by other investigators (Ross et al., 2008) (Bartlett et al., 2010). Mammostrat contains five IHC markers -P53, NDRG1 (N-myc downstream-regulated gene 1), SLC7A5 (solute carrier family 7 cationic amino acid transporter), CEACAM5 (carcinoembryonic antigen cell adhesion molecule 5) and HTF9C (HpaII tiny fragments locus 9C) - which stratify ER-positive tamoxifen-treated BC patients into three risk categories. Prognostic index 0 represents the low-risk group; prognostic index >0 and ≤0.7 reflects the moderate-risk group; and prognostic index >0.7 represents the high-risk group (Ring et al., 2006) (Ross et al., 2008).

1.3.4.2. IHC4

IHC4 is a prognostic score developed by Cuzick et al. which evaluates the levels of four widely assessed proteins in BC (HER2, ER, PR and Ki67). They compared it to the Oncotype DX to assess its feasibility on 1125 ER-positive cases that had GHI-RS data and whether it can improve a prognostic/predictive value to the classical prognostic variables (lymph node status, tumour burden and histological grade) in early stage BC patients. IHC4 score is recognised as an independent prognostic tool along with the existing classical variables. Importantly, results presented by IHC4 score are found to be akin to those produced by Oncotype DX. In addition, the IHC4 prognostic value was validated on an independent cohort of 786 patients: their outcomes were the same as evaluated by both Oncotype DX and IHC4 assays (Cuzick et al., 2011).

Although previous prognostic assays are affordable, robust and IHC-dependent, they address the question regarding treatment use in hormone receptor-positive, node-negative BC only. The available prognostic tools vary in terms of compatibility to fit patients with different characteristics. At present, no particular tool based on validated biomarkers can provide a comprehensive overview of the molecular and pathological characteristics relevant to all BC patients. The combination of both clinical and molecular variables to fit all patients, regardless of ER positivity or nodal status, is necessary to meet precise management and to avoid overtreatment. Modernisation of the existing NPI through inclusion of relevant biological data to provide personalised decision-making tools for BC (NPI+), has occurred and details have been recently published.

1.3.4.3. Nottingham Prognostic Index plus (NPI+)

Although the NPI is still valid and extensively used, it does not predict the full heterogeneity of outcome currently observed in BC patients. NPI+ is the updated version used to combine molecular phenotype information with clinical and pathology information, to provide a comprehensive test suitable for all BC patients. Recently, the updated NPI+ has been published as a combination of two steps. The initial step is based on the determination of the biological class of the tumour, by evaluating the expression level of a panel of biomarkers. The second step which is the analysis of traditional clinicopathologic prognostic parameters producing the structured NPI-like formulae for each class (Green et al., 2013) (Rakha et al., 2014). The biomarkers included in the biological classification are ER, PR, HER2, CK5/6, CK7/8, P53, HER1, HER3, HER4 and Mucin 1, and play a significant role in BC, especially those markers used in routine clinical practice. These biomarkers were selected from a large group of biomarkers as the minimum number of markers needed to preserve molecular classification using 1,076 BC samples (Abd El-Rehim et al., 2005).

The molecular classes (**Figure-7**) that resulted in the first step of the NPI+ are three luminal classes, generally identified by high expression of both CK7/8 and ER. The luminal classes are divided into luminal A, luminal B and luminal N subgroups. While luminal A and luminal B have a high level of expression of ER, CK7/8, HER3 and HER4, luminal N is negative for HER3 and HER4 but positive for ER and CK7/8. Both luminal A and B can be separated by a different level of expression of PR. Two basal classes were recognised by the negativity for the essential clinical biomarkers: ER, PR and HER2, low luminal cytokeratin and high expression of basal marker CK5/6. The basal subclasses are further allocated into two groups according to p53 status: basal–p53 normal, where the expression of p53 is low, and basal–p53 altered, in which there is a high expression. Lastly, the HER2 class is recognised by HER2 overexpression, while demonstrating heterogeneity in the hormonal receptors' expression. Hence, they

are separated into two subgroups depending on ER expression: the HER2-positive/ER-positive group, in which both HER2 and ER are expressed; and the HER2-positive/ER-negative group, where there is no expression for ER. In these studies, the basic classification is identical to the genotypic biological classes; however, NPI+ contains novel subtypes of basal and luminal BC. These distinctive classes of BC show a significant correlation with patient outcome. The advantage of NPI+ compared to IHC4 and Mammostrat is that it has the ability to be applied to all women diagnosed with invasive BC and aged 70 or under. On the other hand, IHC4 and Mammostrat address the question regarding risk of recurrence using hormone receptor- positive, lymph node-negative BC cases only. Furthermore, IHC4 compatibility is restricted to tumour size, which has to be less than 5cm and to postmenopausal patients.

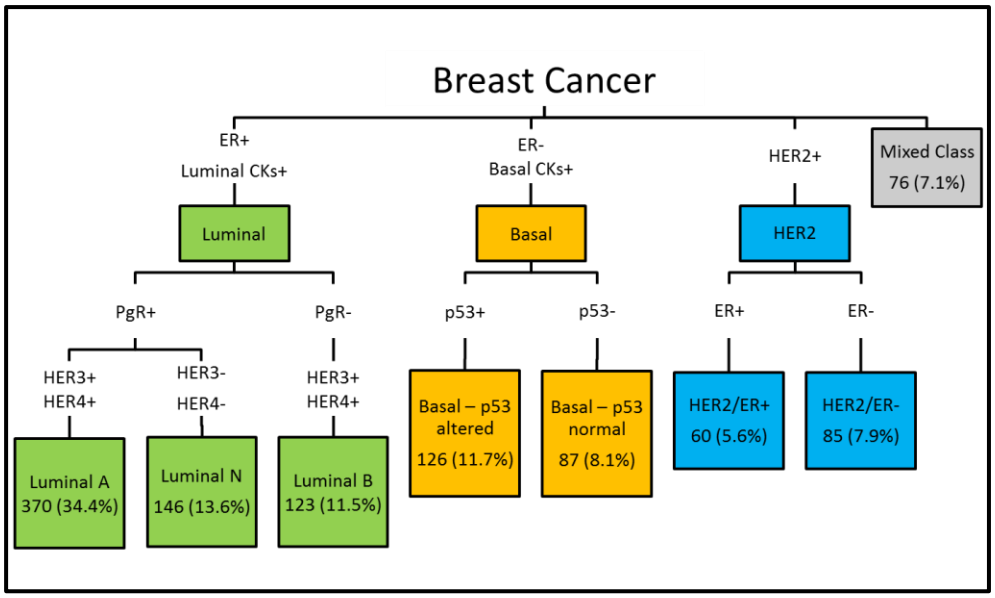


Figure-7: NPI+ biological classes
, classification and proportions of cases (Green et al., 2013)

The NPI+ prognostic score was calculated using modified NPI-like formulae that were previously established in each NPI+ biological class of the Nottingham series. These formulae originated from the biological classes (Green et al., 2013) and were consequently refined using the updated biological classification (Soria et al., 2013). For these calculations, the available clinicopathological variables were used (Rakha et al., 2014). These were determined by using the β-values

produced by a Cox regression model in predicting BC-specific survival (BCSS) with the confirmed histopathologic prognostic variables. These consist of: nodal ratio, lymph node stage, pathological tumour size, tubule formation, nuclear pleomorphism and mitotic counts. These indicators are the most significant variables in the Nottingham series that impact on survival, depending on their β-values in the Cox regression, and represent the magnitude of the effect of the hazard. NPI+ prognostic groups were assigned using the categorical cut-points derived from the Nottingham series in each of the NPI+ biological classes (Rakha et al., 2014). However, the NPI+ does not incorporate proliferation which would complement the robust biomarker panel of the index, giving rise to further novel subclasses or the refinement of the existing classification. Moreover, although IHC is robust and affordable, easily implemented into standard pathology workflow, and has the advantage of providing information about the cellular localisation of the specific protein in tumour samples, it is a semi-quantitative technique and cannot detect subtle changes compared with novel emerging technologies. The growing application of new proteomic technologies can possibly facilitate further discoveries in BC research. A huge amount of work that has been fulfilled to develop and improve the BC molecular prognostic assays, molecular testing remains in development.

1.4. Hypothesis

There are increasing evidence that current prognostic parameters have limited value in assessing the diversity and heterogeneity of BC behaviour, and are insufficient in providing appropriate personalised management for BC patients. Tumours sharing different clinical parameters respond differently to treatment. In this study we hypothesised that using modern applicable techniques and incorporation of additional molecular biomarkers into existing clinical classification systems can refine prognostic taxonomy of BC and improve personalisation of treatment.

1.5. Aims of this study

This study aimed to investigate some relevant potential prognostic markers that can improve BC prognostic classification in the context of combined molecular and morphological prognostic BC taxonomy and further to focus on assessment and updating the performance of the existing prognostic indices to the recent advancement in molecular classification of BC. In this study, NPI and NPI+ indices were used as a testing tool for this purpose.

The specific aims were:

- To evaluate Ki67 expression and its integration into the prognostic indices, NPI and NPI+.
- To evaluate the use of a high-throughput proteomic technique (RPPA) in quantification of protein expression using FFPE for BC classification.
- To assess vital biomarkers using large cohorts on both proteomic and transcriptomic levels.
- To determine the biological relevance of BC with low ER expression using both protein and transcriptome techniques.
- To investigate the incorporation of a novel CSC factor in BC prognostication.

2. Material and Methods

2.1. Patient series

2.1.1. The first cohort (Primary Series, version one)

This series comprised 1954 patients diagnosed between 1988 and 1998 whose breast cancer tissues were prepared as TMA as part of Nottingham Tenovus Primary Breast Carcinoma Series. This is a consecutive well-characterised series of early stage primary operable breast cancer patients aged 70 years or less. The TMA was previously constructed using retrieved FFPE archival tissue blocks and has been extensively evaluated with a wide range of biological markers. In this cohort, patients' outcome, including regional and distant events, survival and time to the event, were recorded and updated regularly. This included breast cancer specific survival (BCSS), which was defined as the time from the date of the primary surgical intervention to the time of death from breast cancer. Moreover, this cohort includes the clinical details of the patients as age and menopausal status and also includes the tumour details: tumour size, tumour grade, nodal stage, lymphovascular invasion (LVI) and NPI, **Table-3**.

2.1.2. The second cohort (Core Biopsy ER series)

This comprised a consecutive series of symptomatic and screen detected breast cancer patients who had ER status assessed on a preoperative core needle biopsy (CNB) at Nottingham City Hospital in routine practice between March 2008 and November 2014 (n=3649 cases). CNBs were fixed, processed and stained following a standardised protocol, as previously described (Hodi et al., 2007).

Table-3: Clinicopathological characteristics of the first cohort

Parameter	Number (%)
Patient Age	
≤ 50 years	669(34.2)
> 50 years	1285(65.8)
Menopausal Status	
Premenopausal	740(38.0)
Postmenopausal	1209(62.0)
Tumour Grade	
1	359(18.5)
2	652(33.7)
3	925(47.8)
Pleomorphism	
1	52(2.8)
2	746(40.1)
3	1063(57.1)
Tubule Formation	
1	119(6.4)
2	620(33.2)
3	1127(60.4)
Mitotic Figures	
1	701(37.6)
2	347(18.6)
3	818(43.8)
Tumour Size	
≤ 2 cm	1010(52.0)
> 2 cm	931(48.0)
LN status	
1 (0 Positive nodes)	1085(63.4)
2 (1-3 Positive nodes)	514(30.0)
3 (>3 Positive nodes)	113(6.6)
NPI	
Good NPI (<3.4)	638(34.3)
Moderate NPI (3.41-5.4)	970(52.1)
Poor NPI (≥5.4)	253(13.6)
LVI	
Negative	1327(68.8)
Definite	602(31.2)
ER status	
ER Negative	474(25.1)
ER Positive	1418(74.9)
PR status	
PR Negative	744(40.6)
PR Positive	1090(59.4)
HER2 status	
HER2 Negative	1614(87.6)
HER2 Positive	229(12.4)

2.1.3. The third cohort (Primary Series, version two)

The third cohort comprised 1825 BC cases diagnosed between 1998 and 2006 which prepared as TMA. This is a consecutive well-characterised series of early stage primary operable breast cancer patients aged 70 years or less. In this cohort, the clinical details of the patients including age, tumour grade, tubular formation, pleomorphism, mitotic frequency, tumour type, stage, LVI, NPI, ER, PR and HER2, **Table 4**.

2.1.4. The fourth cohort (METABRIC data set)

In this cohort, total RNA was extracted using fresh-frozen BC cases. RNA integrity and DNA were analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The extracted and purified DNA probes were hybridised to Affymetrix SNP 6.0 arrays (Affymetrix) at the QC criteria established by AROS Applied Biotechnology (Aarhus). Total RNA was biotin-labelled using the Illumina TotalPrep RNA Amplification kit (Ambion). For each hybridisation, Biotin-labelled RNA (1.5 μg) was used on Sentrix Human-6 BeadChips (Illumina) as described in Curtis et al. (Curtis et al., 2012b). The resulted gene expression data was containing 47,293 transcripts from the microarray experiments and were statistically analysed by applying the Linear Models for Microarray Data (LIMMA) inclusive software package that is compatible with Affymetrix data. LIMMA software package comprised Student's t- test, base 2 logarithms of fold changes between the normal and tested samples, log2FC, average expression measurements, t-test values, *p*-values and log-odds of differential expression, B-values as its primary statistical outputs. In this series, the survival data and the ER status defined using immunohistochemistry was available for 262 cases from Nottingham. **Table-5** showed the clinicopathological characteristic of this series.

Table 4: Clinicopathological characteristics of the third cohort

Parameter	Number (%)
Patient Age	
≤ 50 years	524(29.3)
> 50 years	1264(70.7)
Tumour Grade	
1	284(15.9)
2	723(40.5)
3	778(43.6)
Pleomorphism	
1	25(1.4)
2	533(30.1)
3	1214(68.5)
Tubule Formation	
1	131(7.4)
2	530(29.9)
3	1111(62.7)
Mitotic Figures	
1	886(50.1)
2	332(18.8)
3	552(31.2)
Tumour type	
Ductal (including mixed)	1537(86.2)
Lobular	150(8.4)
Medullary-like	13(0.7)
Miscellaneous	11(0.6)
Special type	72(4.0)
Stage	
1	1124(63.0)
2	482(27.0)
3	179(10.0)
NPI	
Good NPI (<3.4)	622(34.9)
Moderate NPI (3.41-5.4)	881(49.4)
Poor NPI (≥5.4)	279(15.7)
LVI	
Negative	1298(72.7)
Definite	488(27.3)
ER status	
ER Negative	352(20.9)
ER Positive	1334(79.1)
PR status	
PR Negative	699(42.4)
PR Positive	950(57.6)
HER2 status	
HER2 Negative	1570(91.3)
HER2 Positive	149(8.7)

Table-5: Clinicopathological characteristics of the four cohort

Parameter	Number (%)
Patient Age	
≤ 50 years	424(21.9)
> 50 years	1511(78.1)
Menopausal Status	
Premenopausal	436(22.1)
Postmenopausal	1533(77.9)
Tumour Grade	
1	170(9.0)
2	770(40.7)
3	952(50.3)
Tumour Size	
≤ 2 cm	858(43.8)
> 2 cm	1102(56.2)
LN status	
1 (0 Positive nodes)	1035(52.5)
2 (1-3 Positive nodes)	622(31.5)
3 (>3 Positive nodes)	316(16.0)
NPI	
Good NPI (<3.4)	680(34.3)
Moderate NPI (3.41-5.4)	1101(55.6)
Poor NPI (≥5.4)	199(10.1)

2.1.5. The fifth cohort (Nottingham prospective series)

The series of 143 cases of primary early invasive breast cancer FFPE tissues were collected from patients treated at Nottingham City Hospital between 2004 and 2011. This subset was selected randomly. Evaluation of this series included reassessment of the blocks and substitution with the more suitable ones depending on reviewing of H&E staining slides.

The first four cohorts are independent cohorts. These cohorts used to validate the results according to each chapter hypothesis. However, regarding the fifth cohort, it was chosen separately because of the limited availability of resources of the other series.

2.2. TMA construction

The TMA was selected as a suitable technique to assess large numbers of samples. If the diameter of the core is 0.6 mm, more than 100 samples can be embedded in a single multicore paraffin block; these cylindrical core samples are extracted from the most representative tumour areas of the donor blocks. The advantage of this method is that it can offer the same experimental conditions and subsequent analysis for large numbers of tissue. Moreover, it saves time, money and materials as well as donor blocks.

2.2.1. Steps of TMA construction

2.2.1.1. Collection and selection of donor blocks

Donor blocks were retrieved from the blocks-storage area, then arranged sequentially depending on identification number. The thickest donor block with the large invasive tumour area was selected for TMA. However, if there was no sufficient material in the block, the area of the invasive tumour was inadequate, or the sample was a core biopsy samples, then the block was discarded. If there was only one block, then it was preserved as a diagnostic block.

2.2.1.2. Selection of tumour areas for TMA

One section was taken for every selected block, then each slide was stained using haematoxylin and eosin stain (H&E) to select the most suitable areas for TMA. The slides were reviewed under light microscope for these representative areas (the invasive borders of the tumour were selected while the areas of necrosis, fibrosis, haemorrhage and DCIS were neglected). Subsequently, each slide was matched with the corresponding paraffin block; then, the slides were sent for digital scanning.

2.2.1.3. Preparation of recipient blocks

Cassettes were well-mounted with melted regular paraffin. After cooling, the cassettes were assessed for any air bubbles or striations within the blocks. Trimming of block borders using microtome was essential to parallel the block phase to the base of the cassette; all blocks had to be the same height. The recommended thickness for the recipient block was 5 to 10 mm.

2.2.1.4. Digital scanning and marking of the tumour areas

Digital marking was done using the Pannoramic Digital Viewer 1.15.4 (3DHISTECH). The H&E slides were scanned using the Pannoramic 250 Flash II Scanner. After uploading the slides and displaying information relating to them, the TMA master marker was created. The diameter of the marker was chosen to correspond to the size of the puncher in the TMA machine. The slides containing master markers were used with the TMA marker application to decide which areas on the donor blocks would be transferred to the recipient blocks. Digital marking of the slides was mainly done by Dr Mohammed Aleskandarany.

2.2.1.5. TMA block creation using TMA grand master machine

After starting the TMA control software, drilling was performed, the puncture was inserted and the TMA project was created. The donor blocks' IDs were imported and then the donor blocks as well as the empty recipient blocks were inserted into the machine. Thereafter, TMA layout was designed: the number of columns and rows were specified, including orientation. Consequently, the tissue cores were selected for extraction guided by the matched digital slides. Finally, the data was exported, including the TMA maps. In every run, 72 blocks were (60 donor blocks and 12 recipient blocks) were loaded.

2.3. Immunohistochemistry

IHC was used as a semi-quantitative assessment of protein expression in FFPE tissue. FFPE sections of 4µm thickness were used throughout.

2.3.1. Antibody selection

The selection of the antibody used was based on online search of the available antibodies using manufacture information sheets. Furthermore, it was depending on review of published articles and relevant websites including Human Protein Atlas (http://www.proteinatlas.org/).

2.3.2. Antibody optimisation

Optimisation was performed using FFPE sections from tissues known to express the epitope or using TMA breast tumour sections. Sections were prepared 24 hours before use. Antibody dilution, incubation time, antigen retrieval (AR) were all optimised to ensure assessable staining.

2.3.3. Controls

Positive and negative controls were included in every experiment. Rabbit polyclonal β_2-microglobuline (Dako) was used as a positive control of the IHC experiment at the concentration of 1:2000 while a negative control was included where only Tris buffered saline (TBS) was used.

2.3.4. Steps of IHC procedure

2.3.5. Manual staining

2.3.5.1. Slide heating

Slides were placed on a 60˚C hotplate for 10 minutes, to melt paraffin and affix tissue sections and left to air cool for 5 minutes.

2.3.5.2. Deparaffinisation and rehydration

Slides were passed through two baths of xylene for 5 minutes each followed by washing with industrial methylated spirit (IMS 100% concentrations- the product code: M/4450/25) (Thermo Fisher Scientific) 3 times for 2 minutes each using Leica strainer. Slides were rehydrated for 5 minutes three times in water.

2.3.5.3. Antigen retrieval

Slides were submerged in citrate at pH 6.0. Testing of the pH was performed in each experiment using the pH meter, different pH buffer values including 4.0 and 10.0 were tested for higher accuracy. Whirlpool JT359 Microwave (1000 W) heat assessed retrieval for 20 minutes at 98˚C was applied with citrate buffer, followed by immediate cooling under running tap water at room temperature for 5 minutes.

2.3.5.4. Peroxidase and protein blocking

Manual IHC staining was conducted using the Novolink Max Polymer Detection Kit (Leica). Blocking of endogenous peroxidase activity was performed by applying peroxidase block (Novolink kit) for 5 minutes. Slides were washed by TBS twice 5 minutes each. After that, protein blocking (Novolink kit) was added for 5 minutes, followed by washing in TBS for 5 minutes twice.

2.3.5.5. Application of a primary antibody and Novolink Polymer

Antibody was applied to the sections. Afterwards, sections were washed by TBS for 5 minutes twice. Then, 100 µl of the post primary block (Novolink kit) was added for 30 minutes. Followed by 100 µl Novolink polymer for 30 minutes, before and after the polymer application two washes by TBS for 5 minutes each was done.

2.3.5.6. DAB application

The DAB (Novolink kit) working solution was prepared in the dark area by adding 1:20 DAB chromogen to DAB substrate buffer. 100 µl of DAB working solution was added followed by two TBS washings each for 5 minutes. Following this step, nuclear staining was achieved using 100 µl of Novolink hematoxylin for 6 minutes.

2.3.5.7. Dehydration and coverslipping

In this step, the slides were loaded to the Leica machine. The slides were passed through IMS for 2 minutes three times, then in xylene 2 times for 5 minutes each. Finally, the slides were removed from xylene and coverslips were mounted using DPX (BDH, Leica Microsystems, Newcastle, UK), and left overnight to dray.

2.3.6. Ventana Autostainer

The staining procedure using Roche Ventana Ultra machine was used, in collaboration with Histopathology- Queen's Medical Centre, Nottingham. Tissue thickness of 4µm was cut and placed on the bottom of the microslide which contains the identified positive tissue controls on the top of the slide for the selected antibody. Initially, three cycles of dewaxing 4 minutes each at 72 °C. Then, in the antigen retrieval step, the slides were processed in Cell Conditioning solution 1 (CC1) for 64 minutes at 95°C. CC1 is a pre-diluted tris-

based buffer with a slightly basic pH (Roche Ventana). The next step was using peroxidase inhibiter (3% H_2O_2) twice for 4 minutes at 36°C, preceded and followed by washing. Thereafter, the slides were incubated in the primary antibody and after wash, in HRP Multimer for 8 minutes. Blocking step (Roche, antibody diluent with casein) and DAB were added for 8 minutes each with wash before and after. Copper sulphate was used for 4 minutes and lastly Haematoxylin (Roche, Ventana Haematoxylin II) for 12 minutes.

2.3.7. Assessment of IHC staining

Different methods were used in this study to evaluate immunoreactivity: 1) semi-quantitative methodology using light microscopy, and (2) digital pathology viewing software, using a NanoZoomer slide scanner (Hamamtsu Photonics, Welwyn Garden City, UK) to produce digital images with high resolution (0.45 μm/pixel). Digital slides were viewed using a web based interface (Distiller, Leica Ltd, Newcastle, UK). In addition, Aperio ImageScope viewing software was used.

2.3.7.1. Scoring of IHC staining

In this study, two methods of scoring were used: 1) The percentage scoring: positive immunostained cells were counted relative to the total cells in each selected area (Fedchenko and Reifenrath, 2014), 2) The modified histochemical score (H-score) which includes a semi-quantitative assessment of the percentage of positive cells and the intensity of staining scored as 0–3 corresponding to negative, weak, moderate and strong staining. The H-score was calculated by multiplication of the intensity with the percentage (McCarty et al., 1985) (Abd El-Rehim et al., 2004).

2.3.7.2. Identification of Cut-Off point

One method to determine an appropriate dichotomisation for biomarkers, the cut-off point between populations, was dependent on the distribution of the

protein expression using IHC data where either the mean or median were chosen based on normally or not normally distributed data respectively. Additionally, frequency histograms were used for visualisation of the distribution and for detection of apparent cut-off points. Furthermore, cut-off points from the published literature were considered. An alternative method for dichotomisation was derived using X-tile software (version 3.6.1, 2003-2005, Yale University, USA). X-tile randomly splits the data into two groups and the optimal cut-off point, based on prediction of BCSS is determined in a training group and validated in the second group.

2.4. Western blot

WB was performed on cell lysates of different human cell lines (MCF7, MDA-MB-231, SKBR3, MDA-MB-468 and HeLa) to validate the specificity of the antibodies used in IHC. The specificity were reflected as a specific band.

2.4.1. Cell lysis preparation

Cell lysates were previously prepared by a technical colleague (Maria Diez Rodriguez). Cells were grown to about 80-90% confluency. After removing growth media and washing with PBS, the cells trypsinised by adding trypsin (Sigma) and incubated at 37°C for less than 5 minutes to dissociate cells. Detached cells were neutralised by adding 5 ml fresh media and centrifuged at 1,000 rpm for 5 minutes. The pellet was then re-suspended in 5-10 ml of media. Fresh media were then used to suspend the cells at a concentration of $1x10^6$ cells per ml and centrifuged for 5 minutes then placed on ice. For each 10^6 cell pellet (on ice) 1ml of RIPA buffer (Thermo Fisher Scientific) and protease inhibitors were added and the pellet re-suspended. After wards, the cell lysis was incubated for 15 minutes on ice with frequent gentle shaking and then centrifuged at 13,000 rpm for 4 minutes. The supernatant was removed and stored either in -20°C or -80°C for long term storage.

2.4.2. Sample preparation for electrophoresis

The cell lysates were thawed and prepared by adding 10-20 µg of the sample, 1 µl of NuPAGE Reducing Agent (10x) (Thermo Fisher Scientific) 2.5 µl of NuPAGE LDS Sample Buffer (4x) and distilled H_2O to make up 10µl of the total volume. After mixing the samples, heating at 100° for 5 minutes and pulse spinning, the samples were placed back on ice for denaturation.

2.4.3. Preparation of gel and running buffer

After removing of gel from NuPAGE gel package and drainage-off of storage solution, the coomb was carefully removed without disturbance of the wells. Thereafter, gel was placed into running tank and the inner chamber was filled with the running buffer. Thenceforth, 10-20 µl of the samples were loaded into NuPAGE gel as well as 5µl of the Novex Sharp marker for electrophoresis. The gel was run at 150v for 90 minutes. The running buffer was prepared using 950 ml distiller H_2O and 50ml 20x MOPS SDS Running Buffer as 1L in total.

2.4.4. Electro-blotting transfer step

The transfer buffer was prepared using 50ml 20x Transfer Buffer, 849 ml distiller H_2O, 100 ml Methanol and 1 ml antioxidant. Nitrocellulose membrane, gel, filter papers and sponges were soaked in running buffer and used to form a membrane sandwich (**Figure-8**). It was then placed into the transfer tank, covered by transfer buffer and protein transferred at 30v for 1 hour.

2.4.5. Primary and secondary antibodies incubations

Primary and secondary antibodies were prepared in the blocking solution with the recommended concentration. The blocking solution was made by adding 5% Marvel milk (2.5g) to PBS-Tween-20 (0.1%). After blocking of the membrane for 1 hour in room temperature, it was incubated over night with the primary antibody at 4°C. The membrane was washed 3 times, 5 minutes each in wash solution using PBS-Tween-20 (0.1%) (Sigma). Afterwards, it was incubated in an appropriate secondary antibody (Licor) for 1 hour at room temperature followed by washing in PBS-Tween-20, once for 15 minutes then 3 times 5 minutes each. Ultimately, to visualise the proteins bands, the fluorescence was detected using the Licor Odyssey Fc machine with wavelengths of 600, 700 and 800 nm.



Figure-8: Representative diagram for preparation of membrane sandwich , electro-blotting transfer step

2.5. Protein extraction from FFPE sections

This was performed using Qproteome FFPE tissue kit (Qiagen) as per manufacturer's instructions.

2.5.1. Deparaffinisation and sampling of FFPE tissue sections

Tissue sections of 10 µm thickness dependent on tumour burden were cut and transferred directly to 1.5 ml collecting tubes and 1 ml xylene was added. The samples were vortexed for 10 seconds, incubated for 10 minutes and centrifuged at full speed for 2 minutes. The supernatant was discarded without disturbance to the pellet. This step was repeated twice. Then, 1 ml of 100% ethanol was added. The samples were vortexed and incubated for 10 minutes, then, centrifuged for 2 minutes at maximum speed. The supernatant was removed with careful attention not to disturb pellets. This step was repeated. The previous steps were repeated using graded ethanol 96% and 70%.

2.5.2. Extraction of protein from FFPE tissue section

100 µl of extraction buffer EXB Plus with β-mercaptoethanol was added to each sample, vortexed and incubated on ice for 5 minutes followed by mixing. After that, the samples were placed in thermomixer without agitation at 100°C for 20 minutes. Following by 2 hr at 80°C with agitation at 750 rpm followed by cooling for 1 minute at 4°C. The samples were centrifuged at 4°C at 14,000 rpm for 15 minutes. The supernatant was removed to a new collecting tube and stored at -80° C until use.

2.6. Laser Capture Microdissection (LCM)

LCM is an automated sample preparation method that enables isolation of subpopulation of tissue cells from a mixed population in FFPE tissue sections guided by microscopic visualisation. The non-contact laser instrument

(PALMRobo, version 4.3, - Carl Zeiss) at Nottingham Trent University was used in the preparation of laser microdissected samples.

2.6.1. Optimisation of LCM

In order to determine the minimum adequate area to be microdissected for sufficient protein expression, an optimisation stage was done. Pure populations of invasive tumour cells were microdissected from whole sections of FFPE tissue representative of each case. At the optimisation step, different areas had been used.

2.6.2. Deparaffinisation and sampling of FFPE tissue sections

FFPE tissue sections of 10 µm thickness were cut and placed on an uncovered membranous coated slide. The samples were incubated in xylene thrice at room temperature for 10 minutes. The slides were dehydrated in graded fresh ethanol series (100%, 96% and 70 %) twice for each concertation at room temperature for 10 minutes each. Finally, slides were transferred to the slide dish containing distilled water. The uncovered membranous coated slides were removed from the distilled water and gently the slides were tapped on a paper towel without disturbing the tissue. The slides were stained with haematoxylin for 10 seconds. Then, non-contact laser instrument (palmRobo version 4.3. - Carl Zeiss) was used. Once the slide was fixed in microscope stage, the adhesive cap 500 clear was attached to the cap handle. The electronic marking of pure invasive area was defined using the software. The microdissected tissue areas were then collected automatically to the adhesive cap. The minimum area microdissected per case was 15×10^6 µm2 (**Figure-9**).

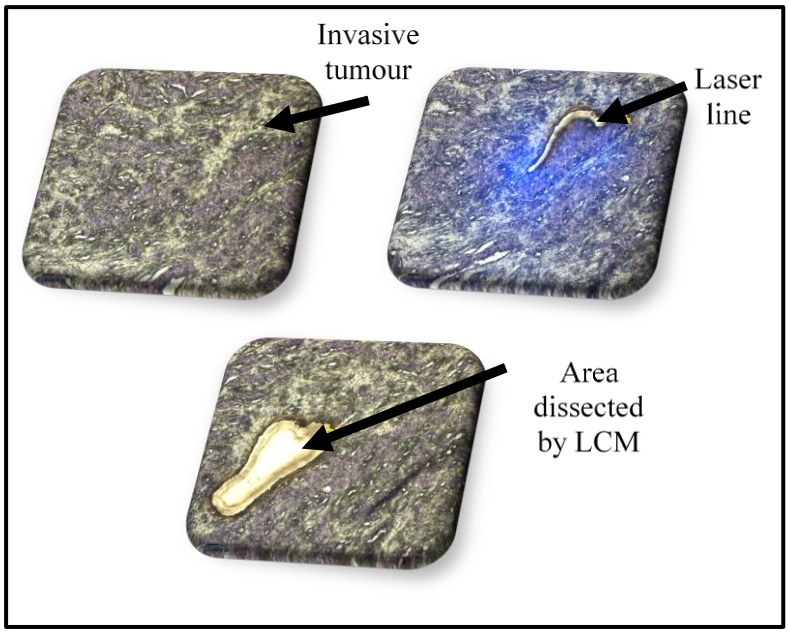


Figure-9: Laser capture microdissection
image shows marking of pure invasive breast cancer area and automatic cutting along defined line followed by catapulting

2.7. Reverse phase protein array

2.7.1. Validation of primary antibodies for microarray experiments

One of the challenges in RPPA is the availability of specific antibodies. The specificity was tested using a strip western device and it was revealed by bands of the expected sizes. Proteins from the lysates were solubilised in 4x SDS loading buffer and were heated for 5 minutes at 100 ° C. Equal amounts of protein (50µg) were separated on Novex 12% tris glycine polyacrylamide gels (Invitrogen life technologies) and transferred to nitrocellulose 0.1% Tween-20 and 5% non-fat dry milk powder, and incubated with primary antibodies [diluted in 5% Bovine serum albumin (BSA)/PBST]. Blots were incubated overnight at 4 °C then washed with PBST, and incubated with swine anti-rabbit or anti-mouse immunoglobulin-HRP conjugate (Dako) diluted 1:2000 in blocking buffer for 1 hour at room temperature. Following an additional PBST wash, membranes were incubated with Enhanced Chemiluminescence Detection reagent (ECL; Amersham, UK), exposed to film (Kodak, Sigma) (This step conducted by Dr

Paddy Tighe and Dr Ola Negm from the Post-genomics Facility, University of Nottingham), **Figure-10.**

2.7.2. Lysates array

The concentration of protein in the lysates prepared from FFPE samples using the Qproteome FFPE tissue kit was quantified using Fast Green assay (Loebke et al., 2007). Protein concentrations from each lysate were normalised and an equal amount of protein was used for the microarray. First, the proteins from the cell lysate were solubilised in 4x SDS sample buffer at a ratio of 1:3 and heated at 95°C for 5 minutes. Then, after prepared samples of 10 µl were loaded onto a 384-well plate (Genetix). The next steps were conducted by Dr Paddy Tighe and Dr Ola Negm from the Post-genomics Facility, University of Nottingham.

Samples were robotically spotted in triplicate onto nitrocellulose-coated glass slides (Grace Bio-labs) using a microarrayer (MicroGridII). Slides were incubated overnight in blocking solution using Super G blocking buffer (Grace Bio-labs) at 4°C with shaking. After washing three times 5 minutes each, the slide was incubated with the primary antibodies (Section-4.4.3). In addition mouse β-actin (Sigma), diluted 1:1000 was used as a housekeeping protein to control protein loading. Slides were incubated overnight at 4°C with shaking. Following washing, the slides were incubated with diluted infrared (1:5000 in washing buffer) secondary antibodies that 800 CW anti-rabbit antibody and 700 CW anti-mouse antibodies for 30 minutes at room temperature in dark with shaking. Then, slides were washed and dried by centrifugation at 500xg for 5 minutes and scanned with a Licor Odyssey scanner at 21 µm resolutions at 800 nm (green) and 700 nm (red). The resultant TIFF images were processed with Axon Genepix Pro-6 Microarray Image Analysis software (Molecular Services Inc.) to obtain fluorescence data for each feature and generate gpr files. Protein signals were finally determined with background subtraction and normalisation to the internal housekeeping targets using RPP analyzer, a module within the R statistical language on the CRAN (http://cran.r-project.org/).

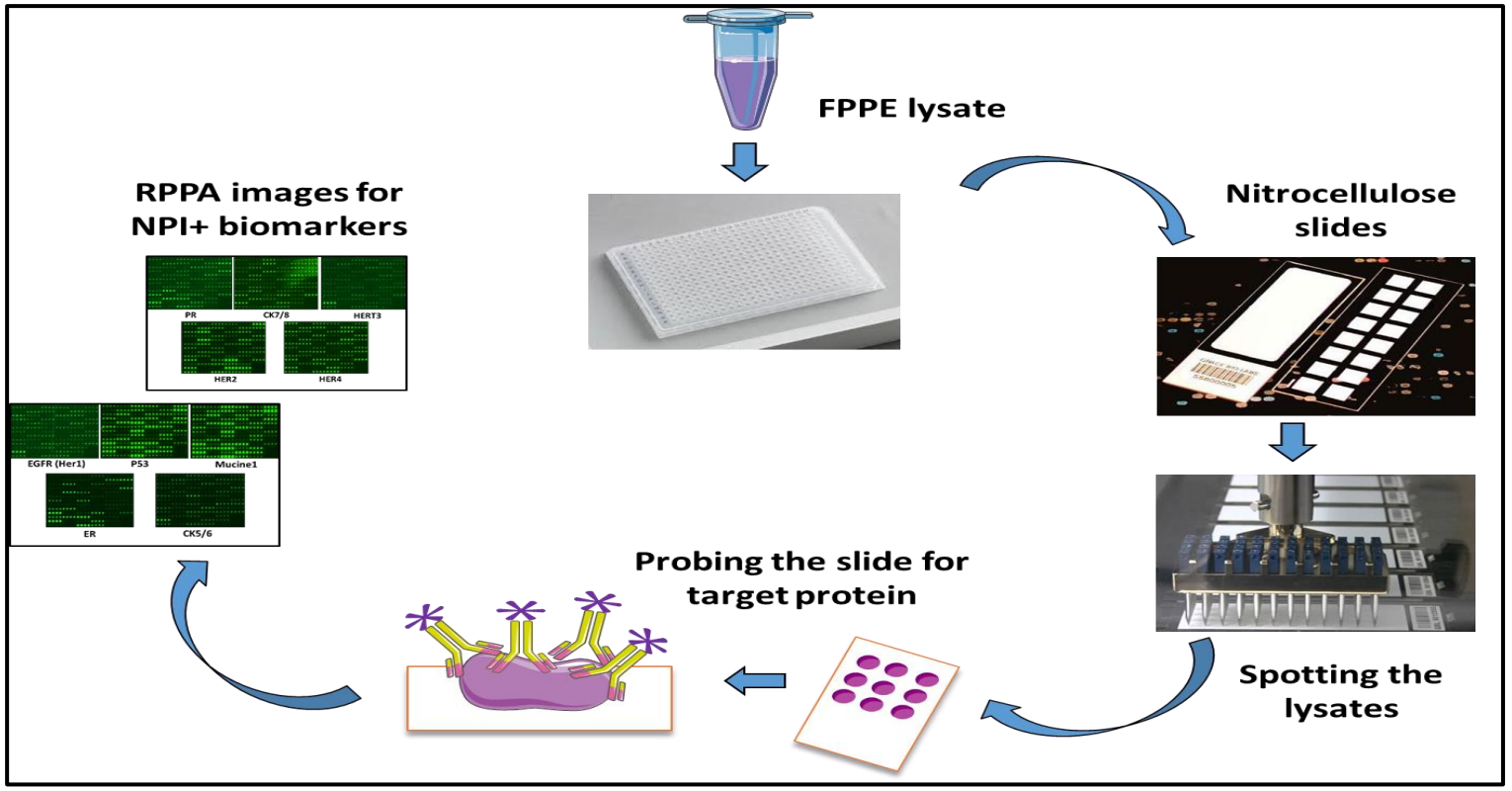


Figure-10: Representative diagram showing steps of reverse phase protein array.
After extraction of protein from FFPE, lysate were printed or spotted on nitrocellulose slides using robotic microarray. The slides probed with the antibodies against the protein of interest and infra-red secondary antibodies were used for detection of binding. Slides were scanned, and analysed

2.8. Real time- polymerase chain reaction (RT-PCR)

2.8.1. RNA extraction from FFPE tissue sections

Great care was taken when handling RNA samples, both during and after the isolation procedure to avoid unwanted changes in the expression profile. Benches and all involved equipment were cleaned with RNase-Away to prevent RNase contamination. Real-Time PCR System (Applied Biosystems 7500/7500 Fast) at Queen's Medical Centre, Nottingham, Molecular Pathology Group was used.

2.8.2. RNA extraction using RNeasy FFPE Kit (Qiagen)

Macrodissection of FFPE tissue sections was performed for invasive tumour components. In each case, four serial sections were cut onto glass slides under RNase free conditions. One serial section 4 µm was stained for H&E and used as a guide for gross manual dissection of the invasive breast carcinoma component using disposable surgical blades in the unstained 10 µm thickness tissue sections. Immediately, the sections were placed into a microcentrifuge tube. After adding 160 µl of deparaffinisation solution (Qiagen) and mixing for 10 seconds, the samples were centrifuged. The samples were incubated for 3 minutes at 56°C. After cooling at room temperature, buffer PKD (part of the RNeasy FFPE Kit) was added, vortexed and centrifuged for 1 minute at 11,000 xg. Next, proteinase K (part of the RNeasy FFPE Kit) was added and mixed gently by pipetting up and down to the lower clear phase. Afterwards, samples were incubated at 56°C for 15 minutes, then at 80°C for 15 minutes with mixing. Consequently, the lower uncoloured phase was transferred into a new

microcentrifuge tube. Then, the samples were incubated for 3 minutes on ice and centrifuged for 15 minutes at 20,000 xg.

After transferring the supernatant to a new microcentrifuge tube, 16 µl of DNase booster Buffer and 10 µl DNase were added, mixed by gently inverting the tubes and incubated for 15 minutes. Following this, 320 µl of buffer RBC (part of the RNeasy FFPE Kit) then, 720 µl of ethanol were added. Immediately 700 µl of the samples including any precipitate were transferred to an RNeasy MinElute spin column placed in collection tube and centrifuged for 15 seconds at ≥ 8000 xg then, Discard the flow-through. This step was repeated until the entire sample has passed through the RNeasy MinElute spin column. Afterwards, 500 µl of Buffer RPE were added to the RNeasy MinElute spin column and centrifuged for 2 minutes at ≥8000xg to wash the spin column membrane. Then, the collection tubes were discarded with the flow-through. The RNeasy MinElute spin columns were placed in a new collection tube, which were opened and centrifuged at full speed for 5 minutes. The collection tubes were discarded with the flow through. Finally, after placing the RNeasy MinElute spin columns in a new collection tubes, 14–30 µl RNase-free water was directly added to the spin column membrane and centrifuged for 1 minute at full speed to elute the RNA.

2.8.3. Testing quality and quantity of RNA.

RNA levels were quantified by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). To establish the blank an appropriate buffer was used. After that, 1-2 µl of sample was pipetted onto the lower measurement pedestal of the instrument. Then, a spectral measurement was initiated using the software on the PC. According to NanoDrop 2000/2000c Spectrophotometer V1.0 user manual; samples having A260/A280 ratio between (1.8-2.00), indicating adequately pure RNA were considered appropriate for subsequent steps. The eluted RNA was stored at -80°C until used.

2.8.4. Complementary DNA (cDNA) Generation

Complementary deoxyribonucleic acid (cDNA) was produced by reverse transcription of RNA. Reverse transcription of RNA performed with RT+ and RT- reactions for every sample. After adding 2.5 µl of 0.5 µg random hexamers (Thermo Fisher Scientific, 200 ng/µl) to 1 µg RNA, the samples were made up to a total volume of 15 µl of H2O and were incubated at 70°C for 5 minutes. Samples were immediately placed on ice for 5 minutes and then centrifuged. A master mix was prepared as follows; 5 µl M-MLV RT Buffer 5x (Promega), 1.25 µl dNTP (10 mM) (Thermo Fisher Scientific), 1 µl M-MLV Reverse Transcriptase (Promega 200 U/µl) and 5.25 µl water. 10 µl of appropriate master mix was added to each sample to reach the total volume 25 µl. Next, samples were incubated at 37°C for 1 hour followed by 95°C for 10 minutes. After that, samples were diluted 1:10 and stored at -20°C. RT-negative samples in which RNase-free water were included instead of reverse transcriptase enzyme to ensure the purity of the RNA extracted sample from remains of genomic DNA.

2.8.5. Primers design

Primers for RT-PCR were validated using PubMed and the Primer 3 software (web version 4.0.0) (http://bioinfo.ut.ee/primer3-0.4.0/) and targeted at exon-exon junctions. Forward and Reverse primers were designed to encode for the beginning of the DNA fragment; N-terminus 5‟ and for the end of the DNA fragment; C-terminus 3‟, or to encode a specific fragment of DNA within the gene of study. The sequences were tested using a Basic Local Alignment Search Tool (BLAST) on the NCBI site search for specificity. Additionally, SNP check and Insilico PCR web sites were used.

2.8.6. Optimisation of primers

For optimisation of each primer pair, a running PCR reaction with a 12 temperature gradient at the annealing stage was performed. The reaction was taken in a final volume of 12.5 µl, each reaction contained x1 Hotshot^{TM}HRM, forward and reverse primers pair at 250 nM, 20ng template DNA, LC Green and water to complete the total volume. Initial denaturation at 95°C for 5 minutes subsequently 40 cycles of denaturation at 95°C for 10sec, annealing (gradient PCR 51°C to 69°C) and extension phase at 72°C for 20sec. followed these cycles was a final elongation at 72°C for 5minutes. Finally, PCR products were analysed using Light Scanner. The PCR products were transferred to Light Cycler capillaries (20 ml) (Roche). The products were melted in the High resolution melting (HRM) instrument (Idaho Technology).

2.9. RNAscope technique

The RNAscope assay utilises a novel technique of *in situ* hybridisation (ISH) to detect single RNA molecules per cell in FFPE sections mounted on glass slides. RNA-specific probes are hybridised to specific targeted RNA molecule and engaged to a cascade of signal amplification molecules ended with signal detection (**Figure-11**). This work was performed in collaboration with Advanced Cell Diagnostics (ACD Bio) using their RNAscope kit.

2.9.1. Preparation and pre-treatment of FFPE for the RNAscope assay

Tissue sections of 4µm thickness were cut and mounted on Superfrost plus slides with air drying overnight at room temperature. Then, the sections were deparaffinised in fresh xylene, followed by dehydration in an ethanol series. On slides, a hydrophobic barrier was created by drawing a barrier 2–4 times around each section about 0.75 x 0.75 inch with Immedge hydrophobic barrier pen. Thereafter, 5-8 drops of pre-treat 1 were added to cover the entire section for 10

minutes at room temperature. After removing of pre-treat 1 (ready-to-use endogenous blocker) by tabbing the slides on paper, the slides were washed twice with distilled water by moving the rack from 3 to 5 times. Next, the slides were submerged into the boiling pre-treat 2 (citrate buffer 10 nmol/L, pH 6) for 15 minutes. pre-treat 2 was prepared by mixing 630ml of distilled water to 70 ml of 10x pre-treat 2 solution in a 1 L beaker and then, placing the baker on a hotplate at 100–104°C. Consequently, the slides were washed twice with distilled water and air dried. After drying of slides, 5 drops of pre-treat 3 (10 g/mL protease, Sigma) were added and incubated at 40°C for 30 minutes in a HybEZ hybridisation oven (ACD Bio). Then, slides were washed twice in distilled water.

Thenceforth, in hybridisation step, the tissue were placed at 40°C with the following solutions: target probes in hybridisation buffer A (6x SSC, 0.2% lithium dodecyl sulfate, formamide 25% and blocking reagents) for 3 hours; preamplifier (2 nmol/L) in hybridisation buffer B (20% formamide, 5x SSC, 10% dextran sulfate, 0.3% lithium dodecyl sulfate and blocking reagents) for 30 minutes; amplifier (2 nmol/L) in hybridisation buffer B for 15 minutes at 40°C and label probe (2 nmol/L) for 15 minutes in hybridisation buffer C (5x SSC, 0.3% lithium dodecyl sulfate and blocking reagents). Following every hybridisation step, slides were washed three times with wash buffer (0.1x SSC and 0.03% lithium dodecyl sulfate) at room temperature. Assays were conducted in parallel with negative and positive controls to confirm interpretable results. The endogenous housekeeping gene was used as positive control to evaluate both tissue RNA integrity and assay method. Pre-treatment and RNAscope protocol were optimised for each sample set to reach maximum signal and minimum background. Optimised conditions were used for all following marker assessment assay runs (Wang et al., 2012a).

2.10. Ethical Approval

This study was approved by the Nottingham Research Ethics Committee 2 under the title developed under the ethical approval number (REC202313) and the "Nottingham health Science Biobank (NHSB)" approved by the North West - Greater Manchester Central Research Ethics Committee (15/NW/0685).

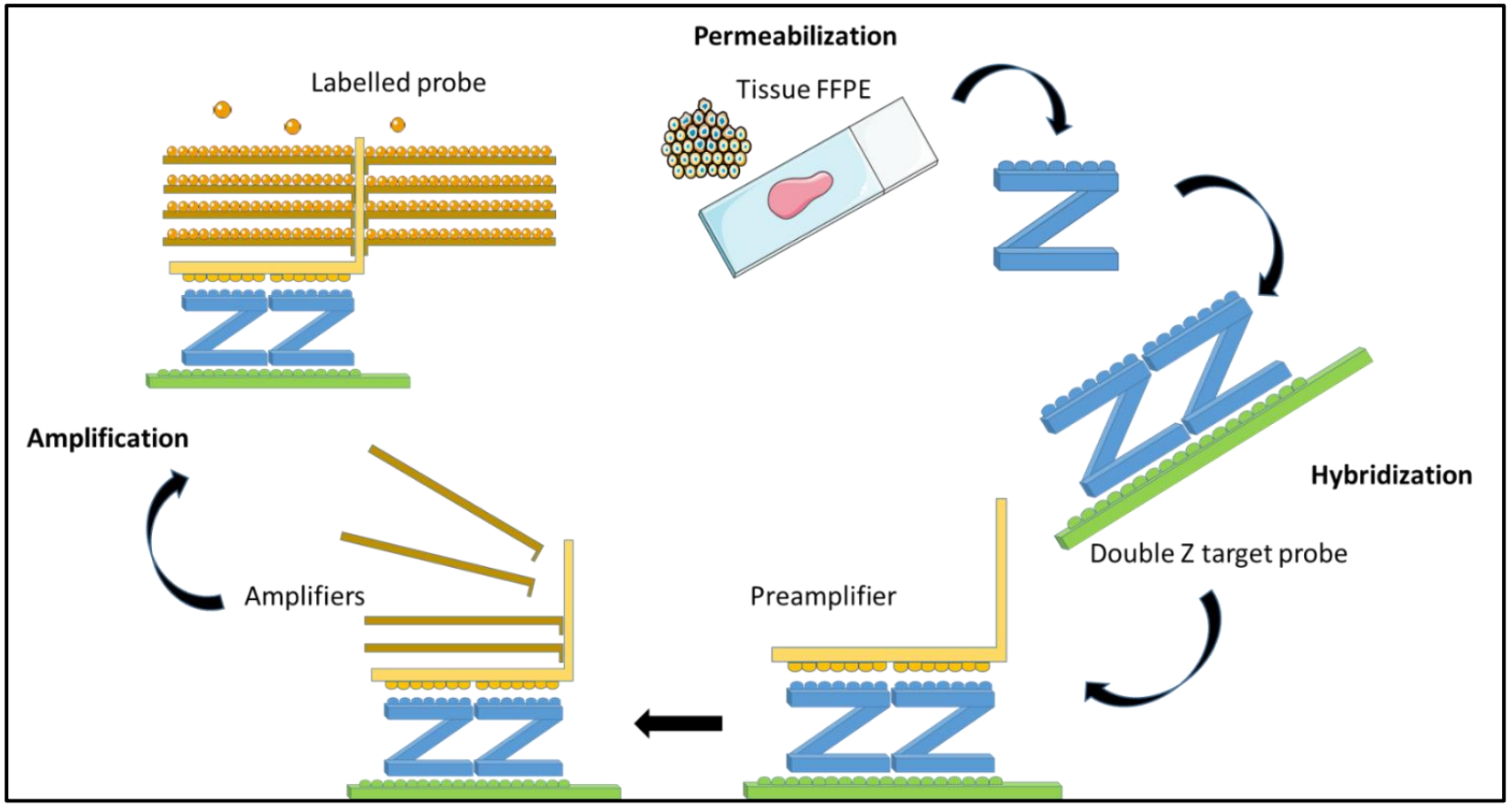


Figure-11: Representative diagram showing steps of RNAscope

3. Evaluation of integration of Ki67 into BC classification

3.1. Introduction

3.1.1. Clinical importance in breast cancer

In last two decades, Ki67 has been extensively examined and documented as a prognostic biomarker in invasive BC showing worse prognosis with higher Ki67 expression (Domagala et al., 1996) (Trihia et al., 2003) (Viale et al., 2008) (Aleskandarany et al., 2010) (Brown et al., 2014) (Ingolf et al., 2014). Additionally, systematic reviews support the role of Ki67 as a prognostic marker and an independent predictive factor for neoadjuvant chemotherapy (Yerushalmi et al., 2010) (Luporsi et al., 2012). In two different meta-analyses published in 2007 and 2008; high expression of Ki67 in both node positive and negative BC revealed a significantly worse overall and disease free survival (de Azambuja et al., 2007) (Stuart-Harris et al., 2008). Furthermore, Ki67 has been recommended by the St. Gallen consensus panel as a marker for the definition of intrinsic subtypes of BC to differentiate between luminal A and luminal B subgroups (Goldhirsch et al., 2011) (Goldhirsch et al., 2013).

On account of the clinical importance of Ki67, the reproducibility of its assessment is of obvious importance. However, the assessment of Ki67 remains challenging. The expression of Ki67 is basically assessed by proving the percentage of Ki67 positive cells in immunostained FFPE using WTS. It is a labour intensive and subjective that requires counting at least 500-1000 cells to achieve an acceptable reproducibility rate and suit the recommendation of the International Ki67 Working Group (Dowsett et al., 2011). However, due to the heterogeneity of Ki67 staining in breast cancer, there are concerns about the reliability of this scoring methodology. Currently, there is no universally recommended method for selecting the appropriate areas to assess; which varies from counting within the area of highest staining (hot spot) in the whole tumour section to the average stained cells within the whole diffuse stained area. There is a wide scope of references with respect to the minimum number of BC cells

that have to be assessed. Additionally, eyeballing as a manual fast estimation and automated counts are different technique used by various research groups to interpret Ki67 staining (Warth et al., 2013) (Laurinavicius et al., 2014).

One of the attractive alternatives for assessment of whole tissue section in the research is the use of the TMA, since its usage can analyse a large number of the tissue samples. Additionally, it is a time and cost effective (Shergill et al., 2004) (Tzankov et al., 2005). Although it remains unclear as to their validity for Ki67 evaluation in breast cancer, different studies have reported the assessment of Ki67 utilising TMA platform. Based on data from a series of TMAs combined with gene expression analysis (Cheang et al., 2009) (Goldhirsch et al., 2011), Ki67 cut-off point of 14% was recommended for treatment decisions by the St Gallen 2011 guidelines. Further, Polley et al. (Polley et al., 2013a) (Polley et al., 2013b), conducted a Ki67 reproducibility study. This study revealed a concerning lack of concordance as a result of scoring approach when a set of stained BC TMAs were evaluated by eight different laboratories across the world experienced in Ki67 immunostaining. Ruiz et al. concluded that associations between molecular features and Ki67 could be reproduced with high statistical significance using a TMA containing only one tissue sample measuring 0.6 mm in diameter per tumour (Ruiz et al., 2006).

Conversely, a recent study evaluated the impact of tissue sampling on accuracy of Ki67 immunostaining evaluation in breast cancer. They concluded that to achieve a coefficient error of 10 %, 5 to 6 TMA cores were needed for homogeneous cases, 11 to 12 cores for heterogeneous cases and in mixed tumour populations 8 cores were required (Laurinavicius et al., 2014). Moreover, recommendations from the international Ki67 in breast cancer working group showed that the deficit of systematic comparisons of Ki67 expression levels between TMA and WTS was noted (Dowsett et al., 2011). The issue of reproducibility and to what extent TMA represents WTS is one of the aspects evaluated in this study.

Moreover, the optimal cut-off point for Ki67 in BC is another controversial issue despite the enormous number of published studies with significant results in BC (Dowsett et al., 2011). In clinical practice the evaluation of prognostic/predictive variables mainly relies on the stratification of the patients into risk groups. The basic approach is to dichotomise the Ki67 which is assessed as a continuous variable into two groups conferring a high and low risk groups based on patients outcome. In 2009, the St. Gallen panel suggested that Ki67 expression in BC have to be classified into three groups: low <15%, intermediate 16-30% and high group >30% (Goldhirsch et al., 2009). This stratification was dependent on univariate analysis of patient outcome implemented with different Ki67 expression cut-off points. The best dividing cut-off points with lowest significant *p*-values according to survival utilising Ki67 status were 15% and 30% (Jalava et al., 2006). However, in 2011, St. Gallen panel endorsed 14% as an alternative cut-point with the purpose of separation of ER-positive tumours into luminal A (<14%) and luminal B (≥14%) (Goldhirsch et al., 2011). This was resulting from comparison with gene array data used as a prognostic factor (Cheang et al., 2009).

In 2013, St Gallen updated their Ki67 threshold to ≥ 20% for high Ki67 expression with an option to also utilise locally defined cut-points (Goldhirsch et al., 2013). More recently, the St. Gallen BC conference accepted a median cut-off estimation of Ki67 within the range of 20–29% to distinguish the luminal B subgroup (Coates et al., 2015). Additionally, in 2011, the International Ki67 Breast Cancer Working Group reported that there was no consistency in an ideal cut-off point for use in daily clinical routine (Dowsett et al., 2011). The recent report on the second phase of the Ki67 ring trial was emphasise the importance of standardising the pre-analytical and analytical features for Ki67 immunostaining including the previous mentioned issues. Therefore, optimisation of Ki67 is the initial step to be incorporated to drive patient treatment decisions in clinical practice (Ingolf et al., 2014).

As reported by the international Ki67 in breast cancer working group, the authors endorsed using of MIB-1 antibody clone in IHC (Dowsett et al., 2011). MIB-1

clone, a mouse monoclonal antibody is regarded as the gold standard and the most robust biomarkers measuring Ki67 expression. It is showing a relatively consistent measurements in specimens across a range of conditions used in routine fixation, tissue processing and immunohistochemical methods (Yerushalmi et al., 2010, Dowsett et al., 2011). However, evaluation of Ki67 expression using MIB-1 has been described to have cytoplasmic/membranous expression in addition to specific nuclear expression (Faratian et al., 2009) (Ekholm et al., 2014).

3.1.2. Ki67 and mitotic count

Both Ki67 and mitotic activity are parameters that can be used to determine tumour proliferation rate. The mitotic score depends on the number of mitoses per 10 high power fields and correct counting of mitosis needs high quality fixation, obtained when fresh specimens are sectioned promptly with tumour blocks of reasonable thickness (Ellis et al., 2016). Mitotic frequency is considered a strong prognostic indicator as a part of NGS which is the modification of Bloom-Richardson grading system for invasive breast carcinoma (Elston and Ellis, 1991). Furthermore, patients with high mitotic activity have an increased risk of death (Stuart-Harris et al., 2008) and recurrence regardless of lymph node status or tumour size (Baak et al., 2005) (Baak et al., 2008).

However, there are limitations to its precision as mitosis rates vary particularly in aneuploid tumours owing to discrepancy in cell cycle duration, particularly the M phase (Sennerstam and Auer, 1990) (van Diest et al., 2004). This may change the proportion of visible mitosis within the same tumour at different times. Thus, the correlation between the proliferation rate and the number of mitosis is not strictly linear (Beresford et al, 2006). Moreover, the instability of mitotic figures throughout specimen processing and fixation, reproducibility issues due to tissue section quality and experience of the pathologists and to some extent long time required for mitotic counting are all described technical problems challenged during mitotic scoring (Salisbury, 2009). Although most

studies of various proliferation assays have revealed significant agreement in outcome predictions for individual patients, there is no consensus on the ideal proliferation rate testing method (van Diest et al., 2004). Therefore, evaluation of other indicators of cell proliferation, such as Ki67 as a reliable tool seems to be important especially different studies have revealed that there is a positive correlation between the Ki67 score and mitotic activity, however, they differ in some cases (Barnard et al., 1987) (Spyratos et al., 2002) (Offersen et al., 2003).

Other studies show percentage of discrepancy as showed by Rossi et al., among 1430 patients, 19.6% had discrepant Ki67 and mitotic activity (Rossi et al., 2015). Therefore, using of Ki67 as a substitution for mitotic frequency may add more biological effect or reclassify the patients into a different prognostic subgroups with more clinical importance. To apply this, in this study, NPI and NPI+ indices were used as a tool for testing this hypothesis. As mentioned earlier, NPI is a widely used index that determines prognosis for BC and is calculated using three main clinicopathological criteria; tumour size, lymph nodes status and the grade (including mitotic count) of the tumour. NPI+ is an updated index, which uses a combination of traditional clinicopathological parameters and the biological features of the tumour determined by immunohistochemistry. NPI+ is a combination of two steps; the first to determine the biological classes using 10 biomarkers by immunohistochemistry. The biomarkers that are used are ER, PR, HER2, CK5/6, CK7/8, P53, HER1, HER3, HER4, and Mucin 1. The second step is the combination of traditional prognostic parameters, including mitotic count, within each biological classes resulting in the structured NPI-like formulae for each class.

As mentioned earlier, the molecular classes in the NPI+ consists of three Luminal classes. The luminal classes are divided into Luminal A, Luminal B and Luminal N subgroups (Green et al., 2013) (Rakha et al., 2014). Mitotic frequency is included in NPI+ as an indicator for the degree of cellular proliferation for both the luminal A and luminal B classes. After successive removal of the least significant parameters during different steps, the final factors with the most significant results, according to their β-value in the Cox regression analysis,

were identified. The proportional hazard ratio Cox regression identified six clinicopathologic prognostic factors of importance within the population in predicting BCSS: mitotic activity is one of these factors. Once these factors were identified, the population was split into the biological classes and Cox regression analyses were performed independently for each class to obtain the most significant clinicopathologic prognostic factors and their β-value in the context of the classes.

Kaplan–Meier analysis showed that using formulae, based on these clinicopathological parameters for each biological BC class provides improved and highly significant patient outcome stratification compared with the traditional NPI. These variables were combined to form formulae that vary among different classes leading to bespoke NPI-like formulae for each of the seven biological classes forming a new biomarker-based prognostic index, NPI+. Mitotic frequency represents a most significant clinicopathological prognostic factors of importance within the population in predicting BCSS in luminal A and luminal B classes. Similarly, Ki67 gave an important in terms of the differentiation between Luminal A and B subclasses, it is the only marker recommended by the St. Gallen consensus panel as a marker for the definition of intrinsic subtypes of BC to differentiate between luminal A and luminal B subgroups, which has so far not been investigated with respect to the NPI+.

3.2. The hypothesis

To what extent can integration of Ki67 expression, determined by immunohistochemistry, in BC enhance existing prognostic indices.

3.3. Aim of this chapter

There are four main aims for this Chapter

1) To assess the matched level of proliferation of Ki67 assessment in BC using TMA or WTS, specifically herein:
    a. To what extent Ki67 protein levels are matched between TMA and WTS.
    b. Whether the cut-off used for WTS is reproducible using TMA in different molecular classes.
    c. To what extent the transcriptomic correlates with protein level within both WTS and TMA.
2) To evaluate the reproducibility of Ki67 scoring
3) To examine the correlation between different cut-offs and both the patient outcome and important clinicopathological variables.
4) To assess non-specific cytoplasmic/membranous staining using:
    a. Different MIB-1 antibody concentrations
    b. Different antigen retrieval time and methods for MIB-1
    c. Manual staining vs fully automated immunostainer using MIB-1
    d. Different anti Ki67 clones (MIB-1, SP6, MKi67 and 30-9).
5) Evaluation of integration of Ki67 into both NPI and NPI+ and substitution of Ki67 to the mitotic count.

3.4. Methodology

3.4.1. Study group

In this study, the first cohort was used (Section-2.1.1). The expression of Ki67 was examined on 707 cases of invasive BC cases with matched WTS and TMA cores (**Table-6, Figure-12**). Molecular classes were defined as luminal (ER-positive and/or PR-negative), HER2-positive (HER2-positive irrespective of the expression of other markers), and triple-negative (HER2-negative, ER-negative and PR-negative). For the analysis of transcriptomic data for Ki67, METABRIC series was used. As mentioned in section-2.1.4, total RNA was extracted using fresh-frozen BC cases (Curtis et al., 2012b). METABRIC data was matched with 197 cases of WTS and 123 of TMA cases. The matched cases between transcriptomic data and both TMA and WTS are 101 cases.

3.4.2. Comparison of Ki67 expression between TMA and WTS

3.4.2.1. Immunohistochemistry:

For TMA staining, the details of methodology in Section- 2.3. Antigen retrieval citrate at pH 6.0 was used in 98˚C for 20 minutes in the microwave. Optimised primary antibody, MIB-1 monoclonal mouse diluted 1:100 (Dako) antibody was applied and incubated for 1 hour at room temperature. For WTS (stained by Dr.MA), the standard streptavidin–biotin complex method was used. Primary mouse monoclonal antibody against Ki67 antigen (clone MIB-1; Dako) diluted 1:100 in normal swine serum was applied to each slide and incubated for 60 minutes at room temperature. 3, 30-Diaminobenzidine tetrahydrochloride (Dako liquid DAB plus) was used as a chromogen. The sections were counterstained with haematoxylin. Human tonsillar sections were used as a positive control while negative controls were performed by omitting the application of primary antibody.

Table-6: Summary of the characteristics of the study cohort

Parameter	Number (%)
Patient Age	
≤ 50 years	271(38.4)
> 50 years	436(61.6)
Menopausal Status	
Premenopausal	283(40.1)
Postmenopausal	423(59.9)
Tumour Grade	
1	93(13.2)
2	216(30.6)
3	397(56.2)
Pleomorphism	
1	10(1.5)
2	233(33.9)
3	444(64.6)
Tubule Formation	
1	34(4.9)
2	211(30.7)
3	443(64.4)
Mitotic Figures	
1	203(29.5)
2	132(19.2)
3	353(51.3)
Tumour Size	
≤ 2 cm	407(57.6)
> 2 cm	300(42.4)
LN status	
1 (0 Positive nodes)	405(57.3)
2 (1-3 Positive nodes)	235(33.2)
3 (>3 Positive nodes)	67(9.5)
NPI	
Good NPI (<3.4)	179(25.3)
Moderate NPI (3.41-5.4)	389(55.0)
Poor NPI (≥5.4)	139(19.7)
LVI	
Negative	444(63.1)
Definite	263(36.9)
ER status	
ER Negative	189(26.8)
ER Positive	516(73.2)
PR status	
PR Negative	291(42.0)
PR Positive	402(58.0)
HER2 status	
HER2 Negative	588(85.2)
HER2 Positive	102(14.8)
Molecular Subtype	
Luminal	515(68.8)
HER2 positive	102(13.6)
Triple Negative	132(17.6)

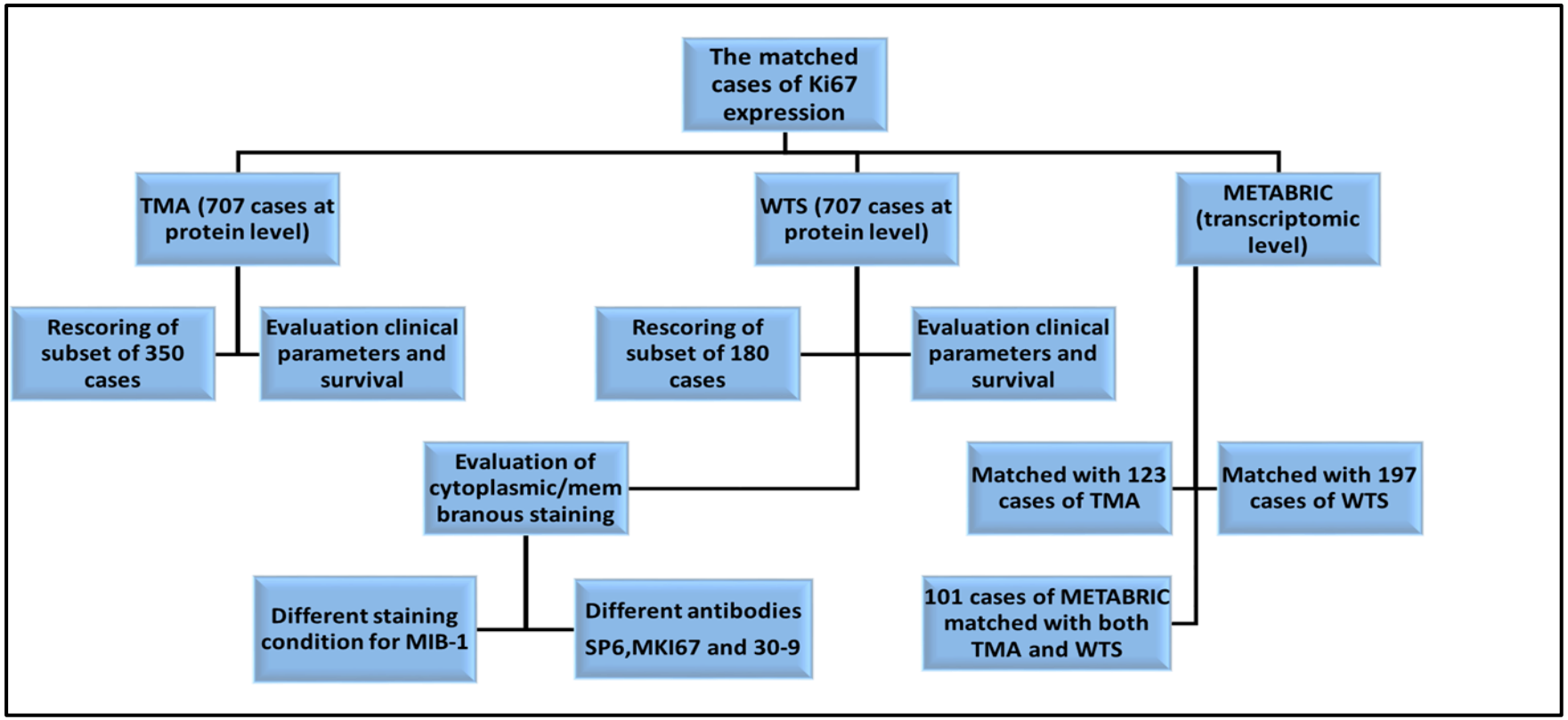


Figure-12: A flowchart representing the cohorts and steps were used in the comparison between TMA and WTS of BC tissue

3.4.2.2. IHC scoring

All of the invasive BC cells present in the TMA cores were evaluated for Ki67 expression and scored as a percentage of Ki67 positive stained nuclei. All tumour cell nuclei with homogenous, multiple speckled, granular staining or even nucleolar staining were considered as positive staining irrespective of their staining intensity. A subset of TMA cases (n=350) was re-scored by another observer, Dr.MM to assess for inter-observer concordance. WTS scoring was previously performed by Dr.MA in the hot spot areas (areas with the highest number of positive nuclei within the invasive tumour) (Aleskandarany et al., 2010). Hot spots were recognised by scanning the section for immunostaining assessment, utilising a light microscope at low power magnification (x100). Ki67 was expressed as the percentage of positive malignant cells in 1,000 malignant cells measured under high power magnification (x400). A subset of cases (n=180) was re-scored by Abir Muftah to test intra-observer reproducibility.

3.4.3. Assessment of cytoplasmic/membranous staining

For this purpose ten BC cases were selected which showed membranous staining using the MIB-1 clone. Different antigen retrieval time and methods, different MIB-1 antibody concentrations, different anti Ki67 clones (MIB-1, SP6, MKi67 and 30-9) and manual staining vs Ventana Autostainer were tested on WTS as shown in **Table-7**.

3.4.3.1. Optimisation using different MIB-1 antibody concentrations

Same methodology described in Section-2.3 was used except the antibody concentration. Different dilutions for MIB-1 antibody were tried at 1:100, 1:200, 1:300, 1:400 and 1:500.

3.4.3.2. Optimisation of MIB-1 using different antigen retrieval methods

Except for the antigen retrieval and the protein blocking steps, the same methodology described in Section-2.3 was followed. In the antigen retrieval step, the antigen retrieval solution (bond epitope retrieval solution 2 at pH 9.0) was used in the water bath at 100˚C for 35 minutes, followed by TBS at 50˚C for 10 minutes. After that running tap water for 5 minutes. Additionally, in the protein blocking step instead of using the Novolink kit protein blocking, ultra V block for 5 minutes was used instead.

3.4.3.3. Optimisation of MIB-1 using different antigen retrieval time

For this purpose, three antigen retrieval times were set, while the other parameters remained as previously described. In the antigen retrieval step, the antigen retrieval solution (bond epitope retrieval solution 2 at pH 9.0) was used in the water bath then the first, second and the third sets placed in 100˚C for 25, 35, and 45 minutes respectively. After that the slides were cooled in the water bath for 5 minutes.

3.4.3.4. Comparison between manual staining and fully automated immunostainer using MIB-1 clone

Roche Ventana Ultra machine was used was used. Ki67 (MIB-1) with dilution 1:200 was incubated for 32 minutes at 37°C. The retrieval time was 36 minutes in pH 9.0 at 98°C. Ventana DAB working solution was used for 8 minutes at 37°C. Hematoxylin was added for 8 minutes (details in section-2.3.6).

3.4.3.5. Using different anti Ki67 clones (MIB-1, SP6, MKi67 and 30-9)

Different antibodies clone were used with different dilution and incubation period. SP6, rabbit monoclonal antibody (Thermo Fisher Scientific) with 1:30 dilution and was incubated for 1 hour. MKi67, rabbit polyclonal (Sigma) with 1:100 dilution and was incubated for 1 hour. The clone 30-9, rabbit monoclonal antibody (Roche) ready to use. The dilution was 1:25 for 16 minutes and antigen retrieval time was 64 minutes with CC1 with basic pH at 95C° using Ventana Autostainer. More details in **Table-7**.

Table-7: Different antibodies and different staining conditions for Ki67 staining

Antibody Clones (supplier)	Manual immunostaining				Automated staining
	Dilution	Incubation time	Antigen retrieval (AR) time	Antigen retrieval method	
MIB-1 (Dako)	Different concentration was tested 1:100, 1:200, 1:300, 1:400 and 1:500.	60 minutes	Different antigen retrieval time was used: 25, 35 and 45 minutes using water-path or 20 minutes using microwave.	Different methods was tested 1) Using citrate at pH6, Microwave for 20 minutes. 2) Using BERS2 at pH9, water-path at 100C° for 35 minutes.	Ventana Autostainer was used with dilution 1:200, incubated for 32 minutes at 37C°
SP6 (Thermo Fisher Scientific)	1:30	60 minutes	20 minutes, microwave	Citrate at pH6	/
MKi67 (Sigma)	1:100	60 minutes	20 minutes, microwave	Citrate at pH6	/
30-9 (Roche)	1:25	16 minutes	64 minutes	CC1 with basic pH at 95C°	Ventana Autostainer

3.4.4. Integration of Ki67 into NPI

NPI scores were available (rely on previously recorded grading from histopathology reports) for 1,375 cases in the first series (details are given in Section-2.1.1). NPI was calculated using the equation:

lymph node (LN) stage (1–3) + histological grade (1–3) + (0.2) x tumour size (cm)

where LN stage 1 means lymph node-negative; stage 2 means three or fewer nodes involved in metastatic carcinoma; stage 3 is four or more nodes, or apical node, or any axillary plus internal mammary node invaded by carcinoma. Regarding the histological grade or Nottingham Grading System, three separate scores were included:

1) Gland (acinus) formation
    a. Score 1: more than 75% of the whole carcinoma has formed acini.
    b. Score 2: 10-75% of the whole carcinoma has formed acini.
    c. Score 3: less than 10% of the whole carcinoma has formed acini.
2) Nuclear pleomorphism:
    a. Score 1: nuclei only slightly larger than benign breast epithelium (<1.5 x normal area); normal variation in size, shape and chromatin pattern.
    b. Score 2: nuclei distinctly enlarged (1.5-2 x normal area), often vesicular, nucleoli visible; may be distinctly variable in size and shape but not always.
    c. Score 3: markedly enlarged vesicular nuclei (>2 x normal area), nucleoli often prominent; generally marked variation in shape and size.
3) Mitotic counts:

Mitotic counts were measured as the number of mitoses/10 high power field. After scanning the sections to find the areas of greatest mitotic activity, definite mitoses in ten consecutive fields were counted. Depending on the field diameter of the microscope, the mitotic frequency scored 1, 2 or 3 (Haybittle et al., 1982) (Blamey et al., 2007a).

Final grading was calculated according to the previous scores:
Total score 3, 4 and 5 = grade 1
Total score 6 or 7 = grade 2
Total score 8 or 9 = grade 3 (Ellis et al., 2016).

For integration of Ki67 into NPI index, immunostaining for Ki67 WTS was used. Initially, the correlation between both mitotic activity and Ki67 was evaluated. Then, for the substitution of Ki67 instead of mitotic counts, Ki67 was categorised into three groups using 10% and 70% as cut-offs, as previously published (Aleskandarany et al., 2010) (Aleskandarany et al., 2011). Group 1 included cases that scored between 0% and 9% Ki67 expression. Groups 2 and 3 represented cases between 10-69% and 70-100% of Ki67 expression, respectively. These groups substituted scores 1, 2 and 3 of mitotic score in NPI. The comparison involved plotting time to death from each prognostic group into different curves. BCSS was used as an end-point to compare both the traditional NPI and the Ki67-NPI. The reported *p*-values indicate the degree of significance in difference between the groups' categories of prognostic tools. In the first comparison, three traditional groups of NPI were used: good prognostic group (GPG), moderate prognostic group (MPG) and poor prognostic group (PPG). Then, six groups of NPI were used: excellent prognostic group (EPG), GPG, MPG M1 and MPG M2, PPG and very poor prognostic group (VPG).

3.4.5. Integration of Ki67 into NPI+

The integration of Ki67 in replacing mitotic activity in the BC prognostic classifier NPI+ was investigated. Mitotic counts appeared in the formulae for luminal A and B classes only, **Table-8**. Using Nottingham series (first series, details given in Section-2.1.1), the NPI+ biological classes were previously determined (Green et al., 2013) (Soria et al., 2013) and the NPI+ prognostic groups were calculated (Rakha et al., 2014). For the luminal A and B classes, 234 and 141 cases were available for assessment.

Table-8: NPI+ formulae for the Biological classes (Green et al., 2016)

Class	NPI+ formula
Luminal A	(0.8 x Mitosis) + (0.5 x Size) + (1.8 x Nodal ratio*)
Luminal N	(0.8 x Tubules) + (0.6 x Stage)
Luminal B	(0.7 x Mitosis) + (1.0 x Nodal ratio)
Basal p53 altered	(1.4 x Nodal ratio) + (0.4 x Size)
Basal p53 normal	(0.6 x Stage) + (1.8 x Pleomorphism)
HER2+/ER+	(0.5 x Size) + (0.9 x Stage)
HER2+/ER-	(0.9 x Stage) – (0.6 x Nodal ratio)

*Number of nodes positive / total number of nodes.

3.4.5.1. Development of NPI formulae for each biological class after integration of Ki67:

A Cox regression analysis was performed for all cases for selecting available and traditional histopathologic prognostic factors. The variables tested were coded numerically, categorically or continuously. The variables coded as categorical variables included tumour size, nodal stage, tubule formation and nuclear pleomorphism while nodal ratio and Ki67 scores were continuous variables. Grade was excluded as mitotic count is one of the variables included into the grade. After successive removal of the least significant parameters, the final factors with the most significant results were identified, depending on their β-value in the Cox regression analysis. The proportional hazard ratio Cox regression identified nodal ratio and tumour size as clinicopathologic prognostic factors of importance within the population in predicting BCSS. The new formulae for Luminal A and Luminal B classes using Cox regression models with Ki67 replacing mitosis was performed by Dr. Daniele Soria (Computer Science, University of Nottingham).

1) The new NPI+ formula for Luminal A class:

   NPI+ = (2.322 * NodalRatio) + (0.018 * Ki67)

2) The new NPI+ formula for Luminal B class:

   NPI+ = (0.329 * Size) + (0.019 * ki67)

The new NPI+ values were computed, and Kaplan-Meier curves produced and compared with the original NPI+ in each class. In addition, to improve outcome prediction using NPI+ and to provide more clinically relevant stratification, each class was split into two or three groups. X-tile software was used to subdivide the classes into prognostic subgroups.

3.5. Statistical analysis

The statistical analysis was performed using IBM SPSS software version 22 (SPSS Inc., Chicago, IL, USA) (PAWS statistics). *P*-value was considered as a significant value when it was <0.05. Intraclass Correlation Coefficient (ICC), Kappa statistic and Spearman's correlation test were used to test the reproducibility and the correlation between the TMA and WTS Ki67 stained cases. In Kappa, complete agreement is revealed by a value of 1.0 while only by chance alone results is reflected by a value of zero. Although in the literature there is no agreed standard criteria for Kappa-value that indicates adequate agreement, Landis and Koch proposed the following agreement values for categorical data: Kappa-value <0.00 signifies poor agreement, 0.00-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial and 0.81-1.00 almost perfect agreement (Landis and Koch, 1977). Accordingly, mikami, Y., et al. proposed that based on the similarity to the Kappa coefficient, ICC between 0.41-0.60 is considered as moderate correlation; 0.61-0.80 as substantial correlation and >0.80 as a perfect correlation (Mikami et al., 2013). Chi-square tests were applied to reflect the correlation with the essential prognostic biomarker and standard clinicopathological parameters of BC patients. Cox regression models with Ki67 replacing mitosis was used. Kruskal Wallis test was

used to analyse the correlations of proliferation marker with mitosis. Survival curves were generated by the Kaplan-Meier method and differences in survival were assessed by the Log Rank test. BCSS were used as end points. X-tile software was used to subdivide the classes into prognostic subgroups.

3.6. Results

3.6.1. Comparison of Ki67 expression between TMA and WTS

3.6.1.1. Comparison of Ki67 expression between TMA and WTS at protein level

Using WTS, the expression of Ki67 was not normally distributed. The upper and lower limit were 0% and 99% respectively, the mean percentage was 34.8% while the median percentage was 20%. Similar to the distribution observed with WTS, Ki67 expression using TMA was not normally distributed. The range was between 0% and 95%, the mean percentage was 21.9 % while the median percentage was 10%.

When the concordance of Ki67 expression was compared between both TMA and WTS as a continuous variables, there was a significant correlation (*p*-value <0.001), the Spearman's correlations was 0.50 with an r^2 value of 0.254, **Figure-13**. In addition, when ICC was assessed, substantial correlation was resulted (ICC=0.61, *p*-value <0.001, 95% CI=0.45-0.71). The reproducibility of Ki67 expression between WTS and TMA was assessed using different cut-offs, initially, similar cut-points from both TMA and WTS were compared. The highest concordance was achieved when WTS is 10%, **Figure-14**. Therefore, for subsequent analysis, the data was dichotomised using 10% as a fixed cut-off for WTS. Additionally, it was previously chosen as an optimal cut-off (Aleskandarany et al., 2011). However, when dichotomised data was used to evaluate the concordance of Ki67 expression, low concordance was observed;

Kappa-value =0.3 when 10%, 15%, 20% cut-offs were used while Kappa-value lower than 0.3 with higher cut-offs.

Different cut-offs of TMA were tested to evaluate the reproducibility of the WTS Ki67 expression; 10%, 15%, 20%, 25% and 30%, **Table-9**. As shown in, **Figure-15**, higher cut-off values resulted in misclassification of high percentage of TMA cases (62.8%) into the low proliferation group compared with WTS. In contrast, lower cut-off values resulted in a higher number of cases matching between the positive cases, however, there was a high percentage of false positive TMA cases compared with WTS (49.8%). Although it is difficult to choose an optimal cut-off point, 20% cut-off for Ki67 determined using TMA seems to give the highest concordance in both positive and negative groups with less both false positive and negative Ki67 expression at 10% determined using WTS.

To assess for the impact of cut-offs determination on patient outcome as an end point, cases were categorised based on their Ki67 expression, whether it is low or high, on both TMAs and WTS. Hence, using 10% as a cut-off for Ki67 on WTS and 20% as a cut-off for on TMAs, four groups were resulted. Group one included cases with low Ki67 expression on TMAs and their corresponding WTS cases, group two consisted of cases with high Ki67 status on TMAs with low Ki67 status on their matched WTS, group three contained cases with low Ki67 status on TMAs with high Ki67 expression on their matched WTS and group four comprised high Ki67 expression on TMAs with high Ki67status on their matched WTS. Statistically significant differences were resulted between these groups with respect to the patients outcome, *p*-value <0.001 and Log Rank =31.79, **Figure-16**. Interestingly, there was no significant difference between the groups comprising high Ki67 expression on WTS (LR =0.39, p = 0.52), therefore, they appears to have more or less similar poor outcome.

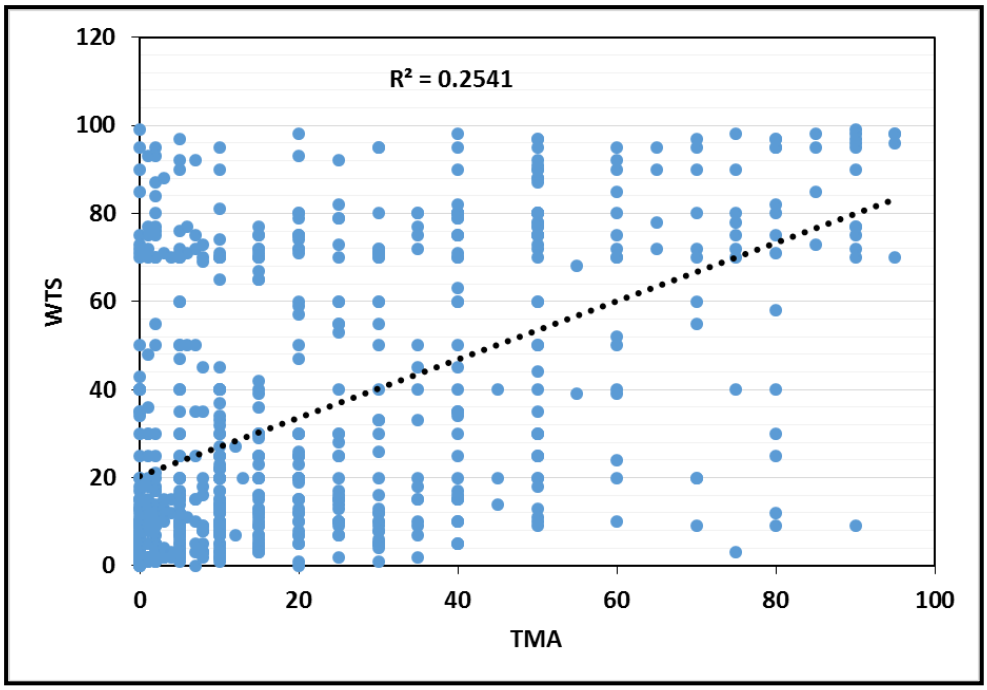


Figure-13: Correlation between Ki67 expressions evaluated on matched cases on WTS and TMA (Muftah et al., 2017)

Table-9: Demonstrates the results of the different cut-off points (TMA) with cut-off 10% (WTS)

Ki67 on TMAs (%)	Ki67 on WTS (at 10%)	
	Negative (%)	Positive (%)
10% low	135(47.9)	147(52.1)
High	68(16.0)	357(84.0)
15% low	157(43.9)	201(56.1)
High	46(13.2)	303(86.8)
20% low	170(41.8)	237(58.2)
High	33(11.0)	267(89.0)
25% low	177(39.1)	276(60.9)
High	26(10.2)	228(89.8)
30% low	181(37.2)	305(62.8)
High	22(10.0)	199(90.0)

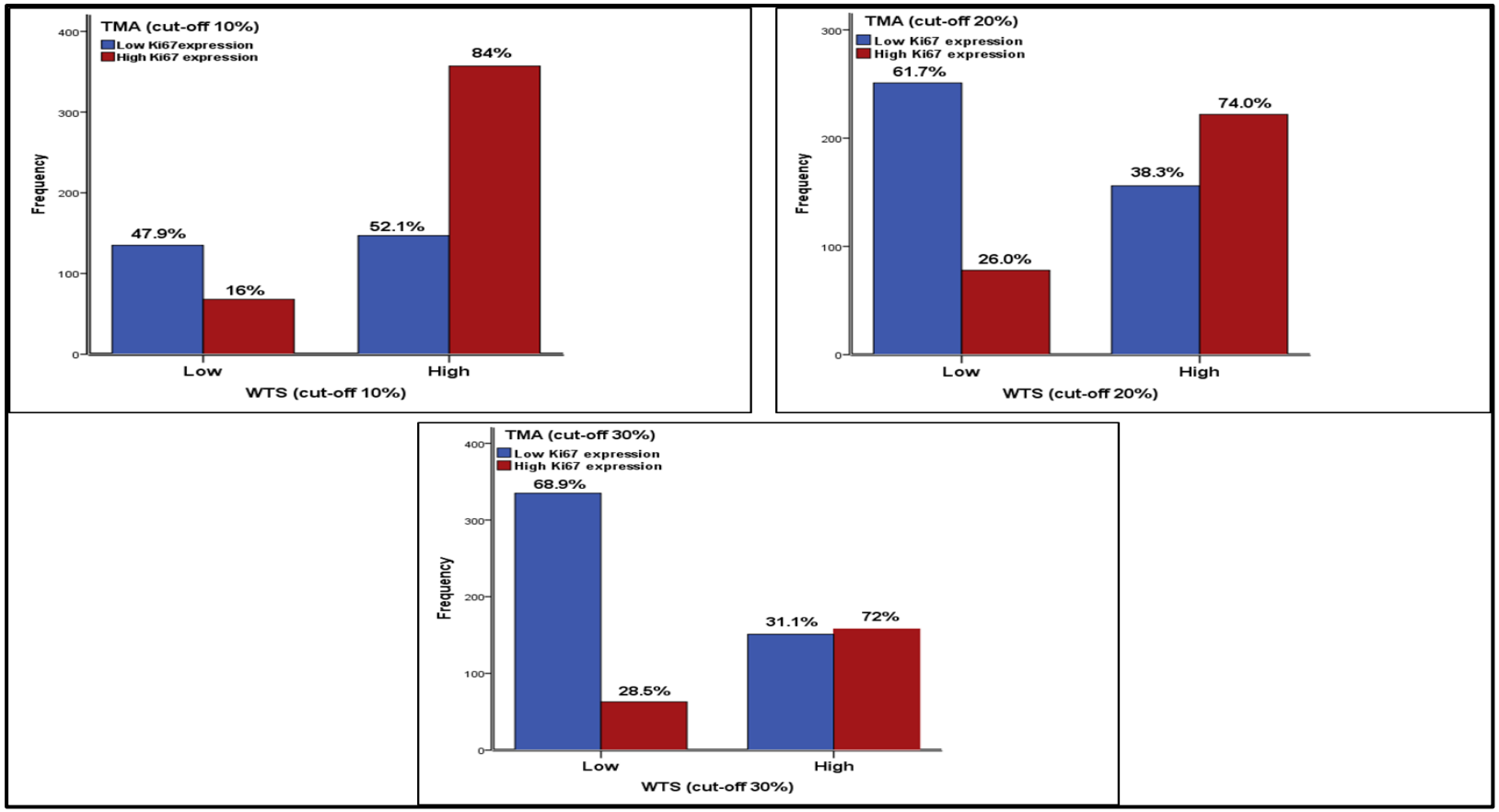


Figure-14: The bar chart demonstrates the percentage of matched cases between TMA and WTS of the whole cohort using similar cut-off points

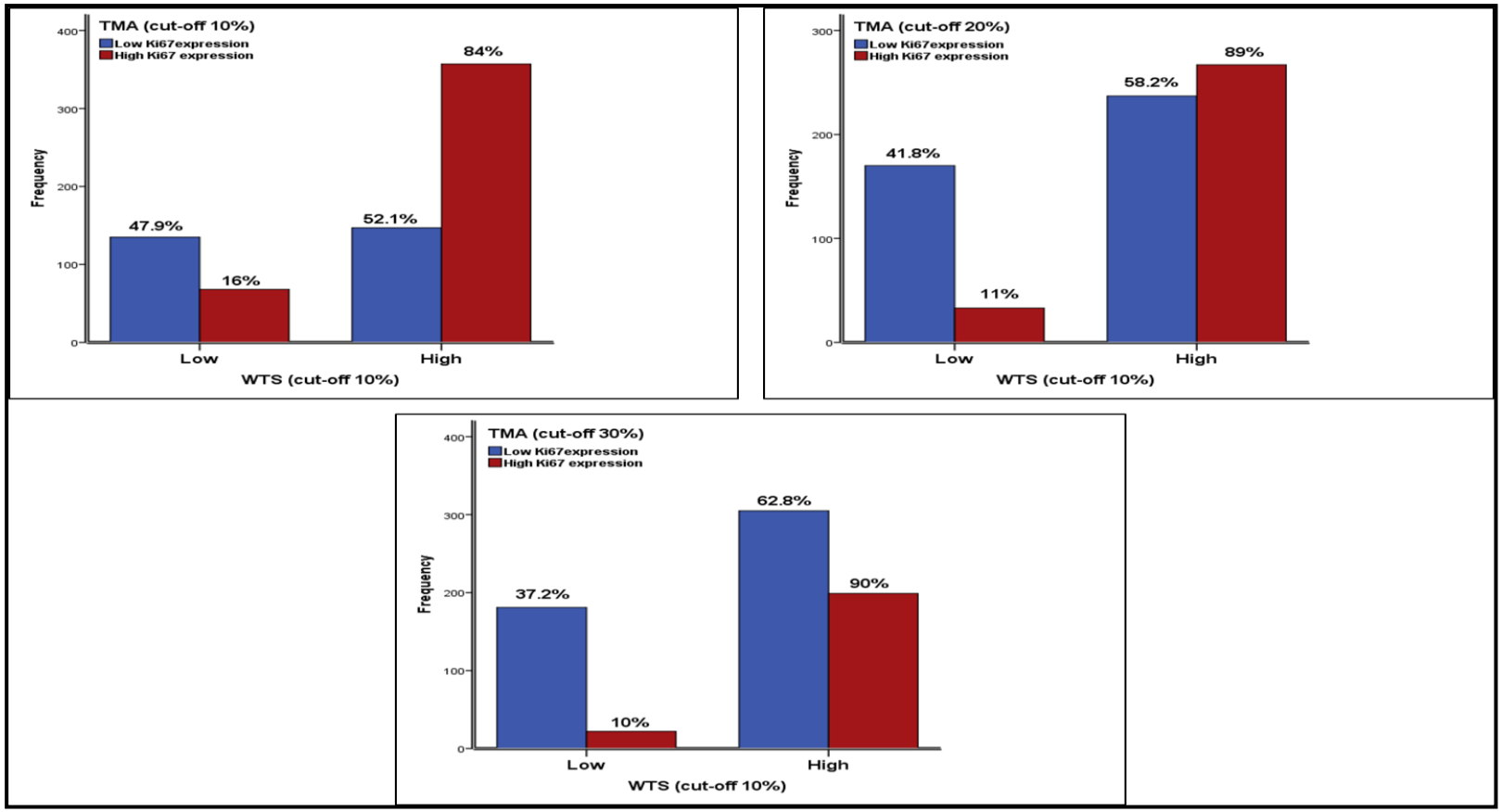


Figure-15: The bar chart demonstrates the percentage of matched cases between TMA and WTS of the whole cohort using different cut-off points

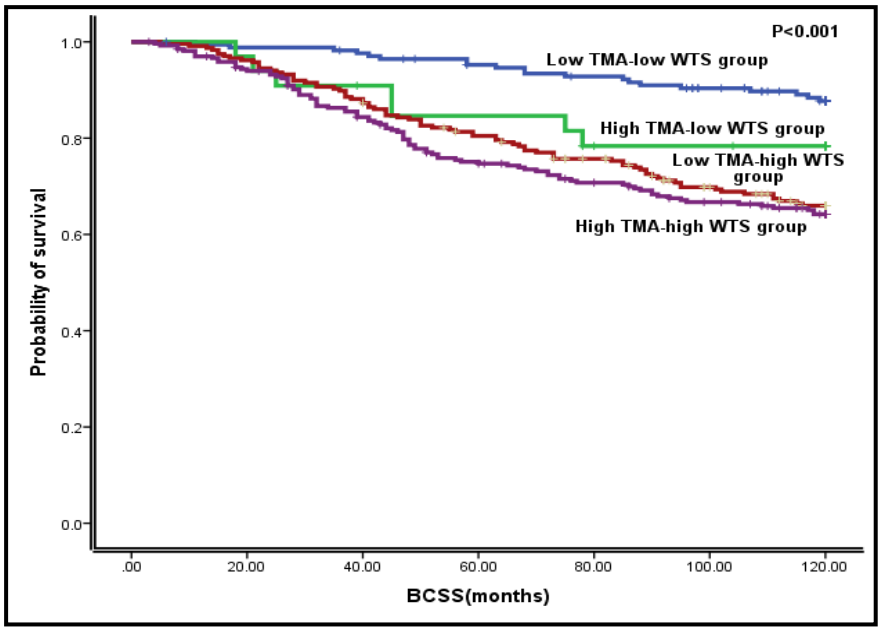


Figure-16: BCSS for Ki67 high/low groups as determined by Ki67 status on WTS and TMA at 10% and 20%, respectively (Muftah et al., 2017)

3.6.1.2. Comparison of Ki67 expression between TMA and WTS in different molecular subclasses

The studied cases were defined regarding their molecular class as luminal, which contains ER-positive cases; HER2-positive class and triple-negative (TN), which comprises ER-negative, PR-negative and HER2-negative cases. Evaluation of the concordance between Ki67 expression on TMA and WTS in molecular subclasses utilising different cut-off values presented that; in TN and HER2-positive BC, a cut-off of 20% appeared to give a highest concordance between WTS and TMA. In TN classes, the highest concordance between the positive cases (91.5%) with lowest false positive (8.5%) was shown with 20% cut-point using TMA, **Figure-17**. Along the same lines, for HER2-positive 20% seemed to produce the most suitable cut-off to classify the patients using TMA. 96.2% concordance for the positive cases and 3.8% represented false positive cases **Figure-18**. In the luminal class, there seemed to be no optimal cut-off for Ki67 determined using TMA which was reproducible to WTS scoring, **Figure-19**.

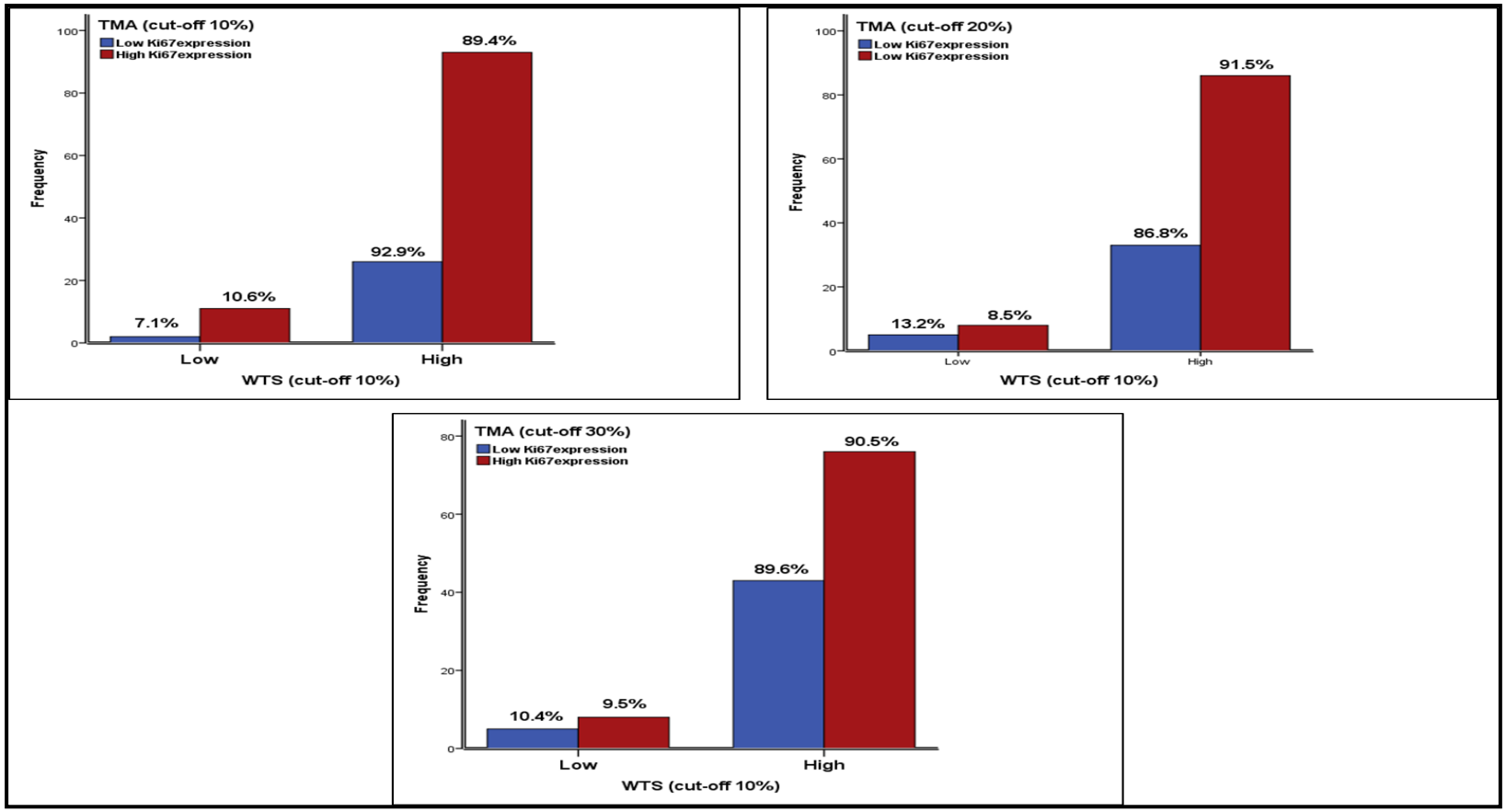


Figure-17: The bar chart demonstrates the percentage of matched cases between TMA and WTS using Triple-negative class

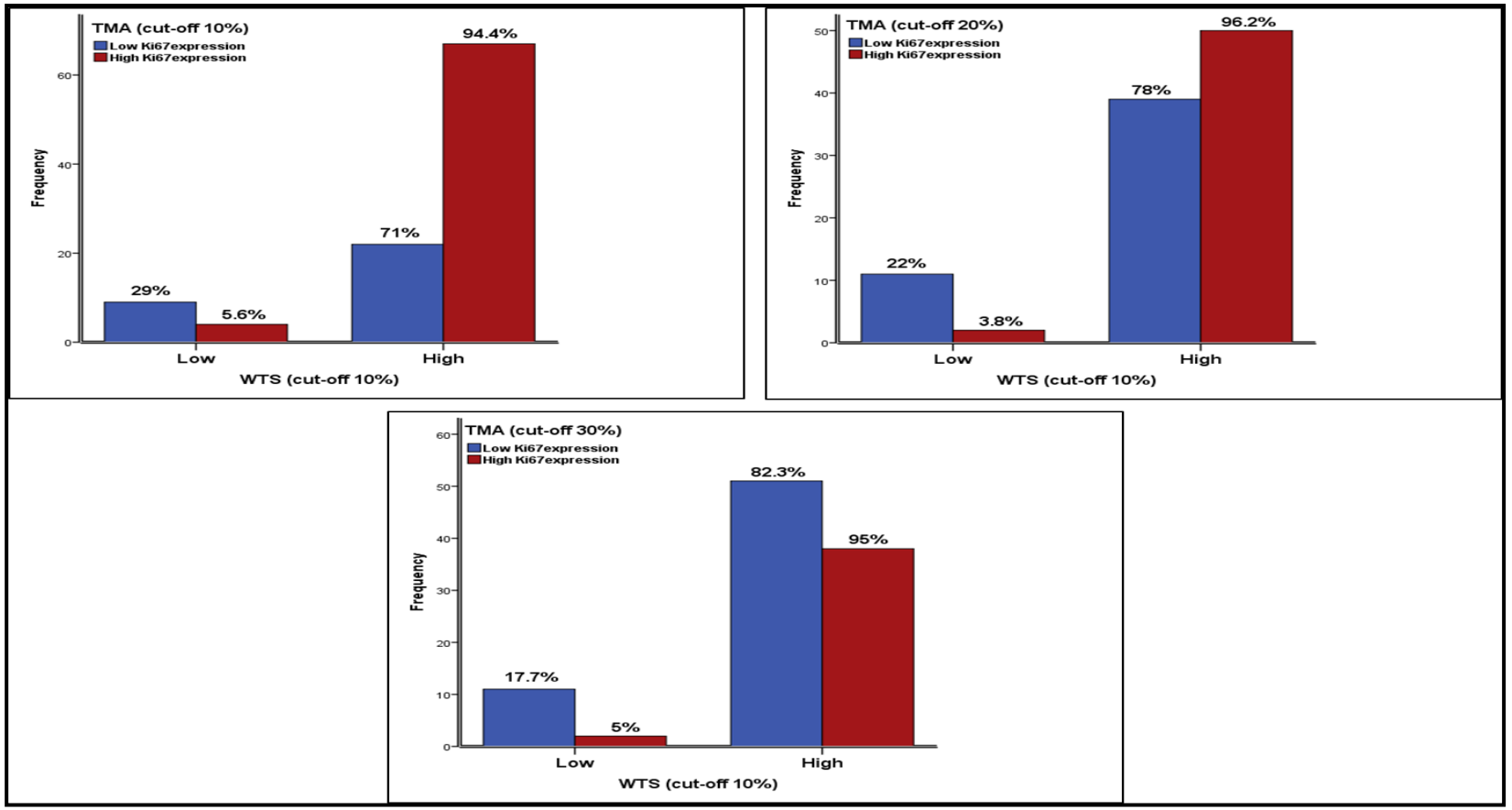


Figure-18: The bar chart demonstrates the percentage of matched cases between TMA and WTS using Her2-positive class

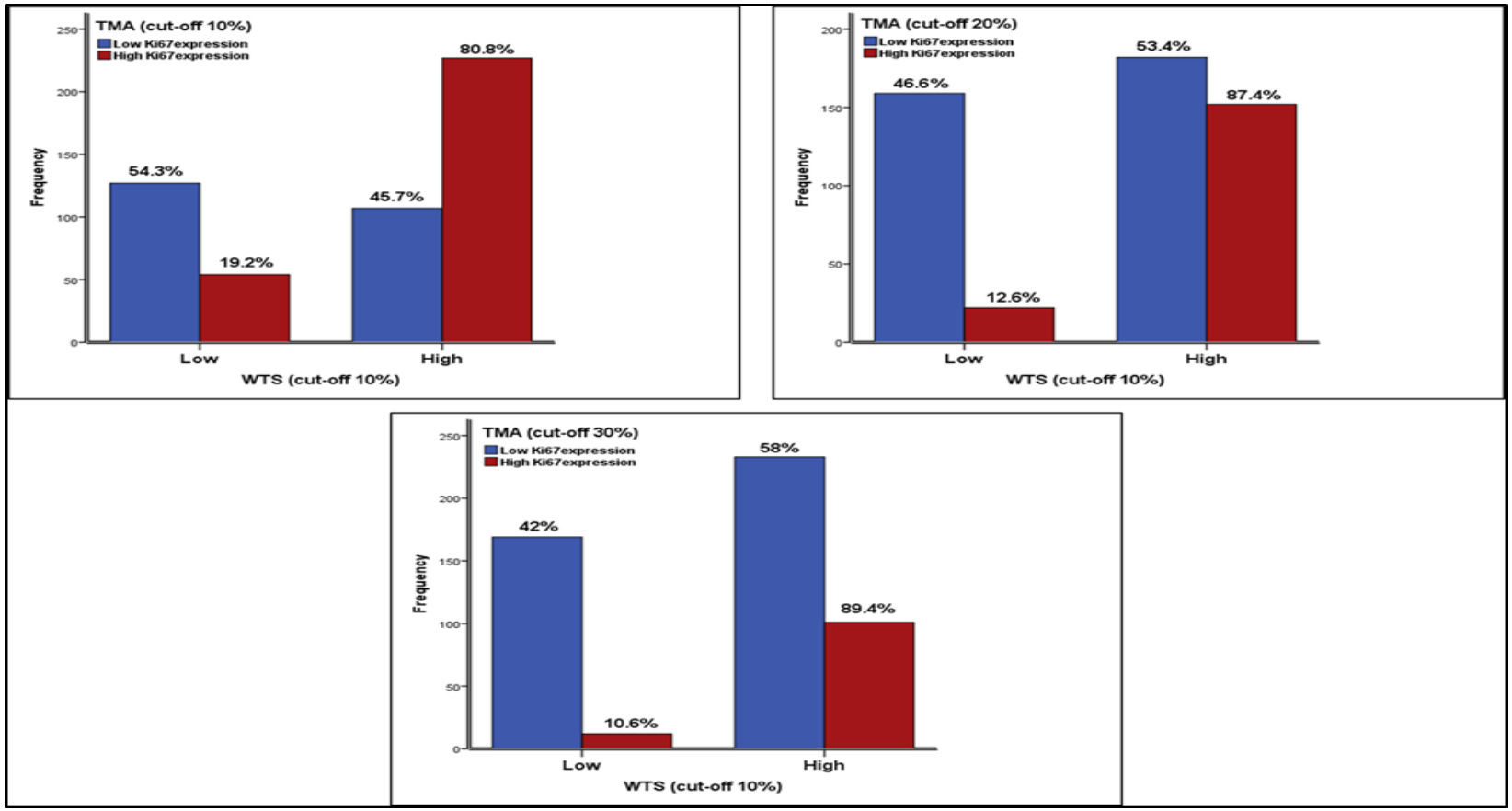


Figure-19: The bar chart demonstrates the percentage of matched cases between TMA and WTS using Luminal class

Comparison of Ki67 expression on TMA and WTS with *MKI67* mRNA expression

Using METABRIC cohort, *MKI67* data was matched with 197 cases of WTS and 123 of TMA cases. The correlation between *MKI67* mRNA and WTS (using 197 cases) was significant, *p*-value <0.001 and Spearman's correlation coefficient =0.587. Although the correlation between *MKI67* mRNA and Ki67 assessed on TMAs (using 123 cases) was significant (*p*-value <0.001), Spearman's correlations was less significant compared with WTS =0.343. **Figure-20** showed the correlation between MKI67 mRNA with WTS ($r^2 = 0.31$) and TMA ($r^2 = 0.17$). In matching 101 BC cases examined on both TMA and WTS, *MKI67* mRNA expression was assessed. Higher significance was shown for Ki67 evaluated on WTS (Spearman's correlation coefficient =0.529 and *p* <0.001) than on TMA (Spearman's correlation coefficient =0.341 and *p*<0.001).

3.6.2. The reproducibility of Ki67 scoring

To assess inter-observer variability, ICC was calculated with a 95% CI. The assessment of reproducibility of the Ki67 TMA scoring revealed that there is a significant correlation (*p*-value <0.001) between the two scorer, ICC = 0.870 (95% CI = 0.838-0.896) which is considered as an almost perfect concordance. On the other hand, Ki67 scoring on WTS showed substantial concordance, ICC = 0.75 (95% CI = 0.670-0.815).

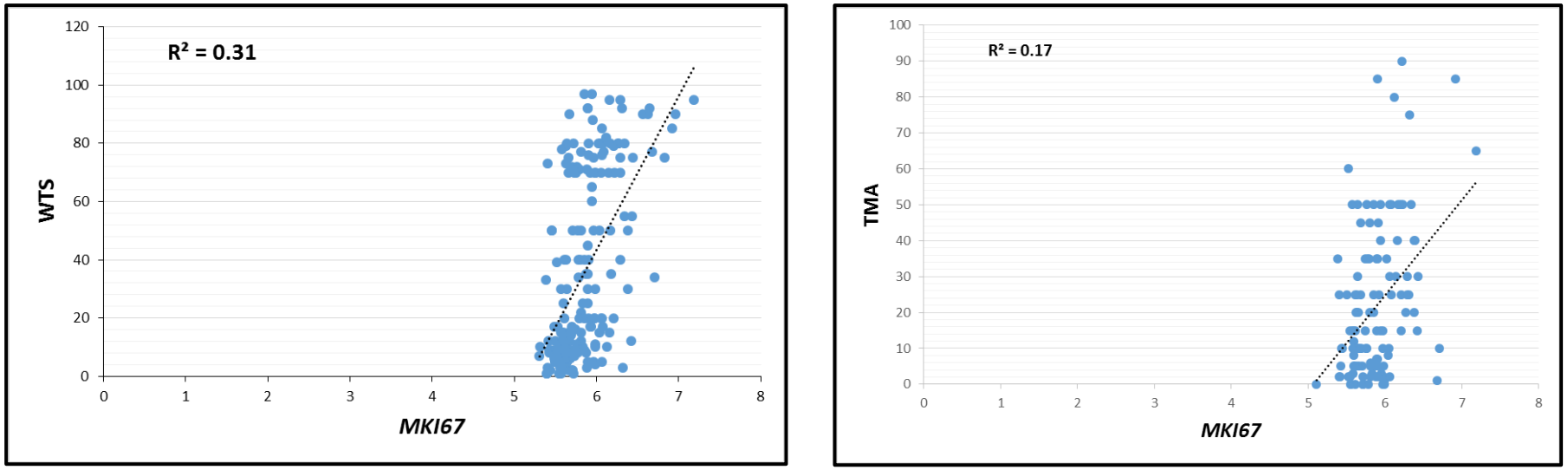


Figure-20: Correlation between *MKI67* mRNA and immunostained Ki67 evaluated on WTS and TMA (Muftah et al., 2017)

3.6.3. The association with the important clinicopathological variables

When Ki67 was evaluated using WTS, the high expression of Ki67 had a significant (*p*-value <0.05) positive association with larger tumour size, nuclear pleomorphism, higher grade, more higher mitotic scores, positive axillary nodal status and less tubule formation with a wide range of cut-offs. Similarly, when TMA was used, high expression of Ki67 was significantly associated with larger tumour size, higher grade, more nuclear pleomorphism, higher mitotic scores and less tubule formation with a wide range of cut-offs (*p*-value <0.05). Interestingly, using TMA, the highest X^2 values were associated with the cut-off 20%, which was suggested as the optimal cu-off to reproduce WTS scores. Although both WTS and TMA were significantly correlated with these clinicopathological variables, WTS has higher correlation with higher X^2 values compared with TMA, **Table-10** and **Table-11**.

Table-10: The correlation between the different cut-off values with the important clinicopathological parameters (TMA)

TMA (%)		Tumour Size (cm)				Grade				
		< 2.0	>= 2.0	*p*-value	X2	1	2	3	*p*-value	X^2
10%	**low**	154 (54.6)	128 (45.4)	0.001	11.4	69 (24.5)	114 (40.4)	99 (35.1)	<0.001	97.1
	high	177 (41.6)	248 (58.4)			24 (5.7)	102 (24.1)	297 (70.2)		
15%	**low**	192(53.6)	166(46.4)	<0.001	13.5	80(22.3)	142(39.7)	136(38.0)	<0.001	108.3
	high	139(39.8)	210(60.2)			13(3.7)	74(21.3)	260(74.9)		
20%	**low**	218(53.6)	189(46.4)	<0.001	17.5	84(20.7)	165(40.6)	157(38.7)	<0.001	124.2
	high	113(37.7)	187(62.3)			9(3.0)	51(17.1)	239(79.9)		
25%	**low**	238(52.5)	215(47.5)	<0.001	16.5	88(19.5)	175(38.7)	189(41.8)	<0.001	110.6
	high	93(36.6)	161(63.4)			5(2.0)	41(16.2)	207(81.8)		
30%	**low**	255(52.5)	231(47.5)	<0.001	19.945	89(18.4)	186(38.4)	210(43.3)	<0.001	107.3
	high	76(34.4)	145(65.6)			4(1.8)	30(13.6)	186(84.5)		

TMA (%)		Pleomorphism					Mitosis				
		1	2	3	*p*-value	X^2	1	2	3	*p*-value	X^2
10%	**low**	9 (3.3)	133 (48.7)	131 (48.0)	<0.001	59.23	142 (51.8)	61 (22.3)	71 (25.9)	<0.001	136.35
	high	1 (0.2)	100 (24.2)	313 (75.6)			61 (14.7)	71 (17.1)	282 (68.1)		
15%	**low**	9(2.6)	161(46.5)	176(50.9)	<0.001	59.42	169(48.7)	76(21.9)	102(29.4)	<0.001	155.66
	high	1(0.3)	72(21.1)	268(78.6)			34(10.0)	56(16.4)	251(73.6)		
20%	**low**	9(2.3)	180(45.8)	204(51.9)	<0.001	65.63	182(46.2)	89(22.6)	123(31.2)	<0.001	165.10
	high	1(0.3)	53(18.0)	240(81.6)			21(7.1)	43(14.6)	230(78.2)		
25%	**low**	9(2.1)	191(43.6)	238(54.3)	<0.001	56.25	187(42.6)	97(22.1)	155(35.3)	<0.001	136.33
	high	1(0.4)	42(16.9)	206(82.7)			16(6.4)	35(14.1)	198(79.5)		
30%	**low**	10(2.1)	200(42.5)	261(55.4)	<0.001	56.53	193(40.9)	106(22.5)	173(36.7)	<0.001	137.35
	high	0(0.0)	33(15.3)	183(84.7)			10(4.6)	26(12.0)	180(83.3)		

TMA (%)		Tubular formation					LN status				
		1	2	3	*p*-value	X^2	1(0LN)	2 (1-3LN)	3(>3LN)	*p*-value	X^2
10%	**low**	23 (8.4)	95 (34.7)	156 (56.9)	<0.001	17.29	150 (61.7)	75 (30.9)	18 (7.4)	0.120	4.23
	high	11 (2.7)	116 (28)	287 (69.3)			201 (53.9)	145 (38.9)	27 (7.2)		
15%	**low**	28(8.1)	116(33.4)	203(58.5)	<0.001	19.36	186(60.4)	101(32.8)	21(6.8)	0.231	2.92
	high	6(1.8)	95(27.9)	240(70.4)			165(53.6)	119(38.6)	24(7.8)		
20%	**low**	30(7.6)	138(35.0)	226(57.4)	<0.001	26.10	210(59.7)	120(34.1)	22(6.3)	0.235	2.89
	high	4(1.4)	73(24.8)	217(73.8)			141(53.4)	100(37.9)	23(8.7)		
25%	**low**	31(7.1)	152(34.6)	256(58.3)	<0.001	24.16	232(59.3)	135(34.5)	24(6.1)	0.177	3.46
	high	3(1.2)	59(23.7)	187(75.1)			119(52.9)	85(37.8)	21(9.3)		
30%	**low**	32(6.8)	163(34.5)	277(58.7)	<0.001	25.19	248(58.8)	145(34.4)	29(6.9)	0.410	1.78
	high	2(0.9)	48(22.2)	166(76.9)			103(53.1)	75(38.7)	16(8.2)		

Table-11: The correlation between the results of the different cut-off points with the important clinicopathological parameters (WTS)

WTS (%)		Tumour Size (cm)				Grade				
		< 2.0	>= 2.0	*p*-value	X^2	1	2	3	*p*-value	X^2
10%	**low**	119(58.6)	84(41.4)	<0.001	15.93	75(36.9)	93(45.8)	35(17.2)	<0.001	220.29
	high	212(42.1)	292(57.9)			18(3.6)	123(24.5)	361(71.9)		
15%	**low**	157(55.7)	125(44.3)	<0.001	14.77	82(29.1)	134(47.5)	66(23.4)	<0.001	223.46
	high	174(40.9)	251(59.1)			11(2.6)	82(19.4)	330(78.0)		
20%	**low**	181(55.0)	148(45.0)	<0.001	16.60	86(26.1)	155(47.1)	88(26.7)	<0.001	228.11
	high	150(39.7)	228(60.3)			7(1.9)	61(16.2)	308(81.9)		
25%	**low**	204(55.1)	166(44.9)	<0.001	21.56	88(23.8)	170(45.9)	112(30.3)	<0.001	218.76
	high	127(37.7)	210(62.3)			5(1.5)	46(13.7)	284(84.8)		
30%	**low**	215(54.0)	183(46.0)	<0.001	18.97	90(22.7)	175(44.1)	132(33.2)	<0.001	200.47
	high	116(37.5)	193(62.5)			3(1.0)	41(13.3)	264(85.7)		

WTS (%)	Tubular formation					LN status				
	1	2	3	*p*-value	X^2	1(0LN)	2 (1-3LN)	3(>3LN)	*p*-value	X^2
10% low	29(14.6)	77(38.9)	92(46.5)	<0.001	72.97	128(69.2)	50(27.0)	7(3.8)	<0.001	16.99
high	5(1.0)	134(27.3)	351(71.6)			223(51.7)	170(39.4)	38(8.8)		
15% low	31(11.4)	111(40.8)	130(47.8)	<0.001	72.25	165(65.2)	78(30.8)	10(4.0)	0.001	14.58
high	3(0.7)	100(24.0)	313(75.2)			186(51.2)	142(39.1)	35(9.6)		
20% low	32(10.1)	133(42.0)	152(47.9)	<0.001	80.68	190(64.8)	91(31.1)	12(4.1)	<0.001	17.34
high	2(0.5)	78(21.0)	291(78.4)			161(49.8)	129(39.9)	33(10.2)		
25% low	34(9.5)	146(40.9)	177(49.6)	<0.001	82.11	213(65.1)	100(30.6)	14(4.3)	<0.001	22.00
high	0(0.0)	65(19.6)	266(80.4)			138(47.8)	120(41.5)	31(10.7)		
30% low	34(8.9)	154(40.2)	195(50.9)	<0.001	77.08	224(64.0)	110(31.4)	16(4.6)	<0.001	19.46
high	0(0.0)	57(18.7)	248(81.3)			127(47.7)	110(41.4)	29(10.9)		

WTS (%)	Pleomorphism					Mitosis				
	1	2	3	*p*-value	X^2	1	2	3	*p*-value	X^2
10% low	10(5.1)	127(64.1)	61(30.8)	<0.001	148.86	149(75.3)	21(10.6)	28(14.1)	<0.001	282.69
high	0(0.0)	106(21.7)	383(78.3)			54(11.0)	111(22.7)	325(66.3)		
15% low	10(3.7)	159(58.7)	102(37.6)	<0.001	146.66	171(62.9)	49(18.0)	52(19.1)	<0.001	260.86
high	0(0.0)	74(17.8)	342(82.2)			32(7.7)	83(20.0)	301(72.4)		
20% low	10(3.2)	173(54.7)	133(42.1)	<0.001	132.61	178(56.2)	73(23.0)	66(20.8)	<0.001	252.47
high	0(0.0)	60(16.2)	311(83.8)			25(6.7)	59(15.9)	287(77.4)		
25% low	10(2.8)	184(51.7)	162(45.5)	<0.001	119.90	184(51.5)	85(23.8)	88(24.6)	<0.001	233.15
high	0(0.0)	49(14.8)	282(85.2)			19(5.7)	47(14.2)	265(80.1)		
30% low	10(2.6)	186(48.7)	186(48.7)	<0.001	97.18	187(48.8)	90(23.5)	106(27.7)	<0.001	211.69
high	0(0.0)	47(15.4)	258(84.6)			16(5.2)	42(13.8)	247(81.0)		

3.6.4. The association with survival using different Cut-off values

Regarding patient outcome, univariate survival analysis of Ki67 defined using TMA cohort showed that a cut-off of 20% was the most significantly correlated with BCSS (Log Rank =8.76, *p*-value =0.003, **Table-12**). Furthermore, using Cox regression, variant cut-offs points and the association with BCSS were assessed, **Table-13**. Interestingly, 20% gave the best cut-off (hazard ratio (HR) =1.52, 95% CI=1.17-1.95, *p*-value =0.001). In univariate survival analysis of the WTS cohort using Kaplan-Meier test presented that Ki67 at a cut-off of 10% was significantly associated with BCSS (Log Rank =30.1, *p*-value <0.001, **Table-14**). Furthermore, using Cox regression, a cut-off of 10% expression of Ki67 showed the best cut-point (HR = 2.95, 95% CI = 1.96-4.43, *p*-value <0.001, **Table-15**).

Table-12: Different cut-off values and the relations with BCSS (TMA)

Ki67 cut-points (%)	BCSS (*p*-value)	Log Rank
10	0.008	7.135
15	0.011	6.526
20	0.003	8.766
25	0.026	4.979
30	0.033	4.547

Table-13: Cox proportional hazards analysis for predictors of BCSS (TMA)

Ki67 cut-points (%)	*p*-value	HR	95% CI
10	0.008	1.496	1.111-2.016
15	0.011	1.440	1.086-1.908
20	0.003	1.519	1.149-2.009
25	0.027	1.379	1.038-1.832
30	0.034	1.370	1.024-1.832

Table-14: Different cut-off values and the relations with BCSS (WTS)

Ki67 cut-points (%)	BCSS (*p*-value)	Log Rank
10	<0.001	30.144
15	<0.001	19.176
20	<0.001	23.292
25	<0.001	22.436
30	<0.001	18.821

Table-15: Cox proportional hazards analysis for predictors of BCSS (WTS)

Ki67 cut-points (%)	*p*-value	HR	95% CI
10	<0.001	2.953	1.967-4.432
15	<0.001	1.984	1.451-2.714
20	<0.001	2.050	1.521-2.763
25	<0.001	1.970	1.479-2.623
30	<0.001	1.843	1.392-2.442

3.6.5. Assessment of different immunostaining conditions

When using the MIB-1 clone, cytoplasmic and membranous staining were noted in both WTS and TMA. Upon further investigation, it was noticed that the maximum dilution that could be used was 1:200. However, there was no change in the cytoplasmic/membranous staining. The dilution of antibody (MIB-1) concentration to 1:300 or more, in order to reduce cytoplasmic and membranous staining, also affected the intensity of the nuclear staining, **Figure-21**. Using MIB-1, there were no obvious differences that could be appreciated when different retrieval times were applied, **Figure-22**. Regarding the use of various antigen retrieval methods with different pH levels (pH6 or pH9), it appeared that there was a slight variation when the pH was changed. There was insignificant reduction in the membranous/cytoplasmic staining, with MIB-1, when pH9 was

used, instead of pH6; however, nuclear staining was also reduced, **Figure-23**. Additionally, there remained no difference in detection of membranous staining between manual and automated staining when MIB-1 was used. Similar cytoplasmic/membranous staining has been seen with MIB-1 and MKi67 clones. However, using SP6 and 30-9 clones resulted in nuclear staining with obvious negative cytoplasmic/membranous reactivity, **Figure 24**.

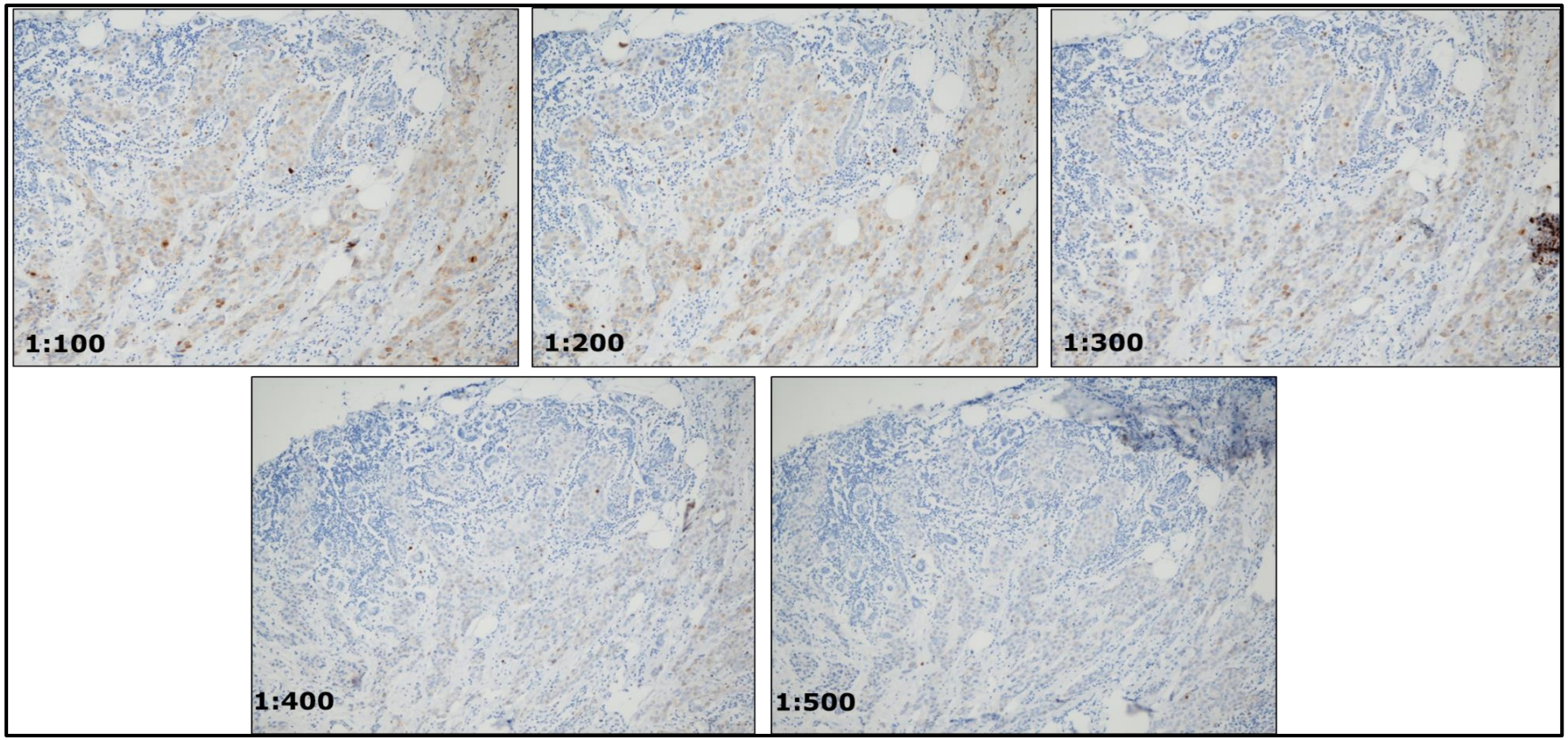


Figure-21: Optimisation using different MIB-1 antibody concentrations using breast cancer tissues ,with higher dilution, there was reduction in cytoplasmic/membranous staining and the intensity of the nuclear staining was affected

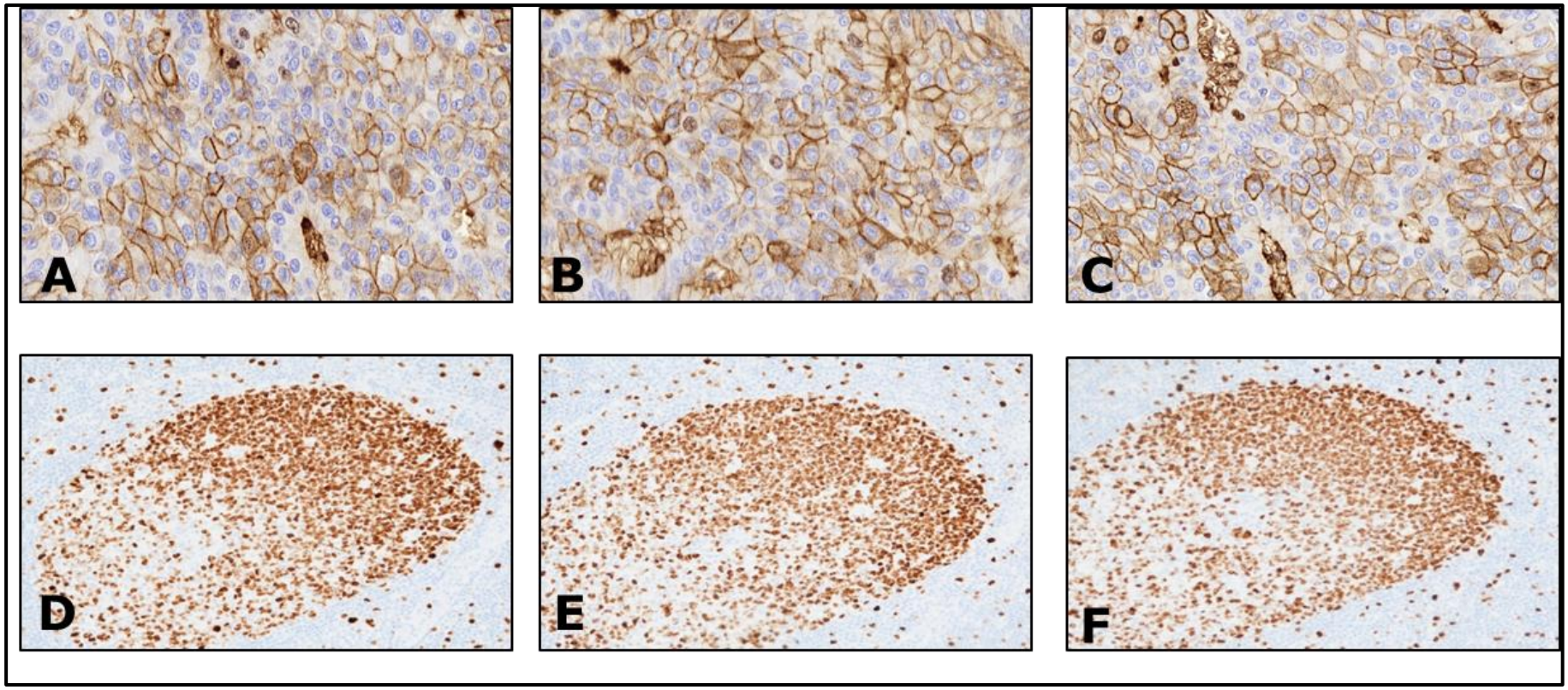


Figure-22: Optimisation of MIB-1 using different antigen retrieval time
Where A, B and C represent breast cancer tissues with different retrieval time for 25, 35, and 45 minutes respectively. D, E and F represent lymph node tissues with different retrieval time for 25, 35, and 45 minutes respectively

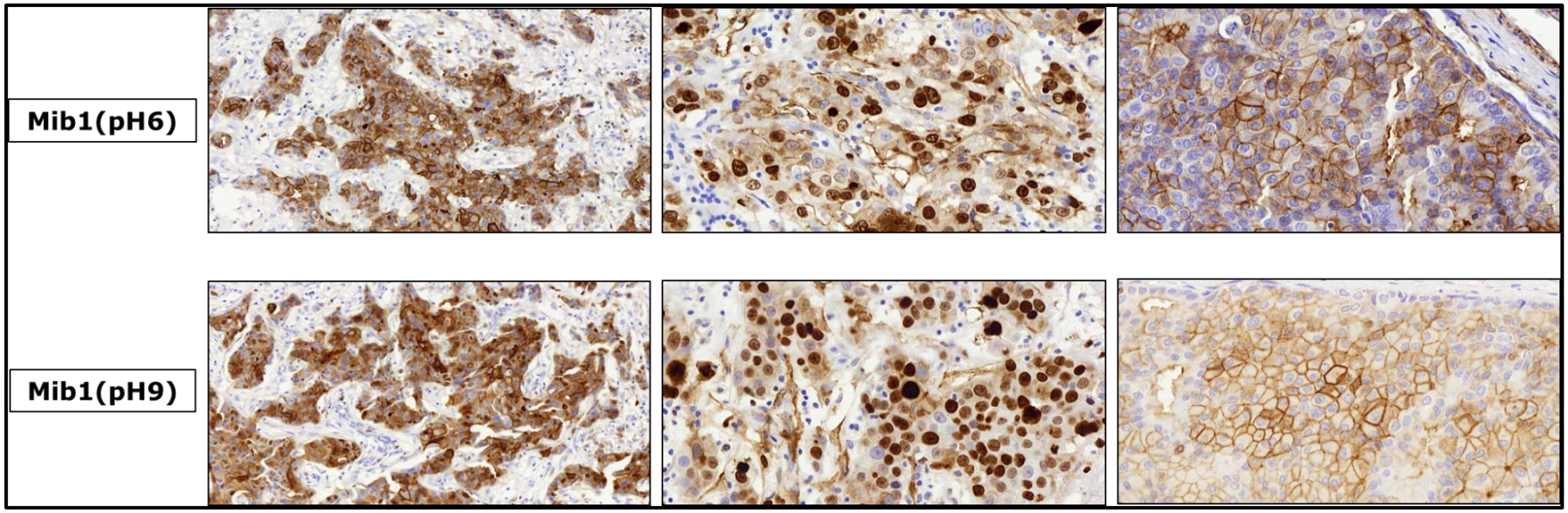


Figure-23: Optimisation of MIB-1 using different antigen retrieval methods using breast cancer tissues
Where the upper images represent different breast cancer tissues using citrate buffer at pH 6.0 using microwave. The bottom images represent the same breast cancer cases using bond epitope retrieval solution 2 at pH 9.0 in the water bath at 100˚C for 35 minutes

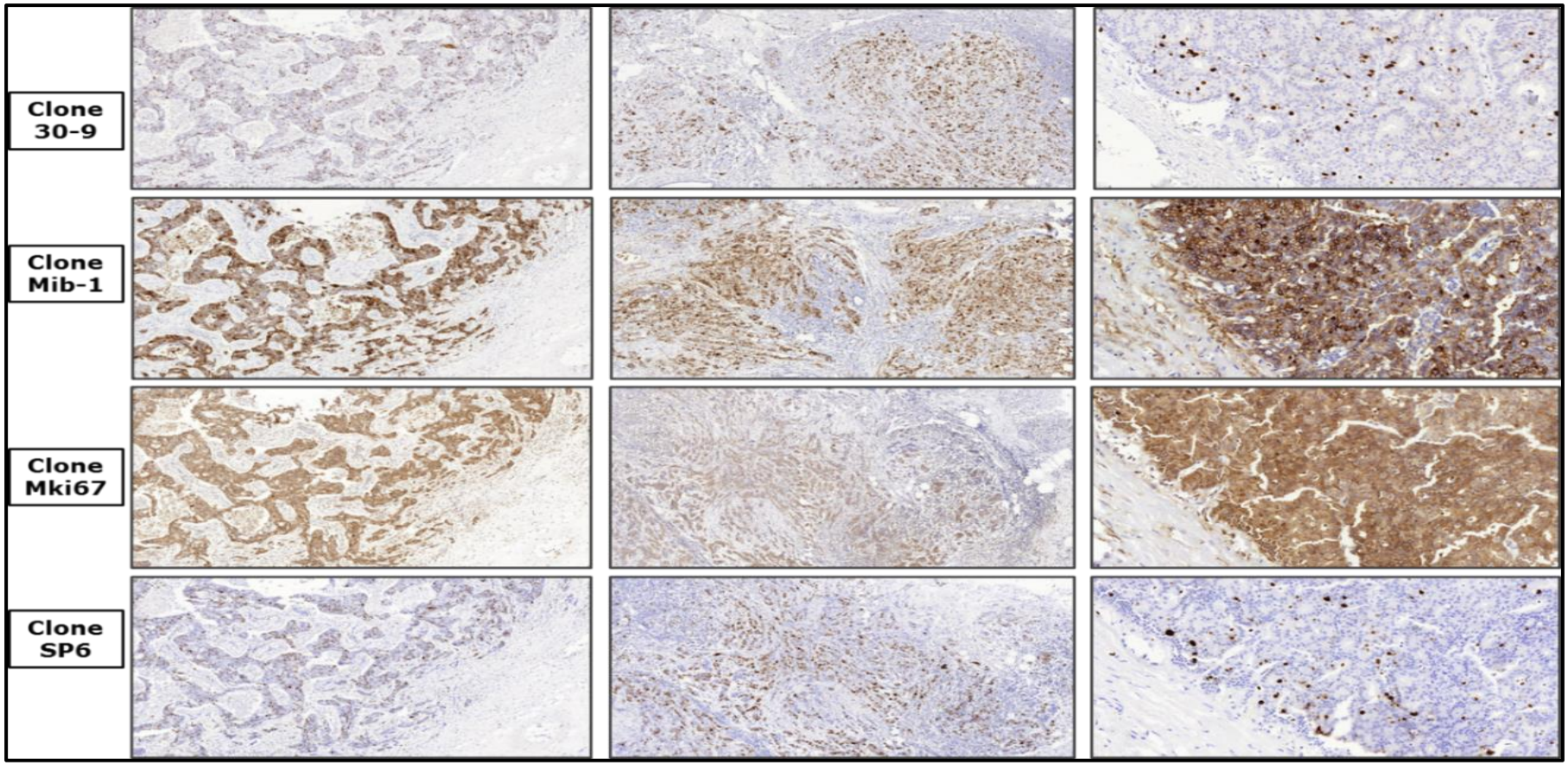


Figure 24: Using different anti Ki67 clones (MIB-1, SP6, MKI67 and 30-9), each column representing the same breast cancer case stained with different clones

3.6.6. Evaluation of integration of Ki67 to NPI and NPI+

Ki67 was integrated into NPI as a categorical variable. Initially, the correlation between mitosis and both continuous and categorical Ki67 expression was evaluated. For the continuous variable, the Kruskal-Wallis test was used and resulted in significant correlation with *p*-value <0.001. Similarly, using the categorical variables, there was a significant correlation; *p*-value <0.001 with X^2 value =847.9. As shown in **Figure-25**, the highest concordance was in the group of mitosis with score 1, where about 80% of these cases had low expression of Ki67.

For the purpose of comparing Ki67 against mitosis in both traditional NPI and the updated version (NPI+), stratified Kaplan-Meier curves were used as a comparison tool. The reported *p*-values indicate the degree of significance in difference between groups. Using the traditional NPI, Ki67 can successfully stratify the patient into groups of three and six. However, using mitosis in NPI seems to be superior to Ki67, as Log Rank values were higher in the three and six groups that used the NPI mitosis variable, **Figure-26** and **Figure-27**. Conversely, in NPI+, applying Ki67 to the patients stratified according to time to death appears superior to mitosis. In the luminal A class (n=234), Ki67 split the class into three clinically significant groups (*p*-value <0.001) with higher Log Rank = 67.0, **Figure-28**. In the luminal B class (n=141), Ki67 split the class into two clinically significant groups (*p*-value <0.001) with higher Log Rank = 28.4, **Figure-29**. Interestingly, Ki67 was superior for patients' stratification when used as a continuous variable, while it was less significant when used as a categorical variable.

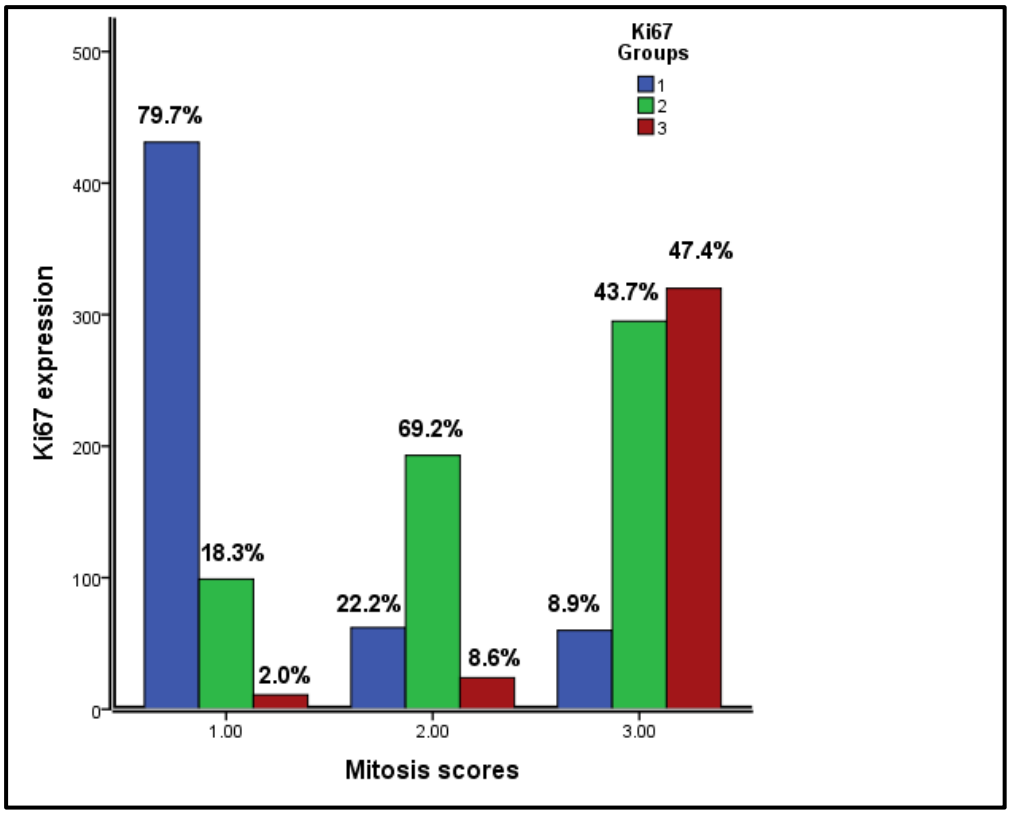


Figure-25: Correlation between Ki67 and mitosis
The X access represents the basic three groups of mitosis in Nottingham Grading System (NGS). The Y access represent the different Ki67 expression levels (groups) where blue color represents low Ki67 expression, green color represents intermediate group and red color represents high Ki67 expression group

3.7. Discussion

Immunostaining for Ki67 detection has been used to evaluate cancer proliferation for many years. However, the clinical application of this biomarker as a routine marker in BC remains controversial. The main issues include lack of a standardised procedure for Ki67 assessment. In this chapter different aspects relating to Ki67 assessment were evaluated, including its ability to stratify patients into different prognostic groups using both the traditional NPI and the NPI+.

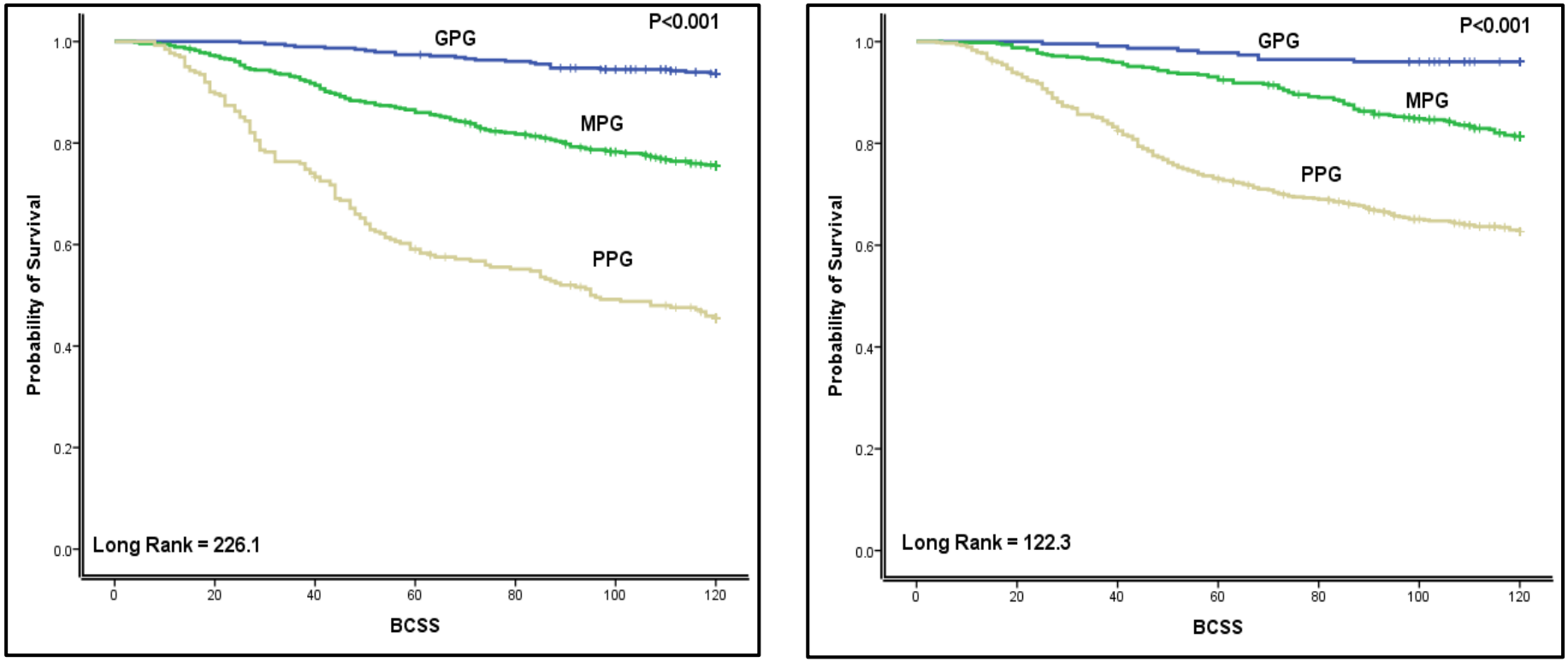


Figure-26: Survival curve of the three NPI groups
The original NPI with mitosis (Left) and after integration of Ki67 (Right)

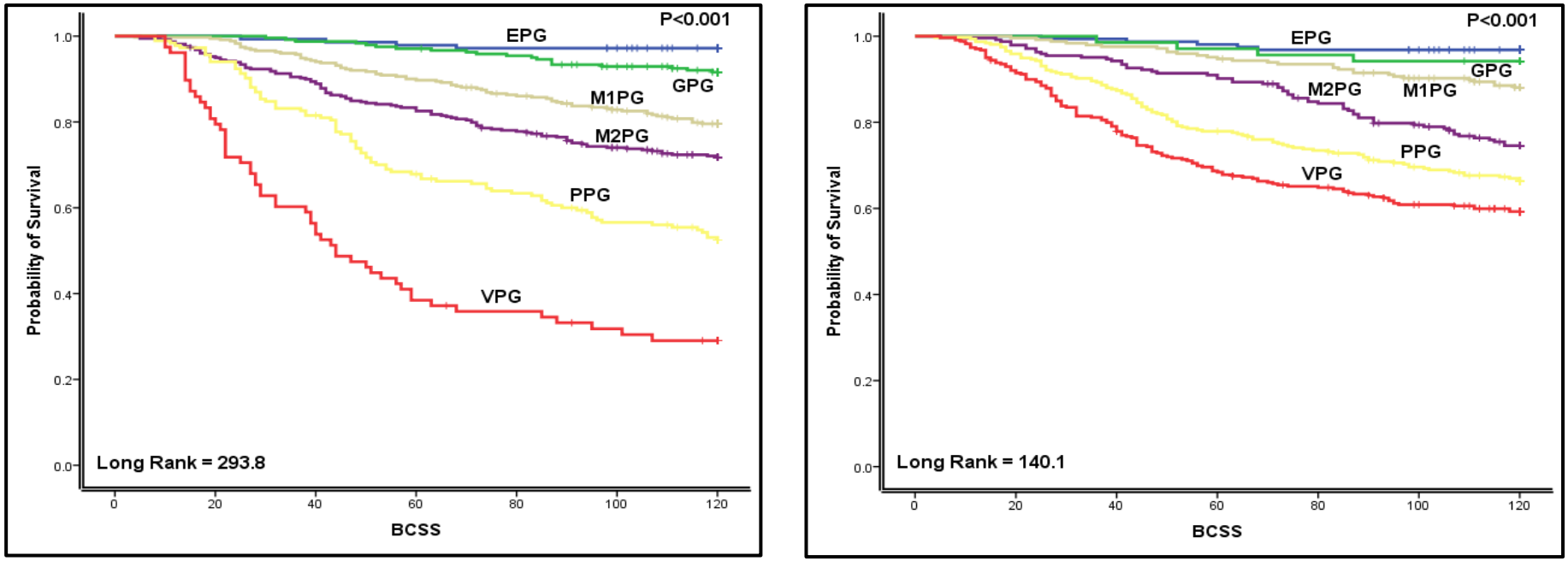


Figure-27: Survival curve of the six NPI groups
The original NPI with mitosis (Left) and after integration of Ki67 (Right)

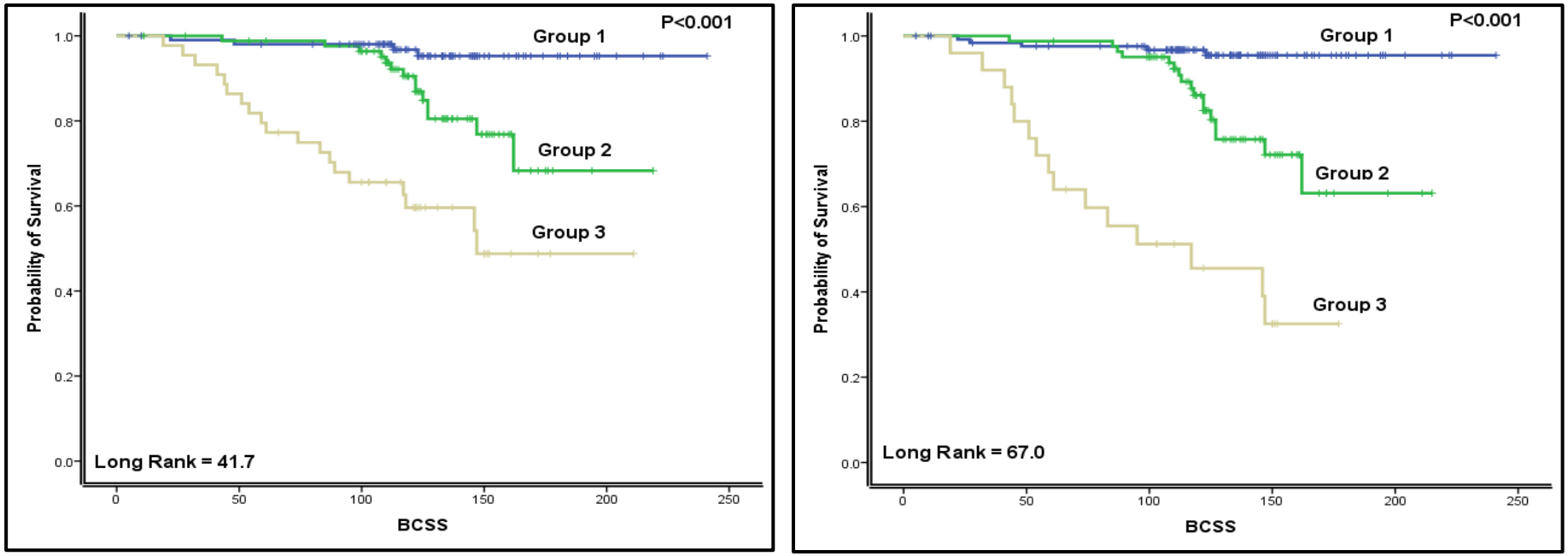


Figure-28: Survival curve of the three luminal A prognostic groups of original NPI+ (Left) and after integration of Ki67 (Right)

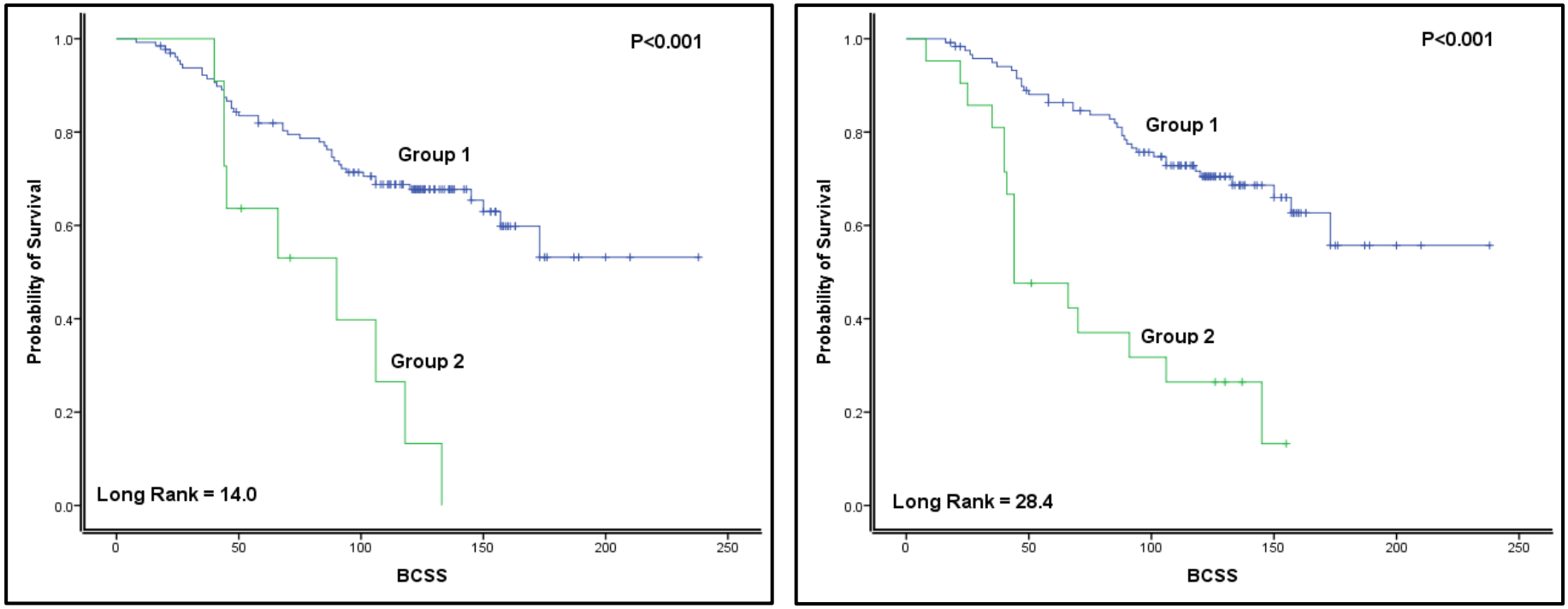


Figure-29: Survival curve of the two luminal B prognostic groups of original NPI+ (Left) and after integration of Ki67 (Right)

One of the issues regarding the assessment of Ki67 in research is the reliability of using TMA versus the whole sections; this has yet to achieve sufficient consensus. This study assessed different aspects regarding a comparison between the WTS and TMA, demonstrating that, in relation to reproducibility, although there was a significant concordance between the Ki67 expression in the WTS and TMA when a continuous data is used, ICC reflected substantial correlation and much lower when dichotomised. One important explanation is the effect of the diameter of the TMA core (Kobierzycki et al., 2012). Apart from intratumoural heterogeneity, this could be owing to using one core, which is a relatively small area compared to the WTS, and the greater tendency of TMA to give lower Ki67 estimates than the whole sections. Although there is evidence to demonstrate the usefulness of TMA for assessment of Ki67 with a high concordance rate between TMA and WTS, even when using only one core of TMA per tumour (Ruiz et al., 2006), others have proposed the use of multiple TMA cores per tumour for Ki67 analysis (Sapino et al., 2006). Recently, Laurinavicius et al. presented a novel tissue sampling simulation model and assessed this model using Ki67 in BC tissue. They showed that optimal tissue sampling for the IHC biomarker is at least five to six cores for homogeneous cases while eight TMA cores were required in a mixed tumour population (Laurinavicius et al., 2014).

The inter-observer reproducibility of Ki67 assessment is another important issue. In both WTS and TMA scoring, ICC showed a highly significant correlation. This result is consistent with the results published by an international Ki67 reproducibility study in which intra-laboratory reproducibility was high (Polley et al., 2013b). However, the same study demonstrated only moderate reproducibility between various laboratories, which requires a standardised scoring methodology. The almost perfect concordance for TMAs could be due to the small area and then less heterogeneity than WTS.

The comparison between the Ki67 assessed in both WTS and TMA with regard to the clinicopathological parameters produced comparably akin associations. However, WTS analysis performs better than analysis of Ki67 expression

utilising TMAs. As expected, high Ki67 was significantly associated with higher tumour grade, larger tumour size and additional nuclear pleomorphism and mitotic scores. However, statistical significance was higher using WTS. Further, Ki67 in WTS was significantly associated with patients' LN status while in the TMA it was not.

Currently, consensus is lacking concerning an ideal cut-off for Ki67 expression both in the clinical and research settings. This influences the comparison of Ki67 status in different clinical trials (Polley et al., 2013b). In this study, the cut-off point (10%) was chosen for WTS as previously published. Additionally, a large retrospective Australian study has emphasised that the highest sensitivity and specificity Ki67 cut-off is 10% (Pathmanathan et al., 2014). The same cut-off was used in several reports especially in the recent studies (Domagala et al., 1996) (Bago-Horvath et al., 2011) (Alba et al., 2012) (Kobayashi et al., 2013) (Honma et al., 2015), and reveals a significant association with BCSS, which is in line with our results: the most significant cut-off in WTS with patients' outcome is 10%. However, a wide-range of cut-offs were evaluated and remained significant with BCSS as the end-point. This is typical of Urruticoechea et al, who demonstrated, after assessing 18 studies, the wide range of Ki67 cut-offs from 1% up to 29%. Accordingly, they concluded that this different Ki67 cut-off might be the reason for its limited clinical use (Urruticoechea et al., 2005). Therefore, Ki67 has a significant prognostic value over a broad range of cut-offs and the optimisation of one cut-off is controversial. This has prompted Denkert et al. to suggest that Ki67 should be used as a continuous marker (Denkert et al., 2013) The 4th St. Gallen International Consensus Meeting agreed that there seems to be no optimal cut-off point; despite that, the high and low values are reproducible and clinically useful (Coates et al., 2015). Although our results suggest 20% as an optimal cut-off for TMA and 10% for WTS, they may support the view that Ki67 has to be assessed as a continuous variable.

One possible explanation for the wide spectrum of cut-offs could be the lack of standardisation in the pre-analytical tissue handling, in terms of time to fixation, duration of ischaemia, dilution and pH of formalin used in tissue fixation and

procedures of antigen retrieval which is largely based on the pre-analytical phase. The pathologist's scoring of the immunohistochemical staining might be also having a minor role (Colozza et al., 2010). Therefore, standardised methods in the pre-analytical tumour handling, especially sufficient fixation, is essential for reliable proliferative fraction assay. Interestingly, in this study, Ki67 is superior for patient stratification when used as a continuous variable (integration of Ki67 in NPI+ classes), while it is less significant when used as a categorical variable (integration of Ki67 in traditional NPI).

One of the noticeable issues in the assessment of this cohort was the cytoplasmic/ membranous staining signal. Although Ki67 is a nuclear expressed protein, cytoplasmic and membranous expression has been noticed in different tissues. For instance, pleomorphic adenoma of the salivary gland (Tashiro et al., 2002), sclerosing haemangioma of the lung (Hattori, 2002), hyalinising trabecular adenoma of the thyroid gland (Hirokawa and Carney, 2000) and invasive BC (Faratian et al., 2009). In addition, Faratian et al. reported that there are unusual cell membrane and cytoplasmic expression of Ki67 in a group of invasive BC which could be a distinctive group with importance in BC prognosis (Faratian et al., 2009). This may be one of the explanations of the unusual cytoplasmic and membranous staining. Another explanation could be related to the antibody used (Rossi et al., 2005). Interestingly, this is consistent with our results when a different clone (30-9) or SP6 was used; there was only nuclear staining with no cytoplasmic or membranous staining. This can be supported by, more recently, the increasingly rising use of rabbit monoclonal anti–Ki67 ready-to-use 30-9 antibody in different tissues, including BC using the Ventana system (Aigelsreiter et al., 2013) (Shaikhibrahim et al., 2014, Shaikhibrahim et al., 2015, Gil-Sanchis et al., 2015) (Lemound et al., 2015) (Mas et al., 2015) (Christgen et al., 2015).

3.8. Conclusion

Ki67 expression can be assessed utilising WTS and TMA, however, the latter need to be interpreted with caution. Therefore, using of multiple or larger TMA cores should be explored as an alternative to WTS. However, for practical purposes, Ki67 assessment in whole sections should be preferred when available. Evaluation of Ki67 as a continuous variable may be more applicable than the dichotomised values currently used in clinical practice. A cut-point of 20% in BC when assessing Ki67 on TMAs appears to be optimum both at concordance with WTS as well as with patients' outcome. Using MIB-1 clone with different optimisation conditions is associated with cytoplasmic/membranous reactivity. In this context, it is suggested alternative anti-Ki67 clones as clone (30-9) could be used for clearer staining. It seems that Ki67 could be used as a replacement of mitotic count in NPI+ with superior stratification of patients into prognostic groups.

4. Clinical utility of RPPA for molecular classification of BC

4.1. Introduction

Currently, high-throughput molecular techniques provide an opportunity to challenge the complexity of BC. GEP approaches as a large-scale technique applied to the oncology field can identify biologically relevant molecular subtypes related to different clinical presentations and outcomes (Sorlie et al., 2001) (van de Vijver et al., 2002). The available molecular signatures will probably be more reliable and sensitive than using a single biomarker to reflect the real heterogeneity of BC. However, proteins-rather than nucleic acids- are the final effector molecule of cellular processes and relevant therapeutic targets; addition the genomics approach has its limitations. The levels of mRNA do not necessarily associate with corresponding protein abundance (Gygi et al., 1999) (Ginestier et al., 2002). Moreover, transcriptomes do not capture the post-translational modifications that affect protein function, interaction and stability, such as phosphorylation, glycosylation and acetylation (Tyers and Mann, 2003) (Amerik and Hochstrasser, 2004) (Deribe et al., 2010). Furthermore, gene expression microarrays are relatively expensive and not easily available.

On the other hand, for a long time, quantitative protein measurements have been considered challenging. In a biological sample, protein abundance varies from a limited number to several million. Accordingly, an accurate quantitative assessment needs both high sensitivity and accuracy, especially with the restricted availability of material in clinical samples. These requirements have been partially achieved, during the last few decades, several techniques have been developed with promising results in the clinical setting, and which have improved diagnosis, prognosis and prediction of treatment responses. IHC is one of these techniques and has long been used to reliably examine the expression status of various prognostic and predictive biomarkers in BC. Further, in practice, IHC is widely used for molecular prognostic classification of BC. However, one of the limitations of using IHC is that it is a semi-quantitative technique and observer-dependent. Additionally, it does not facilitate quantification of the subtle changes in protein expression levels. NPI+, as

described earlier, is a novel immunohistochemistry-based prognostic test that has been developed by our group. It incorporates the measurement of biological and clinicopathological variables to stratify patients into distinct prognostic groups (Green et al., 2013) (Rakha et al., 2014) with a panel of ten biomarkers relevant to BC: ER, PR, HER2, HER1, HER3, HER4, Cytokeratin 5/6, Cytokeratin 7/8, p53 and Mucin 1. However, the initial NPI+ formulae did not include Ki67 biomarker, the importance of Ki67 in BC classification has been highlighted in the previous chapter. Moreover, it would be more efficient for a panel of biomarkers to be assessed via a high-throughput technique using one experiment rather than the current requirement of assessing all ten biomarkers on individual tissue sections using immunohistochemistry.

RPPA is an alternative and promising approach; it represents a high-throughput technique that enables simultaneous measurement of protein expression levels in a large number of biological samples in a quantitative manner. RPPA allows analysis of multiple samples, so the expression of proteins of interest can be studied under the equal experimental conditions. Moreover, it has the potential to replace standard IHC, particularly as it less subjective proteomic technique and it farther can help in the detection of subtle protein changes in terms of the relative abundance or down-regulation of key cellular signalling proteins in the studied sample, therefore giving portraits of the protein cascades of individual patients. Depending on whether the analyte is captured from the solution phase or bound to the solid phase, protein microarrays include two major classes of array methods capable of generating quantitative protein assessment: forward phase protein array (FPPA) and RPPA. FPPA utilises an array of immobilised multiple antibodies spotted on a nitrocellulose-coated microarray slide, which is incubated with a test sample. Alternatively, FPPA can use a sandwich array, where there are two antibodies: one is required to bind the protein of interest to the base of the array; and the second antibody binds a different epitope onto the same molecule (Templin et al., 2002) (Espina et al., 2003). In RPPA the lysate, rather than the antibody, is printed robotically onto a slide as individual spots, which are then incubated with a single specific antibody to detect expression of the target (Paweletz et al., 2001, Brase et al., 2010). Although one of the

limitation of RPPA technique is the availability of validated and highly specific antibodies (Boellner and Becker, 2015), RPPA has unique advantages over other array formats for protein profiling. FPPA is complicated by the issue of detection, and direct protein labelling may affect the binding of an analyte. Because all proteins are labelled, high-abundance proteins might bind non-specifically to other particles, producing high background. Moreover, the FPPA approach can be viewed as an array of miniaturised immunoprecipitation (IP) reactions that occur proximally to the array surface; only a low percentage of commercially available antibodies can target intracellular proteins for IP experiments. Conversely, a high percentage of these antibodies would bind with high affinity to their antigen when the antigen is denatured and immobilized on a membrane - as in RPPA, in which the samples are immobilized on the array surface (Chan and Utz, 2005). Further, one of the major advantages of RPPA is that the sample proteins do not need to be labelled.

The feasibility of RRPA in determining cancer protein profiles was first shown by Paweletz et al through analysis of lysates prepared from a laser capture microdissection (LCM) of 70% ethanol fixed and paraffin-embedded prostate cancer tissues (Paweletz et al., 2001). They expanded the number of analytes and performed proteomic profiling of the 60 human different carcinoma cell lines set (Nishizuka et al., 2003). Byers et al. described the proteomic profiling using RPPA of 193 proteins and phospho-proteins in a panel of 108 small-cell lung cancer and non-small-cell lung cancer cell lines (Byers et al., 2012). RPPA as previously used successfully to quantitate proteomic profiles in BC cell lines, which demonstrated a higher level of expression of KU70/KU80 in BRCA1-deficient HeLa cell lines (Alshareeda et al., 2013). Moreover, the Cancer Genome Atlas (TCGA) defined four BC subgroups: luminal A and B, HER2-enriched, and basal-like. This study used the RPPA platform along with other genomic DNA techniques and each group exhibited significant molecular heterogeneity (Network, 2012).

However, one of the challenges at high-throughput proteomics is using FFPE tissues. Although these can be stored for decades while still preserving cellular

morphology and proteins, the use of formalin as a fixative induces protein cross-links and modifications, creating subsequent problems for separating and characterising individual proteins (Palmer-Toy et al., 2005). Despite this, proteins isolated from FFPE tissue have been used for proteomic studies (Martinet et al., 2004) (Becker et al., 2007). Few studies comparing the efficacy of different protein extraction protocols have been published in the literature, and these studies have found that the quantity and quality of the recovered proteins are determined by the method and extraction buffer used (Luebker et al., 2015) (Broeckx et al., 2016). The primary conclusion drawn from these studies is that there is no current agreement on the optimal protein extraction protocol from FFPE tissue; neither are there any established standards for the quantitative assessment of extracts (Shi et al., 2013).

The heterogeneity of tumours, particularly in BC, is another challenge for proteomics analyses. In any pathological processes, only specific cell types can be affected, while the other cells compromise the results by diluting or increasing the differences in the amount of protein expression. Regarding molecular studies of cancer, it is preferable to separate tumour cells from its surrounding stroma, where malignant and non-malignant cells are mixed. Practically, sampling is of highest importance for obtaining reliable and reproducible results that can be translated into clinical practice. Various studies have shown the possibility of mapping proteomic signalling using RPPA in limited fine-needle aspiration samples (Rapkiewicz et al., 2007) (Ibarra-Drendall et al., 2012) and in LCM using frozen tissue specimens (Wulfkuhle et al., 2008). However, the isolation of a specific cell population is challenging, labour-intensive, and time-consuming. Applying non-microdissected human tumour material to RPPA possibly makes the technology more widely practical in the clinical setting. Hennessy et al recognised kinase and steroid signalling proteins in macrodissected frozen human BC tumours using RPPA, after intervals of up to 24 hours on the benchtop at room temperature, following surgical resection (Hennessy et al., 2010). Moreover, Becker et al. successfully quantified protein analysis from different FFPE tissue, including 17 BC samples, using macrodissection (needle microdissection) (Becker et al., 2007).

4.2. Hypothesis

Molecular features of BC determined by a quantitative high-throughput method. It was hypothesised that RRPA allow more accurate determination of protein levels using the NPI+ 10 biomarkers. Moreover, the possibility to realign the NPI+ classification based on the expression of Ki67.

4.3. Aim of this chapter

The aim of this study is to therefore evaluate the use of RPPA in improving molecular classification of BC in order to stratify patients according to the NPI+ decision making tool. Further, to examine the effect of Ki67 on this classification, **Figure-30**.

In particular the specific aims are:

- To determine the most suitable protocol for protein extraction from FFPE tissues.
- To determine the most suitable method for protein extraction from FFPE tissues by testing three different platforms.

- To evaluate and compare the expression of proteins used in the NPI+: using RPPA and immunohistochemistry.
- To evaluate the correlation between expression of Ki67 in both IHC and RPPA.

- To evaluate the effect of re-classification of RPPA data using Ki67 on NPI+ classification.

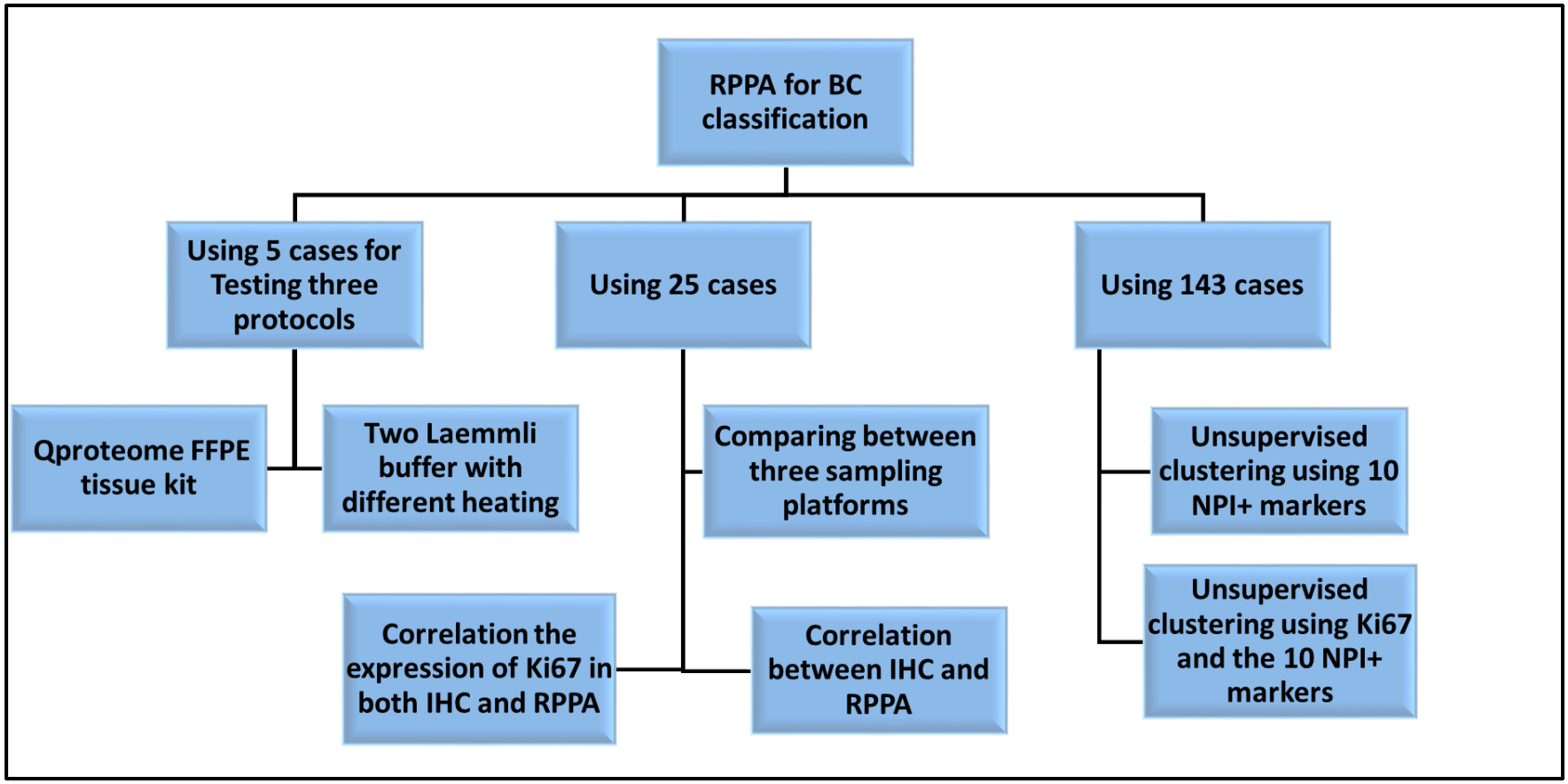


Figure-30: A flowchart representing the steps of study in this Chapter

4.4. Methodology

4.4.1. Patient cohorts

In this study, the fifth cohort (143 cases) was used (Section-2.1.5). An initial optimisation step to compare the three protein extraction protocols was piloted using five randomly chosen invasive BC cases from this series. Next, a comparison between three different tumour tissue sampling techniques (full-face, macrodissection and laser capture microdissection) from FFPE tissue was then conducted using a total of 25 cases of the same series. These cases were selected based on the availability of at least 50 % invasive tumour tissue burden of the whole section area in paraffin blocks. Finally, the most suitable protocol and sampling method were used using the whole cohort (143 cases) to evaluate the utility of RPPA in classification of BC and to examine the role of Ki67 in this classification.

4.4.2. Optimisation of protein extraction from FFPE tissues

Three serial sections, 10 µm thickness each of five randomly chosen invasive BC were used.

4.4.2.1. Protocol 1:

Protein was extracted using the Qproteome FFPE tissue kit as per manufacturer's instructions (Becker et al., 2007). Deparaffinisation and protein extraction in summarised Section-2.5.

4.4.2.2. Protocol 2:

Tissue sections deparaffinisation was performed as described in Section-2.5.1 After deparaffinisation, protein extraction was performed using Laemmli buffer

(Laemmli, 1970), where 150 µl of Laemmli buffer (100 mM Tris–HCl, pH 6.8; 20% v/v glycerol; 2% w/v SDS; and 4% v/v β-mercaptoethanol) was added to the deparaffinised tissue where incubated on heating block for 20 minutes at 105˚C. Samples were cooled in ice for 5 minutes. Finally, after centrifugation 14,000 ×g for 15 the supernatant stored at -80˚C until use.

4.4.2.3. Protocol 3:

An alternative method used in Protocol 2 was used where 100 µl laemmli buffer and 6 µl β-mercaptoethanol was added and incubated for 20 minutes at 100˚C followed by 2 h at 80°C in thermomixer. Finally, centrifugation was done for 15 minutes at 4˚C and the supernatant was kept at - 80˚C until use.

4.4.2.4. Validation of the protein extracts

Protein concentrations from the FFPE extracted lysates were measured using Fast Green stain (Loebke et al., 2007). Two fold dilutions of 1.0 mg/mL BSA (Sigma) standard were spotted onto the array as a total protein standard. Printed arrays were stained with Fast Green stain (0.005 % FCF (Sigma) in 30 % ethanol, 10 % acetic acid and 60 % water) at room temperature for 45 minutes on a rocking platform at 40 rpm. Destaining was applied twice for 15 minutes each with 30 % ethanol, 10 % acetic acid and 60 % water. The slide was washed then, air dried and scanned at 700 nm utilising an Odyssey scanner (Licor). Additionally, further validation of the quantity of the protein was measured using β-actin levels. Each sample was printed 20 times onto a nitrocellulose slide using the MicroGridII arrayer. Protein lysates from samples were spotted onto nitrocellulose-coated glass slides (Grace Bio-labs) in quadruplicate with a microarraying robot (MicroGrid 610, Digilab). The quality and immunoreactivity of the extracted proteins within lysates were evaluated by western blotting (Section-2.4). Primary anti-β-actin antibody (Sigma) was applied diluted at 1:1000 in 5 % BSA in TBS 0.05 % Tween-20, pH 7.5 and incubated overnight at 4 °C. Blots were incubated with swine anti-rabbit immunoglobulin- HRP conjugate (Dako) diluted 1:2000 in blocking buffer for 1

hour at room temperature (This step conducted by Dr Paddy Tighe and Dr Ola Negm from the Post-genomics Facility, University of Nottingham).

4.4.3. Comparison between three sampling methods

Three different sampling techniques (full-face section, macrodissection and LCM) were compared using 25 invasive BC cases of FFPE tissues to evaluate which method is most suitable for RPPA. A full-face section refers to sectioning the whole tissue of the paraffin block including the tumour area using a microtome. However, macrodissection refers to the gross manual dissection of invasive tumour area from FFPE sections placed on normal glass slides using disposable surgical blades. Macrodissection was guided by H&E slides from the same FFPE block. Slides of 10 µm thickness were prepared and stained with H&E; then, the area of invasive tumour for each case was marked using special pen (more than 50%). For LCM, tissue sections of 10 µm thickness were placed on membranous coated slides and tumour areas dissected using non-contact laser instruments, details of the LCM described in Section-2.6. The clinicopathological criteria of the 25 BC cases used in this step of the study demonstrated in **Table-16**.

Table-16: Clinicopathological criteria of the 25 BC cases used in this step of the study (Negm et al., 2016)

Clinical parameters		Number of cases (%)
Tumour grade	2	5 (20)
	3	20 (80)
Tumour size	≤2cm	6 (24)
	>2cm	19 (76)
Axillary lymph node stage	1	16 (64)
	2	6 (24)
	3	3 (12)
Nottingham prognostic index (NPI)	GPG	1 (4)
	MPG	4 (16)
	PPG	20 (80)
NPI+ Cluster membership	Luminal	7 (29.1)
	Basal	14 (58.3)
	HER2-positive	3 (12.6)
	Unclassified	1

GPG: Good Prognostic Group, MPG: Moderate Prognostic Group, PPG: Poor Prognostic Group.

4.4.3.1. Deparaffinisation, protein extraction

For the detailed methodology, please referred to Section-2.5.1.

4.4.3.2. LCM from FFPE samples

Before using of LCM in the comparative step, both the optimal area needed for LCM and the effect of staining were evaluated: Firstly, Different sized areas of tumour cells were microdissected from sections: 1, 5, 10 and 15 million µm² to determine the optimal area needed to yield sufficient protein for RPPA. The concentration of the extracted protein from the microdissected cases was compared with two other sections: one was prepared by using the

macrodissection method, the other by the full-face sectioning method. The comparison performed using Fast Green stain as previously described. Secondly, In LCM, haematoxylin staining was used as describe in Section-2.6. Compression between both stained and unstained sections were performed to investigate the effect of staining on the amount of protein yield. For this purpose, 10 cases were used following the same steps in Section-2.6, except that the slides did not stained with haematoxylin. The comparison performed using the level of β-actin was yield.

4.4.3.3. Immunohistochemistry

The cases were previously immunostained (group data) using the NPI+ biomarkers. The details of antibodies that used in this step were shown in, **Table-17**. The H-score method was used to score the antibody as it is representative for the intensity and the percentage as previously described in Section-2.3.7. The primary antibodies were validated using strip western blotting, Section 2.4.

4.4.3.4. Revere phase protein array

The details of RPPA are described in Section 2.7. The reproducibility of RPPA using FFPE extracts was evaluated. For intra-slide validations, FFPE extracted lysates (using Q-proteome FFPE Tissue kit) from five BC tissues were used and printed in ten replicates onto 16-pad nitrocellulose slides. The inter-slide reproducibility was assessed by printing the slides on two different days to produce identical RPPA slides needed for the profiling of BC samples. To confirm the reproducibility of achieved results, representative slides of each run were reserved with primary antibodies which directed the housekeeping β-actin.

Table-17: Antibodies used in RPPA and the original NPI+ using IHC
their clones, dilutions, incubation times and antigen retrieval methods used for IHC staining (Negm et al., 2016)

Antibody	Product Code	Clone	RPPA dilution	IHC dilution	Incubation Time	Antigen retrieval Method, buffer solution & duration
ER	M3634	SP1	1:200	1:150	30 minutes	Microwave (MW), Citrate, pH6.0 for 20 minutes
CK 7/8	345779	CAM 5.2	1:50	1:125	30 minutes	MW, Citrate, pH6.0 for 20 minutes.
CK5/6	M7237	D5/16B4	1:400	1:750	60 minutes	MW, EDTA, pH8.0 for 20 minutes.
HER1	Mouse	31G7	1:100	1:50	60 minutes	Proteinase K (8minutes at 37°C on cytomation hotplate)
HER 2	A0485	Rabbit polyclonal	1:50	1:1	30 minutes	NONE
HER 3	NCL-c-erbB-3	RTJ1	1:200	1:400	15 minutes	Hot water bath (95°C), Leica ER1 buffer, for 25minutes
HER 4	RB-9045-P	Rabbit polyclonal	1:50	1:30	30 minutes	MW, Citrate, pH6.0 for 20 minutes.
Mucin 1	NCL-MUC-1	MA695	1:50	1:50	30 minutes	MW, Citrate, pH6.0 for 20 minutes.
p53	NCL-p53-DO7	DO-7	1:25	1:30	30 minutes	MW, Citrate, pH6.0 for 20 minutes.
PR	M3439	PR 637	1:200	1:100	30 minutes	MW, Citrate, pH6.0 for 20 minutes.

4.4.4. Correlation between IHC and RPPA

Statistical analyses were performed using GraphPad prism version 6.5 and SPSS version 22 (PAWS statistics). Statistical correlation between IHC (using H-score) and RPPA was tested as continuous values using linear regression test. A *p*-value <0.05 was considered significant. Unsupervised one-way hierarchical cluster analysis of the expression of the 10 biomarkers used in NPI+ using RPPA based on Pearson distance combined with complete linkage rule was performed using MeV (Multiple Experiment Viewer) version 4_8.

4.4.5. Analysis of RPPA data using consensus clustering and classification techniques including the role of Ki67

4.4.5.1. Correlation between expression of Ki67 in both IHC and RPPA

Statistical analyses were performed using SPSS version 22 (PAWS statistics). Statistical correlation between IHC (using percentage score) and RPPA was tested as continuous values using linear regression test and Spearman's correlation, *p*-value <0.05 was considered significant.

4.4.5.2. Evaluation the effect of re-classification of RPPA data using Ki67 on NPI+ classification

Clustering allows the examination of data defining its intrinsic structure by organising data variables into similar clusters or groups. Expression of biomarkers was quantified, in triplicate, using RPPA from 143 BC tumour cell

populations, which isolated using macrodissection. The methodology of classification was performed in two steps: firstly by considering the 10 NPI+ protein biomarkers, secondly with the addition of; Ki67. The RPPA data for 143 patients was collected in triplicate. The average of all the three replicates was taken. Data was processed using min-max normalisation (Shalabi LA, 2006). The analysis was performed by Dr. Daniele Soria and Utkarsh Agrawal (Computer Science, University of Nottingham) and consisted:

1) Hierarchical methods: Agglomerative methods were used. The two most similar clusters were joined and performed repetitively until all subjects were in one cluster. Then, the optimum number of clusters was chosen from all of the cluster solutions.
2) Non-hierarchical methods: two clustering algorithms namely Partitioning Around Medoids (PAM) and K-means with centres from pruning of Hierarchical cluster tree were run on the dataset.
3) A consensus of both clustering algorithms was performed (clusterfusion).
4) Three clusters were created, chosen through cluster validity indices as the optimal number of indices, and similar steps were repeated for creating seven clusters.
5) The results are reported for the ten NPI+ biomarkers and subsequently including Ki67.

Boxplots were used to visualise the differential expression of the biomarkers in three and seven BC classes. The vertical bars represent the expression of the biomarker and the bold horizontal line in the box revealed the median. Further, the correlation between expression of Ki67 in both IHC and RPPA was evaluated.

4.5. Results

4.5.1. Protein extraction from FFPE tissues

The protein yield from FFPE breast tissue utilising the three various protocols was evaluated with regard to the quantity and quality. The results of protein extraction from FFPE samples showed the commercial Q-proteome FFPE Tissue Kit gave the best extraction method compared to the other two protocols used where the concentration of the extracted protein (2mg/ml) was double obtained using the other two protocols as shown in, **Figure-31 A and B**. Western blot analysis for β-actin showed preservation of protein extracted from the FFPE sections using all extraction protocols where a single band at the accurate molecular weight (42 kDa) was identified. However, the signal expression of β-actin was higher for extracts using Q-proteome FFPE Tissue Kit protocol, **Figure-31 C.** RPPA also demonstrated that the level of β-actin was higher in lysates extracted by Q-Proteome FFPE tissue kit compared with the other two protocols, **Figure-32**.

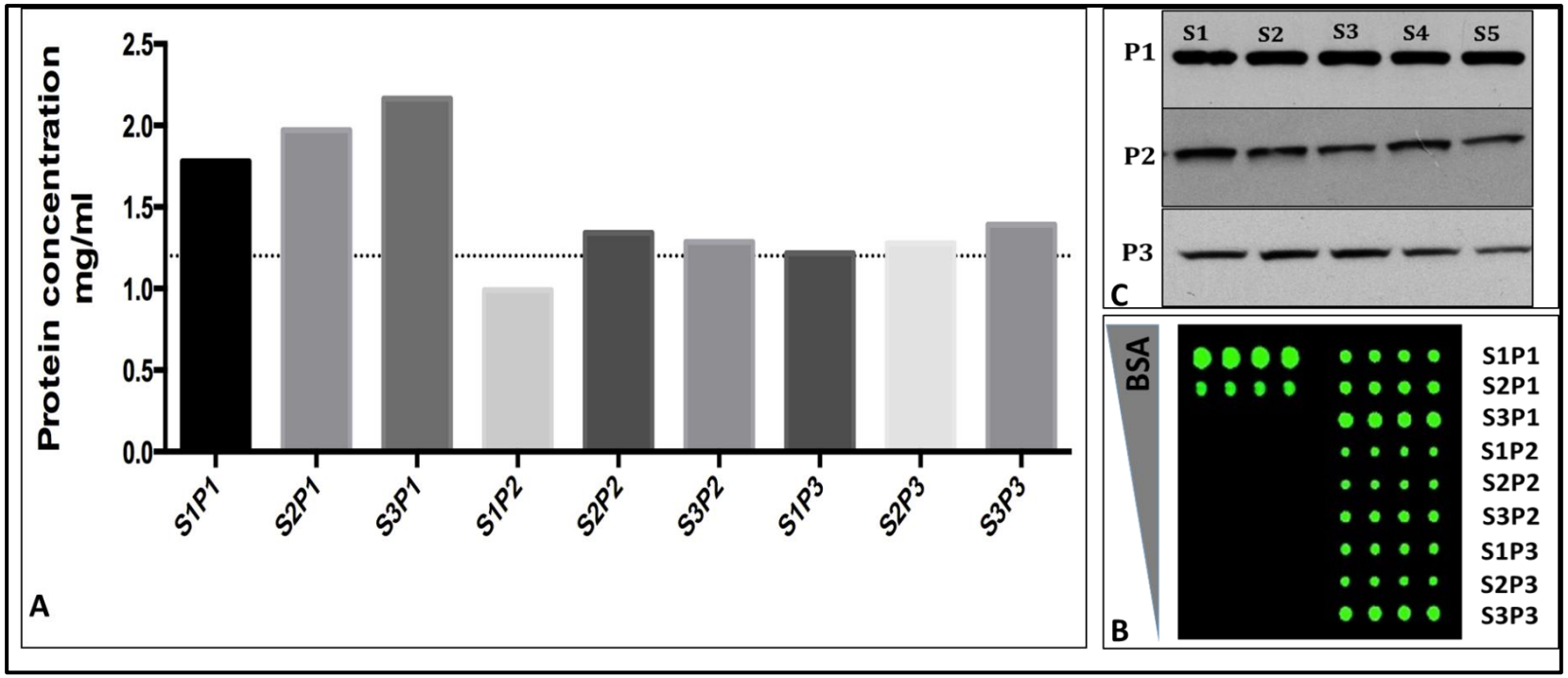


Figure-31: Quantification and reproducibility of protein extraction from FFPE samples
(A) Quantification of β-actin from FFPE prepared using three different protocols (P =Protocol, S=Sample), (B) Quantification of protein using Fast green stain, (C) Western blotting for β-actin detection (42 kDa) in lysates prepared from the tested protocols (Negm et al., 2016)

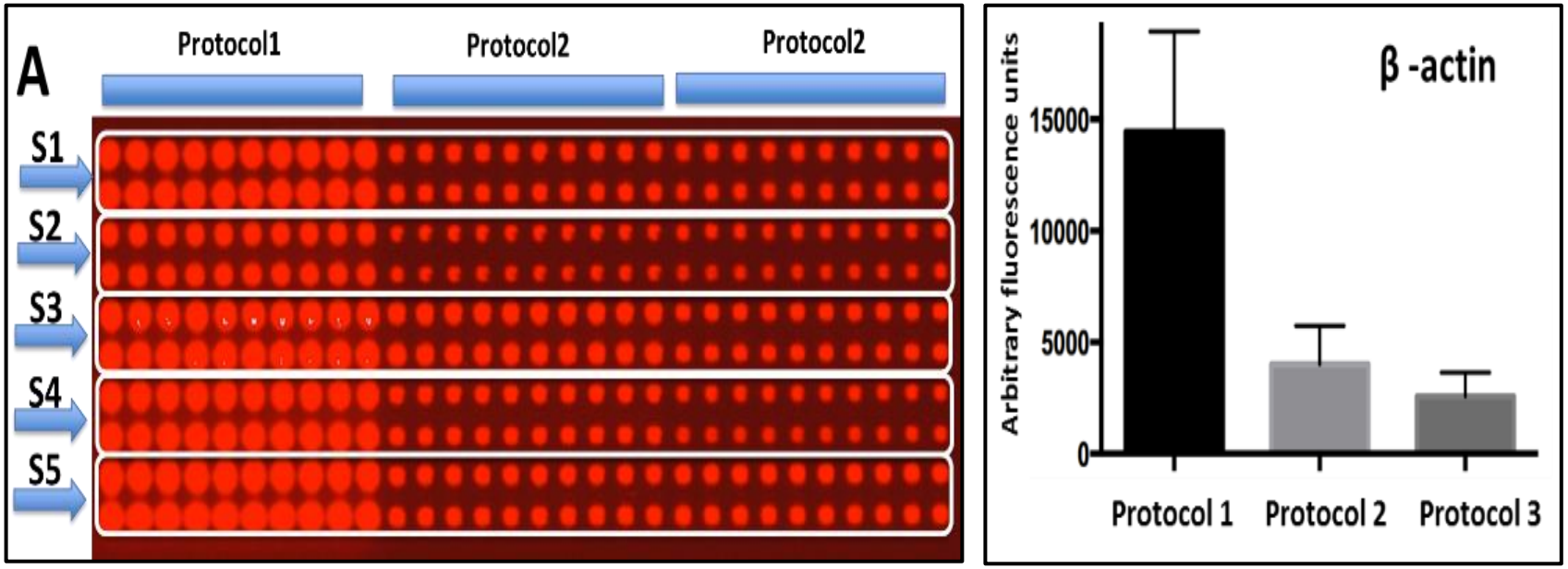


Figure-32: Representative microarray image
showing higher signals with the protocol 1 using Q-proteome kit (Left) and bar graph demonstrates the quantitative representation of the signal intensities of β-actin (Right)

4.5.2. Comparison between three sampling methods

4.5.2.1. Protein extraction and optimisation of LCM

As it is a challenge to extract protein from the microdissected FFPE sections, initially the minimum microdissected area that yield protein can be detected using RPPA was evaluated. 15 million µm² was the minimum microdissected area required for protein detection using RPPA, **Figure-33.** 1, 5 and 10 million µm² dissected areas revealed respectively with low level of protein yield. However, when the three sampling techniques were compared macrodissection and full-face gave higher levels of β-actin expression compared to the LCM samples. As shown in **Figure-34** the amount of protein yield using the macrodissection method was higher than both full-face and LCM methods. There was no significant difference between the amount of protein yield of the stained and unstained sections used in LCM, **Figure-35.**

4.5.2.2. Optimisation of antibodies

The result of the western blot of the antibodies used this study where a specific band of the predicted molecular weight for each protein were illustrated in, **Figure-36**. Regarding the reproducibility, a statistically significant (*p*-value <0.05) high correlation was also detected between signals of matching samples printed on the two different slides from the two experiments (day 1 and day 2) (*p*-value <0.001, r^2 =0.61), **Figure-37**. The coefficient of variation (CV %) within the 10 intra-slide replicates was as low as 7% for all the analysed samples.

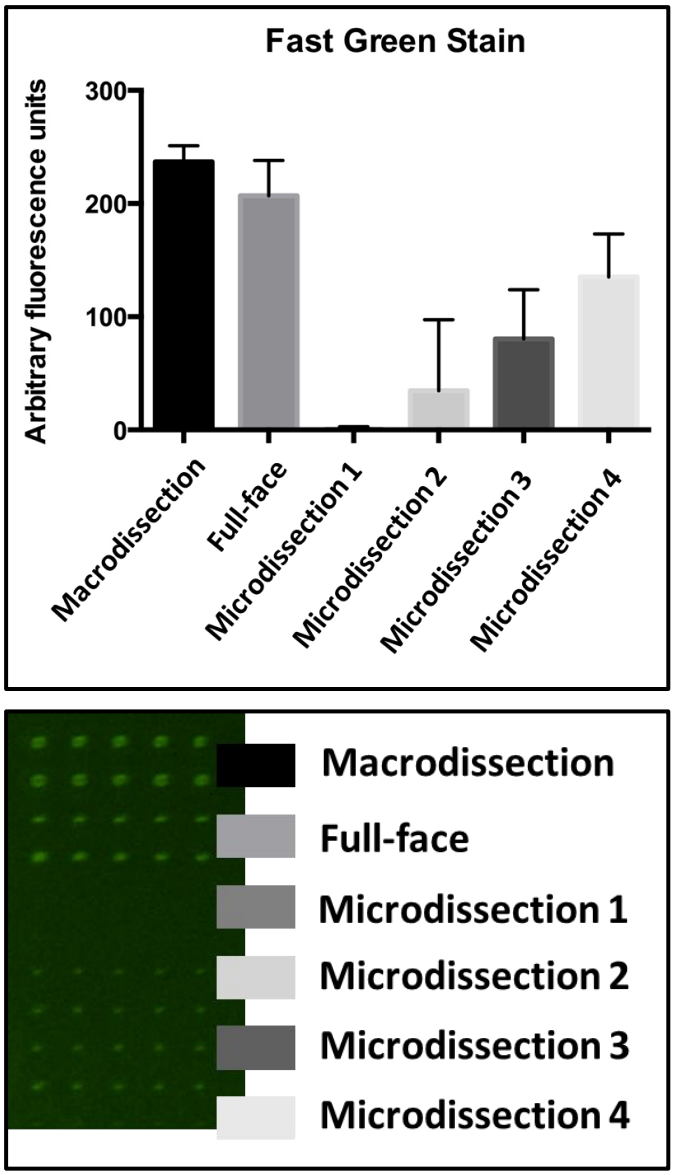


Figure-33: Comparison between the three sampling methods macrodissection and full-face methods produce a higher protein level than LCM. The microdissected area (15 million µm^2) represented by sample 4 have a relatively higher level of protein as tested by Fast Green staining. 1, 2, 3, 4 represent the areas 1, 5, 10, 15 million µm^2 respectively

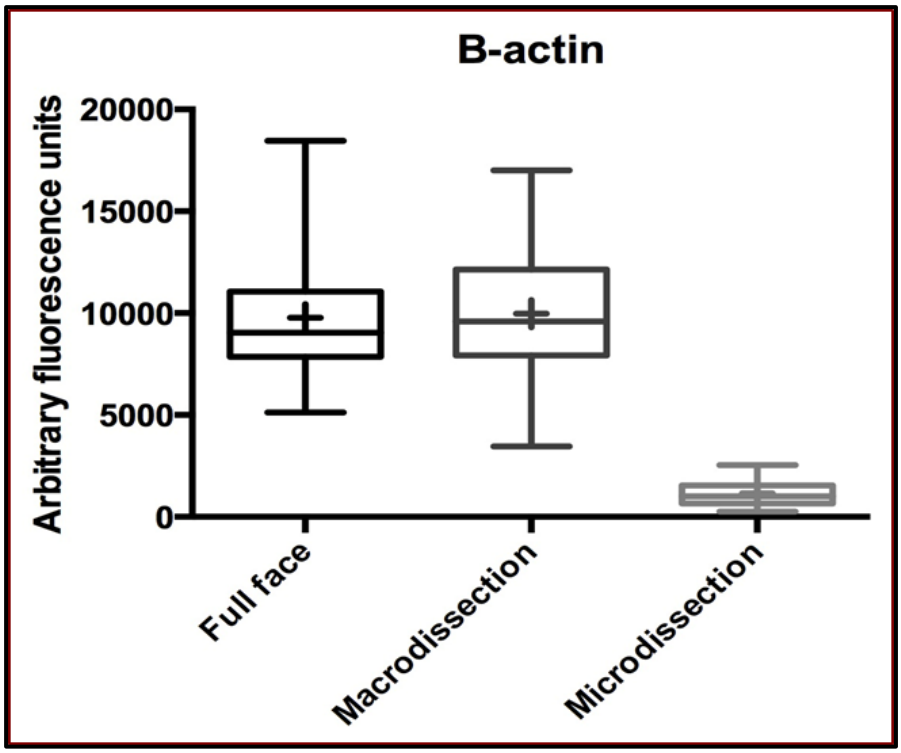


Figure-34: The boxplot shows that full-face and macrodissection gave higher level of β-actin expression compared to LCM samples

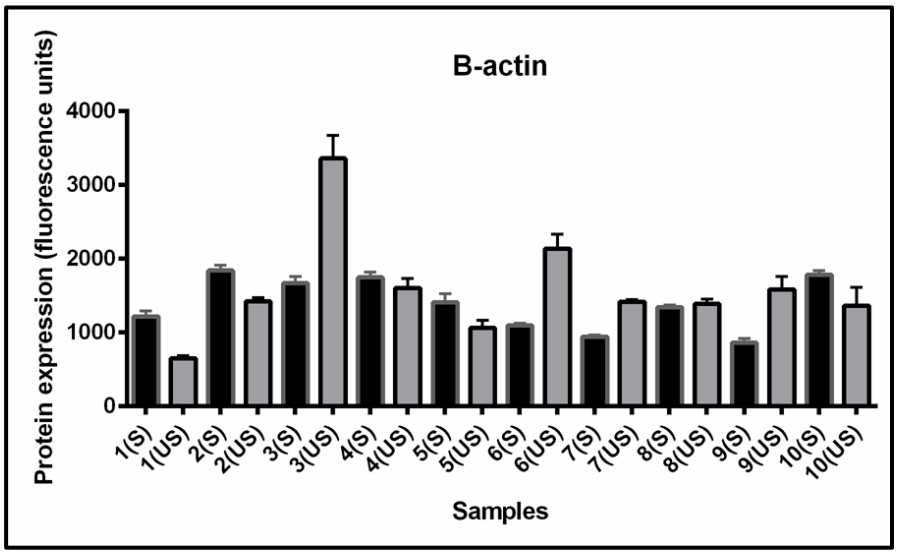


Figure-35: The bar graph shows the stained (S) and unstained (US) BC cases using LCM

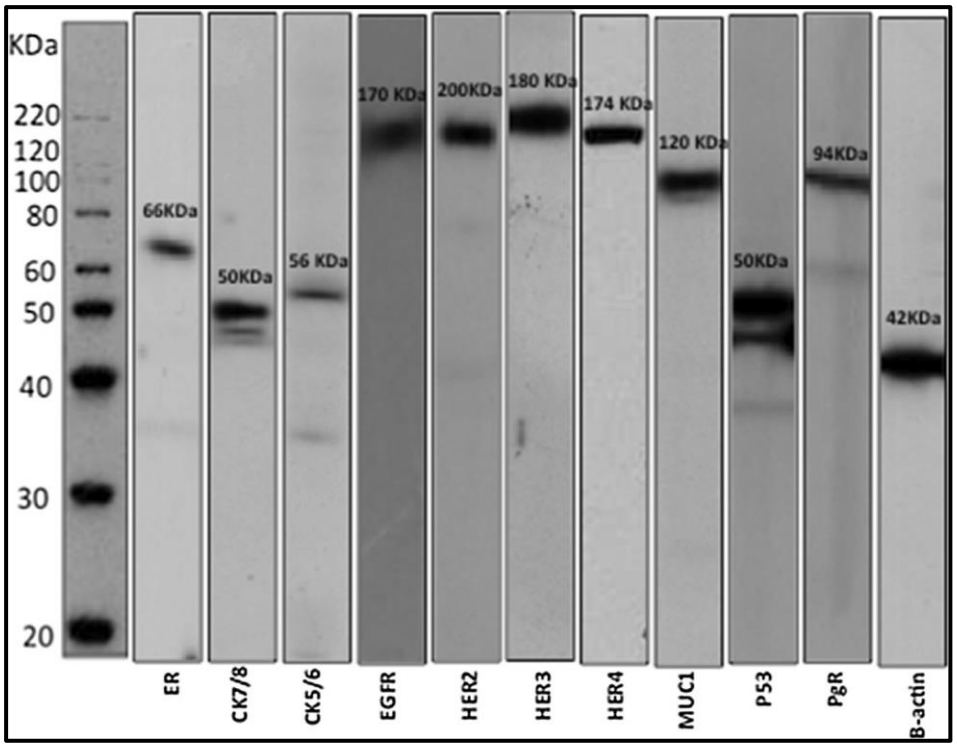


Figure-36: Western blotting validation of the primary antibodies , a single band was obtained of each protein at the predicted molecular weight (Negm et al., 2016)

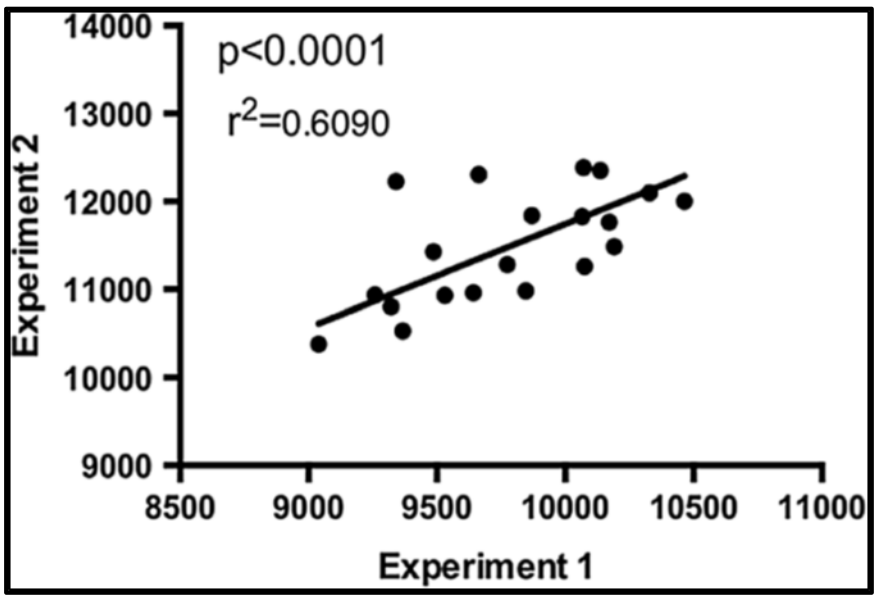


Figure-37: Inter-slide reproducibility
protein extracts from FFPE breast cancer tissues were prepared and were printed on two different days to produce identical RPPA slides. Slides of each run were reserved with primary antibodies which directed the housekeeping β-actin. High correlation between the signals from the two experiments (day 1 and day 2) is seen (r^2 = 0.6090) (Negm et al., 2016)

4.5.3. Correlation between IHC and RPPA

RPPA-resulting expression levels of the 10 NPI+ markers from macrodissection and full-face methods were assessed and compared as a continuous variable to the IHC continuous scoring values (H-score), IHC images for the expression of NPI+ markers represented in **Figure-38**. Regarding the LCM, due to the low level of protein yield, it was difficult to compare the results. The results as shown in **Table-18**, where using the macrodissection method 8 out 10 biomarkers (ER, PR, HER2, HER1, HER4, CK5/6, CK7/8 and Mucin1) were significantly correlated using RPPA and IHC while only 6 markers (ER, PR, HER2, HER4, CK7/8 and Mucin1) utilising the full-face sections were significantly correlated (*p*-value <0.005).

Although the number of the cases in this study is limited, a hierarchical cluster examination of the used biomarkers was applied for RPPA data acquired from macrodissected BC tissue samples to molecularly classify BC into specific classes. Interestingly, the 25 samples were clustered into six groups. To classify and characterise each group, these groups were compared with the NPI+ biological class membership determined using immunohistochemistry. **Figure-39** and

Table-19 show that RPPA Groups 1, 2, 3 and 5 represent basal tumours and including all the cases that belonging to basal class (NPI+). Interestingly, Group 5 was separated from the remaining groups via low/high P53 expression. This was remarkably consistent with the NPI+ basal class subgroup that were divided by p53 subgroup, **Figure-40**. Further, Group 6 primarily represents the luminal group where it contained 5/7 luminal tumours while HER2-positive class was represented in Group 4. Only 3 cases were misclassified to other classes.

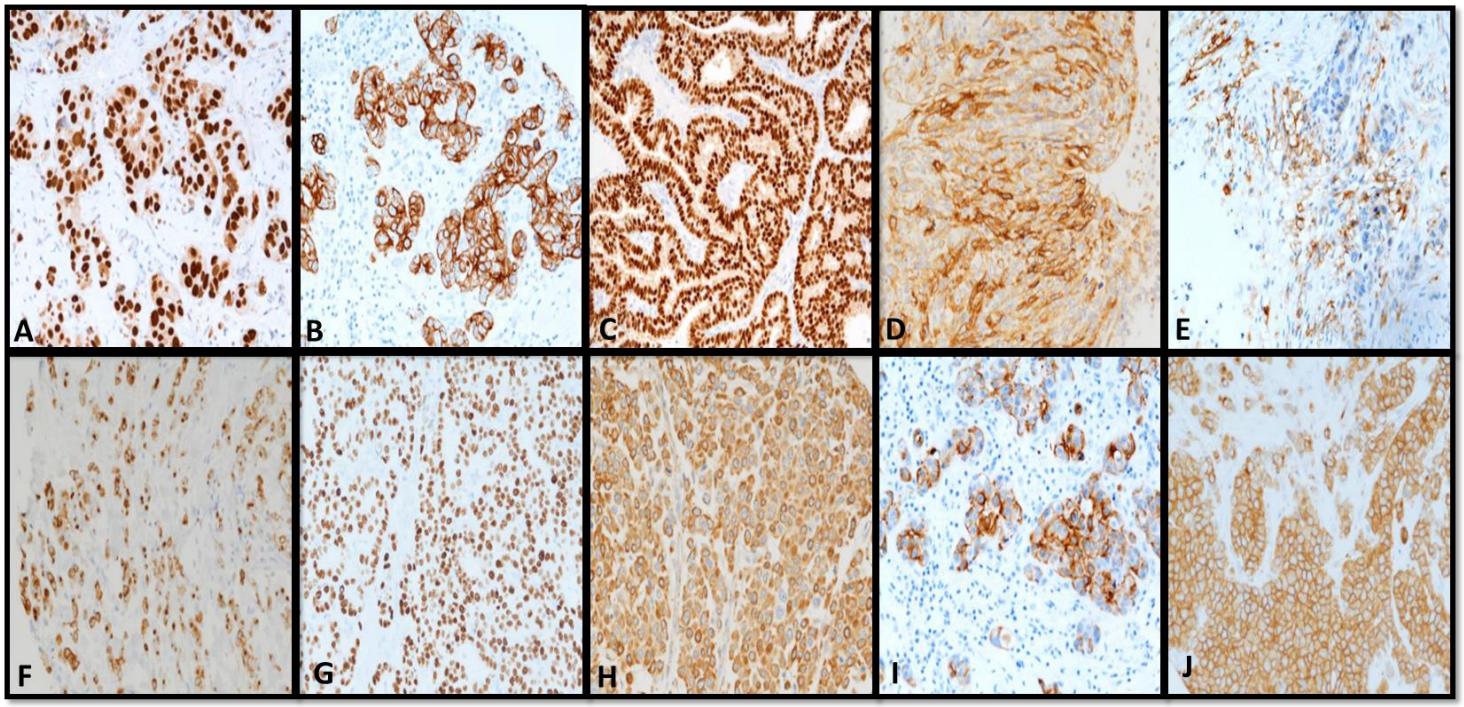


Figure-38: Immunostaining expression of NPI+ markers in breast cancer
(A) P53, (B) CK7/8, (C) ER, (D) HER1, (E) HER4, (F) Mucin1, (G) PR, (H) HER3, (I) CK5/6 and (J) HER2 respectively

Table-18: Correlation between RPPA and IHC results using NPI+ biomarkers (Negm et al., 2016)

NPI+ markers	Full-face (*p*-value)	Macrodissection (*p*-value)
PR	**< 0.0001**	**< 0.0001**
CK7/8	**0.0092**	**0.0460**
CK5/6	0.3738	**0.0107**
HER3	0.0813	0.8083
HER1	0.2761	**0.0328**
P53	0.8286	0.9607
Mucin1	**0.0126**	**0.0022**
ER	**< 0.0001**	**< 0.0001**
HER2	**< 0.0001**	**< 0.0001**
HER4	**0.0001**	**0.0456**

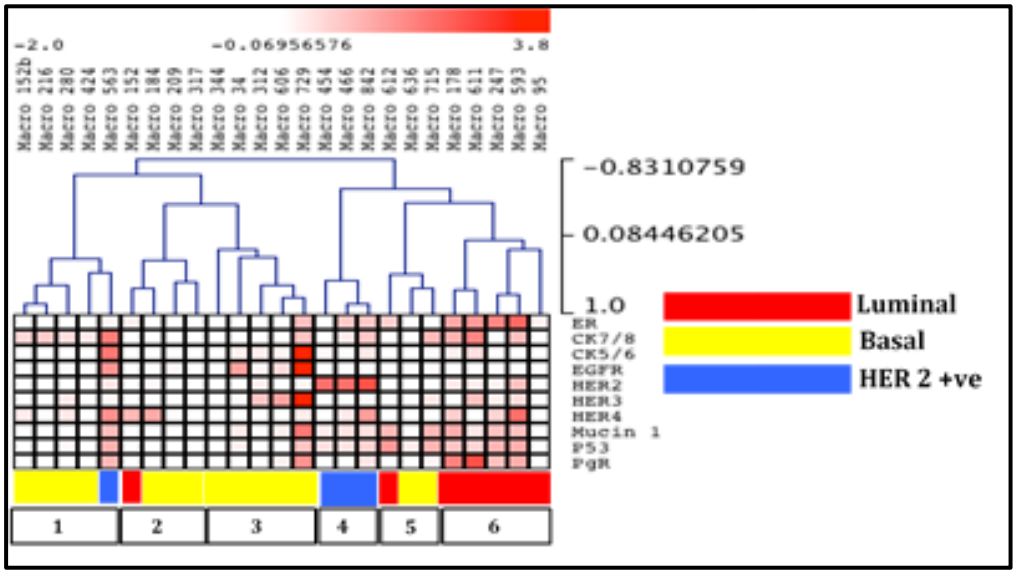


Figure-39: Hierarchical cluster analysis of NPI+ biomarkers
Heat map illustrating an unsupervised hierarchical cluster analysis of the expression of the 10 biomarkers used in NPI+ using reverse phase protein arrays. Patient samples are oriented on the horizontal axis and the different studied target molecules are oriented on the vertical axis. Clustering of the 25 tumour samples was achieved using RPPA into 6 different group-derived protein expression data of NPI+ biomarkers. Colour code of the heat map: white low expression, red high expression (Negm et al., 2016)

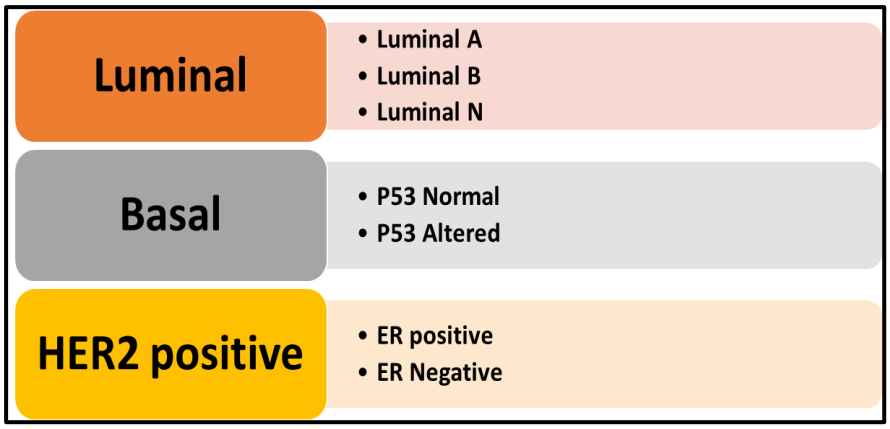


Figure-40: Nottingham prognostic index plus (NPI+) classes (Green et al., 2013)

Table-19: RPPA classification of BC according to NPI+ , where group 6 represents luminal class, group 4 represents HER2-positive class and groups 1, 2, 3 and 5 represent basal classes p53-normal and p53-altered respectively (Negm et al., 2016)

RPPA Groups	Number of cases	IHC classes			
		Luminal	Basal	HER2-positive	Unclassified
Group 1	5		4	1	
Group 2	4	1	3		
Group 3	5		5		
Group 4	3			2	1
Group 5	3	1	2		
Group 6	5	5			

4.5.4. Analysis of RPPA data using consensus clustering and classification techniques including the role of Ki67

4.5.4.1. Correlation between expression of Ki67 in both IHC and RPPA

Ki67 expression using IHC was significantly correlated with RPPA data with *p*-value <0.002, **Figure-41**.

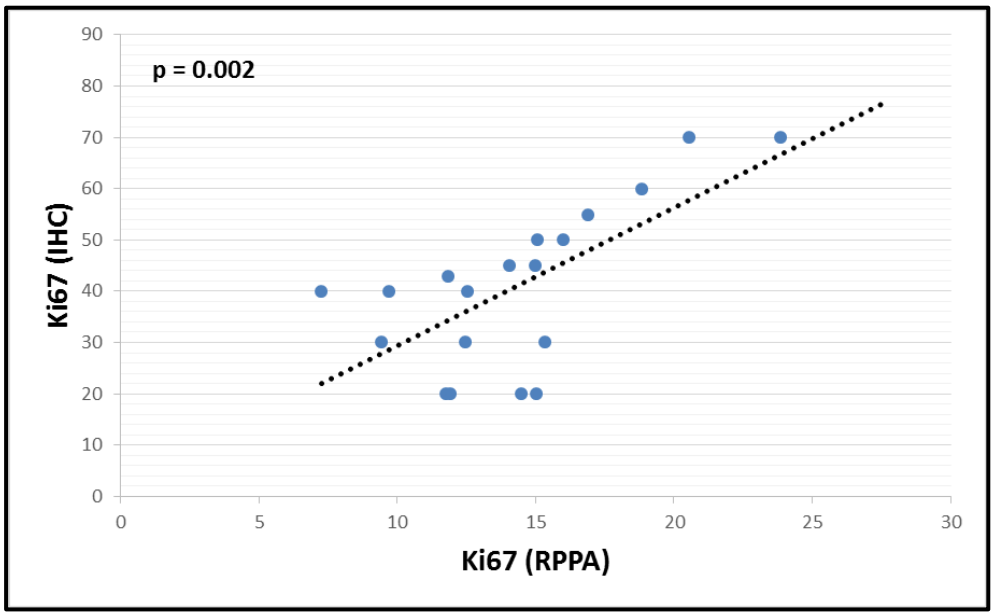


Figure-41: Correlation between Ki67 expressions using immunostaining (IHC) and RPPA

4.5.4.2. Evaluation the effect of re-classification of RPPA data using Ki67 on NPI+ classification

When 10 biomarkers were considered (NPI+ biomarkers)

The distribution of each variable in the class was compared with its distribution in the total sample using a boxplot. Initially, boxplot of the three consensus classes were produced (**Figure-42**) for all cases and for those cases assigned to classes 1-3. The boxplots of the consensus after the clusterfusion of the K-means and the PAM (**Table-20**). 99/143 cases were divided into three classes.

Table-20: A total of 99 BC cases are consensus among K-means and PAM

Method	Class 1 (n)	Class 2 (n)	Class 3 (n)	Total
K-means	51	46	46	
PAM	58	32	53	
Consensus	**35**	**23**	**41**	**99**

Figure-43 shows the boxplots of RPPA data classified into seven classes, a total of 92/143 cases showed consensus between K-means and PAM, **Table-21**. Classes six and seven differ from each other in terms of the biomarkers CK7/8, CK5/6 and HER1. These are on the higher side in the class six, on the lower side in class seven. The major difference among the classes; one, three and five have higher values for PR than class one. Among classes three and five, class three has a relatively lower value for ER and high values of MUC1, and vice-versa for both in class five, i.e. high values of ER and low values of MUC1. Classes two and four had a very similar distribution, except for the MUC1 biomarker. Class two have higher values as compared to the class four.

Table-21: A total of 92 BC cases are consensus among K-means and PAM

Method	NC	C*1 (n)	C2 (n)	C3 (n)	C4 (n)	C5 (n)	C6 (n)	C7 (n)	Total
K-means	14	24	25	6	28	6	33	7	
PAM	11	21	28	6	23	17	25	12	
Consensus	**0**	**11**	**19**	**6**	**19**	**5**	**25**	**7**	**92**

*C=Class

When 11 biomarkers were considered

Figure-44 shows the boxplots of the consensus after the clusterfusion of the K-means and the PAM, 122 cases were divided into three classes, **Table-22**. Further division was attempted by sub-classifying classes one and three. This was performed as these classes were large in number and the outliers were distribution was dense.

Table-22: A total of 122 BC cases are consensus among K-means and PAM

Method	Class 1 (n)	Class 2 (n)	Class 3 (n)	Total
K-means	63	18	62	
PAM	52	31	60	
Consensus	**52**	**18**	**52**	**122**

Figure-45 shows the boxplots of the seven classes. Classes one and five differed from each other in terms of PR and HER1 expression. These were higher in class one and lower in class five. The major difference amongst classes three, four and six were the high values of MUC1 for the class four. Class three had high values of ER and HER2. Classes two and seven had very different distributions. A total of 99/143 cases showed consensus between K-means and PAM, **Table-23**.

Table-23: A total of 99 BC cases are consensus among K-means and PAM

Method	C1 (n)	C2 (n)	C3 (n)	NC (n)	C4 (n)	C5 (n)	C6 (n)	C7 (n)	Total
K-means	33	24	12	4	11	24	22	13	
PAM	25	23	10	19	11	22	21	12	
Consensus	**25**	**16**	**7**	**0**	**6**	**18**	**16**	**11**	**99**

*C=Class

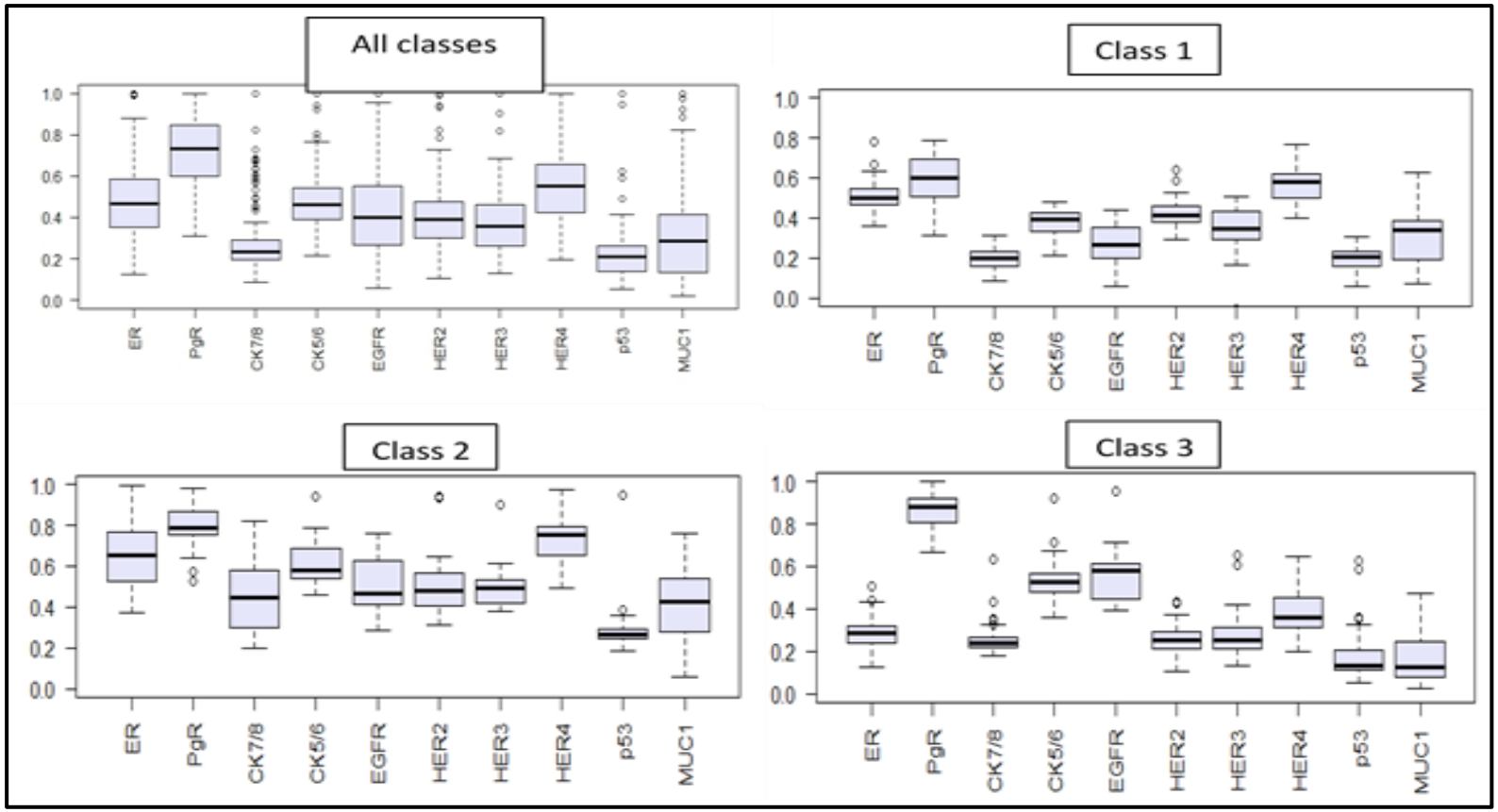


Figure-42: Boxplots of RPPA- breast cancer classification using NPI+ biomarkers to the main three classes

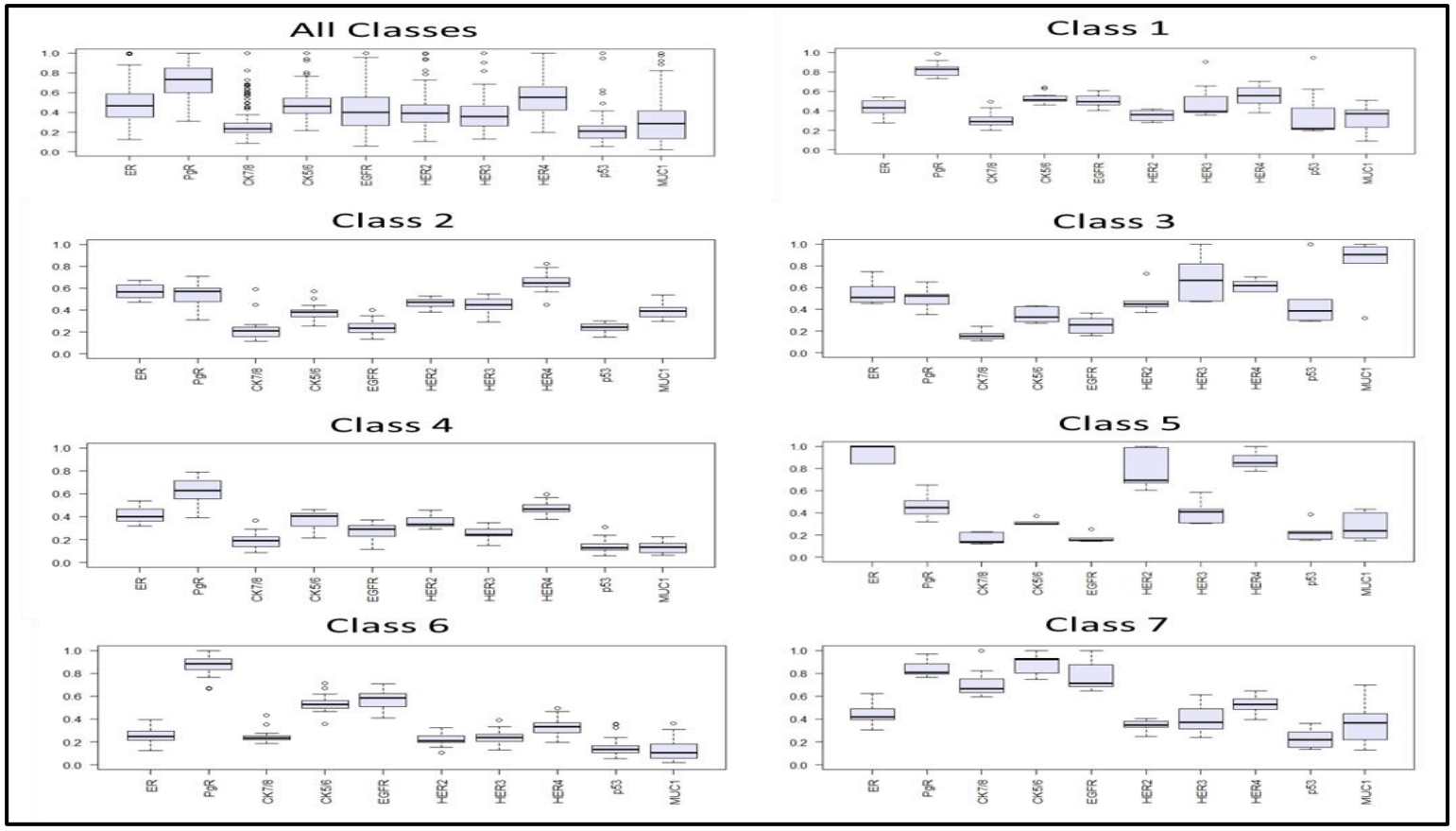


Figure-43: Boxplots of RPPA- breast cancer classification using NPI+ biomarkers for seven classes

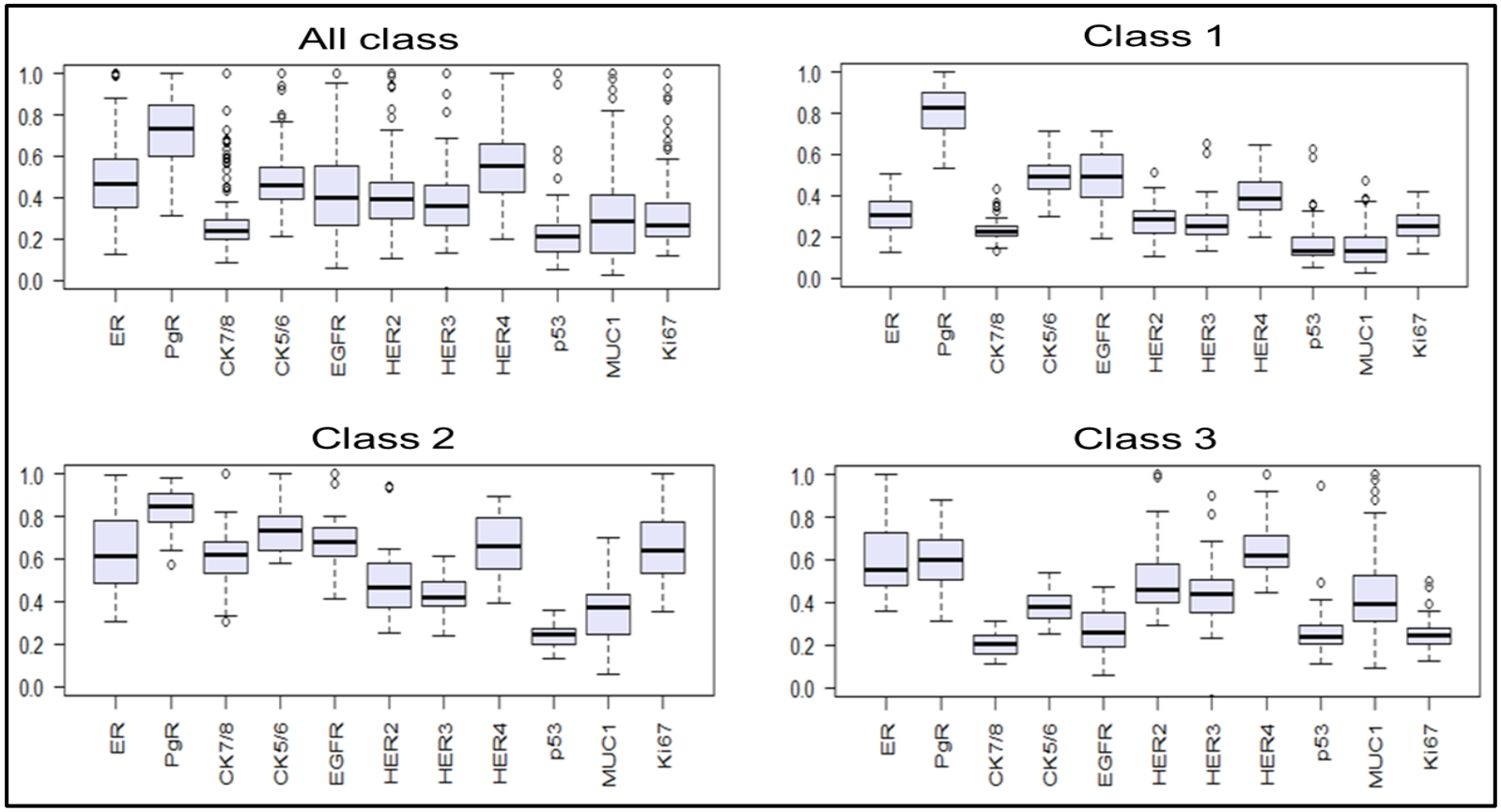


Figure-44: Boxplots of RPPA- breast cancer classification using NPI+ biomarkers to the main three classes after adding Ki67

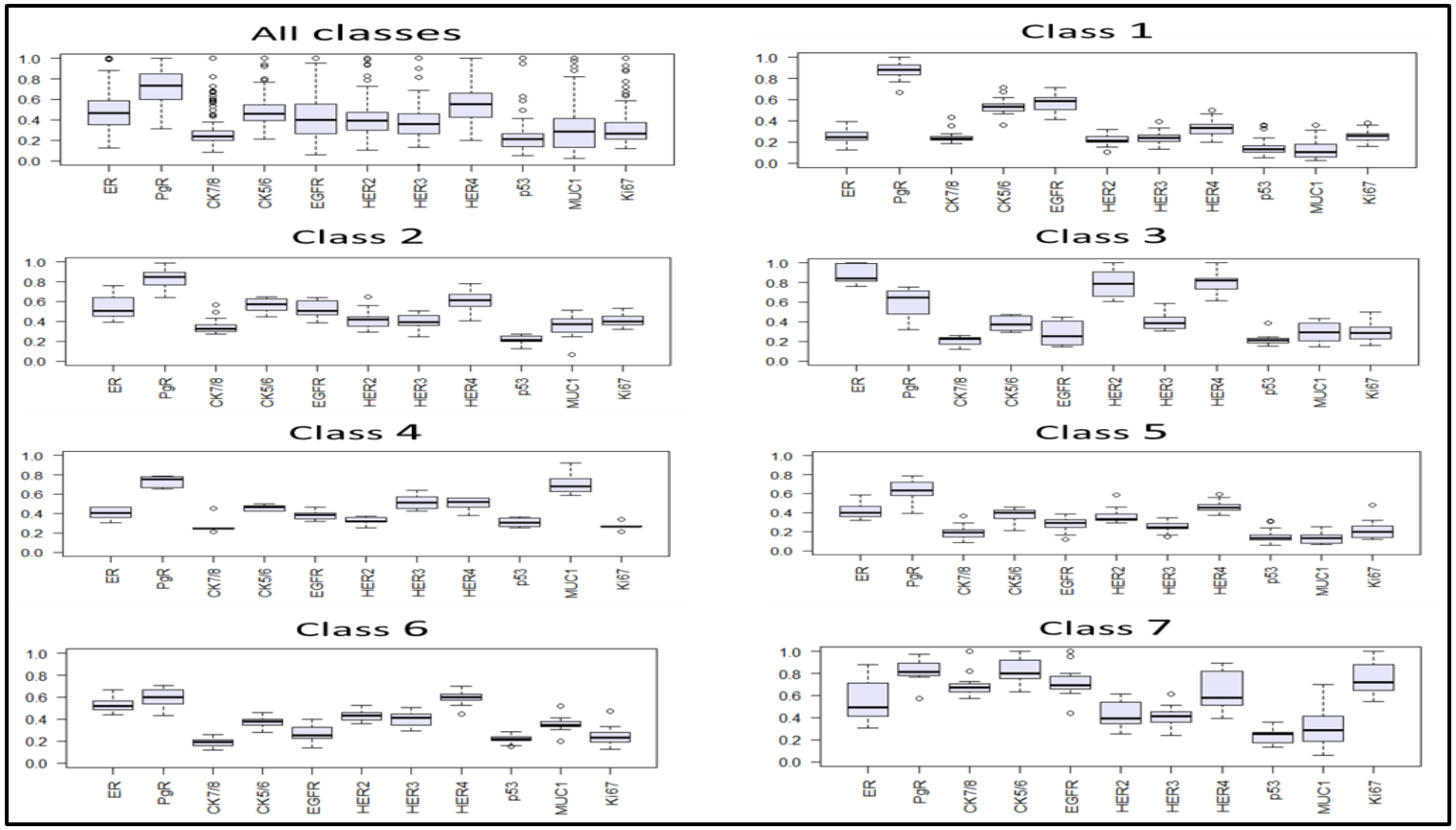


Figure-45: Boxplots of RPPA- breast cancer classification using NPI+ biomarkers for seven classes after adding Ki67

4.6. Discussion

The most critical concern at proteomic studies using clinical samples is the extraction and preparation of the protein lysate. This issue is more challenging when protein is extracted from FFPE tissue, which is associated with extensive molecular cross-linking. Based on the theory of heat-induced antigen retrieval method for immunohistochemistry, Ikeda et al. described successful protein extraction from FFPE tissue (Ikeda et al., 1998). Furthermore, there are reports of protein extraction from FFPE tissue using gel-based and mass spectrometry proteomic approaches (Addis et al., 2009) (Tanca et al., 2011). Although there are several published FFPE tissue proteomic studies of successful protein extractions, there has yet been no agreement on the optimal extraction protocol for protein from FFPE tissue or established standards for quantitative assessment of these extracts (Shi et al., 2013).

In this Chapter, the samples were systematically evaluated, with regard to both quantity and quality, protein extraction methods from FFPE tissue for optimal performance. This step is essential for consequent proteomic application on extracts of protein from FFPE tissue. It is well-acknowledged that the quantity and quality of protein produced from FFPE tissue is affected by the FFPE process, and the differences in time taken for preparation compared to frozen equivalents. The efficiency of protein extraction from BC tissues were compared utilising three different protocols; the commercially accessible kit and an in-house protocols. An efficient and successful extraction of proteins from FFPE samples was obtained using a Qproteome FFPE tissue kit, which was first used by Becker et al. in conjunction with RPPA technology in BC (Becker et al., 2007). Consistent with the results, the same kit used in that study gave an efficient protein extraction amount from samples that had been fixed and stored for a long time (Wolff et al., 2011b).

In order to examine the effect of tissue heterogeneity, lysates were prepared from FFPE tumour samples utilising three isolation platforms; full-face,

macrodissection and LCM. However, the protein yield by LCM was much less than the other two methods, while macrodissection produced the highest amount of protein. Becker et al. evaluated the influence of histological stains on proteins isolated from FFPE BC tissues and found that they may affect the amount of protein yield; however, this was representative of only two BC cases (Becker et al., 2008). To avoid the effect of staining, both stained and unstained sections were compared where there was no difference in protein yield between either stained or unstained tissue conditions. It may appear paradoxical to obtain low protein using LCM; yet it is not, as the results are in line with Scicchitano et al. (Scicchitano et al., 2009). They published a comparison between protein yields from controlled non-LCM and LCM using FFPE liver samples and demonstrated that protein yields from LCM samples are much less than those from non-LCM samples. Furthermore, frozen tissue is usually used with LCM and, as expected, results in a higher amount of extractable protein (Jiang et al., 2007) (Grantzdorffer et al., 2010) (Fowler et al., 2012).

RPPA is a high-throughput technology and a powerful tool in protein profiling, signalling pathway analysis and biomarker discovery. It has been successfully applied in quantifying proteins extracted from FFPE materials (Becker et al., 2007) (Wolff et al., 2011a). One of the technical aspects of RPPA-based tumour profiling is the reproducibility of target protein expression measurements. In order to evaluate the benefits of applying RPPA to archival FFPE BC tissues for the molecular classification of NPI+, this study was performed comparing the panel of NPI+ biomarkers' determination with IHC and RPPA. Interestingly, there was a consistency between the results of IHC and RPPA using both macrodissection and full-face methods. However, there was more concordance between IHC and RPPA in the macrodissection method than in the full-face method. Assadi et al. conclude that there is a good assignment of RPPA signals to IHC scores using FFPE tissue (Assadi et al., 2013). However, this study analysed only one biomarker (HER2) in BC tissue. In the current study, ten biomarkers were used: using macrodissection, eight markers (80%) showed significant agreement between IHC and RPPA, with six markers (60%) in the full-face method. The protein profiling differences observed are likely to have

arisen from the input samples, as full-face sections incorporate the tumour microenvironment and other tissues (Hennessy et al., 2010). The discordance at the level of some of these proteins determined by both IHC and RPPA is perhaps due to the heterogeneity of breast tissue staining, or might be due to their expression and subcellular distribution; as in our study, the membranous IHC staining of HER3 was only deemed positive regardless of cytoplasmic staining, if at all. However, RPPA can identify subtle differences in membranous/ cytoplasmic protein expression levels. In addition, some antibodies fail to recognise the expected membranous localisation of particular markers by IHC and this in sequence will affect reproducibility (Anagnostou et al., 2010). RPPA analysis of FFPE tissue lysates gives rise to biologically meaningful data. Initially, using a set of 25 cases, patients were clustered into subgroups using RPPA broadly based on the applied expression of these ten biomarkers. These clustered subgroups are strongly consistent with the IHC subclasses of NPI+; for instance, the groups of RPPA that represent the basal class in NPI+ classification were separated depending on P53 consistent with NPI+, as basal-p53 normal or altered.

Interestingly, when unsupervised classification with a larger cohort (n=143 cases) was used, the results were promising and encouraging for the application of RPPA in BC clustering and classification. Consensus clustering, which used in this study, which offers a method to represent the agreement through multiple runs of a clustering algorithm and to evaluate the stability of the discovered clusters (Soria et al., 2010). The more the achieved clusters are robust to sampling variability, the more one can be assured that these clusters characterise real structure. The idea of combining and comparing the results of different clustering algorithms is important with the purpose of assess the stability of a proposed classification. The reasons for using consensus classes are to get the highest number of patients assigned to any class and to consider all the clustering techniques analysed. Instead of using a comparison approaches, this consensus methodology was used to pool results obtained by different clustering algorithms. The best number of clusters to consider is another important issue that arises when cluster analysis is conducted. Several validity indices have been

suggested in recent years. For K-means and PAM algorithms, cluster validity indices were used to direct the selection of the optimal number of clusters, in the study were three and seven classes, based on NPI+. Although the class assignments were performed blind to all clinical and outcome data, the initial cluster using three classes showed 69% of data was assigned to these core classes. Interestingly, 85% of data was assigned when Ki67 added. In this study, the previous proposed methodology (Soria et al., 2010) for reaching consensus from the various results obtained from clustering algorithms were applied using RPPA BC data. This consensus clustering and classification to reproduce NPI+ using RPPA should be considered as an exploratory study and further work linking these biological classes to the patient outcome have to be considered. However, the results reflect the ability of RPPA and could be beneficial for the stratification of patients for therapy regimens and classification of BC. Furthermore, it emphasised the importance of Ki67 in BC classification using RPPA data.

4.7. Conclusion

In conclusion, the analysis of FFPE tissue lysates using RPPA is demonstrated as a reliable method for protein quantification. Data generated through this high-throughput technique can be used in simultaneous analysis of protein portraits in a large number of clinical samples. Accordingly, RPPA could be reliably used in molecular classification of BC, such as the NPI+, through fast and reliable proteomic quantification of multiple proteins in BC samples allowing patient stratification according to BC progression, risk of recurrence and response to therapy. Further, Ki67 is an essential marker for BC classification.

5. Evaluation of tumours expressing low ER in the molecular classification of BC

5.1. Introduction

The detection of ER resulted in the development of the first effective targeted therapy approach in BC and to the determination of two basic molecular classes of this disease: ER-positive and ER-negative. The correlation between the hormones and BC growth was documented when George Beatson discovered that exposing an advanced BC patient to oophorectomy (the surgical removal of the ovaries) leads to remission of the tumour (Beatson, 1899). Until the 1950s and 1960s, oophorectomy was the standard therapy of choice. However, only one-third of the patients responded to this ablation surgery. Therefore, to avert inadequate management, there was a necessity for predictive markers. The discovery of ER and its recognition in BC produced targeted-strategies for BC therapy (Jensen and Jacobson, 1962) (Jensen and Jordan, 2003).

Certainly, accurate evaluation of biological markers in BC is needed to have clear indicators of effects from targeted therapies. Studies have revealed that ER is the key differentiating marker of molecular signature classification in BC, supporting the fact that BC is a heterogeneous disease and ER-positive and -negative breast carcinomas are biologically distinct entities. Clinically, BCs are classified as either ER-positive or ER-negative and different studies investigated the prognostic value of ER negative status and its association with decreased survival in node negative BC (Crowe et al., 1982) (Fisher et al., 1988) (Pichon et al., 1996) (Railo et al., 1997). ER expression in BC has a high clinical significance as a predictive factor and benefit of hormonal therapy. In a meta-analysis, tamoxifen as well as other selective oestrogen receptor modulators have offered an overall improvement of patient survival by 38 %. The reduction was higher (42%) in the first 5 years than in years 5-10 (25%) (Cuzick et al., 2013). Moreover, Early Breast Cancer Trialists' Collaborative Group demonstrated that treatment of patients with tamoxifen as anti-oestrogen therapy in BC had an effect on 5-year survival and recurrence in BC (2005). Numerous studies endorse the relation between hormonal receptor status and response to chemotherapy where the ER negative group (21%–33%) had a better response

to neoadjuvant chemotherapy than the ER positive group (7%–8%) (Ring et al., 2004) (Colleoni et al., 2004) (von Minckwitz et al., 2011).

ER is expressed in 60–80% of BC cases and most of these cases express PR as well (Dobrescu et al., 2011) (Cui et al., 2005). However, the discrepancy between PR and ER expression can occur (Cui et al., 2005) (Rakha et al., 2007). As PR is an ER-regulated gene, the expression of this receptor is highly related to the presence of the ER. ER−/PR+ is exceptional and represents less than 1 % of all BCs (Hefti et al., 2013). The PR level is increased in normal reproductive tissues and in MCF-7 BC cells depending on treatment with 17β-estradiol (E2). Moreover, *PGR* mRNA and protein reach their highest measures after MCF-7 cells have been treated with E2 for 72 hours (Nardulli et al., 1988, Wei et al., 1988, Read et al., 1988). Although no classical oestrogen responsive element was identified in the corresponding sequences, the PR promoter regions A and B are oestrogen inducible (Kastner et al., 1990). Schultz et.al suggests that ER is involved in mediating oestrogen responsiveness to the PR gene by stimulating Sp1 interaction with the Sp1 site of the human PR gene (Schultz et al., 2003) (Petz et al., 2004).

5.1.1. The pattern of ER distribution

The association between the level of ER expression and response to endocrine treatment has been reported by several studies and clinical trials showing that only 50% of tumours scored 4–6 with Allred score respond to endocrine therapy compared to 75%–80% of cases scored 7 or 8 (Harvey et al., 1999) (Elledge et al., 2000) (2005) (Davies et al., 2011). Initially, ER expression was tested by radiolabelled ligand binding assays (LBA) using fresh-frozen tumour samples which exhibited a continuous scale of values. For a period of over two decades, IHC has been utilised for assessing ER expression to predict the response to endocrine therapy. Some researchers published that the linear distribution of ER expression demonstrated with LBA was also observed with IHC (Harvey et al., 1999, Elledge et al., 2000). However, other authors have challenged this concept providing evidence that ER expression is basically bimodal. Based on reviewing

825 BC cases, Collins et al. showed that 99% of tumour cases showed negative staining or staining in 70% or more concluding that ER expression is essentially bimodal (Collins et al., 2005). Furthermore, using a series of 5,993 ER immunostained BC cases, Nadji et al. supported an all or none phenomenon for the ER expression (Nadji et al., 2005). Our study added further evidence to support bimodality of ER expression in BC (Muftah et al., 2016).

At the transcriptomic level, recognising genes with bimodal expression patterns from large-scale gene expression profiling data has presented new insights into the distribution of expression of important genes. Two different classes of genes have been defined. The first class is represents genes with a Gaussian or continuous distribution, which can be explained where two small groups of tumours have very high expression and very low expression and the rest, which represent the majority, falling somewhere in the middle. The second class is consistent of genes with a bimodal distribution of expression. This class has the majority of tumours with two peaks either high levels of gene expression or no expression and only a relatively small number of tumours fall in between. Previous studies have indicated significant correlations between *ESR1* expression and the clinical ER status. *ESR1* has a high bimodality index score and it can be used to categorise samples into two distinct expression levels (Wang et al., 2009) (Ertel, 2010). The useful point of genes with bimodality at the transcript level is that they may be easily translated to the protein level and detected using IHC. A variability of available assays may need to be evaluated at the gene or protein level before a satisfactory reliability is obtained. This might demonstrate how far the relation between transcripts based assays and IHC is and to what extent these RNA assays can support manually scored IHC.

5.1.2. Luminal/ER-positive class and low ER subgroup

Gene expression profiling studies in BC have demonstrated the presence of distinct biological classes with clinical and therapeutic importance. ER has been considered to be a key marker of the molecular signature. These molecular profiling studies have recognised a well characterised classes: luminal (ER-positive group), HER2-enriched, the basal-like and normal-like classes (Perou et al., 2000). The luminal class is characterised by high level of ER expression and other genes identified to be characteristic of normal luminal cells of the breast. This class is the largest of the GEP defined molecular subclasses reflecting the importance of ER. The luminal class can be divided into at least two subtypes; namely luminal A and luminal B (Hu et al., 2006). Furthermore, subsequent studies supported the presence of further luminal subgroups (Naderi et al., 2007) (Teschendorff et al., 2006). Based on protein level, luminal subgroups had been determined by using simple routine IHC technique, including both ER and PR, on tissue microarrays preparations of 1,076 BCs (Abd El-Rehim et al., 2005).

Interestingly, Brandon et al reported intrinsic subtype classification based upon the PAM50 gene expression signature on a cohort of weakly positive ER BCs (Sheffield et al., 2016). Weakly positive ER breast is the group lies between 1% and 10% of ER expression in invasive BC where 10% was previously used as the cut-off point of positivity whereas 1% is the current threshold of positivity adopted in 2010 according to the ASCO/CAP guidelines (Hammond et al., 2010b). Evidence shows that a response to tamoxifen hormonal therapy may be detected with as little as 1% of ER status (Harvey et al., 1999) (Hammond et al., 2010b). However, from the Early Breast Cancer Trialists' Collaborative Group, the results of a meta-analysis showed tamoxifen was ineffective with low ER expression (<10 fmol/mg) using ligand-binding assay (Davies et al., 2011). Lin et al. highlighted a need to evaluate decision making for the use of adjuvant hormonal therapy might differ in low ER- or PR-positive patients (Lin et al.,

2013). Little is known about how weakly ER-positive BCs behave and how they are related to the ER-positive BC. Therefore, there is necessitate to assess this critical subgroup of weakly positive ER and to what extent a recommended threshold can reflected the real positivity of these cases.

5.2. Hypothesis

Although ER positivity (luminal class) is currently defined as >1% ER expression, it is hypothesised that BC expressing low ER (i.e. between 1-10%), clinically behave differently to those BC with high ER expression (>10%), and whether the low ER group belongs to the ER-positive or ER–negative group.

5.3. Aim of this chapter

This chapter aims to assess the clinical and molecular features of tumours with low ER expression, to evaluate the definition of positivity (cut-off point) and to examine to what extent this group is related to the ER-positive subcategory in BC classification.

To achieve these aims the following will be assessed:

1. The pattern of ER expression, both at the protein (IHC) and transcriptome (microarray-based gene expression) levels using four large patient' cohorts.
2. The molecular features of the low ER group examining specific cohort:
    a. At the protein level, assessing ER and PR using two different IHC platforms (manual vs automated techniques)
    b. At the transcriptomic level, assessing ER and PR using two different techniques RT-PCR and RNAscope.

5.4. Methodology

The first part of this study assessed the pattern of ER expression using four different cohorts at both the transcriptome and protein levels. In the second part, a subset was used to further evaluate those cases in the low ER subcategory as initially diagnosed using CNB in routine clinic, at Nottingham City Hospital. This subset contains 71 WTS cases where 56 cases as low/intermediate ER expression patient samples, 5 cases as a highly ER positive expression and 10 cases as a negative control. These cases were evaluated at the protein level using two different staining platforms of IHC, initially with IHC manual staining then the staining was validated using the routine automated platform. For further validation of the protein levels results, the same subset was evaluated at the transcriptomic level using two different techniques: initially RT-PCR and further the new *in situ* RNAscope technique, which has an advantage of allowing the examination of biomarker status within the histopathological context of clinical specimens. Therefore, RNAscope can compared easily with IHC, **Figure-46**.

5.4.1. Assessment of ER expression pattern

5.4.1.1. Patient cohorts

This study was based on four different BC cohorts. In the first, second and third cohorts, ER was assessed at the protein level using IHC while in the fourth cohort ER expression was examined on transcriptomic level (RNA level). (Details of these series were previously described in Section 2.1). At protein level using IHC, ER weakly positive were defined as those scored between 1 and 10% expression while intermediate ER expression was defined as those with percentage staining up to 69%. ER negative cases were those cases scored <1%. Highly positive ER cases showed 70% or more of the tumour cells expressing ER.

1) The first cohort (Primary Series) consisted of 1954 patients diagnosed between 1988 and 1998 whose BC tissues were prepared as TMA as part of the previously published Nottingham Tenovus Primary Breast Carcinoma Series. ER expression was available for 1892 cases. TMA sections of this series were stained with the ER SP1 clone antibody (Dako), using a dilution 1:150 and incubated for 30 minutes. Levels of immunohistochemical ER expression were evaluated by microscopic assessment as the percentage of positively stained ER invasive tumour cells.

2) The second cohort consisted of a consecutive series of symptomatic and screen detected BC patients at Nottingham City Hospital. These cases had ER status assessed on a preoperative CNB in routine practice between March 2008 and November 2014 (n=3649 cases). CNB were fixed, processed and stained according to standardised protocol as previously published (Hodi et al., 2007). The primary antibodies used for staining were either 1D5 (Dako), diluted 1/100 or a pre-diluted SP1 clone (Roche). A subset of 71 WTS cases from this cohort was used to evaluate the positivity of the low ER subgroup.

3) The third cohort comprised 1825 BC cases diagnosed between 1998 and 2006 which were prepared as TMA. TMA sections of this series were stained with the ER SP1 clone antibody. Levels of immunohistochemical ER expression (n=1686) were evaluated by microscopic assessment as the percentage of positively stained ER invasive tumour cells.

4) The fourth cohort comprised 1980 BC cases that were included in the METABRIC cohort (Curtis et al., 2012a). In this cohort, the survival data and the ER immunostaining were available for 262 cases from Nottingham.

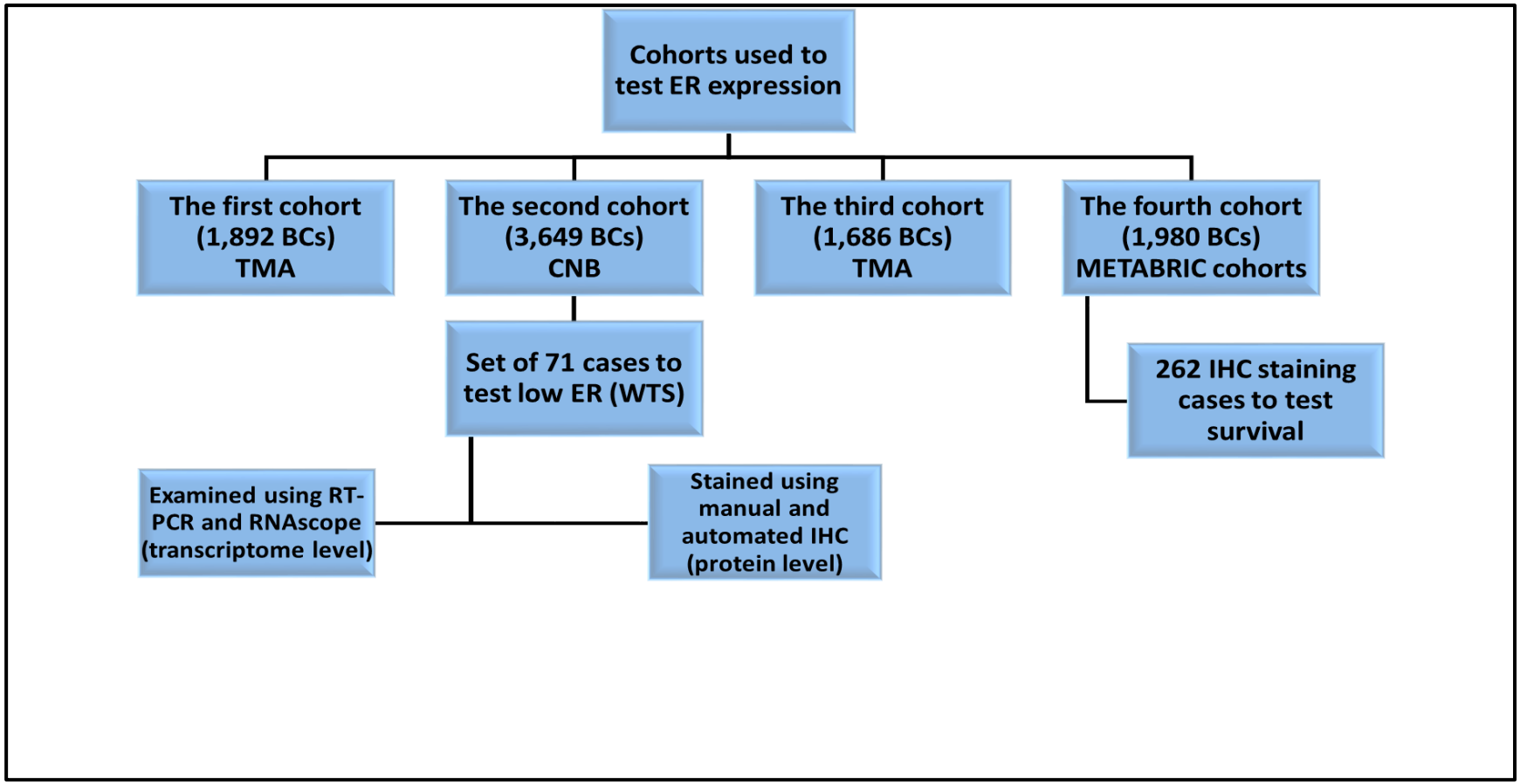


Figure-46: A flowchart representing the cohorts were used in this Chapter

5.4.2. Evaluation of clinical and molecular features of low ER group

5.4.2.1. Immunohistochemistry

To further assess the existence and frequency of low ER expressing tumours, a specific set of 71 cases of the second cohort was used where 56 cases showed a low/ intermediate ER expression on CNB where weakly positive ER staining (scored <10%) was observed in 33 cases and intermediate ER expression (scored ≥10) in 23 cases. The rest of the cases represented a negative ER expression (10 cases) or highly positive ER expression, >70% (5 cases). Surgical excision (WTS) of 56 cases that expressed a low/intermediate ER on CNB were immunostained to validate the level of ER expression. Two different platforms of IHC technique was used for staining WTS for ER, manual staining and Ventana Autostainer (described in Section-2.3). In manual staining, Optimised primary antibody, EP1 anti ER Rabbit monoclonal antibody (Dako) was applied and incubated for 30 minute at room temperature. In addition, the manual staining of PR was performed. The same steps was used for PR staining using anti PR antibody with dilution of 1:125 and incubated for 30 minutes. Assessment of nuclear ER expression was performed using percent and H-score system.

In automated immunostaining, the slides were processed a pre-diluted tris-based buffer with a slightly basic pH (Roche, Ventana) for 64 minutes at 95°C (AR step). Anti ER Rabbit monoclonal antibody, SP1 clone (Roche) was applied and incubated for 16 minute. Blocking solution (Roche, antibody diluent with casein) and DAB were added for 8 minutes each with wash before and after. Haematoxylin (Roche, Ventana Haematoxylin II) was applied for 12 minutes.

5.4.2.2. Western blot

The specificity of anti-oestrogen receptor antibodies (both SP1 and EP1 clones) was validated by western blot utilising MCF7 cell line lysate. The specificity was determined by detecting a particular antigen (ER) reflected as a specific band (**Figure-47**). All steps of western blot and cell lysate preparation described in Section-2.4.

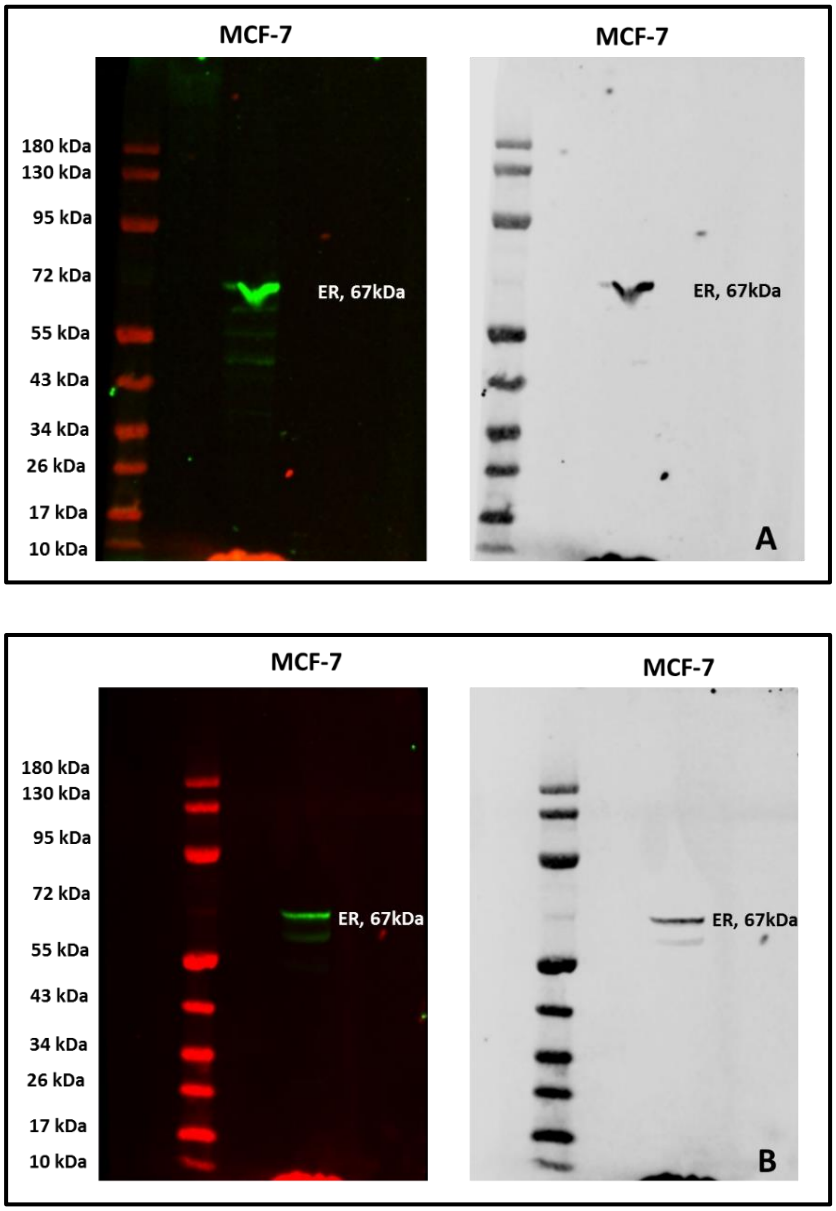


Figure-47: Western blot demonstrates a specific band for ER antibodies: (A) SP1 clone and (B) EP1 clone

5.4.2.3. RT-PCR

The same subset of cases (n=71) that were stained using IHC were assessed at transcriptomic level using RT-PCR to evaluate the expression of *ESR1* and *PGR* genes. Applying a combined transcriptomics and proteomics approach can provide a comprehensive overview of the molecular characterization of this subgroup. Before extraction of RNA, two different RNA extraction kits were compared in obtaining high nucleic acid yields and purity from archival paraffin samples.

Comparison between two protocols for RNA extraction from FFPE tissue sections

Full-face FFPE tissue sections, randomly selected BC samples (n=3) were used. These were primarily chosen based on the adequacy of invasive tumour tissue burden at least 50 % of the overall area of the tissue section, assessed using H&E tissue sections. In the first protocol RNeasy FFPE Kit (Qiagen) was used, the details of the method illustrated in Section-2.8.2, which is the method was chosen later in this study.

In the second protocol, BiOstic FFPE tissue RNA isolation kit was used, in which 300µl of solution FR1 (a high activity dissolving buffer) and 20µl of solution FR2 (Proteinase K) was added to tissue sections (three sections were used each 10µm thickness). After mixing for 20 seconds, the samples were heated at 60C° for 15 minutes. The samples were subsequently heated at 70C° for 15 minutes. Next, the samples were centrifuged at 13,000xg for 1 minute. 300µl of solution FR3 (it is a salt buffer required for RNA binding to the silica filter membrane), 600µl of 100% absolute ethanol was added to the lysates and loaded onto low elution spin filter. After centrifugation at 13,000xg for 30 seconds and discarding the flow-through, the remaining volume of lysate was loaded. Then, 600µl of solution FR4 (isopropanol based wash buffer) was loaded and followed by centrifugation twice at 13,000xg for 30 seconds and at 16,000xg for 2 minutes. The flow-through was discarded. Following that, 50µl of RTS DNase/solution FR5 (is 4µl RTS DNase added to 46µl solution FR5) was added.

Then, 400µl of solution FR6 (wash buffer) was loaded to low elution spin filter and centrifuged at 13,000xg for 30 seconds. After discarding a flow-through, 600µl of solution FR4 and 600µl of 100% absolute ethanol were added respectively with centrifugation at 13,000xg for 30 seconds and discarding a flow-through after each step. Following that, the samples were centrifuged at 16,000xg for 2 minutes and the low elution spin filter were transferred into a clean collecting tubes. Finally, 20µl of solution FR7 (RNase-free water) was added with incubation for 3 minutes at room temperature. Then, the samples were centrifuged at 16,000xg for 2 minutes, low elution spin filter were discarded and the eluted RNA in the collected tubes stored at -80C° after evaluation of the quantity and the quality of the samples. To assess RNA quantity and purity, a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) was used (section 2.8.3).

RNA extraction from FFPE tissue sections using a suitable protocol

RNA extraction from FFPE tissue sections was performed on 71 cases following the steps in Section-2.8. Quality and the quantity of RNA was assessed using NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific), Section-2.8. Prior to RT-PCR, RNA was reverse transcribed to generate cDNA (Section-2.8). The appropriate primers designed for HPRT, ERα, PR-A and PR-B (**Table-24**) were used (Green et al., 2009) .

cDNA Generation and Quantitative RT-PCR

cDNA was produced by reverse transcription of RNA. For more detail about this step please refer to Section-2.8.4.

Each PCR reaction included: 5µl of GoTaq® qPCR Master Mix, 6.25µM of forward primer, 6.25µM of reverse primer, 2µl cDNA and water. The following PCR programme was then used for amplification of the cDNA:

95°C/2minutes		1 cycle
Denaturation	95°C/15sec	45 cycle
Annealing	60±2°C/45sec	
Extension	72°C/20sec	
Final extension	95°C/15minutes	1 cycle

Table-24: Primers sequences for genes used in RT-PCR

Genes	Oligonucleotide primers
HPRT	(Forward primer) AAA TTC TTT GCT GAC CTG CTG
	(Reverse primer) TCC CCT GTT GAC TGG TCA TT
ERα	(Forward primer) TGGGCTTACTGACCAACCTG
	(Reverse primer) CCTGATCATGGAGGGTCAAA
PR-B	(Forward primer) GCCAGACCTCGGACACCTT
	(Reverse primer) CAGGGCCGAGGGAAGAGTAG
PR-A	(Forward primer) AGAGCACTGGATGCTGTTGCT
	(Reverse primer) TGGCTTAGGGCTTGGCTTT

RT-PCR data analysis

RT-PCR data can be reported by several methods including absolute or relative expression levels. Relative quantification or comparative quantification detects the changes in RNA levels of a gene across multiple samples and expresses it relative to the levels of another RNA. The comparative CT method which was used in this study, is one of the method presenting RT-PCR data as a relative quantification (Livak and Schmittgen, 2001) (Schmittgen and Livak, 2008). This method is also known as ΔΔCD method and has been commonly used (Sun et al., 2008) (Tan et al., 2007) (O'Rourke and Ness, 2008). The advantages of this method include the possibility to present data as fold change in expression where the fold change equal to $2^{-\Delta\Delta CT}$. Importantly, when RT-PCR data is presented as individual data points it is preferred to be presented as ΔΔCD instead of the raw CT value (Livak and Schmittgen, 2001), which is consistent with the aim of this study. Additionally, a normaliser is important to correct expression data for differences in RNA quality and RT efficiency between samples. Therefore, hypoxanthine ribosyltransferase (HPRT) gene was selected as the housekeeping genes, which is identified as the single best reference gene that could be used when the expression patterns of 13 frequently used housekeeping genes were tested in normal and tumour samples from colorectal, breast, prostate, skin, and bladder tissues with RT-PCR (de Kok et al., 2005).

The equation full form is:

$2^{-\Delta\Delta ct}$= [(C_T gene of interest –C_T of internal control) sample A – (C_T gene of interest –C_T of internal control) sample B] (Schmittgen and Livak, 2008). This form of the equation used to compare the gene expression in two different samples and each sample is related to an internal control gene. Sample A is a tested sample while sample B is the calibrator, which is here is a pooled of negative ER expression cases. The cut-off point was selected depending on the average of the fold changes of the negative cases relative to the pooled calibrator (**Figure-48**). The cut-off point was 1.00 where the standard deviation was 1.006± 0.100.

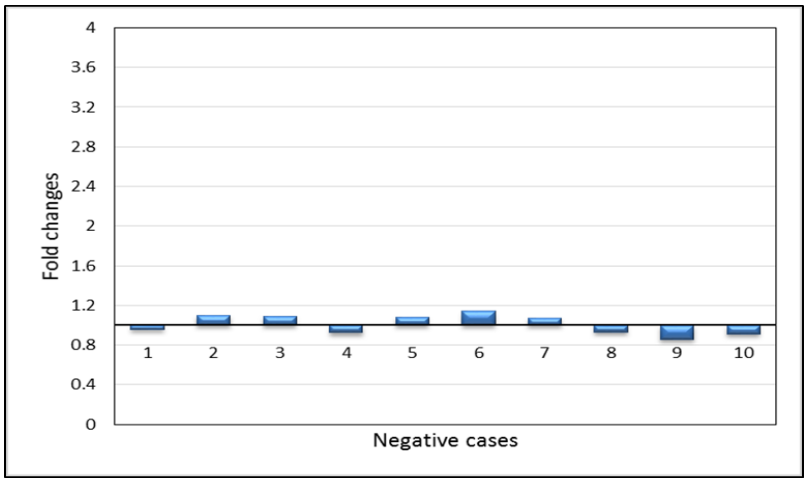


Figure-48: Fold changes of the negative cases relative to the pooled calibrator which used in RT-PCR. The cut-off point was 1.00 where the standard deviation was 1.006± 0.100

5.4.2.4. RNAscope technique

RNA *in situ* hybridisation (ISH) technique was applied to the subset of cases (71 cases of invasive BCs) previously used for IHC and RT-PCR. For each case, a minimum of 4 slides of invasisve BC were evaluated for *ESR1*, *PGR* mRNA expression as well as positive and negative controles.

RNAscope Optimisation

Pre-treatment, RNAscope protocol, was optimised for each sample set to produce maximum signal and minimum background. Optimised conditions were

utilised for all subsequent marker assessment assay runs. Samples had to pass QC (quality control) in Phase I to be included for marker evaluation in Phase II. Sample QC was performed on all samples using Cyclophilin B, peptidyl prolyl isomerase B (PPIB) representing housekeeping genes (positive) and Bacterial dapB (negative). Samples were deemed to be QC Pass if: PPIB ≥2, with relatively uniform PPIB signal throughout the samples and dapB score <1, no non-specific signal from dapB proves that all identified signal with a target-specific probe will correctly reflect target RNA expression. Samples that passed QC were encompassed in Phase II for marker assessment and whereas samples that did not pass QC underwent repeated rounds of optimisation. However, samples that eventually failed QC were not included in Phase II. A TNBC sample were used as a ER-negative control and BC tissue ER-positive control were also used (ACD Bio). The samples were run on Leica BOND Rx system, using ACD Bio standard brown chromogenic single-plex assay. The *ESR1* and *PGR* status information of each sample was blinded. The details of the technique is summarised in section-2.9, (RNAscope staining was performed by ACD Bio).

RNAscope quantification

Semi-quantitative scoring criteria for whole-tissue evaluation of RNAscope staining was used. Each section was visually (initially with eyeballing and then using of the high magnification especially for the hetrogenous slides) scored based on the average number of dots per cell throughout the entire sample using a specific criteria (**Table-25**). Scoring is performed based on the number of dots per cell rather than the signal intensity. Images representing semi-quantitative scoring criteria for RNAscope are shown in **Figure-49**.

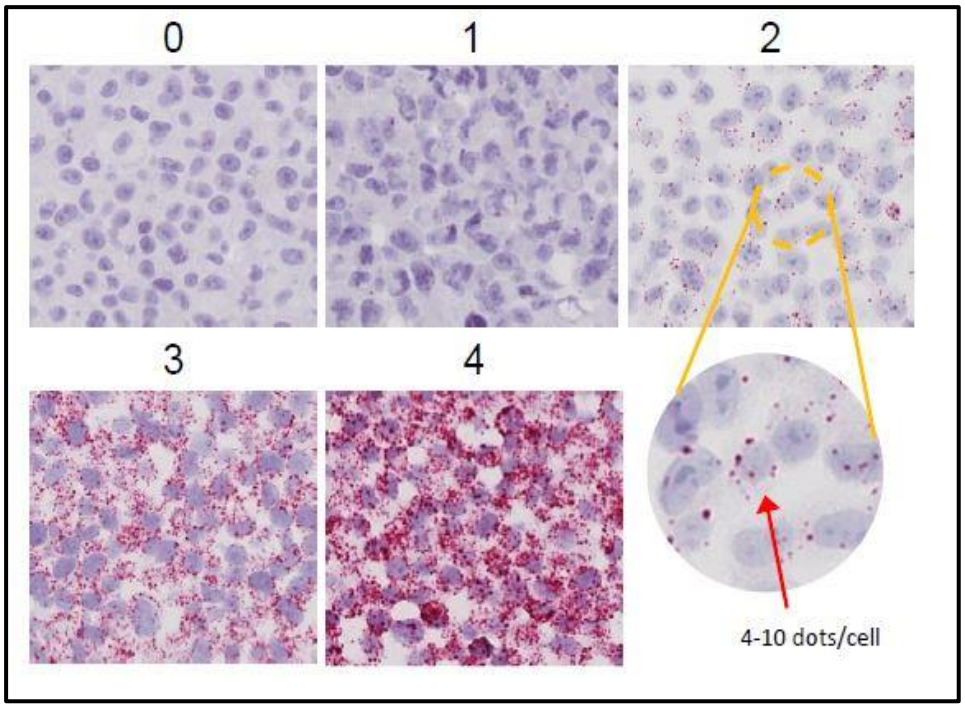


Figure-49: Semi-quantitative scoring criteria for RNAscope (https://acdbio.com/technical-support/solutions)

Table-25: Semi-quantitative scoring criteria for RNAscope staining

Score	Criteria
0	No staining or <1 dot/10 cells
1	1-3 dots/cell
2	4-9 dots/cell, no or very few clusters
3	10-15 dots/cell and <10% dots are in clusters
4	>15 dots/cell and >10% dots are in clusters

(https://acdbio.com/technical-support/solutions)

5.5. Statistical analysis

The statistical tests were mainly performed using IBM SPSS software version 22 (SPSS Inc., Chicago, IL, USA) (PAWS statistics). For the assessment of ER expression pattern on different cohorts, the IHC data was classified into three groups, the first group comprised cases with ER expression scored as <1%. The second group contained cases up to 69% because of the small number of cases in this category the low and intermediate groups were combined together. The third group contained cases with ER-positive ≥70% which is similar to the cut-off point previously used by Collins et al (Collins et al., 2005). For the assessment of ER expression pattern we used wider cut-off to define upper and lower expresser for ER, however, for the low ER evaluation because we are focusing on the optimum cut-off we used cut-off (≤10%) for the low ER group.

In the first cohort, survival curves were analysed by the Kaplan–Meier statistical test with significance determined by the Log Rank. Additionally, the association between subcategories and the clinicopathological variables as well as PR and HER2 as a key prognostic and predictive markers were assessed using the Chi square test. When *p*-value was <0.05, it was taken into consideration as a significant value. For transcriptomic data of the fourth cohort, 1980 BC cases were used to illustrate the *ESR1* gene expression distribution pattern. However, patient outcome and immunostaining of ER were only available for 262 cases of this cohort. These cases were subdivided into three subgroups (at transcriptomic level) according to the changes in the distribution of the curve using the SPSS software (PAWS statistics). The first cut-off point was 6.5 while the second was 8.6 (Muftah et al., 2016).The high expression group included those cases with *ESR1* expression level of >8.6 and the low expression group included cases <6.5. Gene expression was compared with protein expression and patient outcome. Survival curves were analysed using Kaplan–Meier method with significance determined by the LR test and a *p*-value <0.05 was considered significant. In addition, Chi square test, Intraclass Correlation Coefficient (ICC), spearman's correlation were used. Spearman's correlation between 0.30-0.50 is considered

as low correlation, 0.50-0.70 as moderate correlation, 0.70-0.90 as high correlation ad 0.90-1.00 as very high or perfect correlation (Mukaka, 2012).

5.6. Results

5.6.1. Assessment of the ER expression pattern

5.6.1.1. Protein level assessment

Based on the percentage of ER-positive tumour cells, the frequency distribution of ER staining in the first cohort of 1892 cases is shown in **Figure-50**. This series showed bimodality of expression with the strongly positive and completely negative cases representing 89.2% of the cases. Weak and intermediate expression of ER was infrequent representing 1.6% and 9.2% of cases, respectively. As a result of the small number of cases in the weak positive and intermediate expression group, the data was combined and analysed accordingly; negative group (<1%), intermediate group (1-69%) and highly positive group (≥70%). The assessment of the frequency of expression of PR as an ER dependent protein in the TMA series using the previously published IHC stained PR (Rakha et al., 2007) and which showed a bimodal distribution similar to ER (**Figure-51**). Outcome analysis revealed significant association between ER expression groups and survival (**Figure-52**; *p*-value <0.001, LR=23.5). However, further analysis exhibited no significant survival difference between the intermediate ER expression group and the ER negative group (*p*-value =0.324 and LR=0.97), while there was a significant difference between the intermediate group and the highly positive group (*p*-value =0.022 and LR=5.21). Expectedly, the difference in patient outcome between the negative and highly positive groups was highly significant (*p*-value <0.001 and LR=22.05).

Analysis of the second cohort (n= 3649) which evaluated ER on CNBs showed the majority of cases (92.2%) were either negative (<1%) or strong positive (≥70%) for ER expression while weak ER-positive cases (2.7%) and the

intermediate-expression group (5.1%) were infrequent. **Figure-53** illustrates the bimodality of ER expression where those with <1% expression constituted 22.4 % of cases and strong ER accounted 69.8 % of the total cases. **Figure-54** illustrates negative, low, intermediate and high expression of ER.

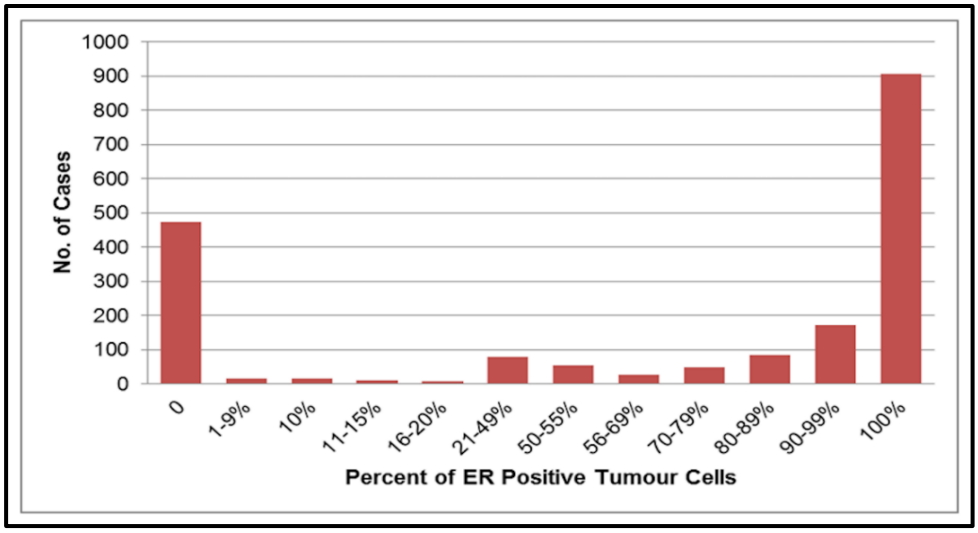


Figure-50: Bimodal distribution of ER
IHC staining scores in the first cohort (1892 cases) assessed on TMA (Muftah et al., 2016)

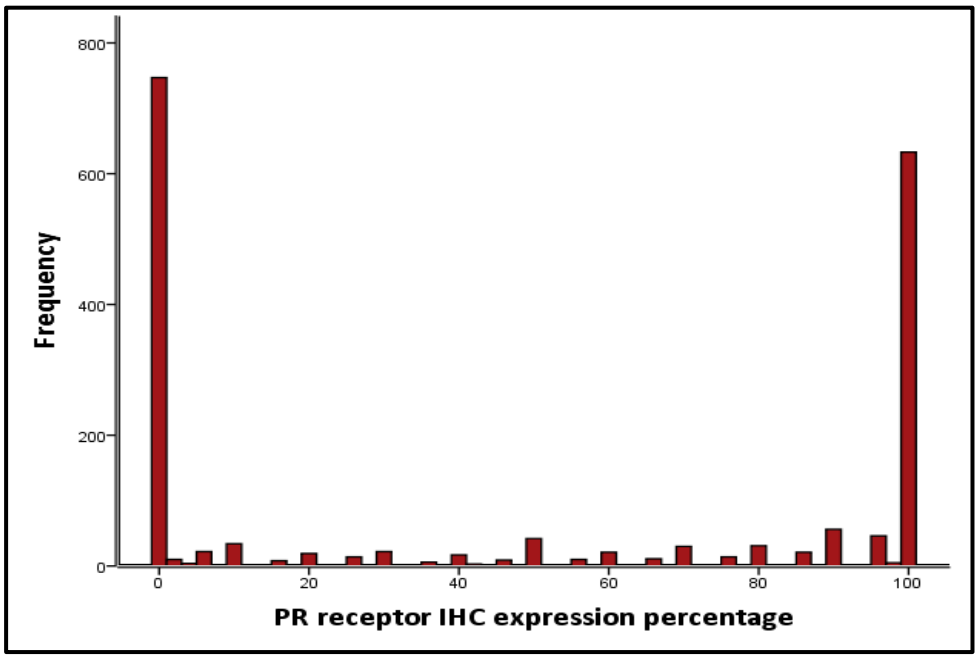


Figure-51: The distribution of PR
IHC expression as assessed on the first cohort representing the bimodal distribution similar to ER expression (Muftah et al., 2016)

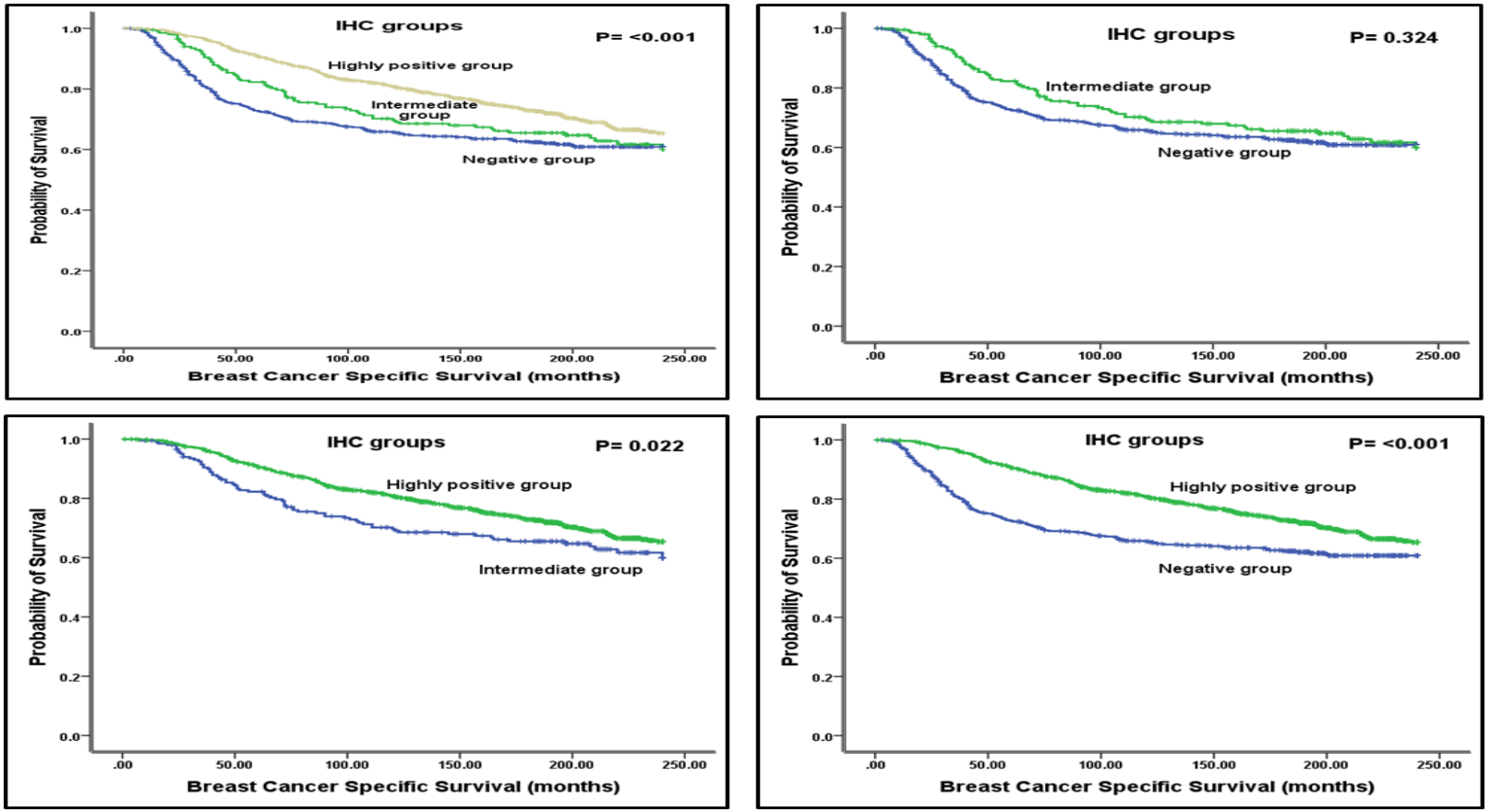


Figure-52: Kaplan Meier plots illustrating BCSS and ER immunostaining subgroups (ER expression negative, intermediate and highly positive groups)

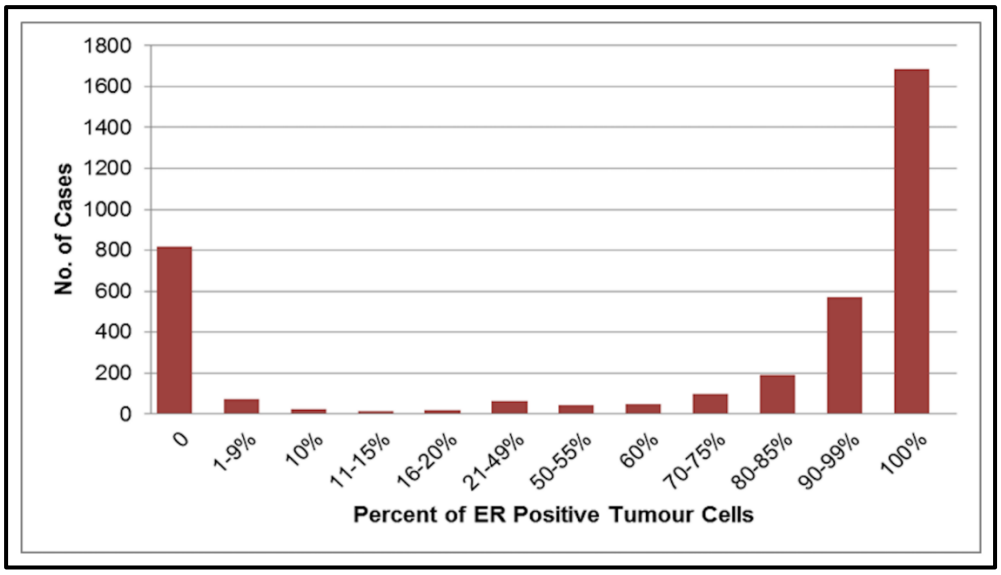


Figure-53: The distribution of ER IHC expression as assessed on CNB (3649 cases) demonstrating the bimodal distribution (Muftah et al., 2016)

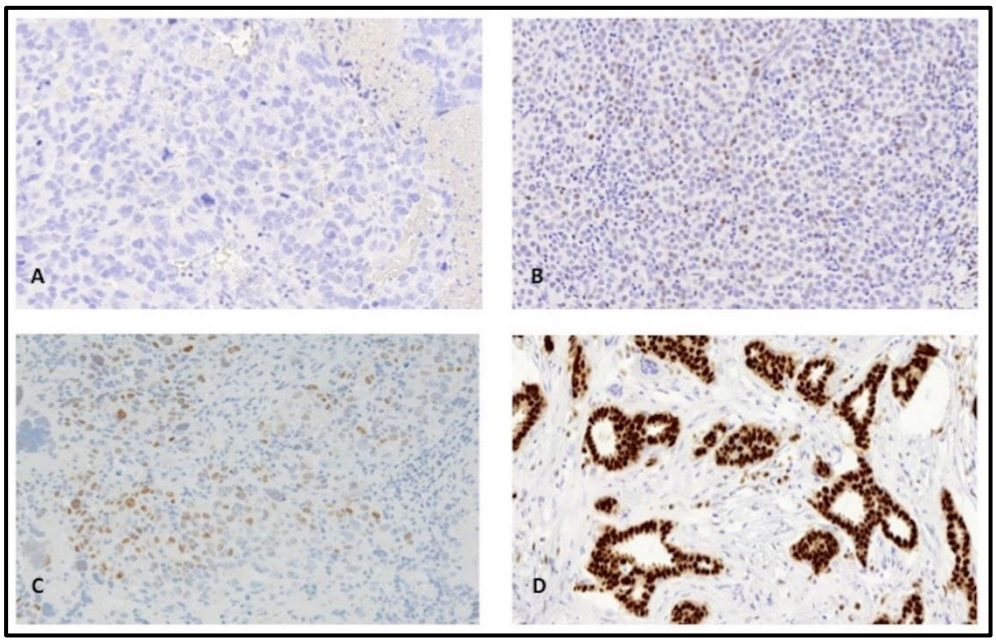


Figure-54: Illustrated different levels of ER expression (A) negative ER immunostaining, (B) Low expression of ER, (C) Intermediate expression of ER and (D) Highly positive expression of ER immunostaining (Muftah et al., 2016)

In the third series, the frequency distribution of ER staining, based on the percentage of ER-positive tumour cells, in the TMA series of 1686 cases is shown in **Figure-55**. Analogous to the previous series, this series demonstrates bimodality of expression with the strongly positive and completely negative cases representing 90.5% of the cases. Weak and intermediate expression of ER was infrequent representing 1.9% and 7.6% of cases, respectively.

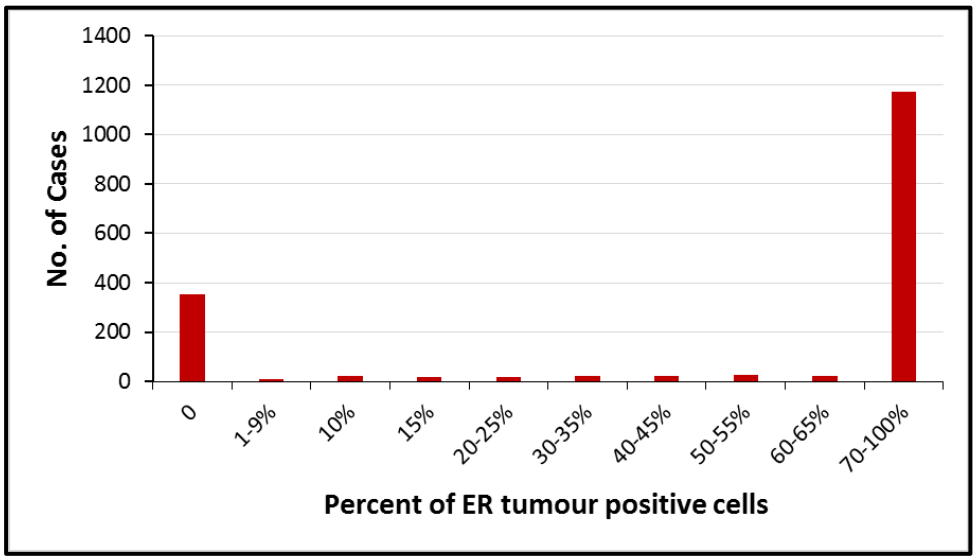


Figure-55: The frequency distribution of ER IHC staining scores in the third cohort (1686 cases) assessed on TMA showing bimodal distribution

5.6.1.2. Transcriptomic level assessment

Consistent with protein level results (IHC), the bimodal distribution of *ESR1* mRNA was observed with the analysis of the METABRIC cases (n=1980) (**Figure-56**). The group with ER-negative expression represented 18.2 % of cases. A total of 72 % cases showed high ER expression whilst the ER intermediate group contained only 9.8% of cases. As mentioned earlier, in this cohort, the survival data and the ER immunostaining were available for 262 cases and these cases were subdivided into three subgroups; ER-negative expression (group-1), ER intermediate group (group-2) and high ER expression (group-3). In this subset, there was a significant positive correlation between the protein expression levels and *ESR1* mRNA (p-value <0.001) (**Table-26**). The majority (79%) of ER IHC negative cases were in the negative *ESR1*

transcriptomic group. Furthermore, 97.1% of the strongly ER IHC positive group belonged to the mRNA highly positive group. Outcome analysis presented significant differences between the groups (**Figure-57**; *p*-value =0.001 and LR =13.28). There was no significant difference between the negative group and the intermediate group (*p*-value =0.74 and LR =0.10). However, there was a significant difference between the highly positive group and the intermediate group regarding the patient outcome (*p*-value =0.01 and LR =6.52).

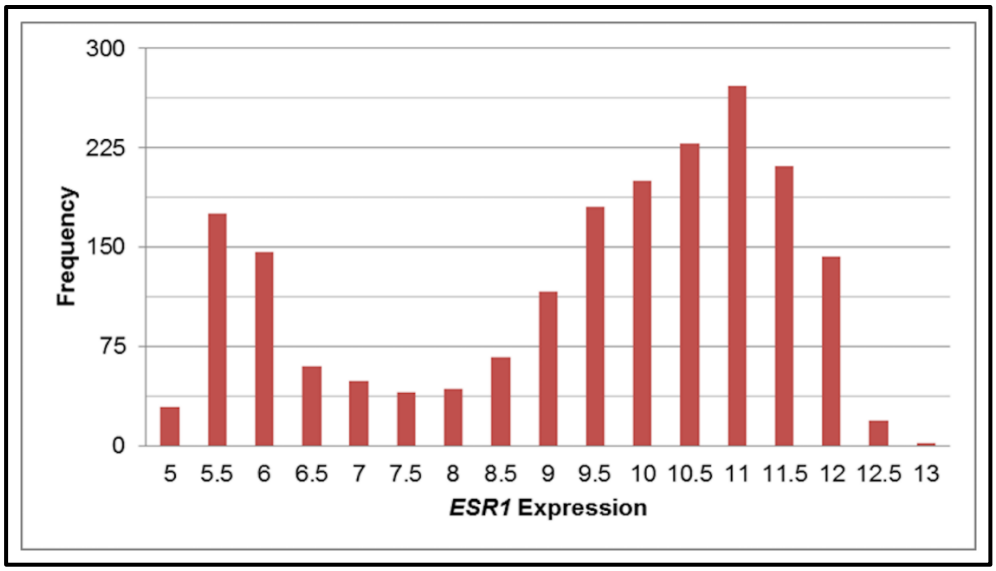


Figure-56: Bimodal distribution of *ESR1* gene (Muftah et al., 2016)

Table-26: Cross-tabulation of immunostaining ER expression groups and *ESR1* transcriptomic groups (Muftah et al., 2016)

Variable	*ESR1* Negative No (%)	*ESR1* Intermediate No (%)	*ESR1* Strong	*p*-value
ER (IHC) Negative (<1%)	49 (79.0)	11 (17.7)	2 (3.2)	
Intermediate (1-69%)	1 (3.7)	6 (22.2)	20 (74.1)	**<0.001**
Strongly positive (≥70%)	4 (2.3)	1 (0.6)	168 (97.1)	

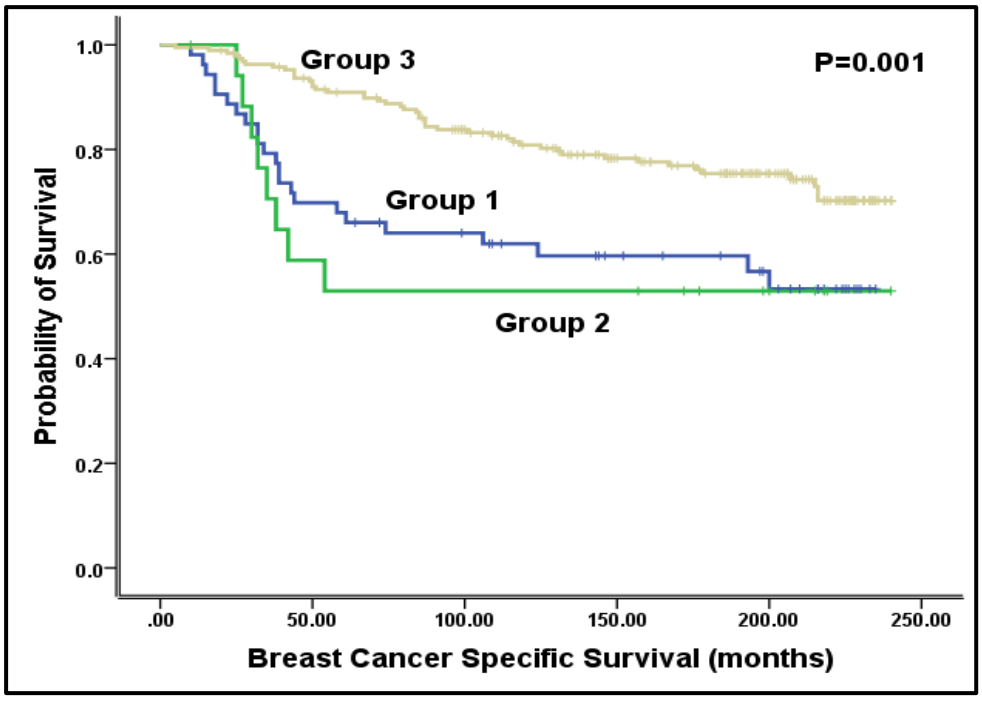


Figure-57: Kaplan Meier plots demonstrating of BC specific survival (BCSS) and transcriptomic subgroups (group 1 is a negative group, group 2 is an intermediate group and group 3 is a highly positive group) (Muftah et al., 2016)

5.6.2. Evaluation of clinical and molecular features of low ER group

A subset of 71 invasive BC cases including 56 cases with low/intermediate ER expression and 15 cases as positive and negative controls were analysed.

5.6.2.1. IHC results

In this study, the agreement between CNB and WTS scoring and between the manual and the automatic IHC staining was assessed. In addition, the evaluation of the relation to important clinincopathological parameters, the percentage of cases shifted to negative ER expression and the lower cut-off at which these cases still express ER was evaluated. The initial results showed low correlation between ER expression in the CNB and WTS using different platforms in this subgroup. The Spearman's correlations between CNB and both manual and automated staining WTS were 0.491 and 0.460 respectively. However, the

correlation between both manual and automated staining was perfect; 0.980. **Figure-58** shows scatterplot diagrams including r^2 values to demonstrate the relationship between CNB, automated and manual staining. R^2 between CNB and both manual and automated staining were 0.097 and 0.099 respectively while between both manual and automated was 0.991. In details, for the group scored <10% (low ER), correlation between CNB and both manual and automated staining WTS were considered low (0.325 and 0.317 respectively). R^2 between CNB and both manual and automated staining were 0.010 and 0.011 respectively (**Figure-59**). However, for the intermediate group (≥10%), the correlation was slightly higher but still fell within the category of low correlation. Correlation between CNB and both manual and automated staining WTS were considered low (0.391 and 0.402 respectively). R^2 between CNB and both manual and automated staining were 0.033 and 0.029 respectively (**Figure-60**).

In the low ER cases (cases scored <10%), out of 33 cases, 19 cases (57.6%) remained in the low category including 4 cases which were scored 1% (at CNB) (**Table-27**). The rest of the cases 14 (42.4%) resulted in negative ER expression on WTS. Different cut-offs are demonstrated in **Table-28**. In the intermediate group, most cases resulted in positive ER expression on excision. The details of clinicopathological parameters for a low/intermediate ER expression on WTS was illustrated in **Table-29** and for the low ER only **Table-30**. Regarding the evaluation of PR, re-staining of these cases using WTS resulted in 75% of the low/intermediate cases being PR-negative and which increased when only low ER cases were analysed, 82% of the cases (ER expression <10%) were negative for PR expression. The comparison between WTS in both ER and PR expression showed that all the negative ER expression cases were negative for PR except three cases; their PR expression being between 1% and 5%.

To validate the IHC results, WTS were further evaluated using RT-PCR and *in situ* RNAscope.

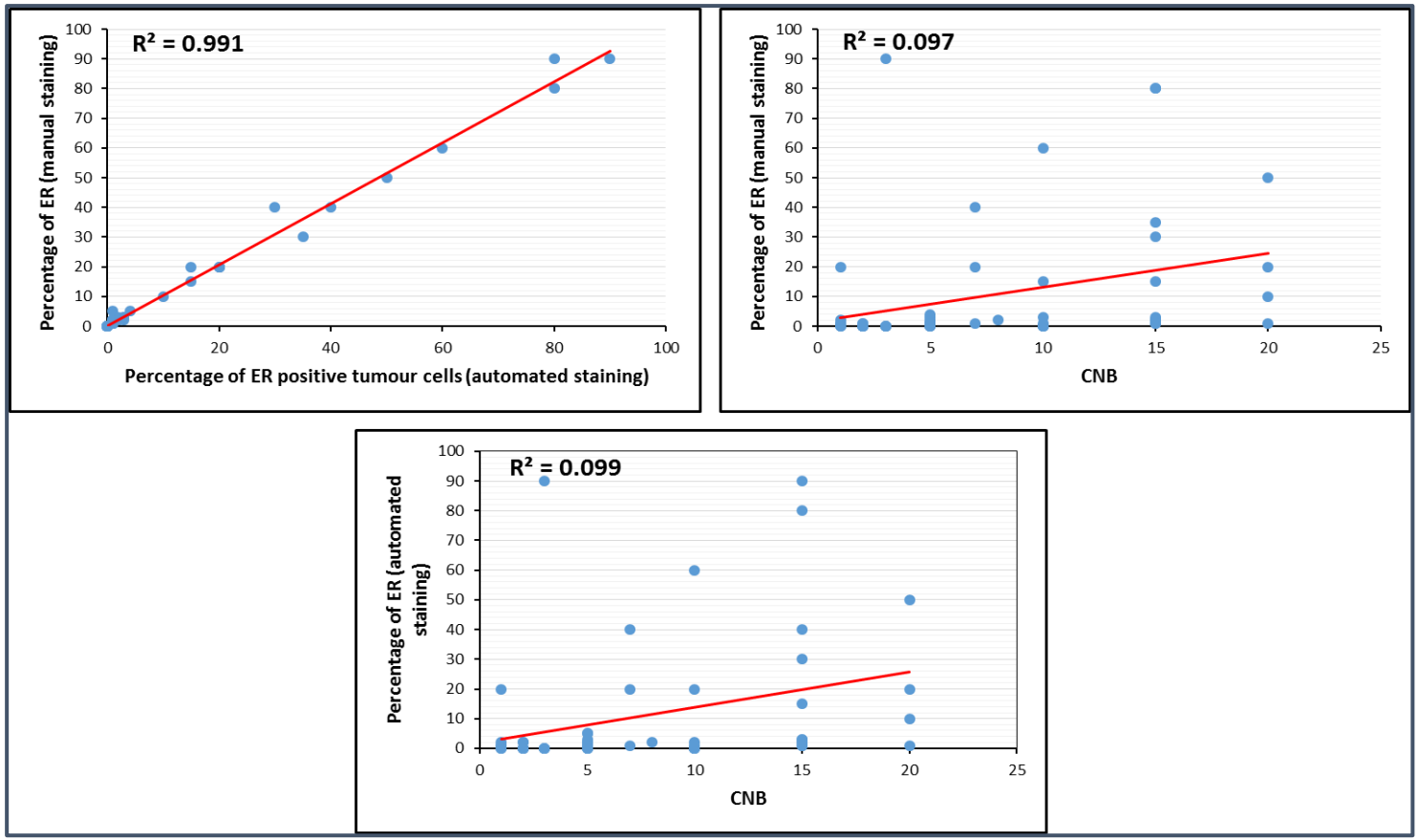


Figure-58: Scatterplot diagrams demonstrate the relation between CNB , automated and manual staining of ER immunostaining expression in all cases

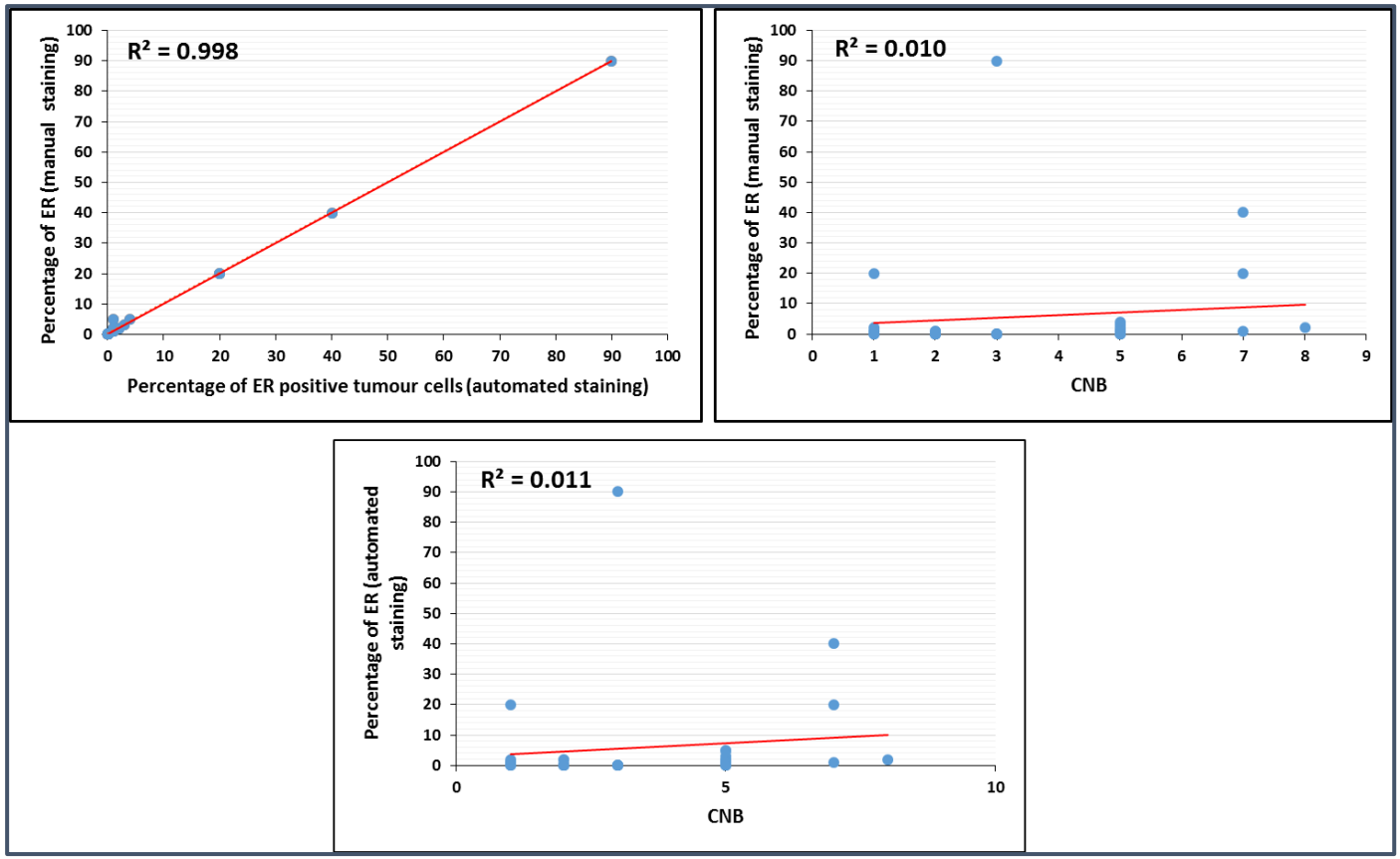


Figure-59: Scatterplot diagrams to demonstrate the relation between CNB , automated and manual of ER immunostaining expression in cases <10%

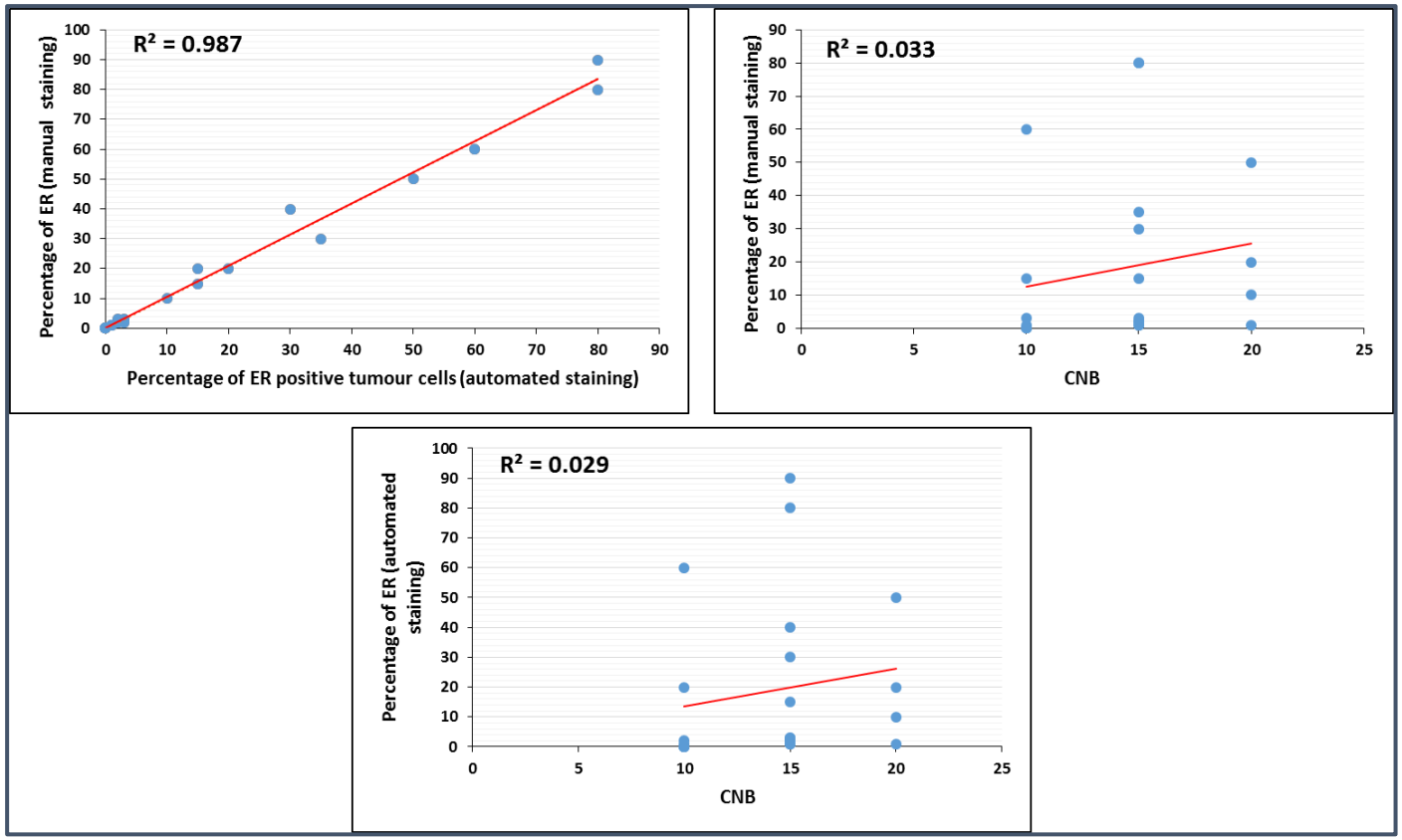


Figure-60: Shows scatterplot diagrams to demonstrate the relation between CNB automated and manual staining of ER immunostaining exoression in cases between 10% and 20%.

Table-27: ER scoring of CNB cases that remain positive on ER immunostaining of WTS

ER Scoring of CNB cases (%)	Total number	Cases positive on WTS (number)	% of expression
1%	6	4	1%, 2% and 20%
2%	7	2	2%
3%	5	1	90%
4%	0	0	/
5%	11	8	1%, 2%,3% and 5%
6%	0	0	
7%	3	3	1%, 20% and 40%
8%	1	1	2%
9%	0	0	/
10%	8	4	1%, 2%, 20% and 60%
Total		23 cases	

Table-28: Comparison between ER expression on CNB and WTS using IHC

CNB(IHC) categories	CNB(IHC) No (%)	WTS(IHC) No (%)	
		Negative	Positive
Positive ER (low/intermediate)	56(100)	18(32.1)	38(67.9)
Cases scored <10%	33(59)	14(42.4)	19(57.6)
Cases scored ≤10%	41(71.9)	18(43.9)	23(56.1)
Cases scored ≥10%	23(41)	4(17.4)	19(82.6)
Cases scored 10% only	8(14.2)	4(22.2)	4(10.5)
Cases scored >10%	15(26.7)	0(0)	15(100)

Table-29: The details of clinicopathological parameters of the low/intermediate ER expression, WTS cases

Variable	WTS (ER expression using IHC) No (%)		*p*-value
	Negative	Positive	
Age			
≤ 50	9(50.0)	14(36.8)	0.208
> 50	9(50.0)	24(63.2)	
Tumour Size (cm)			
≤ 2.0	8(44.4)	12(31.6)	0.348
> 2.0	10(55.6)	26(68.4)	
Pleomorphism			
1	0(0.0)	0(0.0)	0.516
2	3(16.7)	4(10.5)	
3	15(83.8)	34(89.5)	
Mitosis			
1	5(27.8)	7(18.4)	0.689
2	4(22.2)	8(21.1)	
3	19(50.0)	23(60.5)	
Tubular			
1	0(0.0)	2(5.3)	0.371
2	2(11.1)	8 (21.1)	
3	16(88.9)	28(73.7)	
Grade			
1	1(5.6)	2(5.3)	0.836
2	3 (16.7)	9(23.7)	
3	14(77.8)	27(71.1)	
LN status			
Negative	13(72.2)	24(66.7)	0.679
Positive	5(27.8)	12(33.3)	
Stage			
1	13(72.2)	24(66.7)	0.897
2	4(22.2)	9(25.0)	
3	1(5.6)	3(8.3)	

Table-30: The details of clinicopathological parameters of the low ER expression, WTS cases (<10%)

Variable	WTS (ER expression using IHC) No (%)		*p*-value
	Negative	Positive	
Age			
≤ 50	7(50.0)	6(31.6)	0.284
> 50	7(50.0)	13(68.4)	
Tumour Size (cm)			
≤ 2.0	7(50.0)	7(36.8)	0.450
> 2.0	7(50.0)	12(63.2)	
Pleomorphism			
1	0(0.0)	0(0.0)	0.160
2	3(21.4)	1(5.3)	
3	11(78.6)	18(94.7)	
Mitosis			
1	5(35.7)	2(10.5)	0.080
2	4(28.6)	3(15.8)	
3	5(35.7)	14(73.7)	
Tubular			
1	0(0.0)	1(5.3)	0.317
2	2(14.3)	6(31.6)	
3	12(85.7)	12 (63.2)	
Grade			
1	1(7.1)	1(5.3)	0.883
2	3(21.4)	3(15.8)	
3	10(71.4)	15(78.9)	
LN status			
Negative	10(71.4)	12(63.2)	0.618
Positive	4(28.6)	7(36.8)	
Stage			
1	10(71.4)	12(63.2)	0.877
2	3(21.4)	5(26.3)	
3	1(7.1)	2(10.5)	

5.6.2.2. RT-PCR results

Comparison between two protocols for RNA extraction from FFPE tissue sections

The analysis of the comparison between two methods of RNA extraction from FFPE using different kits, BiOstic FFPE tissue RNA isolation kit and RNeasy FFPE Kit (Qiagen) has been done. RNeasy FFPE Kit (Qiagen) showed higher RNA quantity and quality and it was chosen in this study. **Table-31** show the results between the three different samples. The RNA yield using RNeasy FFPE Kit (Qiagen) was higher as illustrated in **Figure-61**. Furthermore, the 260/280 ratio which is used to evaluate the purity of RNA was ≥2.0 when RNeasy FFPE Kit (Qiagen) was used while it was <2 with the other kit as shown in **Figure-62**.

Table-31: The results between the three different samples using different kits

	ng/µl			260/280 ratio		
	Sample (1)	Sample (2)	Sample (3)	Sample (1)	Sample (2)	Sample (3)
RNeasy FFPE Kit (Qiagen)	29.0	194.9	976.8	2.58	2.05	2.05
BiOstic FFPE tissue RNA isolation kit	25.0	35.6	287.4	1.52	1.86	1.89

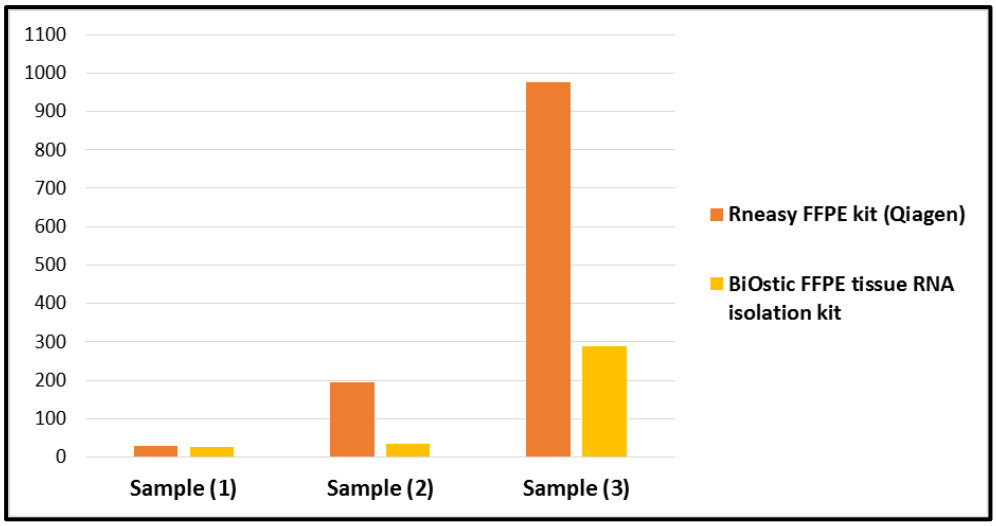


Figure-61: The quantity of RNA yield using different kits
RNeasy FFPE Qiagen Kit (orange) and BiOstic FFPE tissue RNA isolation kit (yellow)

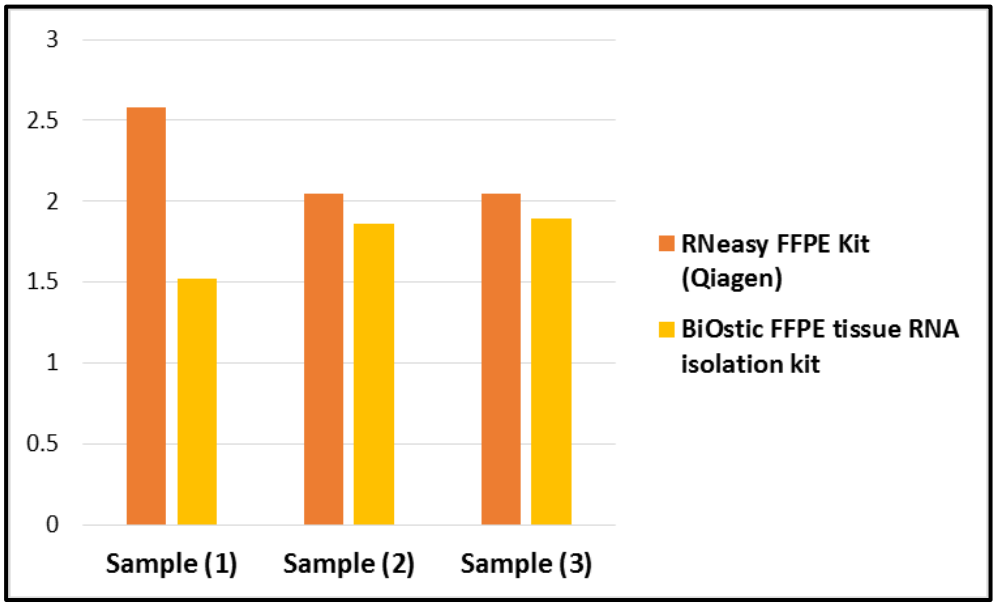


Figure-62: Evaluation of RNA purity using 260/280 ratio
RNeasy FFPE Qiagen Kit (orange) and BiOstic FFPE tissue RNA isolation kit (yellow)

Optimisation of primers

For testing the specificity, initially the sequences were tested using a Basic Local Alignment Search Tool (BLAST) on the NCBI site search for specificity. Additionally, SNP check and Insilico PCR web sites were used. Furthermore, the PCR primers was generated a single amplicon of the correct size for ERα, PR-A, PR-B and HPRT (**Figure-63**). In addition, **Figure-64** shows the melting curves for each PCR product showing a single peak at the correct melting temperate. These melting curves compared for validation with the melting curves produced by uMelt-multing curve prediction software.

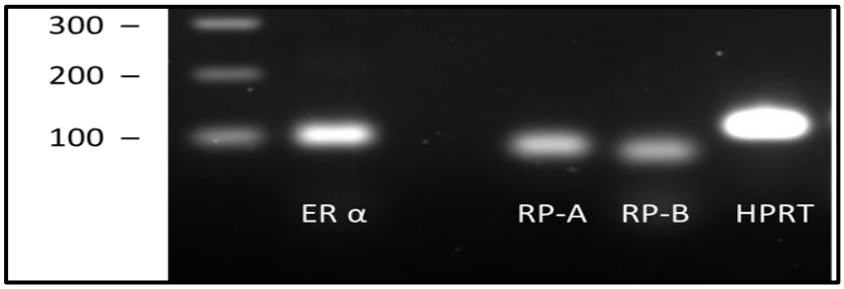


Figure-63: ERα, PR-A, PR-B and HPRT primers produce band PCR amplicon at the correct size

Evaluation of the samples

Using a previous set of BC cases (n=71), the ten negative and five highly positive controls cases showed consistent results between the two techniques; IHC and RT-PCR using WTS. Regarding the 56 cases of low/intermediate cases, the consistency between immunostaining ER expression and *ESR1* expression was 80% in both negative and positive cases. While 76% of cases positive in immunostaining using WTS were positive in *ESR1* expression, 89% of cases negative in immunostaining were negative in *ESR1* expression. The details of clinicopathological parameters has shown in **Table-32**. Furthermore, in general the consistency of *ESR1* and *PR*-B genes was 55%, where the percentage of the cases positive in both is 58% while the negative consistency is 52%. However, the consistency using *PR-A* gene was slightly higher; 62% in both negative and positive cases.

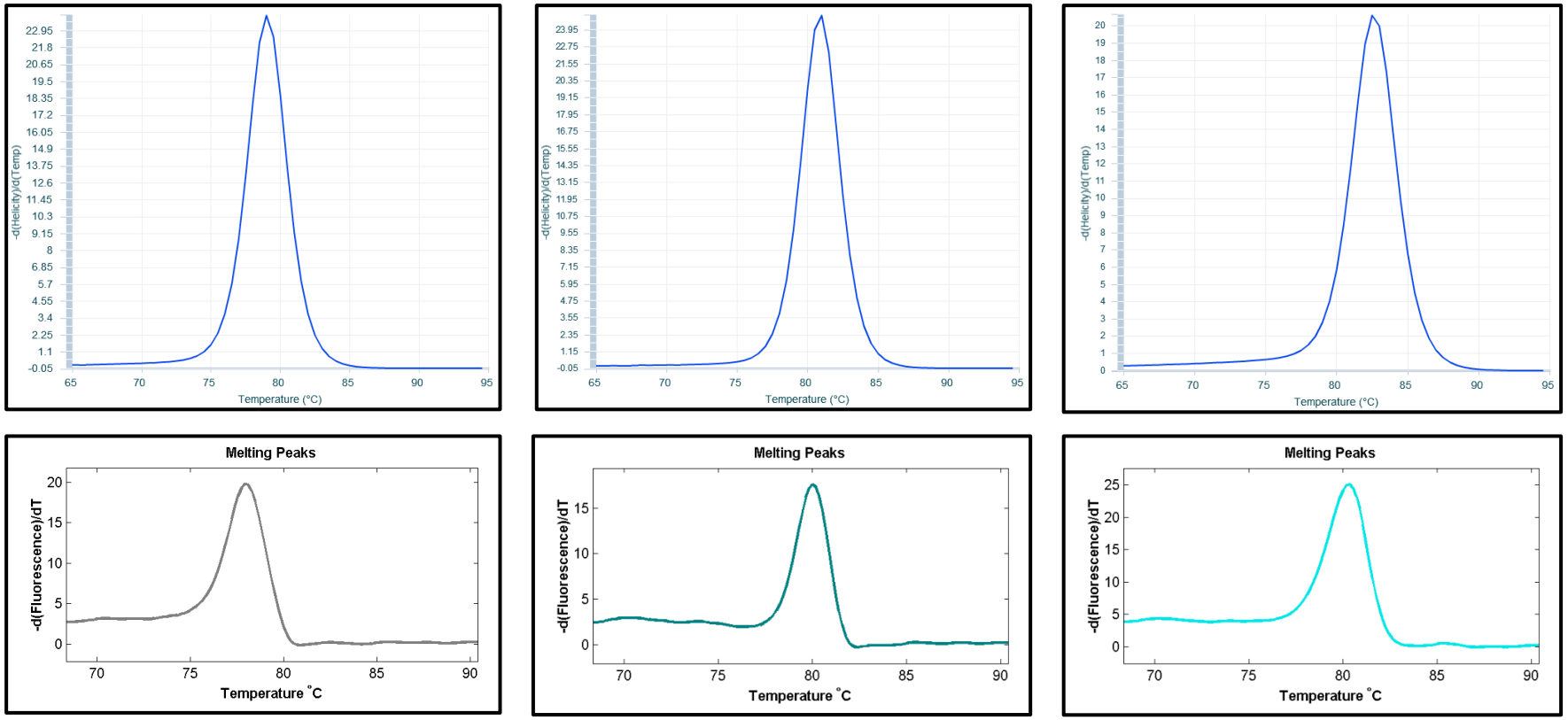


Figure-64: Melting curves for each primer contain a single peak
These melting curves (lower row) compared for validation with the melting curves produced by uMelt-multing curve prediction software (upper row)

Table-32: The details of clinicopathological parameters of the low/intermediate ER expression cases using RT-PCR

Variable	WTS (ER expression using RT-PCR) No (%)		*p*-value
	Negative	Positive	
Age			
≤ 50	12(48.0)	11(35.5)	0.344
> 50	13(52.0)	20(64.5)	
Tumour Size (cm)			
≤ 2.0	9(36.0)	11(35.5)	0.968
> 2.0	16(64.0)	20(64.5)	
Pleomorphism			
1	0(0.0)	0(0.0)	0.477
2	4(16.0)	3(9.7)	
3	21(84.0)	28(90.3)	
Mitosis			
1	5(20.0)	7(22.6)	0.927
2	5(20.0)	7(22.6)	
3	15(60.0)	17(54.8)	
Tubular			
1	1(4.0)	1(3.2)	0.224
2	2(8.0)	8(25.8)	
3	22(88.0)	22(71.0)	
Grade			
1	2(8.0)	1(3.2)	0.253
2	3(12.0)	9(29.0)	
3	20(80.0)	21(67.7)	
LN status			
Negative	18(72.0)	19(65.5)	0.609
Positive	7(28.0)	10(34.5)	
Stage			
1	18(72.0)	19(65.5)	0.666
2	6(24.0)	7(24.1)	
3	1(4.0)	3(10.3)	

5.6.2.3. RNAscope results

Using a previous set of cases, the negative ER expression control cases showed no RNA expression, in contrast the highly positive cases scores by IHC showed high expression of *ESR1*. For the low/intermediate cases (56 cases), there was 98% concordance between ER immunostaining and RNA *in situ* expression of *ERS1*. Only one case showed mRNA expression while being negative at protein level. Moreover, in general the consistency of *ESR1* and *PGR* genes using RNAscope was 53%, where the percentage of the cases positive in both was 36% while the negative consistency was 94%. **Table-33** showes the details of clinicopathological parameters for RNAscope in all cases.

5.6.3. Comparison between the three different techniques

The ER reference cases that were negative using IHC, were re-evaluated using RT-PCR and RNAscope and all were negative for ER expression. For confirmation, the same cases were also examined for the ER related gene PR, and were also negative for PR expression.

Noticeably, all the ER-positive cases >10% were positive (100%) using different clones and IHC methods (manual vs automatic staining and EP1 vs SP1 clones). Remarkably, all these cases were also positive using RNAscope (100%). However, in RT-PCR results only 2 cases (13.3%) were negative. Interestingly, these cases (negative in RT-PCR) had small well defined area of positivity using IHC. Using RNAscope, these small areas were very consistent with the IHC staining (**Figure-65**). One of the explanation is that when there is a bulk of negative tumour this can dilute the positivity of the sample in RT-PCR the positive area is very small compare with the bulk of the tumour. Regarding the remaining 41 cases, 16 were ER-positive using the three different techniques while 17 cases where ER-negative in the three different methods. Therefore, only 8 cases were mismatched between methodology where 7 cases were matched

using IHC and RNAscope (positive using both techniques). However, these showed negative for ER using RT-PCR. On subsequent review, these cases had focal weak staining ER expression by IHC and RNAscope. There was only one case that was negative for ER using immunostaining while positive at RNA level.

Noticeably, the majority of the negative and the low ER expression cases on excision specimens in CNB, even the cases that in early stage with negative lymph node status, were categorised as high grade tumour with high pleomorphic appearance, mitosis, tumour size and low tubular formation. Additionally, a remarkable finding was that all the low ER-positive cases had low staining intensity (**Figure-66**). The majority of cases reflect a heterogeneity for ER staining with focal weak staining.

5.7. Discussion

The concept of bimodality and linear distribution of ER expression in BC still a subject of controversy. This study presents evidence supporting the bimodality distribution of ER expression at both protein and gene expression levels using IHC and gene expression microarrays technology respectively. At the protein level, our results are in line with the results revealed by *Collins* et al., in which 99% of 817 BC cases were totally positive or negative (Collins et al., 2005) and with *Nadji* et al. who reported that the immunostaied ER expression in BC is an all-or-none phenomenon (Nadji et al., 2005). In agreement with our findings, various other researchers have highlighted the bimodality of ER immunostaining expression including *Badve* et al. (776 cases) (Badve et al., 2008), *Bogina* et al. (362 cases) (Bogina et al., 2012), *Khoshnoud* et al. (683 cases) (Khoshnoud et al., 2011) and *Pinto* et al. (360 cases) (Pinto et al., 2013).

Table-33: The details of clinicopathological parameters of the low/intermediate ER expression cases using RNAscope

Variable	WTS (ER expression using RNAscope) No (%)		*p*-value
	Negative	Positive	
Age			
≤ 50	9(52.9)	14(35.9)	0.233
> 50	8(47.1)	25(64.1)	
Tumour Size (cm)			
≤ 2.0	8(47.1)	12(30.8)	0.242
> 2.0	9(52.9)	27(69.2)	
Pleomorphism			
1	0(0.0)	0(0.0)	
2	3(17.6)	4(10.3)	0.442
3	14(82.4)	35(89.7)	
Mitosis			
1	5(29.4)	7(17.9)	0.546
2	4(23.5)	8(20.5)	
3	8(47.1)	24(61.5)	
Tubular			
1	0(0.0)	2(5.1)	0.167
2	1(5.9)	9(23.1)	
3	16(94.1)	28(71.8)	
Grade			
1	1(5.9)	2(5.1)	
2	3(17.6)	9(23.1)	0.900
3	13(76.5)	28(71.8)	
LN status			
Negative	12(70.6)	25(67.6)	0.824
Positive	5(29.4)	12(32.4)	
Stage			
1	12(70.6)	25(67.6)	
2	4(23.5)	9(24.3)	0.953
3	1(5.9)	3(8.1)	

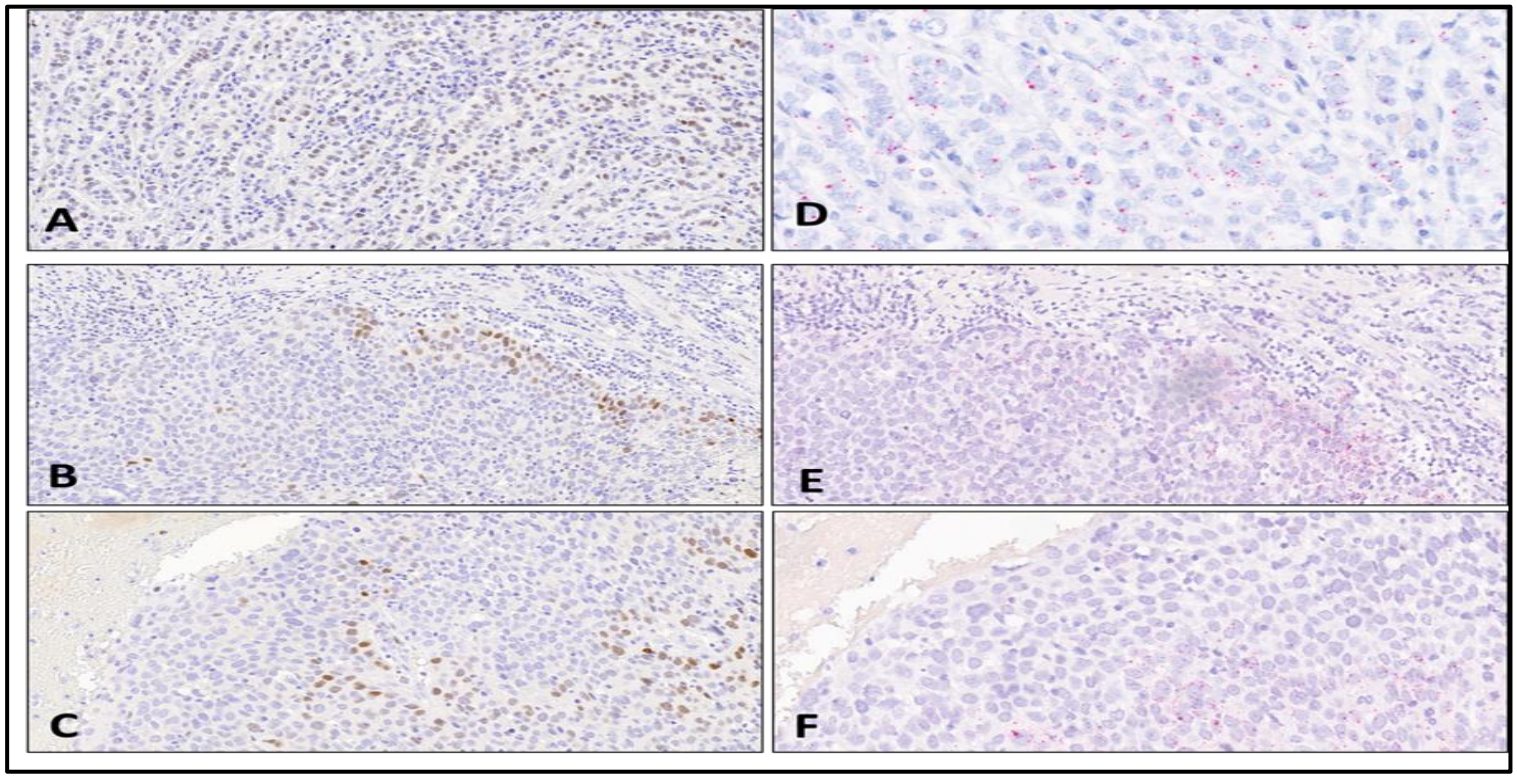


Figure-65:The image shows focal weak IHC staining of breast cancer cases (A, B and C) consistent with RNAscope staining results (D, E and F)

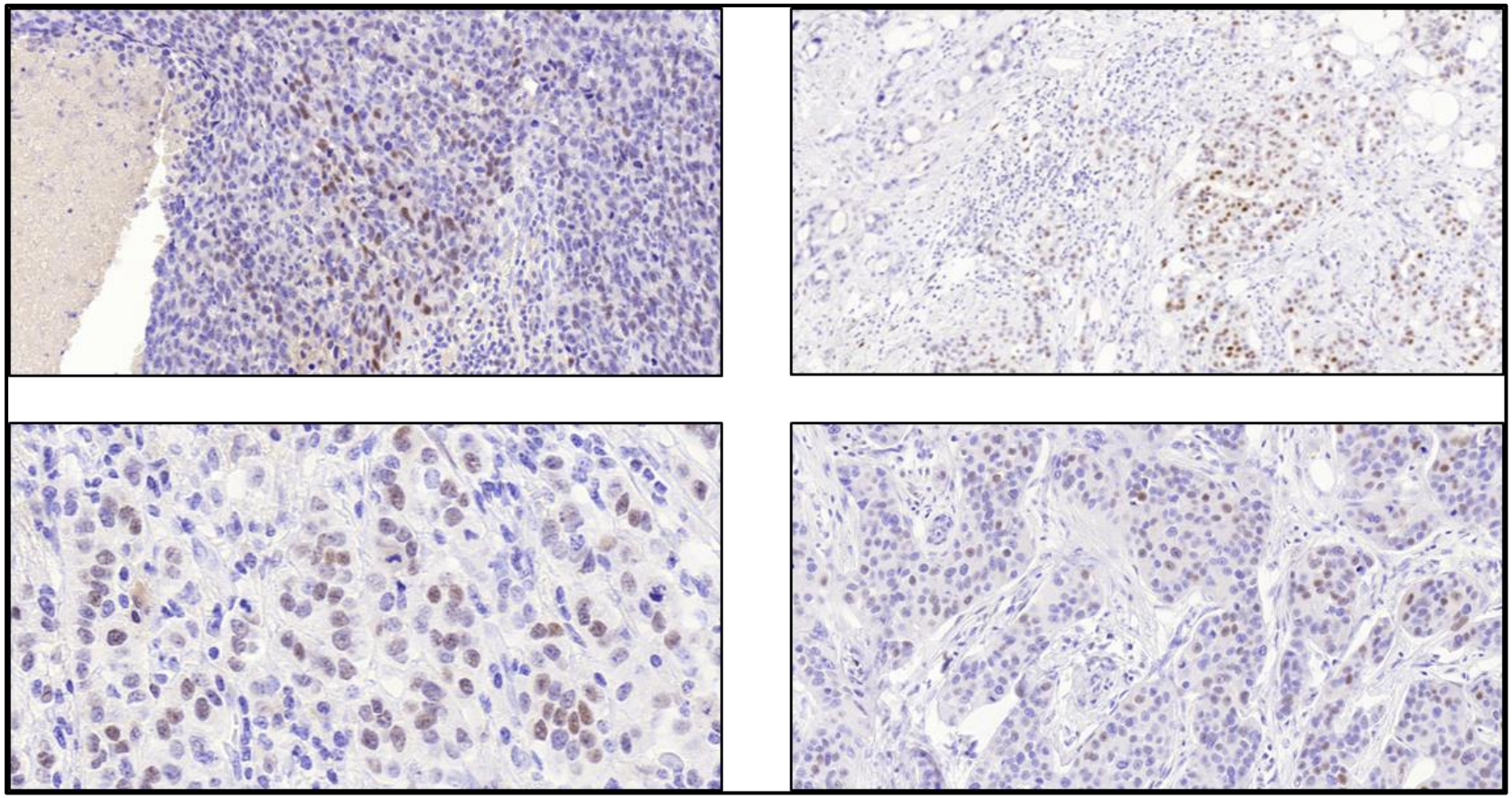


Figure-66: Low ER-positive cases have low staining of breast cancer tissue

The *Schnitt* group provided comments on the reproducibility of the linear distribution of ER expression seen using LBA when a large population-based study was evaluated using IHC (Nadji et al., 2005). They described the association between the apparent amount of ER showed by IHC and the real quantity of ER protein in the malignant cell nuclei as a highly complex process and linked this to the effect of the pre-analytic process on the IHC results (Schnitt, 2006). The pre-analytical factors, such as fixation and processing of tissue and the overall sensitivity of IHC assay were deemed as the main explanation for the bimodal distribution of the ER expression. It has been shown that IHC can vary according to a pre-analytical factors as tissue fixation (Werner et al., 2000, Goldstein et al., 2003) and antigen retrieval. *Umemura* et al studied a positive association regarding a linear correlation between the LBA and a non-highly sensitivity IHC assays. This relationship was decreased using a highly sensitive IHC test with nonlinear association and the biochemical assays resulted in a shift of the low ER cases toward the ER positivity (Umemura et al., 2004). This reason was used to explain the low frequency of the low ER cases. Interestingly, although the assays and the testing level utilised to examine ER distribution is different, in the current study when 1980 cases of METABRIC BC data set was tested, the similarity between the transcriptome and protein levels was considerably high. Both mRNA and IHC levels utilising large data sets revealed the bimodal distribution. Therefore, the explanation that the bimodality of ER expression is a result of the highly sensitive IHC test or the pre-analytical factors has to be re-considered. The bimodality of gene expression for *ESR1* using 123 patients was previously demonstrated by *Wang* et al.(Wang et al., 2009). Their approach was applied to the MDA133 BC microarray dataset (Hess et al., 2006).

This study reflected strong correlation between ER IHC protein level and *ESR*1 mRNA level as examined utilising gene expression microarrays. Interestingly, a high proportion of the IHC intermediate group linked to a strong *ESR1* expresser group. However, this can be explained by using of different cut-offs in each cohort or may be due to differential translation of ER in some cases. In this study, the use of 70% as a dichotomised point of intermediate group may have caused

underestimation of the level of ER positivity defined utilising *ESR1* mRNA. Furthermore, HER2 overexpression was more frequent in the low/ intermediate ER expresser groups compared to the strong ER expresser group. The high level of HER2 overexpression in these cases may support their aggressive tumour behaviour and subsequently their response to treatment may be akin to ER-negative group instead of the strong ER-positive tumours. At the transcriptome and protein levels, our results revealed no significant difference between the intermediate expression groups and the ER-negative and outcome. Therefore, our results raise the clinical question whether patients with low positivity of ER expression are actually ER-positive tumours or related to the ER-negative cases. In this study, approximately 32.1% and 42.4% of the low/intermediate group and only low IHC ER-positive group respectively changed to ER-negative when WTS of excision specimens were immunostained. This may reflect a proportion of cases with false positive ER expression on core biopsies. Rakha et al. experienced some difficulties with the interpretation of 6F11 clone staining patterns and they observed weak granular nuclear staining on CNBs that gave false-positive results. Further, they mentioned that previous studies have demonstrated the sensitivity of detection of ER expression in relation to the quality of fixation, antigen retrieval methods and the clone of antibody used (Rakha et al., 2012). Moreover, the heterogeneity of BC and the presence of the positive ER expression clone in the large ER negative tumour could be another explanation for the cases that positive in CNB and become negative in the excision biopsy. The percentage of the weak positive cases in the first, second and third cohorts is 1.6%, 2.7% and 1.9 respectively, which is as low as what is previously reported by Iwamoto et al. (Iwamoto et al., 2012) and Yi et al.(Yi et al., 2014), 6% and 2.6% respectively.

The cut-off point for ER still a matter of debate previously was 10% and the current one according to ASCO AND ACP is 1% (Harvey et al., 1999) (Hammond et al., 2010a). However, the presence of studies support the poor correlation of the low ER subgroup to the luminal/ ER-positive category raised the question about the current cut-off (Sheffield et al., 2016). In our study, 42.4% of the low ER cases where negative in WTS staining. Moreover, the minor

degree of difference between IHC and RT-PCR in cases scored more than 10% can be explained with that RNA signals from the focal ER-positive cells could be masked by the bulk of the ER-negative tumour cell population or it could be due to the defect in translation of the RNA to protein. This explanation was used previously by Iwamoto et al. when they explained the interpretation of the low ER gene signature expression in weakly ER positive cases by that it may be due to a false-positive weak ER immunostaining or the presence of a truly ER-positive small subpopulation of cells within a larger ER- negative tumour. Then they added, in the latter, mRNA signals from the ER- positive cells could be covered by the bulk of the ER-negative cancer cell group and these two scenarios cannot be differentiated by their analytic methods (Iwamoto et al., 2012). To overcome this issue in our study, we validated the low ER cases using a new RNAscope technique, which has the advantage of possible to map the visible signals to individual cells so these truly positive subpopulation can be detected. Interestingly, all these cases scored positive in transcriptomic level using RNAscope, this may be related to a novel target probe design strategy (a double-Z design) which is used in RNAscope. There is high agreement between RNAscope and IHC, 98% concordance.

Only about 6% of these cases reported as grade I and 93.9% of them categorised between grade II and III, and 75.8% of these cases are high grade tumour. On the other hand, importantly, nearly half of the cases (42.4%) were negative on excision biopsy. These findings may suggest at least to recommend repeating ER staining on excision biopsy as a routine for the low ER cases or further validation using other assay. Identification of ER expressed in BC using IHC combined with advanced molecular technique can be provided the opportunity to establish more specific, sensitive and possibly cost effective ways of identifying a suitable treatment. Although in our study tissue macrodissection had been done, the entrapped DCIS, inflammatory cells and sometimes normal tissue with expression of ER was expected. Therefore, these cases could give positivity in RT-PCR while negative with the other techniques.

In this study, it is observed that the low ER cases expressed weak staining intensity. In consistent with our study, Deyarmin et al. mentioned that out of 26 cases where ER intensity values were available for 17 tumours. The majority 76 % of cases had weak staining tumours (Deyarmin et al., 2013). However, no specific explanation had been mentioned. In the same study, they suggested that at the molecular level, the low ER tumours have to classify as ER-negative rather than ER-positive. Further, they suggested that a more accurate threshold would be to have at least 10 % positive staining cells to define ER positivity which have been approved with our study. Further supporting to our finding was published by Iwamoto et al. they examined 465 cases using IHC and Affymetrix U133A gene chip. The cases of low ER (<10%) were 25 cases only where a minority of these tumours show molecular features similar to those of ER-positive whereas most show ER-negative, basal-like molecular characteristics (Iwamoto et al., 2012).

5.8. Conclusion

ER expression in BC is essentially bimodal with the great majority of cases classified either ER-negative or strongly ER-positive. Furthermore, in our cohort, nearly half of the cases with low ER expression where negative in excision biopsy and most of the cases within high grade tumour category. There is strong agreement between IHC and RNAscope especially in heterogeneous ER staining areas. This study provide scientific evidence that the actual cut-off of ER is seems to be 10%. Additionally, the cases between 1-10% need to repeat on tissue specimen and may need further evaluation using a robust molecular technique.

6. Bmi-1 and a new prognostic subgroups

6.1. Introduction

Increasing the possibility of patient stratification based on prognosis evaluation is an essential step for improving the selection of patients who need additional treatment with new drug targeting. Recent studies are attempting to discover beneficial and reliable prognostic biomarkers for cancer patients with the intention of updating clinical decision-making. Various studies reveal that the expression of Bmi-1 is of clinical significance, and propose that Bmi-1 might be deliberated as a potential prognostic biomarkers in cancer. The special physiological and pathological function of Bmi-1 as a stem cell factor supports its importance in tumour pathogenesis. It has an important role in the organising of stem cells (Lessard and Sauvageau, 2003) (Park et al., 2003). Different mechanisms and pathways contribute to this process: Bmi-1 regulating p16Ink4a/Rb and/or p14ARF/MDM2/p53 tumour suppressor pathways; therefore, it controls the self-renewal, proliferation and cell cycle of CSCs (Jacobs et al., 1999) (Park et al., 2004). Bmi-1 expression can stimulate stem-like properties and epithelial-mesenchymal transition (EMT) (Wu, 2011). In addition, it acts as an oncogene: through suppressing the ink4a locus (containing both the p14ARF and p16Ink4a genes), Bmi-1 with c-Myc was first shown to stimulate lymphomagenesis and to control cell proliferation and senescence (Jacobs et al., 1999).

In clinical research, there is increasing evidence of human carcinomas that expresses a high level of Bmi-1, which reflects that Bmi-1 might have specific roles in cancer initiation and progression. Furthermore, series of studies have examined the association between Bmi-1 and prognosis of different cancers. However, these studies showed inconsistent results as to whether Bmi-1 can be used as a biomarker of favourable or poor prognosis (Shao et al., 2014). Similarly, Bmi-1 has been studied in BC, but the clinical relevance has not been fully elucidated. While some studies indicate that the elevated level of Bmi-1 can be associated with poor prognosis, conversely, other studies support its role

as a good prognostic parameter in BC (Kim et al., 2004) (Arnes et al., 2008) (Choi et al., 2009) (Guo et al., 2011).

6.1.1. Bmi-1 structure and function

Bmi-1 is a common target of the Moloney virus insertion. It was first identified as an oncogene that cooperates with c-myc in the initiation of mouse pre-B lymphoid tumours; for this reason, it was given its name (Haupt et al., 1991) (van Lohuizen et al., 1991). Moreover, Bmi-1 is one of the polycomb-group (PcG) genes, which are epigenetic gene silencers with an essential role in the preservation of cell identity (Raaphorst, 2005). Since its discovery, Bmi-1 has seen to be involved in multiple biological functions involving cell cycle, development and DNA damage response. In last decade, Bmi-1 has been presented to be a key regulator of self-renewal in both normal and BCSCs (Liu et al., 2006), and plays a role in cell immortalisation, and senescence (Siddique and Saleem, 2012). The *BMI1* gene is localised on the short arm of chromosome 10 (10p11.23; the human cDNA is 3203 bp in length). Its protein is consisted of 326 amino acids with molecular weights of 37 kDa (Alkema et al., 1993) (Cao et al., 2011). Bmi-1 protein is mainly characterised by three regions: a conserved RING (Really Interesting New Gene) finger domain at the N-terminal end; a central helix-turn-helix (HTH) domain; and a carboxyl-terminal PEST-like domain at the C-terminal end (**Figure-67**).



Figure-67: Bmi-1 protein structure.
RING – really interesting new gene; HTH – helix-turn-helix; NLS – nuclear localization signal; PEST – motif rich in proline, glutamic acid, serine, and threonine (Bhattacharya et al., 2015)

The RING domain is necessitated for Bmi-1 to localise to DNA strand breaks; thus, it has a crucial role in the cellular response to DNA damage (Ginjala et al., 2011). Both the RING and HTH motifs are required for extending the life span of cells (Itahana et al., 2003). Additionally, the HTH domain is essential in binding Bmi-1 to DNA and offering a scaffold for consequent protein binding (Brennan and Matthews, 1989). PEST domains are polypeptide that have plenty of proline (P), glutamic acid (E), serine (S) and threonine (T) sequences, linked by basic residues that have a tendency to act as proteolytic signals responsible for protein degradation (Rechsteiner and Rogers, 1996). One study noted that the PEST-like domain of Bmi-1 predominantly contained proline and serine residues and deletion of this region caused an increase of Bmi-1 half-life, which can promote cell proliferation and an EMT phenotype in mammary epithelial cells (Yadav et al., 2010). Besides the previous three functional domains, Bmi-1 protein includes two nuclear localisation signals: NLS1 and NLS2. NLS2, of which the latter seems to be fully functional in nuclear localisation of Bmi-1 (Cohen et al., 1996).

Moreover, as a part of the PcG protein group, Bmi-1 has a function with other PcG proteins as an epigenetic repressor, chromatin remodelling through ubiquitination of histone 2A and later, histones methylation (Raaphorst, 2005). There are two types of PcG protein complex: PRC1 and PRC2 (polycomb repressor complexes 1 and 2). Bmi-1 is one of the PRC1 complexes. It reacts with RING1B through its RING domain and promotes the E3 ubiquitin ligase activity resulted in histone H2A ubiquitination (Li et al., 2006). As PcG proteins have a role in transcriptional repression, it is hypothesised that they may be used considerably in stem cell renewal and cancer growth (Sauvageau and Sauvageau, 2010).

Bmi-1 functions in normal development and stem cell conservation have been mainly related to its transcriptional suppression of the Ink4a/Arf locus (Sherr, 2001) (Sharpless, 2004) (Valk-Lingbeek et al., 2004). Different groups have concurrently published on phosphorylation of Bmi-1. Phosphorylated Bmi-1 altered both the dependent and independent Ink4a/Arf functions. Bmi-1 is phosphorylated by 3 pK (MAPKAP kinase 3) and also described to be

phosphorylated by PI3K/ Akt pathway. Bmi-1 and Akt are co-activated in a part of high grade cancers. Akt mediated Bmi-1 phosphorylation, increased its oncogenic potential in an Ink4a/Arf independent pathway. This process also controlled the DNA damage response and affected genomic stability (Nacerddine et al., 2012). Liu et al. revealed that phosphorylation of Bmi-1 by Akt might also be through the Ink/Arf pathway. Phosphorylation of Bmi-1 at Ser316 by Akt decreases its activity by triggering detachment from the Ink4a/Arf locus, which produces decreased ubiquitination of histone H2A and the incapability of Bmi-1 to enhance cellular proliferation and tumour progression (Liu et al., 2012). Bmi-1 modulates the tumour suppressor proteins $p16^{Ink4a}$ and $p14^{Arf}$ (Park et al., 2004). Bmi-1 stimulates CDK4 and CDK6 activity by supressing the Ink4a/Arf locus, which encodes the CDK4/6 inhibitor $p16^{Ink4a}$ and $p19^{Arf}$ that activates p53 (Jacobs et al., 1999). However, Bmi-1 can directly control p53 stability, this is further stressing its role in cellular proliferation and tumorigenesis (Calao et al., 2013). The $p16^{Ink4a}$ protein prevents binding of Cyclin D to CDK4/6 producing suppression of retinoblastoma (Rb) activity and resulting in cell cycle arrest (Jacobs et al., 1999) (Park et al., 2004), p53 induced by $p19^{Arf}$ and causes cell cycle arrest (Molofsky et al., 2005). Bmi-1 stimulates cell proliferation by repressing $p16^{Ink4a}$/Rb and/or $p14^{Arf}$/MDM2/p53 tumour suppressor pathways (**Figure-68**) (Jacobs et al., 1999).

Furthermore, Bmi-1 expression plays a crucial role in invasion and metastasis by modulating the Akt/GSK-3b/Snail pathway and the expression of EMT markers particularly in BC (Morel et al., 2008) (Yang et al., 2010). Bmi-1 has been shown that its overexpression results in activation of human telomerase reverse transcriptase transcription and stimulation of telomerase activity in immortalised mammary epithelial cells (Dimri et al., 2002).

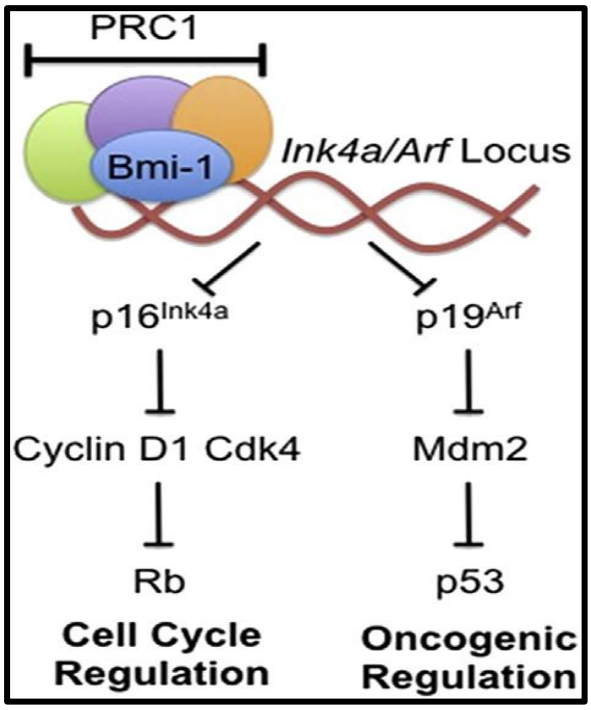


Figure-68: The diagram demonstrating a role of Bmi-1 in Ink4a/Arf signaling Pathway (Bhattacharya et al., 2015)

6.1.2. Bmi-1 and BC

Recently, various studies indicate that the elevated level of Bmi-1 can be associated with poor prognosis in a series of tumour types including BC (Liu et al., 2008b) (Tong et al., 2012) (Wang et al., 2012b) (Yoshikawa et al., 2012). Based on 239 cases with follow-up records, Guo et al. reported that Bmi-1 promotes the invasion and metastasis of BC and predicts poor survival (Guo et al., 2011). Kim et al. concluded that Bmi-1 expression might play a role in the progression and lymph node metastases of BC (Kim et al., 2004).

On the other hand, other studies have published contrary conclusions for different tumour types including BC (Bachmann et al., 2008) (Pietersen et al., 2008) (Choi et al., 2009) (Choy et al., 2012) (Cenci et al., 2012). A study of 960 BC cases revealed that Bmi-1 is significantly correlated with favourable prognosis in BC patients (Choi et al., 2009). Moreover, Shao et al. recently published a large-scale meta-analysis including 40 studies and concluded that Bmi-1 can be considered a prognostic marker with poor survival in Asian

patients with lung, cervical, oesophageal, gastric, colorectal cancer, whereas the prognostic role of Bmi-1 is not strongly evidence-based in BC or hepatocellular carcinomas. However, existing evidence indicates that Bmi-1 can be a positive prognostic marker in BC among Caucasian patients (Shao et al., 2014). Thus, the prognostic value of high Bmi-1 expression remains inconclusive. In this study, the correlation between Bmi-1 and both grade and patient outcome is evaluated at both transcriptome and protein levels in large cohorts of BC.

Although Bmi-1 can individually yield prognostic data in BC, considering its role with other biomarkers, it is possible to classify tumours with promising prognostic importance. CD44 and CD24 are considered the most important stem cell markers in BC. The combination of CD44 and CD24 first allowed Al-Hajj et al. to prospectively isolate a CSC subpopulation from eight of nine BC patients (Al-Hajj et al., 2003). Moreover, Ponti et al. isolated and studied the *in vitro* growth of BC–initiating cells which expressed $CD44^{+}/CD24^{-}$ markers (Ponti et al., 2005). Abraham et al. investigated 136 FFPE tissues for the prevalence of $CD44^{+}/CD24^{-/low}$ tumour cells in BC and they concluded that the occurrence of $CD44^{+}/CD24^{-/low}$ in BC neither be associated with clinical outcome nor survival (Abraham et al., 2005). Another study also found that a $CD44^{+}/ CD24^{-}$ phenotype had no significant impact on patients' prognosis while the phenotype $CD44^{-}/CD24^{+}$ helped identify patients with worse outcome (Mylona et al., 2008). Ahmed et al. demonstrated that the accepted stem cell markers CD44 and CD24 can provide prognostic value in BC individually. Furthermore, it was suggested that when combined together, $CD44^{-}/CD24^{+}$ phenotype had the worst prognosis particularly when compared to tumours with $CD44^{+}/CD24^{-}$phenotype (Ahmed et al., 2012). Different studies have highlighted the expression level of Bmi-1 when CD44/CD24 sub-populations were used and explored different aspect related to BCSCs (Fang et al., 2011) (Barriere et al., 2012) (Yin et al., 2013) (Ma et al., 2015). However, there is a lack of studies that assess the direct correlation between Bmi-1 immunostaining expression and CD44/CD24 sub-populations.

6.1.3. Bmi-1 and Ki67

The expression of both Bmi-1 and Ki67 may correlate to a certain level. However, there are still not many reports on this issue. Lin et al. investigated the correlation and expression significance of Bmi-1 and Ki67 in colorectal carcinoma using IHC in 20 normal colorectal tissues, 30 adenomas and 60 colorectal carcinomas (Lin et al., 2008). The results showed that the expression of both biomarkers were significantly higher in colorectal carcinoma than in normal colorectal tissue and adenomas.

In BC, Arnes et al. investigated the molecular and clinicopathological features of basal-like BC using 190 BC cases and reported that Bmi-1 positivity was correlated with the absence of basal-like features, high ER and low Ki67 expression (Arnes et al., 2008). Moreover, Wang et al. studied the association of ER status with Bmi-1 expression evaluated using IHC in 92 cases of BC. They mention that loss of activity of ER-coupled Bmi-1 pathway leads to an aberrant expression of p16INK4a and, ultimately, a high Ki67 expression (Wang et al., 2014). In a recent study, Gwak et al. evaluated the expression of embryonic stem cell transcription factors, including Bmi-1, in 319 immunostained cases of invasive BC tissues and found that Bmi-1 expression was higher in tumours of low Ki67 expression. Furthermore, the expression of Bmi-1 was inversely correlated with aggressive features of the BC (Gwak et al., 2017).

6.2. Hypothesis

Bmi-1 has clinical value as a stem cell factor in BC and can sub-classify the stem cell $CD44^{-}/CD24^{+}$ phenotype and low expression of Ki67 into prognostic groups.

6.3. Aims

This chapter aims to investigate the incorporation of a novel cancer stem cell (CSC) biomarker as a potential prognostic variable that can refine BC classification, **Figure-69**

1) To study the expression of Bmi-1 transcriptome and protein levels in BC and its relation with pathologic prognostic parameters as well as clinical long-term follow up outcome.
2) To evaluate the correlation with the $CD44^{-}/CD24^{+}$ stem cell phenotype.
3) To assess the correlation with the Ki67, as a proliferative biomarker.

6.4. Methodology

6.4.1. Patient cohorts

At the protein level, Bmi-1 expression was examined based on a well-characterised consecutive series of early stage invasive BC; the valid cases for scoring was 1215 cases (the details of the first cohort in Section-2.1.1). Furthermore, at the transcriptomic level *BMI1* gene expression was evaluated using METABRIC cohort, the valid cases was 1980 cases (the details of cohort in Section-2.1.4). CD44 and CD24 at protein level were previously assessed (Ahmed et al., 2012).

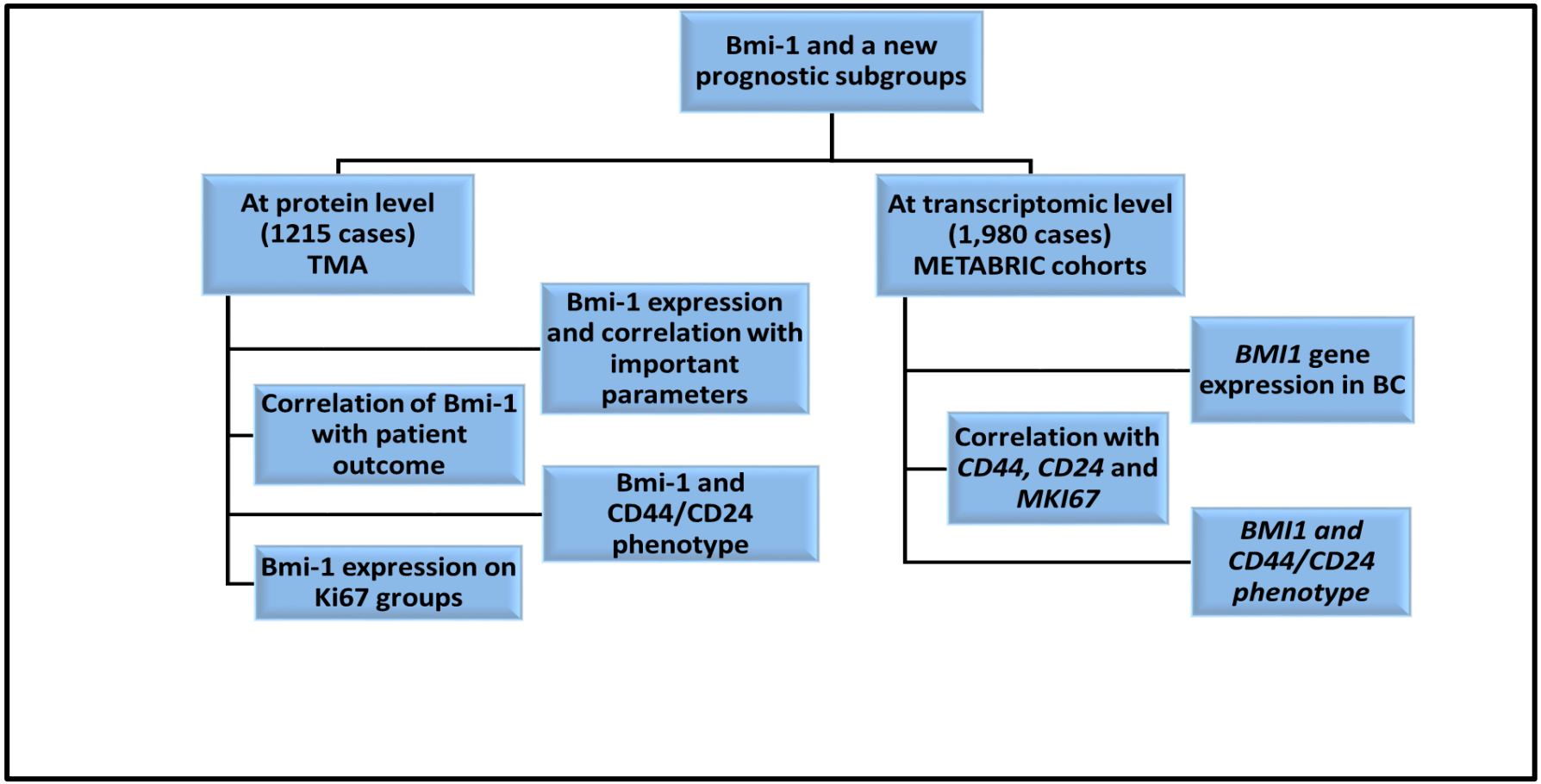


Figure-69: A flowchart representing the cohorts and steps were used in this Chapter

6.4.2. *BMI1* gene expression

BMI1 gene expression was assessed as a continuous and dichotomised variable in relation to patient outcome. The high quality of the microarray gene expression METABRIC data cohort was used. The intrinsic subclasses (luminal A, luminal B, HER2-enriched, normal-like, and basal-like) which are refined according to the PAM50 assay as well as the 10 novel molecular integrative clusters (IntClust 1–10) were used to evaluate the biological and molecular characteristic of *BMI1* in transcriptomic level. The latter are clustering analysis of joint copy number and gene expression changes revealed 10 novel molecular subgroups and associated with distinct clinical features and outcomes (Parker et al., 2009) (Dawson et al., 2013). Furthermore, the correlation between *BMI1* and *MKI67, CD44, CD24* were assessed.

6.4.3. At protein level

6.4.3.1. Western blot

The specificity of the Bmi-1 antibody was validated by western blot. The specificity was determined by detecting a particular antigen reflected as a specific band. The cell lysates used were; MCF7, MDA-MB-231, SKBR3, MDA-MB-468 and HeLa. Rabbit monoclonal to Bmi-1 (Abcam) was used, was incubated overnight with concentration of 1/5000. Western blot and cell lysate preparation is described in section 4.2.

6.4.3.2. Immunohistochemistry

To further assess the existence and frequency of Bmi-1 expression in BC, IHC on TMA was used. The details of methodology described in section 2.3. Anti-Bmi-1 antibody, clone EPR3745 (2) (Abcam) concentration was 1:100 and incubated for 1 hour at room temperature. The details of CD44 and CD24 staining is as previously described (Ahmed et al., 2012). Stained TMAs sections were scored using the semi-quantitative H-score (section 2.3.7.1).

6.5. Statistical analysis

The statistical analysis was performed by using IBM SPSS software version 22 (SPSS Inc., Chicago, IL, USA) (PAWS statistics). *P*-value was taken into consideration as a significant value when it was <0.05. Chi-square tests were applied to reflect the correlation with the essential prognostic biomarker and standard clinicopathological parameters of BC patients. Survival estimation were analysed by Kaplan–Meier method with significance determined using Log Rank test. Multivariate analysis was performed by Cox proportional hazard analysis. ANOVA test was used to evaluate the relation between *BMI1* gen and both PAM50 and IntClust 1–10.

6.6. Results

6.6.1. Transcriptomic level

6.6.1.1. *BMI1* gene expression in BC

BMI1 expression was normally distributed (**Figure-70**). The highest *BMI1* expression value was 12.02 while the lowest was 7.64. The mean and the median were the same; 9.5. Using PAM50 classification, the highest expression of *BMI1* gene appeared in luminal A (mean=9.73, SD=0.63, n=718) and B (mean=9.72, SD=0.72, n=488) classes. The lowest expression was in the basal class (mean=9.02, SD=0.53, n=329). Similarly, using integrative clusters, the highest expression was in IntClust 8 (IntClust 8; mean=10.01, SD=0.70, n=299) and the lowest expression was in IntClust 10 (mean=9.05, SD=0.55, n=226).

IntClust 8 encompasses ER-positive tumours predominately of the luminal A intrinsic subtype. Individuals within IntClust 8 also present with low-grade well-differentiated tumours and this subgroup is associated with a good prognosis and 10-year disease specific survival rates of around 80%. In IntClust 8, the main important clinicopathological character is grade, and then, further analysis was done by testing *BMI1* expression in different grades. ANOVA test was applied

and results revealed that *BMI1* still related to IntClust 8 regardless of grade where the highest expression for *BMI1* was in class 8.

Furthermore, *BMI1* expression was dichotomised using the mean value representing high and low expression. Univariate survival analysis showed that *BMI1* expression was significantly associated with patient outcome (*p*-value= 0.002, LR=9.64) (**Figure-71**) where high expression was associated with better outcome. Multivariate Cox proportional hazards model including *BMI1* mRNA expression, tumour grade, tumour size, lymph node showed that Bmi-1 was an independent prognostic marker where *p*-value= 0.030, (Hazard ratio (HR) = 0.816, 95% confidence interval (CI) = 0.679 - 0.981) (
Table-34).

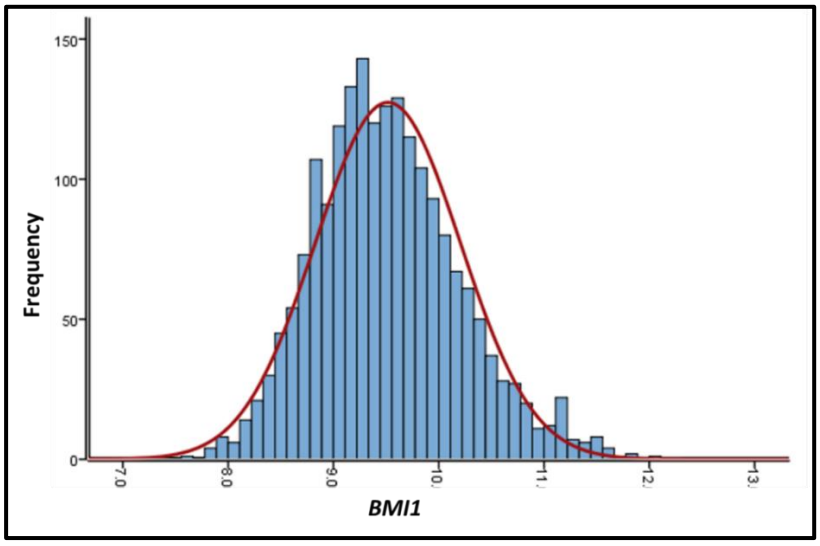


Figure-70: Normally distributed *BMI1* gene expression for the METABRIC cohort

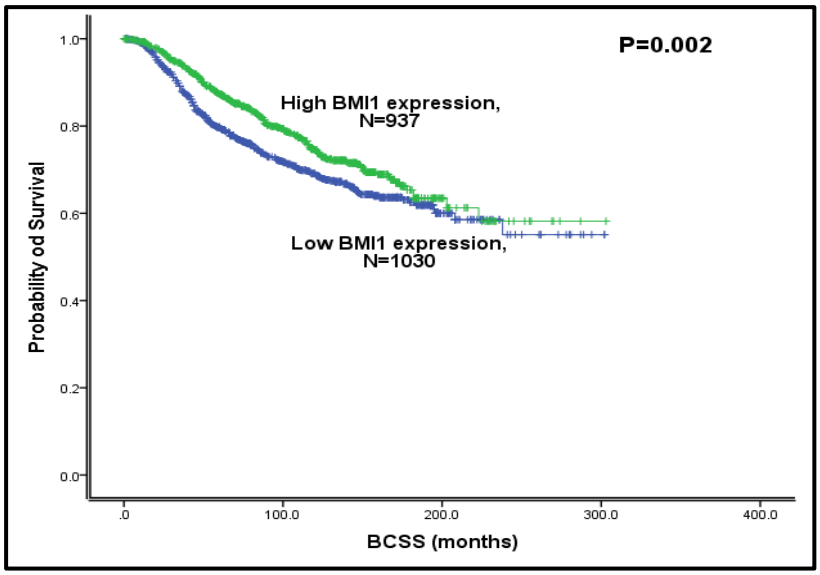


Figure-71: Kaplan–Meier survival plots for *BMI1* expression. The association between *BMI1*expression and patients' BCSS

Table-34: Cox proportional hazards analysis for predictors of BC specific survival. Effect of *BMI1* mRNA expression, tumour grade, tumour size and lymph node stage

Variables	*p*-value	Hazard ratio	95% CI
BMI1	0.030	0.816	0.679 - 0.981
Tumour size	<0.001	1.525	1.251 - 1.859
Tumour grade	<0.001	1.408	1.199 - 1.653
LN stage	<0.001	1.911	1.698 - 2.151

6.6.1.2. Correlation with *CD44, CD24* and *MKI67*

CD44 expression was normally distributed. The highest *CD44* expression value was 11.70 while the lowest was 5.25. The mean and the median were the same; 8.1. *CD24* was not normally distributed, the maximum value was 14.64 while the minimum was 5.34. The mean and the median were 12.8 and 13.1 respectively (**Figure-72**). Similarly, *MKI67* was not normally distributed with the highest value of expression was 7.66 while the lowest was 5.1. The mean and the median were 5.9 and 5.8 respectively. At the transcriptomic level, *BMI1* was

conversely correlated with *CD44*, *CD24* and *MKI67*genes with *p*-value <0.001 (**Table-35**).

Table-35: Correlation of *BMI1* with *CD44, CD24* and *MKI67* genes expression

Variable (gene expression)	*BMI1* expression		*p*-value	*(χ2)*
	Low (%)	High (%)		
CD24				
Low	425(41.4)	548(58.7)		
High	602(58.6)	386(41.3)	<0.001	58.49
CD44				
Low	451(44.0)	519(55.7)		
High	575(56.0)	413(44.3)	<0.001	26.87
MKI67				
Low	421(41.5)	537(57.5)		
High	594(58.5)	397(42.5)	<0.001	49.92

6.6.1.3. Combined analysis of *BMI1* and *CD44/CD24* phenotype

The samples were categorised based on their expression of *BMI1*, *CD44* and *CD24* genes into four main groups. The expression of $BMI1^{\pm}$ were evaluated with $CD44^{+}/CD24^{+}$, $CD44^{-}/CD24^{-}$, $CD44^{+}/CD24^{-}$ *and* $CD44^{-}/CD24^{+}$. The end point was the patient outcomes. *BMI1* has no significant with $CD44^{+}/CD24^{+}$, $CD44^{-}/CD24^{-}$, $CD44^{+}/CD24^{-}$ *and* $CD44^{-}/CD24^{+}$ phenotypes (*p*-values >0.05) (**Figure-73**).

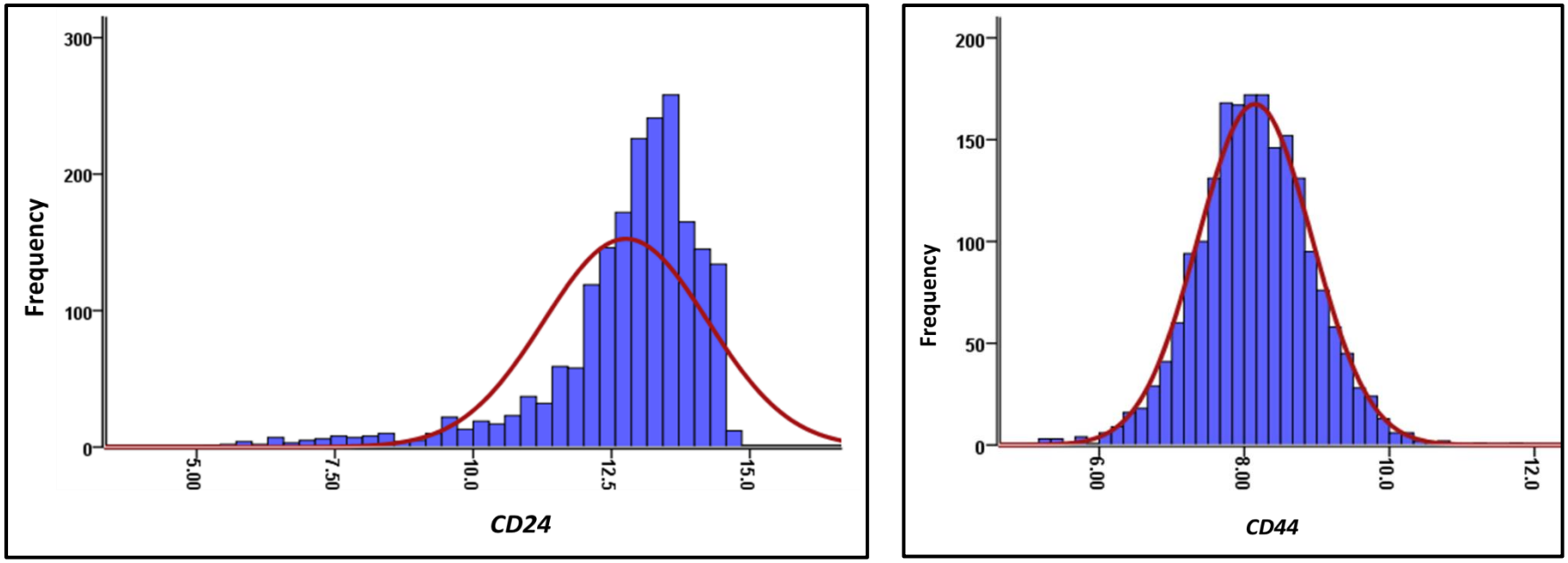


Figure-72: Distribution of *CD24* and *CD44* gene expression
***CD24* was not normally distributed while *CD44* expression was normally distributed**

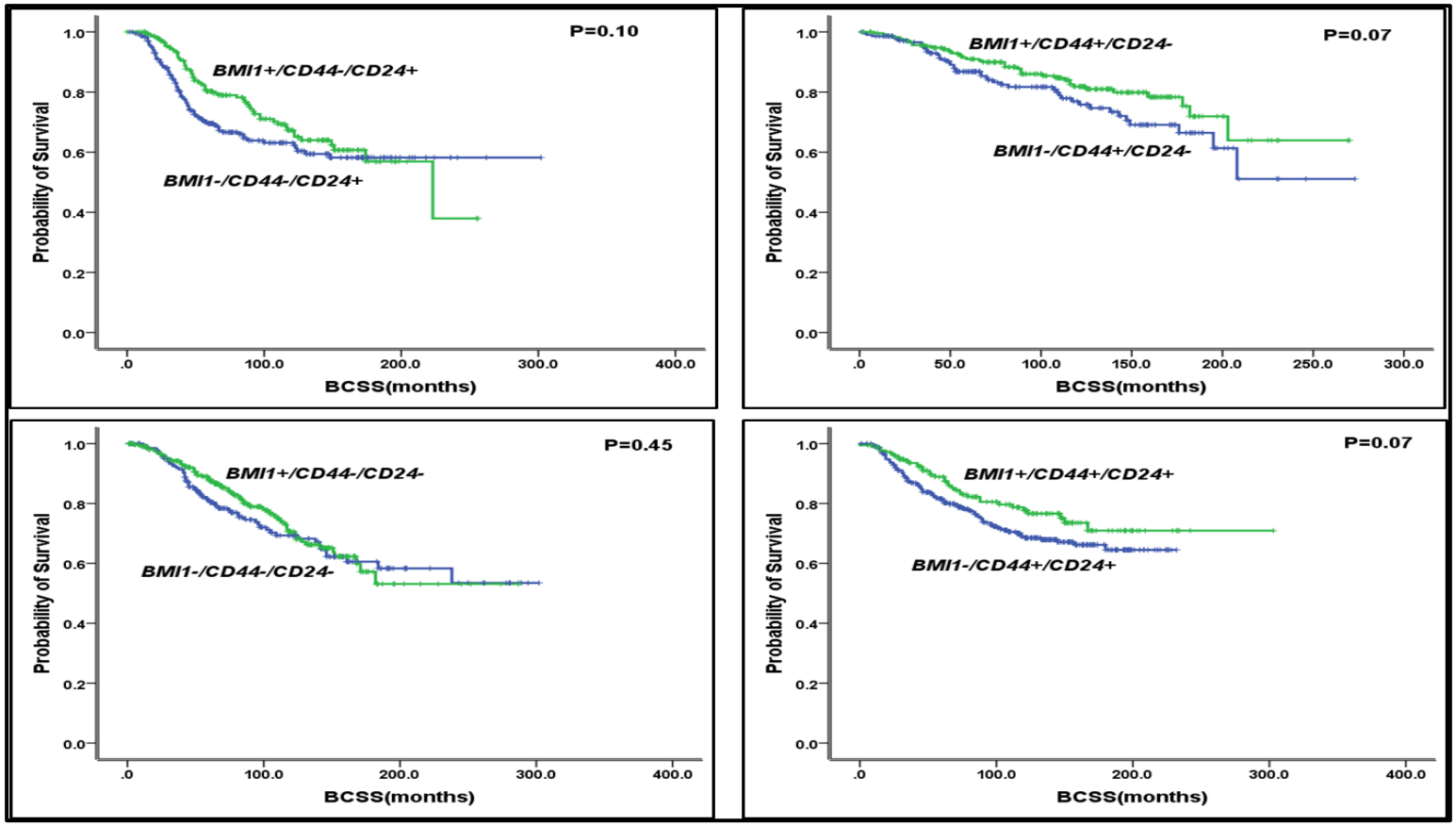


Figure-73: Kaplan–Meier survival plot for combination of *BMI1* and *CD44/CD24* phenotypes

6.6.2. Protein level

6.6.2.1. Bmi-1 protein expression in BC

The specificity of the Bmi-1 antibody was confirmed using western blotting showing a specific band at the correct predicted size, 37 kDa (**Figure-74**).

6.6.2.2. Validation of antibodies using WB

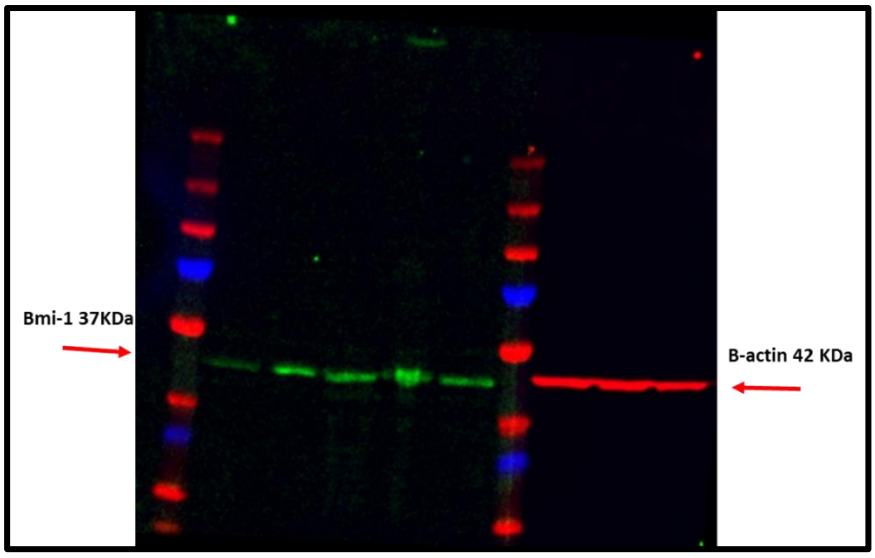


Figure-74: Western blot confirming Bmi-1 antibody specificity and β-actin as a loading control

6.6.2.3. Bmi-1 expression

Bmi-1 showed homogenous nuclear immunoreactivity and of moderate intensity in invasive tumour cells (**Figure-75**). Bmi-1 expression was not normally distributed (**Figure-76**), (H-score range 0 - 270). The cut-off point for Bmi-1 positivity was set at H-score of 80 (median). With this cut-off, 631/1215 (51.9%) cases were classified as positive for Bmi-1 expression whilst the remaining 584/1215 cases (48.1%) were showed low staining.

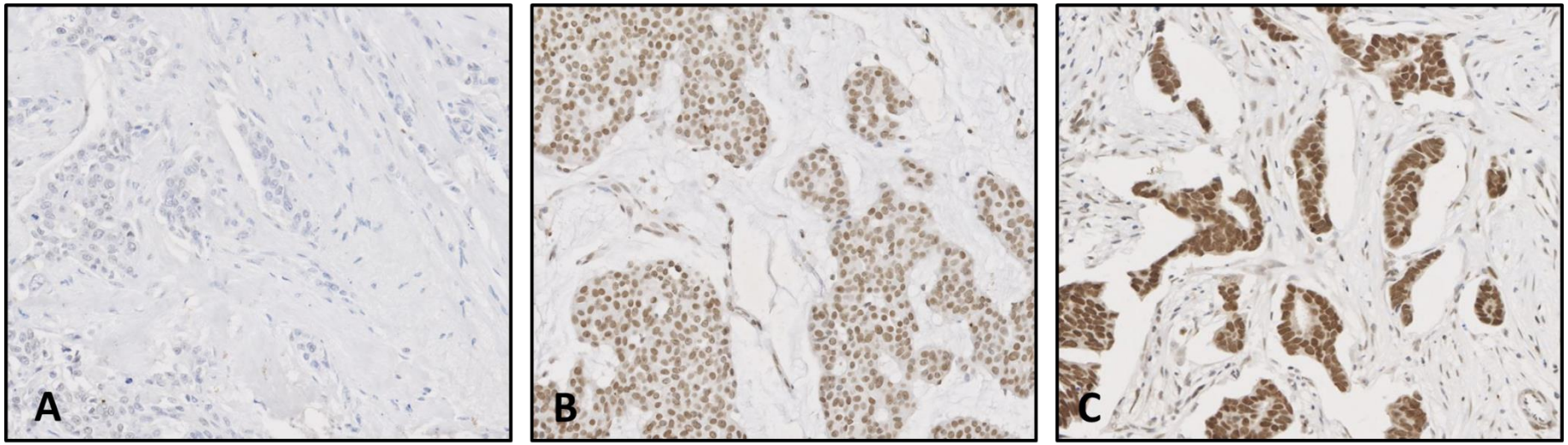


Figure-75: Immunohistochemical expression of Bmi-1 in invasive beast cancer
A, negative, B, moderate nuclear expression and C, strong nuclear expression

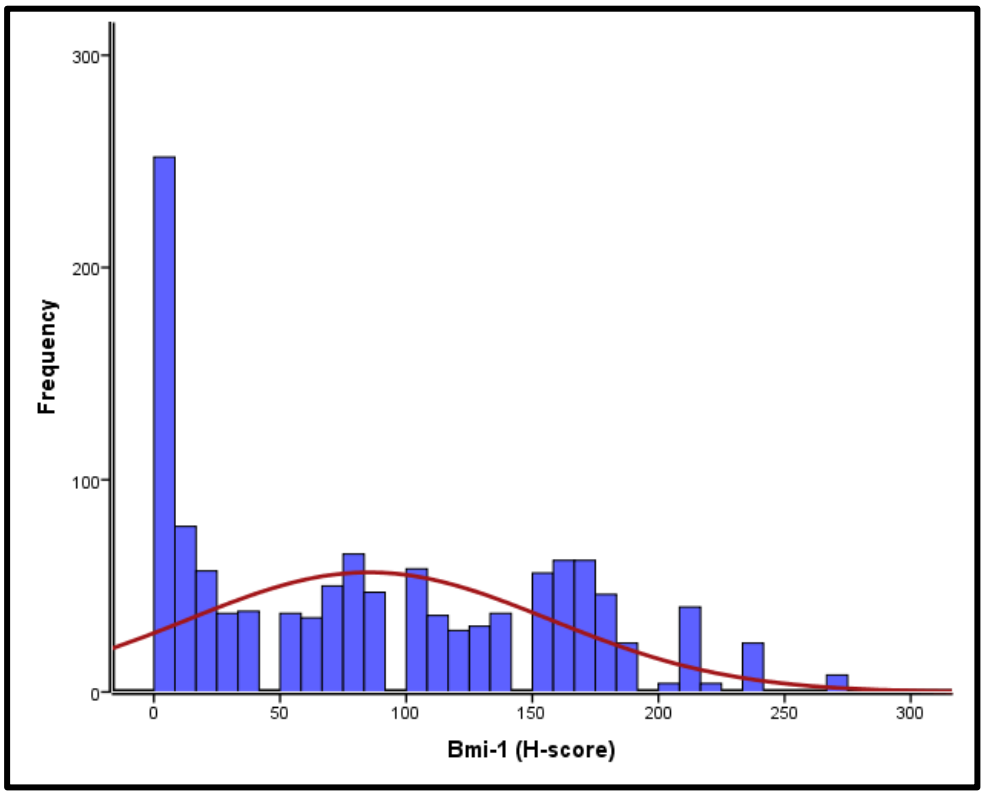


Figure-76: Distribution of Bmi-1 immunostaining expression in the studied cohort

6.6.2.4. Correlation between Bmi-1 and important parameters

The associations between Bmi-1 expression and clinicopathological variables summarises in **Table-36**. Low expression of Bmi-1 was significantly associated with higher tumour grade ($p<0.001$), high pleomorphism ($p<0.001$), higher mitotic count ($p<0.001$), less tubular formation ($p<0.001$) and with the MPG of the NPI ($p<0.001$). Regarding to the other biomarkers, high expression of Bmi-1 showed a positive correlation with ER, PR, and Bcl2 and negative association with HER2, HER1, HER3, HER4, P53, CD24 and Ki67 (**Table-37**).

Table-36: Correlation between Bmi-1 expression and the clinicopathological parameters in the studied cohort

Variable	Low (%) (<80)	High (%) (>=80)	*p-value*	*(χ2)*
Age				
<50	218(37.4)	213(33.8)	0.186	1.75
≥ 50	365(62.6)	418(66.2)		
Menopausal Status				
Pre-	235(40.2)	234(37.3)	0.287	1.13
Post-	349(59.8)	394(62.7)		
Tumour Size (cm)				
< 2.0	262(45.3)	318(50.9)	0.051	3.81
≥ 2.0	317(54.7)	307(49.1)		
Stage				
1	344(59.5)	364(58.2)	0.843	0.342
2	187(32.4)	205(32.8)		
3	47(8.1)	56(9.0)		
Vascular Invasion				
Negative	372(64.5)	406(65.3)	0.771	0.084
Definite	205(35.5)	216(34.7)		
Grade				
1	72(12.5)	125(20.0)	**<0.001**	43.84
2	159(27.5)	243(38.9)		
3	347(60.0)	256(41.0)		
Tubular formation				
1	20 (3.6)	43 (7.1)	**<0.001**	25.75
2	158 (28.1)	234 (38.6)		
3	384 (68.3)	329 (54.3)		
LN status				
1(0 node +ve)	309(59.5)	329(59.2)		
2(1-3 node +ve)	176(33.9)	189(34.0)	0.981	0.03
3(>3 node +ve)	34(6.6)	38(6.8)		
Pleomorphism				
1	7 (1.2)	16(2.6)	**<0.001**	26.76
2	175 (31.1)	269(44.4)		
3	380 (67.6)	321(53.0)		
Mitosis				
1	158(28.1)	231(38.1)	**<0.001**	34.04
2	96(17.1)	146(24.1)		
3	308(54.8)	229(37.8)		
Nottingham Prognostic Index (NPI)				
GPG	139(25.2)	212(35.6)	**<0.001**	15.72
MPG	315(57.1)	305(51.2)		
PPG	98(17.8)	79(13.3)		

Table-37: Correlation between Bmi-1 expression and the important biomarkers in the studied cohort

Variable	Bmi-1 expression		*p*-value	(X^2)
	Low (%) (<80)	High (%) (>=80)		
ER status **Negative** **Positive**	214 (37.0) 365 (63.0)	84 (13.6) 535 (86.4)	<0.001	87.58
PR status **Negative** **Positive**	277 (49.4) 284 (50.6)	194 (32.7) 399 (67.3)	<0.001	33.12
HER2 status **Negative** **Positive**	472(85.0) 83 (15.0)	533 (89.1) 65 (10.9)	0.038	4.294
HER1 status **Negative** **Positive**	425 (76.3) 132 (23.7)	501(83.6) 98(16.4)	9.751	0.002
HER3 status **Negative** **Positive**	37(7.1) 484(92.9)	54(9.8) 499(90.2)	2.453	0.117
HER4 status **Negative** **Positive**	70(12.6) 486(87.4)	79(13.3) 515(86.7)	0.128	0.720
P53 status **Negative** **Positive**	373(66.7) 186(33.3)	447(75.3) 147(24.7)	10.193	0.001
Ki67 status **Negative** **Positive**	128(27.6) 336(72.4)	173(35.1) 320(64.9)	6.245	0.012
Bcl2 status **Negative** **Positive**	114 (35.6) 206 (64.4)	84(24.0) 266(76.0)	10.851	0.001
CD24 status **Negative** **Positive**	36(10.5) 306(89.5)	58(15.9) 307(84.1)	4.407	0.036
CD44 status **Negative** **Positive**	99(35.0) 184(65.0)	129(39.8) 195(60.2)	1.50	0.220

6.6.2.5. Correlation of Bmi-1 with patient outcome

Univariate survival analysis showed a significant positive association between Bmi-1 expression and BCSS (log rank = 7.04, *p*-value= 0.007), **Figure-77**. Multivariate Cox proportional hazard analysis results showed that Bmi-1 was an independent predictor of survival (**Table-38**).

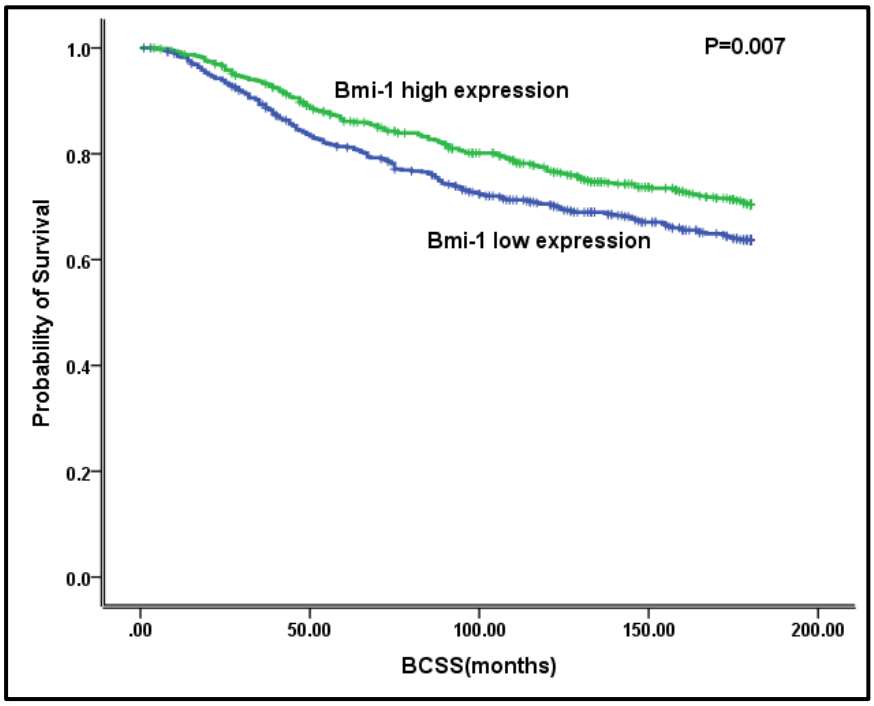


Figure-77: Kaplan–Meier survival plots for Bmi-1 expression. The association between Bmi-1 expression and patients' BCSS

Table-38: Cox proportional hazards analysis for predictors of BCSS. Effect of Bmi-1 IHC expression, tumour grade, tumour size and lymph node stage

Variables	*p*-value	Hazard ratio	95% CI
Bmi-1	0.026	0.771	0.613-0.969
Tumour size	0.013	1.353	1.065-1.719
Tumour grade	<0.001	2.111	1.725-2.583
LN stage	<0.001	2.059	1.739-2.439

6.6.2.6. Combined analysis of Bmi-1 and CD44/CD24 phenotype

Tumours were categorised according to their expression of Bmi-1, CD44 and CD24, resulting in four phenotypic groups: Bmi-1^{+}/$CD44^{-}$/$CD24^{+}$, Bmi-1^{-}/$CD44^{-}$/$CD24^{+}$, Bmi-1^{+}/$CD44^{+}$/$CD24^{-}$ and Bmi-1^{-}/$CD44^{+}$/$CD24^{-}$. The informative cases for each marker were 1512, 1036 and 889 for Bmi-1, CD24 and CD44, respectively. However, when combined few cases were available in Bmi-1^{+}/$CD44^{+}$/$CD24^{-}$ and Bmi-1^{-}/$CD44^{+}$/$CD24^{-}$ phenotypes; in total, there were 37 cases (21 and 16 in each class). For the reason of the small number of cases in this group compared with the other phenotype group, overall survival was not analysed or stratified by this phenotype.

In Bmi-1^{+}/$CD44^{-}$/$CD24^{+}$, Bmi-1^{-}/$CD44^{-}$/$CD24^{+}$ phenotypes, interestingly, Bmi-1 divided $CD44^{-}$/$CD24^{+}$ phenotype into two prognostic groups (log rank = 4.343, *p*= 0.037) (**Figure-78**). Although this phenotype did not correlate with grade, most of the cases were grade 3. Interestingly, analysis reveal that Bmi-1^{+}/$CD44^{-}$/$CD24^{+}$, Bmi-1^{-}/$CD44^{-}$/$CD24^{+}$ phenotypes further split the grade 3 cases into prognostic groups (log rank = 4.817, P= 0.02) (**Figure-79**) where Bmi-1^{+}/$CD44^{-}$/$CD24^{+}$ phenotype tumours displayed the most favourable outcome.

6.6.2.7. Effect of Bmi-1 expression on Ki67 groups

Ki67 data was dichotomised into low and high expression groups using the most optimum cut-point for the WTS, 10%. Bmi-1 successfully classified the Ki67-negative group into two prognostic subgroups (log rank =4.62, p=0.03) while has no effect on the Ki67-positive group (log rank=0.42, p=0.51), **Figure-80**.

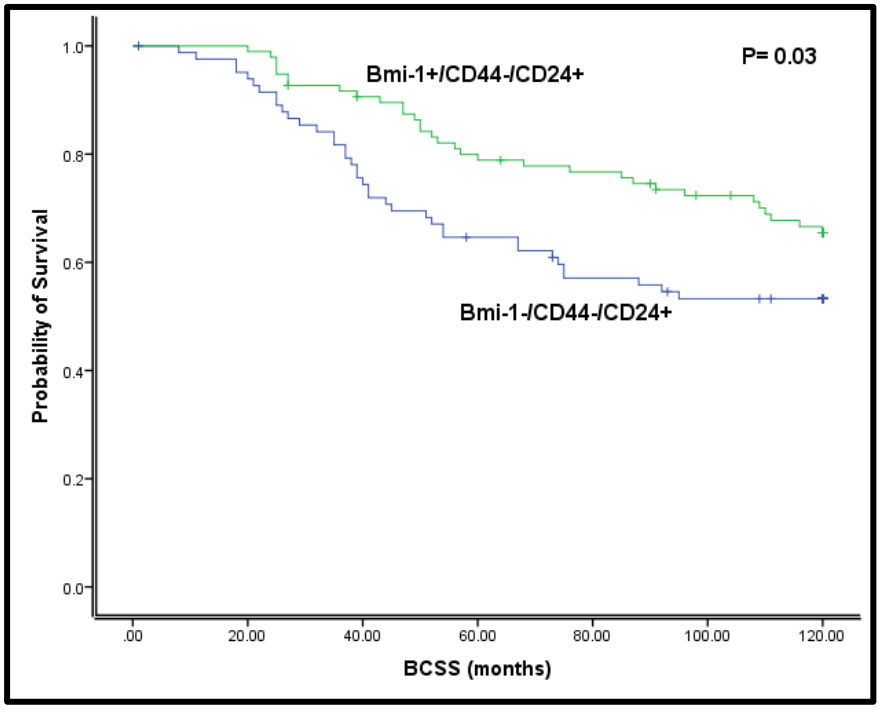


Figure-78: Kaplan–Meier survival plot for combinatorial Bmi-1 expression and $CD44^{-}/CD24^{+}$ phenotypes

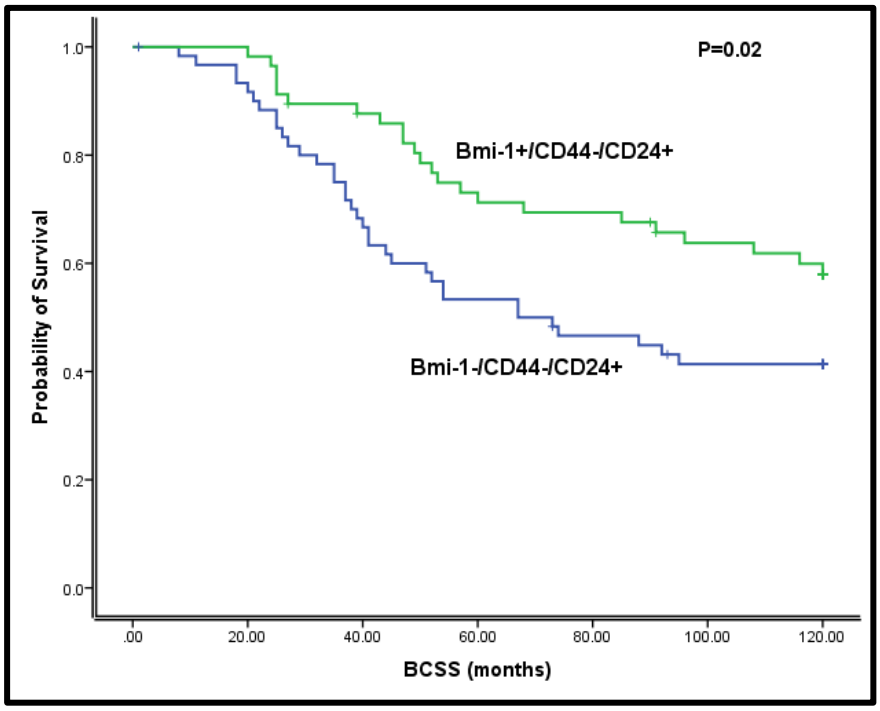


Figure-79: Kaplan–Meier survival plot for combinatorial Bmi-1 expression and $CD44^{-}/CD24^{+}$ phenotypes (Grade III only)

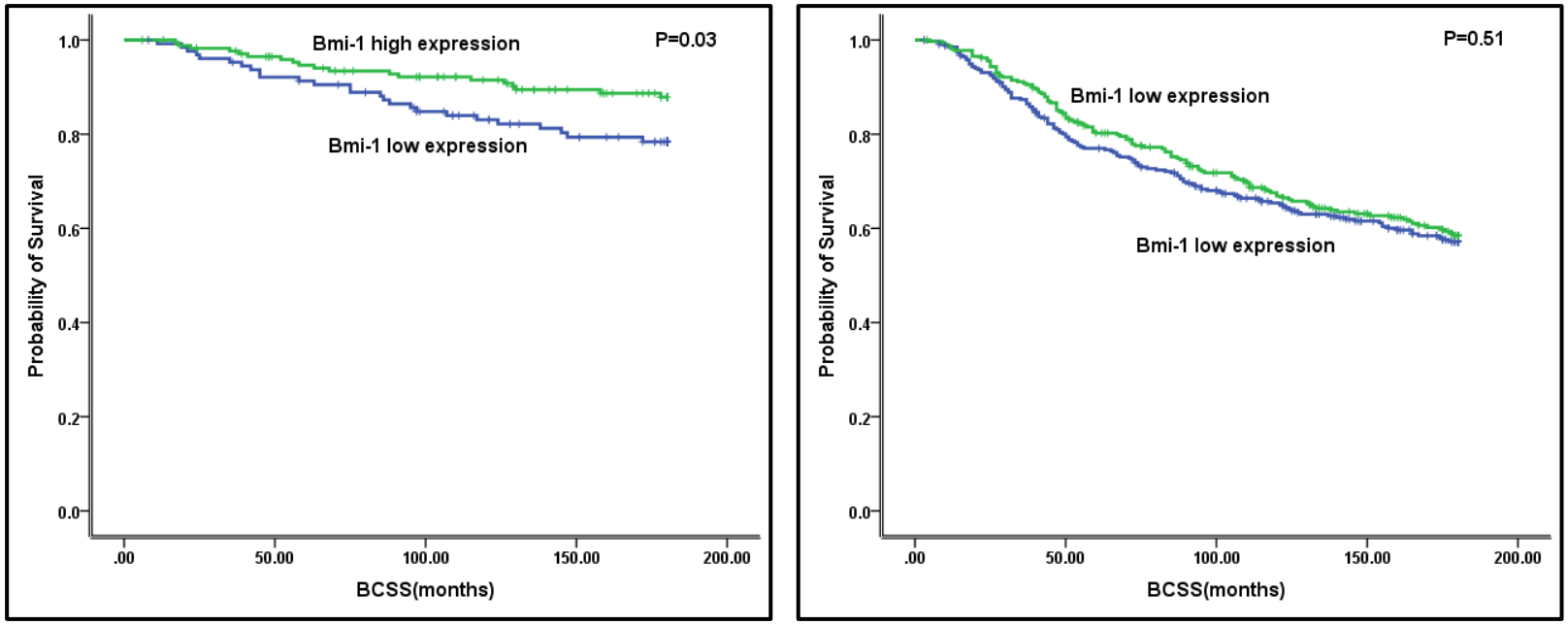


Figure-80: Kaplan–Meier survival plot
for Bmi-1 expression and Ki67-positive group (right) and Ki67-negative group (left)

6.7. Discussion

Bmi-1 expression is observed in different tumour tissue and it suspected to act as a potential prognostic marker. However, the prognostic value of Bmi-1 expression in BC is inconclusive. Some studies have revealed that high expression of Bmi-1 is correlated with cancer progression or poor survival (Guo et al., 2011). Wang et al. evaluated Bmi-1 expression in 171 BC patients and concluded that Bmi-1 is an independent prognostic factor significantly correlated to aggressive behaviour (Wang et al., 2012b). Although Kim et al. suggested the correlation of Bmi-1 expression with axillary LN metastasis using only 71 BC cases, it was found that Bmi-1 expression is also positively correlated with the good prognostic biomarker, ER (Kim et al., 2004). Conversely, the results presented in this chapter, using a large number of cases (n=1215) indicate that high expression of Bmi-1 biomarker is correlated with preferable patient outcome and good prognostic markers; ER, PR and luminal BC but inversely with Ki67 and P53. In line with these results, Choi et al. also support the favourable prognostic value of Bmi-1 in their large cohort (n=960 patients) (Choi et al., 2009) and the small 190 cohort (Arnes et al., 2008). Moreover, a recent meta-analysis showed a high level of Bmi-1 can predict better prognosis in Caucasian patients with BC (Shao et al., 2014).

Further confirmation for the results when *BMI1* transcriptomic data evaluated using PAM50 and integrative clusters. The highest expression of *BMI1* gene was in luminal A intrinsic subtype and IntClust 8 subtype, which contains ER-positive tumours predominately of the luminal A intrinsic subtype (Dawson et al., 2013). Different reports suggest that depending on the cancer context different miRNAs, which are short RNA sequences that post-transcriptionally regulate gene expression, target the Bmi-1 mRNA. In gastrointestinal cancer, miR-30e* was found to suppress Bmi-1 (Sugihara et al., 2013) while miR-135a, miR-203 as well as miR-15 and miR-16 suppress Bmi-1 in pancreatic ductal adenocarcinoma, melanomas and ovarian cancer respectively (Dang et al., 2014) (Chang et al., 2015) (Bhattacharya et al., 2009). This shows that different ways

can be used to suppress Bmi-1 depending on the cancer type and subsequently we can assume the expression can be different depending on the site.

Although the studies have identified the $CD44^+/CD24^-$ phenotype in BC as being the stem cell containing population, this has mainly been conducted in cell cultures or animal models (Al-Hajj et al., 2003) (Ponti et al., 2005). However, Ahmed et al. demonstrated the importance of $CD44^-/CD24^+$ phenotype which had worse prognosis mainly when compared to the $CD44^+/CD24^-$ phenotype (Ahmed et al., 2012). Bmi-1 expression sub-classified $CD44^-/CD24^+$ phenotype into two different prognostic groups. These results further support the value of using of Bmi-1 as a positive prognostic marker in BC.

The correlation between Ki67 and Bmi-1 was investigated by Lin et al. (Lin et al., 2008). Although they showed the expression of both biomarkers were significantly higher in colorectal carcinoma than in normal colorectal tissue and adenomas, the correlation analysis was not statistically significant ($r = 0.224$, *p*-value = 0.086). The researchers gave the reason that the relatively small sample might have been the cause. However, Wang et al. resulted with that loss of activity of the ER-coupled Bmi-1 pathway in BC associated with aberrant expression of p16INK4a and ultimately a high Ki67 expression (Wang et al., 2014). In line with our results, a recent report published by Gwak et al. which investigated the embryonic stem cells markers and revealed that Bmi-1 expression was higher in tumours of low Ki67 expression in 319 BC cases. While this study also supports the view that Bmi-1 is a favourable prognostic marker, the researchers conclude that the expression of Bmi-1 is inversely correlated with aggressive features of the BC (Gwak et al., 2017).

6.8. Conclusion

This chapter has demonstrated that Bmi-1 is a positive prognostic biomarker in BC at both the transcriptome and protein levels. Furthermore, Bmi-1 can provide prognostic value in invasive BC when analysed individually. More importantly

patients with tumours showing Bmi-1^{+}/$CD44^{-}$/$CD24^{+}$ phenotype were found to have the best prognosis in particular when compared to Bmi-1^{-}/$CD44^{-}$/$CD24^{+}$ phenotype. Moreover, Bmi-1 can divided the Ki67-negative group into a significant prognostic subgroups.

7. General Discussion

7.1. Background and cardinal findings of study

BC is a major public health problem, especially in the Western world, and is a complex, heterogeneous disease. The presence of a range of subtypes, with diverse biological characteristics leads to variances in response patterns to different treatment regimens and clinical outcomes. Further, tumours with similar clinical and pathological presentations may behave differently. Therefore, traditional subjective classification systems regarding biological characteristics have limitations for personalised patient treatment modalities. Updating the existing tools and combining clinical with quantitative molecular parameters consequently will provide more accurate prognostic and therapeutic information. This thesis focussed on assessing the performance of the existing prognostic parameters and classifiers and updating them to recent advancements in molecular classification of BC involving vital biomarkers and different techniques for the evaluation/development of new prognostic subclasses for the purpose of personalised patient treatment. This goal was achieved by examining several aims.

In terms of the first research question, this study looked at the issues related to standardisation and assessment of Ki67 for integration into existing prognostication indices. Initially, it assessed the optimal cut-off and compared TMA and WTS in organised steps, starting with the evaluation of Ki67 expression as a continuous variable in both TMA and WTS and, subsequently, as dichotomised data. The comparison between the varieties of tissue specimen type (TMA and WTS) can indisputably increase perception about the stability and the reproducibility of Ki67 as a prognostic parameter. Although TMAs allow the tumours to be screened to detect overall protein expression using large number of samples, rather than smaller cohorts, the international Ki67 in breast cancer working group has presented a lack of systematic comparison of Ki67 expression levels between TMA and WTS (Dowsett et al., 2011).

To my knowledge, our study is the largest to evaluate and compare Ki67 expression in TMA and WTS at both transcriptome and protein levels. It reveals that TMA can be used for research purposes, but with caution, as WTS is superior to TMA at both transcriptome and protein levels.

Although the highest significant correlation of Ki67 with patient outcome were at 10% and 20% for WTS and TMA, respectively, the results remain statistically significant over a wide range of cut-points. This supports the approach of using Ki67 as a continuous biomarker rather than a dichotomised value. This is broadly in harmony with the changeable cut-point recommended by the St Gallen guidelines (Goldhirsch et al., 2009) (Goldhirsch et al., 2011) (Goldhirsch et al., 2013). This is also in line with what the international Ki67 breast cancer working group has reported; that is there is no consistency in an ideal cut-off point for use in daily clinical routine. However, defining a specific cut-off is very critical; not only in the clinical field, when there is a need to dichotomise the low and high levels of Ki67 to determine the patients at risk, but also in determining the prognostic subclasses, which is essential in BC classification. Therefore, if categorisation of Ki67 data is needed, our findings suggest that a cut-point of 20% when assessing Ki67 on TMAs appears to be optimum and in concordance with WTS; while the optimal cut-point for WTS is 10%. These cut-offs (20% using TMA and 10% using WTS) are further supported by the fact that they give the highest significant correlation with survival and other important clinic-pathological parameters (Muftah et al., 2017).
Use of Bonferroni correction depends on the circumstances of the study, and was not used in this study. Armstrong demonstrated that it should not be utilised routinely. In an exploratory context, a researcher might miss a possible effect worthy of further study and therefore, a correction would not be appropriate (Armstrong, 2014).

Integration of Ki67 in BC classification using WTS was successfully achieved, initially by applying the traditional NPI and then by integrating the biological formulae of the NPI+, instead of mitotic count in grade. Interestingly, when Ki67 was used as a continuous variable (in NPI+), it was more superior to mitosis in

patient classification. Although mitotic counts is considered a strong prognostic indicator as a part of NGS, there are limitations to its precision as mitotic figures can be detected morphologically only in mitotic phase. The duration of mitotic phase is unstable and this may change the proportion of visible mitosis within the same tumour at different times. Therefore, the correlation between the proliferation rate and the number of mitosis is not strictly linear (Sennerstam and Auer, 1990) (van Diest et al., 2004) (Beresford et al, 2006). Moreover, the instability of mitotic figures for the duration of specimen fixation, reproducibility issues due to tissue section quality and experience of the scorer and relatively long time required for mitotic counting are considered as technical problems challenged during mitotic counting (Salisbury, 2009).

The second research question concerned the feasibility of RPPA in refining molecular classification of BC with the purpose of stratify patients according to the NPI+ decision making tool. Using RPPA in BC classification can help improve BC clustering and classification. Although the initial work included a small number of samples (n=25), the consistency between IHC and RPPA classes using macrodissection was extremely high at 80%. Additionally, the reproducibility of the same NPI+ classes gives the view that it can, with highly accurate and less subjective proteomic techniques, improve the updated prognostic index, NPI+ (Negm et al., 2016). Although employing FFPE instead of cell lines, which are usually used with RPPA, is challenging, it adds strength to this study as patient tissue used includes a microenvironmental effect and gives the advantage of comparison with a robust IHC technique. Systematic evaluation, in terms of both quantity and quality, of protein extraction methods from FFPE tissue for optimal performance was applied (Negm et al., 2016). An efficient and successful extraction of proteins from FFPE samples was obtained using a Qproteome FFPE tissue kit. Consistent with our result, the same kit was successfully used in BC (Becker et al., 2007) (Wolff et al., 2011b).

Besides evaluation of replacing IHC with RPPA, it seemed interesting and informative to test the possibility of realigning the NPI+ classification based on the expression of Ki67 using RPPA. The unsupervised classification with a

larger cohort (n=143 cases) using RPPA was used with evaluation of the Ki67 role. The clustering demonstrates that Ki67 can classify more patients into the main biological classes, reflecting its ability to improve the existing prognostic classifier, NPI+. Furthermore, adding Ki67 emerged as a necessity for understanding whether Ki67 would affect the unclassified patient group in the framework of proteomic classification.

Consensus clustering, which was used in this study, offers a way to exhibit consensus through several runs of a clustering algorithm and to estimate the stability of the identified clusters (Soria et al., 2010). The more the achieved clusters are efficient in sampling variability, the more one can be convinced that these clusters characterise a genuine structure. The approach of merging and comparing the results of various clustering algorithms is essential with the intention of assessing the stability of an offered classification. The reasons for applying consensus classes are to obtain the highest number of patients assigned to any class and to consider all the clustering techniques investigated. As an alternative to using comparison approaches, this consensus methodology pools the results obtained by various clustering algorithms. Although the class assignments were performed blind to all clinical and outcome data, the initial cluster (using three classes) showed that 69% of data was assigned to these core classes. Interestingly, 85% of data was re-assigned when Ki67 was incorporated. In this study, the previously recommended methodology (Soria et al., 2010) for meeting consensus from the different results that may be obtained from clustering algorithms has been applied using RPPA BC data. This consensus clustering and classification to reproduce NPI+ using RPPA has to be deemed an exploratory study, and further work on these biological classes relating to patient outcome has to be considered. However, the results reveal the capability of RPPA and could be approved for the stratification of patients for therapy regimens and classification of BC and demonstrate the significance of Ki67 in BC classification using RPPA.

The third part of this study evaluated a subgroup of the luminal BC classes, the low ER group (<10%). Initially, the distribution of ER expression in BC and the

debatable bimodality concept was evaluated. Our results provide evidence supporting this concept by evaluating immunostained ER expression in three large at protein level cohorts, which are consistent with other studies (Collins et al., 2005) (Nadji et al., 2005) (Khoshnoud et al., 2011) (Bogina et al., 2012) (Pinto et al., 2013). Moreover, a large transcriptomic cohort was used and the result was consistent with the protein level bimodality results, and revealed a strong correlation between ER IHC protein levels and *ESR1* mRNA (Muftah et al., 2016). Furthermore, the results showed no significant difference between the ER-negative and the low/intermediate group (expressed as ER >1% and <70%) with patient outcome. Therefore, the question raised was whether patients with low positive-ER expression are actually positive cases or fall under a misclassification of ER-negative cancers as borderline weak ER-positive. Our study evaluated these borderline ER-positive cases, raising different issues starting from the consistency between the CNB and WTS in this subgroup. The results show that the correlation between CNB and WTS is low which is apparently seems to be inconsistent with what published before by Hodi et al. (Hodi et al., 2007) where the agreement between core and excision in 334 of 338 tumours was 98.8%. However, our results has a complete agreement with the previous study for the reason that they used H-score cut-off of 10 for positivity. Furthermore, they mentioned the discrepancies were between weakly positive and negative tumours. Similarly, Wood et al. reported that concordant between the core and excision in 95.8% of cases. However, they excluded the cases scored less than 10% with low intensity and considered only intermediate- or high-intensity staining (Wood et al., 2007).

Although ASCO and CAP guidelines recommend that an ER status of 1% or more indicates positive IHC nuclear staining (Harvey et al., 1999) (Hammond et al., 2010a), the low-ER group is still a debatable subclass of the ER-positive tumours and oncologists struggle with treating patients who are weakly positive for ER because of a lack of data about this group (Lin et al., 2013). Many clinicians deem 10% ER nuclear staining as a cut-point for endocrine therapy eligibility. Creighton et al. suggest that the lower ER level tumours have hyperactive GFR/PI3K signalling, which has been correlated with resistance to

endocrine therapy (Creighton et al., 2010). Gloyeske et al. assess the morphologic features of low ER BCs and concluded that this category has morphologic characteristics more similar to the triple-negative BCs than of the usual ER-positive group (Gloyeske et al., 2014). Furthermore, Sheffield et al. applied the PAM50 classifier to ER-weakly positive and ER-negative BC. They compared the intrinsic subtype distributions, with the result that ER-weakly positive BC is more related in clinical behaviour to ER-negative BC (Sheffield et al., 2016).

Our results demonstrated that 42.4% of the low ER cases (ER<10%) were negative in WTS staining using two different immunostaining platforms. Moreover, all the cases that scored >10% for ER expression were positive. For further confirmation, evaluation at the transcriptomic level was used, and all ER cases scored >10% were also positive at the transcriptomic level. Additionally, in cases scored <10%, the consistency in the negative cases in both IHC and *in situ* RNAscope was as high as 93%, and with RT-PCR the consistency was 86%. These results show that nearly half of the cases in the low ER group (<10%), which were initially diagnosed as ER-positive cases using CNB, are negative at both transcriptome and protein levels, while all the cases above 10% are positive at both levels. Therefore, this study provides evidence that this subcategory (<10%) is a distinct subset; careful evaluation seems to be mandatory and suggests adding it as a separate subgroup in ER-positive classification.

In addition to the low ER subgroup, novel prognostic subclasses have been revealed in the final section of this study: Bmi-1 can divide the patients of the $CD44^-/CD24^+$ class into $Bmi\text{-}1^+/CD44^-/CD24^+$ and $Bmi\text{-}1^-/CD44^-/CD24^+$ phenotypes. Importantly, patients with tumours showing the $Bmi\text{-}1^+/CD44^-/CD24^+$ phenotype were found to have the best prognosis in particular, when compared to the $Bmi\text{-}1^-/CD44^-/CD24^+$ phenotype. Moreover, Bmi-1 can subdivided the Ki67-negative group into prognostic subgroups. The prognostic importance of Bmi-1 as a favourable biomarker in BC has been established in this study by using the largest cohort used in BC to evaluate this prognostic effect. For further confirmation, we tested the *BMI1* gene expression in both PAM50 and IntClust classifications using a METABRIC cohort, the results were

consistent with the protein level results. *BMI1* was expressed in Luminal A and IntClust 8, which represent a high ER expression class.

7.2. Conclusion and recommendations

Prognostic tools have become increasingly essential in the diagnosis and treatment of BC. Despite the massive amount of work that has been undertaken to develop and refine BC classification, there is still room for improvement. With the increasing use of panels of biomarkers and high-throughput techniques, new data will emerge, which could result in an updating of the prognostic indices and identification of novel prognostic subclasses and, subsequently, allow more precise classification systems that can predict outcomes and responses to therapy. This thesis concludes that Ki67 can be used as a substitution for mitotic frequency in NPI+ with a superior stratification of patients into prognostic groups. Moreover, using Ki67 as a continuous variable, especially in BC classification, is more applicable than the dichotomised values currently used in clinical practice. However, if a cut-off is required, the suggested optimal cut-points for Ki67 are 20%, when assessing on TMA, and 10% on WTS. Although Ki67 expression can be assessed using WTS and TMA, the latter need to be interpreted with caution and using multiple or larger TMA cores are recommended as an alternative to WTS. Ki67 has a vital role in BC classification using RPPA. Assessment of ki67 using RRPA demonstrates that the analysis of FFPE tissue lysates is a reliable method for protein quantification, especially when using macrodissection. Current clinical guidelines suggest that 1% ER is most appropriate. However the work in thesis provides scientific evidence that it seems to be 10% particularly as ER expression is essentially bimodal, with the vast majority of cases being either ER-negative or strongly ER-positive. Nearly half of the cases with low ER expression were negative in excision biopsies and most of the positive cases were within the high-grade tumour category. These findings suggest at least repeating ER staining on excision biopsies as a routine for the low ER cases or further validation using an effective molecular technique as *in situ* RNAscope. Our results also revealed that Bmi-1 is a good prognostic

biomarker in BC approved at transcriptome and protein levels. Bmi-1 can successfully provide significant prognostic subgroups when it was evaluated with $CD44^{-}/CD24^{+}$ phenotype and low Ki67 proliferative tumours.

7.3. Strengths of the study

- The study samples were from patient samples processed using routine diagnostic instead of cell lines or xenograft.
- Meeting the challenge of using FFPE tissue at the modern proteomic and transcriptomic techniques.
- Five large cohorts were used in this study at both levels: proteomic and transcriptomic.
- Different approaches were used, besides the routine robust techniques: IHC and RT-PCR advanced techniques were used including RPPA and RNAscope.
- The results have direct clinical implications.

7.4. Limitations of the study

- IHC is a semi-quantitative subjective method dependent on the scorers; however, standard methods were used to minimise this limitation.
- Only one representative TMA core from each tumour tissue was available because of the limited availability of resources. It would have been ideal to include more than one core to express tumour heterogeneity.
- In FFPE samples, the pre-analytical variables cannot be completely standardised in such retrospective studies.

7.5. Future work

- Use of multiple TMA cores in future cohorts.

- Validation of the RPPA classification using WTS immunostaining of the studied cohort (n=143 cases) and correlation with patient outcome.
- Expansion of the low ER series for further validation of the results.

8. References

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.

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.

ABD EL-REHIM, D. M., PINDER, S. E., PAISH, C. E., BELL, J. A., RAMPAUL, R. S., BLAMEY, R. W., ROBERTSON, J. F., NICHOLSON, R. I. & ELLIS, I. O. 2004. Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. *Br J Cancer,* 91**,** 1532-42.

ABRAHAM, B. K., FRITZ, P., MCCLELLAN, M., HAUPTVOGEL, P., ATHELOGOU, M. & BRAUCH, H. 2005. Prevalence of CD44+/CD24-/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. *Clin Cancer Res,* 11**,** 1154-9.

ADDIS, M. F., TANCA, A., PAGNOZZI, D., ROCCA, S. & UZZAU, S. 2009. 2-D PAGE and MS analysis of proteins from formalin-fixed, paraffin-embedded tissues. *Proteomics,* 9**,** 4329-39.

AHMED, M. A., ALESKANDARANY, M. A., RAKHA, E. A., MOUSTAFA, R. Z., BENHASOUNA, A., NOLAN, C., GREEN, A. R., ILYAS, M. & ELLIS, I. O. 2012. A CD44(-)/CD24(+) phenotype is a poor prognostic marker in early invasive breast cancer. *Breast Cancer Res Treat,* 133**,** 979-95.

AIGELSREITER, A., RESS, A. L., BETTERMANN, K., SCHAUER, S., KOLLER, K., EISNER, F., KIESSLICH, T., STOJAKOVIC, T., SAMONIGG, H., KORNPRAT, P., LACKNER, C., HAYBAECK, J. & PICHLER, M. 2013. Low expression of the putative tumour suppressor spinophilin is associated with higher proliferative activity and poor prognosis in patients with hepatocellular carcinoma. *Br J Cancer,* 108**,** 1830-7.

AJJ, H., CHESNEL, A., PINEL, S., PLENAT, F., FLAMENT, S. & DUMOND, H. 2013. An alkylphenol mix promotes seminoma derived cell proliferation through an ERalpha36-mediated mechanism. *PLoS One,* 8**,** e61758.

AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. 2003. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A,* 100**,** 3983-8.

ALBA, E., CALVO, L., ALBANELL, J., DE LA HABA, J. R., ARCUSA LANZA, A., CHACON, J. I., SANCHEZ-ROVIRA, P., PLAZAOLA, A., LOPEZ GARCIA-ASENJO, J. A., BERMEJO, B., CARRASCO, E. & LLUCH, A. 2012. Chemotherapy (CT) and hormonotherapy (HT) as neoadjuvant treatment in luminal breast cancer patients: results from the GEICAM/2006-03, a multicenter, randomized, phase-II study. *Ann Oncol,* 23**,** 3069-74.

ALESKANDARANY, M. A., GREEN, A. R., RAKHA, E. A., MOHAMMED, R. A., ELSHEIKH, S. E., POWE, D. G., PAISH, E. C., MACMILLAN, R. D., CHAN, S., AHMED, S. I. & ELLIS, I. O. 2010. Growth fraction as a predictor of response to chemotherapy in node-negative breast cancer. *Int J Cancer,* 126**,** 1761-9.

ALESKANDARANY, M. A., RAKHA, E. A., MACMILLAN, R. D., POWE, D. G., ELLIS, I. O. & GREEN, A. R. 2011. MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. *Breast Cancer Res Treat,* 127**,** 591-9.

ALI, S. & COOMBES, R. C. 2002. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer,* 2**,** 101-12.

ALKEMA, M. J., WIEGANT, J., RAAP, A. K., BERNS, A. & VAN LOHUIZEN, M. 1993. Characterization and chromosomal localization of the human proto-oncogene BMI-1. *Hum Mol Genet,* 2**,** 1597-603.

ALLISON, K. H. 2012. Molecular pathology of breast cancer: what a pathologist needs to know. *Am J Clin Pathol,* 138**,** 770-80.

ALSHAREEDA, A. T., NEGM, O. H., ALBARAKATI, N., GREEN, A. R., NOLAN, C., SULTANA, R., MADHUSUDAN, S., BENHASOUNA, A., TIGHE, P., ELLIS, I. O. & RAKHA, E. A. 2013. Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer. *Breast Cancer Res Treat,* 139**,** 301-10.

AMERIK, A. Y. & HOCHSTRASSER, M. 2004. Mechanism and function of deubiquitinating enzymes. *Biochim Biophys Acta,* 1695**,** 189-207.

ANAGNOSTOU, V. K., WELSH, A. W., GILTNANE, J. M., SIDDIQUI, S., LICEAGA, C., GUSTAVSON, M., SYRIGOS, K. N., REITER, J. L. & RIMM, D. L. 2010. Analytic variability in

immunohistochemistry biomarker studies. *Cancer Epidemiol Biomarkers Prev,* 19**,** 982-91.

ARACTINGI, S., KANITAKIS, J., EUVRARD, S., LE DANFF, C., PEGUILLET, I., KHOSROTEHRANI, K., LANTZ, O. & CAROSELLA, E. D. 2005. Skin carcinoma arising from donor cells in a kidney transplant recipient. *Cancer Res,* 65**,** 1755-60.

ARMSTRONG, R. A. 2014. When to use the Bonferroni correction. *Ophthalmic Physiol Opt,* 34**,** 502-8.

ARNES, J. B., COLLETT, K. & AKSLEN, L. A. 2008. Independent prognostic value of the basal-like phenotype of breast cancer and associations with EGFR and candidate stem cell marker BMI-1. *Histopathology,* 52**,** 370-80.

ARRIAGADA, R., LE, M. G., DUNANT, A., TUBIANA, M. & CONTESSO, G. 2006. Twenty-five years of follow-up in patients with operable breast carcinoma: correlation between clinicopathologic factors and the risk of death in each 5-year period. *Cancer,* 106**,** 743-50.

ASSADI, M., LAMERZ, J., JARUTAT, T., FARFSING, A., PAUL, H., GIERKE, B., BREITINGER, E., TEMPLIN, M. F., ESSIOUX, L., ARBOGAST, S., VENTURI, M., PAWLAK, M., LANGEN, H. & SCHINDLER, T. 2013. Multiple protein analysis of formalin-fixed and paraffin-embedded tissue samples with reverse phase protein arrays. *Mol Cell Proteomics,* 12**,** 2615-22.

AVASTHI, S., SRIVASTAVA, RN, SINGH, A, AND SRIVASTAVA, M 2008. Stem Cell: past, present and future: a review article. *Internet Journal of Medical Update,* Internet Journal of Medical Update**,** 22-30.

BAAK, J. P., VAN DIEST, P. J., JANSSEN, E. A., GUDLAUGSSON, E., VOORHORST, F. J., VAN DER WALL, E. & VERMORKEN, J. B. 2008. Proliferation accurately identifies the high-risk patients among small, low-grade, lymph node-negative invasive breast cancers. *Ann Oncol,* 19**,** 649-54.

BAAK, J. P., VAN DIEST, P. J., VOORHORST, F. J., VAN DER WALL, E., BEEX, L. V., VERMORKEN, J. B. & JANSSEN, E. A. 2005. Prospective multicenter validation of the independent prognostic value of the mitotic activity index in lymph node-negative breast cancer patients younger than 55 years. *J Clin Oncol,* 23**,** 5993-6001.

BACHMANN, I. M., PUNTERVOLL, H. E., OTTE, A. P. & AKSLEN, L. A. 2008. Loss of BMI-1 expression is associated with clinical progress of malignant melanoma. *Mod Pathol,* 21**,** 583-90.

BADVE, S. S., BAEHNER, F. L., GRAY, R. P., CHILDS, B. H., MADDALA, T., LIU, M. L., ROWLEY, S. C., SHAK, S., PEREZ, E. A., SHULMAN, L. J., MARTINO, S., DAVIDSON, N. E., SLEDGE, G. W., GOLDSTEIN, L. J. & SPARANO, J. A. 2008. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol,* 26**,** 2473-81.

BAGO-HORVATH, Z., RUDAS, M., DUBSKY, P., JAKESZ, R., SINGER, C. F., KEMMERLING, R., GREIL, R., JELEN, A., BOHM, G., JASAREVIC, Z., HAID, A., GRUBER, C., POSTLBERGER, S., FILIPITS, M. & GNANT, M. 2011. Adjuvant sequencing of tamoxifen and anastrozole is superior to tamoxifen alone in postmenopausal women with low proliferating breast cancer. *Clin Cancer Res,* 17**,** 7828-34.

BALSLEV, I., AXELSSON, C. K., ZEDELER, K., RASMUSSEN, B. B., CARSTENSEN, B. & MOURIDSEN, H. T. 1994. The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). *Breast Cancer Res Treat,* 32**,** 281-90.

BARNARD, N. J., HALL, P. A., LEMOINE, N. R. & KADAR, N. 1987. Proliferative index in breast carcinoma determined in situ by Ki67 immunostaining and its relationship to clinical and pathological variables. *J Pathol,* 152**,** 287-95.

BARRIERE, G., RIOUALLON, A., RENAUDIE, J., TARTARY, M. & RIGAUD, M. 2012. Mesenchymal and stemness circulating tumor cells in early breast cancer diagnosis. *BMC Cancer,* 12**,** 114.

BARTLETT, J. M., THOMAS, J., ROSS, D. T., SEITZ, R. S., RING, B. Z., BECK, R. A., PEDERSEN, H. C., MUNRO, A., KUNKLER, I. H., CAMPBELL, F. M., JACK, W., KERR, G. R., JOHNSTONE, L., CAMERON, D. A. & CHETTY, U. 2010. Mammostrat as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res,* 12**,** R47.

BASELGA, J., BRADBURY, I., EIDTMANN, H., DI COSIMO, S., DE AZAMBUJA, E., AURA, C., GOMEZ, H., DINH, P., FAURIA, K., VAN DOOREN, V., AKTAN, G., GOLDHIRSCH, A., CHANG, T. W., HORVATH, Z., COCCIA-PORTUGAL, M., DOMONT, J., TSENG, L. M., KUNZ, G., SOHN, J. H., SEMIGLAZOV, V., LERZO, G., PALACOVA, M., PROBACHAI, V., PUSZTAI, L., UNTCH, M., GELBER, R. D. & PICCART-GEBHART, M. 2012. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet,* 379**,** 633-40.

BASTIEN, R. R., RODRIGUEZ-LESCURE, A., EBBERT, M. T., PRAT, A., MUNARRIZ, B., ROWE, L., MILLER, P., RUIZ-BORREGO, M., ANDERSON, D., LYONS, B., ALVAREZ, I., DOWELL, T., WALL, D., SEGUI, M. A., BARLEY, L., BOUCHER, K. M., ALBA, E., PAPPAS, L., DAVIS, C. A., ARANDA, I., FAURON, C., STIJLEMAN, I. J., PALACIOS, J., ANTON, A., CARRASCO, E., CABALLERO, R., ELLIS, M. J., NIELSEN, T. O., PEROU, C. M., ASTILL, M., BERNARD, P. S. & MARTIN, M. 2012. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genomics,* 5**,** 44.

BEATSON, G. T. 1899. On the Etiology of Cancer, with a Note of some Experiments. *Br Med J,* 1**,** 399-400.

BECKER, K. F., SCHOTT, C., BECKER, I. & HOFLER, H. 2008. Guided protein extraction from formalin-fixed tissues for quantitative multiplex analysis avoids detrimental effects of histological stains. *Proteomics Clin Appl,* 2**,** 737-43.

BECKER, K. F., SCHOTT, C., HIPP, S., METZGER, V., PORSCHEWSKI, P., BECK, R., NAHRIG, J., BECKER, I. & HOFLER, H. 2007. Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J Pathol,* 211**,** 370-8.

BERGSMEDH, A., SZELES, A., HENRIKSSON, M., BRATT, A., FOLKMAN, M. J., SPETZ, A. L. & HOLMGREN, L. 2001. Horizontal transfer of oncogenes by uptake of apoptotic bodies. *Proc Natl Acad Sci U S A,* 98**,** 6407-11.

BHATTACHARYA, R., MUSTAFI, S. B., STREET, M., DEY, A. & DWIVEDI, S. K. 2015. Bmi-1: At the crossroads of physiological and pathological biology. *Genes Dis,* 2**,** 225-239.

BHATTACHARYA, R., NICOLOSO, M., ARVIZO, R., WANG, E., CORTEZ, A., ROSSI, S., CALIN, G. A. & MUKHERJEE, P. 2009. MiR-15a and MiR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res,* 69**,** 9090-5.

BHOO-PATHY, N., YIP, C. H., HARTMAN, M., SAXENA, N., TAIB, N. A., HO, G. F., LOOI, L. M., BULGIBA, A. M., VAN DER GRAAF, Y. & VERKOOIJEN, H. M. 2012. Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. *Eur J Cancer,* 48**,** 982-9.

BJERKVIG, R., TYSNES, B. B., ABOODY, K. S., NAJBAUER, J. & TERZIS, A. J. 2005. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat Rev Cancer,* 5**,** 899-904.

BJORNSTROM, L. & SJOBERG, M. 2005. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Mol Endocrinol,* 19**,** 833-42.

BLACKWELL, K. L., BURSTEIN, H. J., STORNIOLO, A. M., RUGO, H. S., SLEDGE, G., AKTAN, G., ELLIS, C., FLORANCE, A., VUKELJA, S., BISCHOFF, J., BASELGA, J. & O'SHAUGHNESSY, J. 2012. Overall survival benefit with lapatinib in combination with trastuzumab for patients with human epidermal growth factor receptor 2-positive metastatic breast cancer: final results from the EGF104900 Study. *J Clin Oncol,* 30**,** 2585-92.

BLAMEY, R. W., ELLIS, I. O., PINDER, S. E., LEE, A. H., MACMILLAN, R. D., MORGAN, D. A., ROBERTSON, J. F., MITCHELL, M. J., BALL, G. R., HAYBITTLE, J. L. & ELSTON, C. W. 2007a. Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. *Eur J Cancer,* 43**,** 1548-55.

BLAMEY, R. W., HORNMARK-STENSTAM, B., BALL, G., BLICHERT-TOFT, M., CATALIOTTI, L., FOURQUET, A., GEE, J., HOLLI, K., JAKESZ, R., KERIN, M., MANSEL, R., NICHOLSON, R., PIENKOWSKI, T., PINDER, S., SUNDQUIST, M., VAN DE VIJVER, M. & ELLIS, I. 2010. ONCOPOOL - a European database for 16,944 cases of breast cancer. *Eur J Cancer,* 46**,** 56-71.

BLAMEY, R. W., PINDER, S. E., BALL, G. R., ELLIS, I. O., ELSTON, C. W., MITCHELL, M. J. & HAYBITTLE, J. L. 2007b. Reading the prognosis of the individual with breast cancer. *Eur J Cancer,* 43**,** 1545-7.

BLAND, K. I., MENCK, H. R., SCOTT-CONNER, C. E., MORROW, M., WINCHESTER, D. J. & WINCHESTER, D. P. 1998. The National Cancer Data Base 10-year survey of breast carcinoma treatment at hospitals in the United States. *Cancer,* 83**,** 1262-73.

BOELLNER, S. & BECKER, K. F. 2015. Reverse Phase Protein Arrays-Quantitative Assessment of Multiple Biomarkers in Biopsies for Clinical Use. *Microarrays (Basel),* 4**,** 98-114.

BOGINA, G., ZAMBONI, G., SAPINO, A., BORTESI, L., MARCONI, M., LUNARDI, G., COATI, F., MASSOCCO, A., MOLINARO, L., PEGORARO, C. & VENTURINI, M. 2012. Comparison of anti-estrogen receptor antibodies SP1, 6F11, and 1D5 in breast cancer: lower 1D5 sensitivity but questionable clinical implications. *Am J Clin Pathol,* 138**,** 697-702.

BOMBONATI, A. & SGROI, D. C. 2011. The molecular pathology of breast cancer progression. *J Pathol,* 223**,** 307-17.

BONNET, D. & DICK, J. E. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med,* 3**,** 730-7.

BOOTH, D. G., TAKAGI, M., SANCHEZ-PULIDO, L., PETFALSKI, E., VARGIU, G., SAMEJIMA, K., IMAMOTO, N., PONTING, C. P., TOLLERVEY, D., EARNSHAW, W. C. & VAGNARELLI, P. 2014. Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *Elife,* 3**,** e01641.

BRASE, J. C., MANNSPERGER, H., FROHLICH, H., GADE, S., SCHMIDT, C., WIEMANN, S., BEISSBARTH, T., SCHLOMM, T., SULTMANN, H. & KORF, U. 2010. Increasing the sensitivity of reverse phase protein arrays by antibody-mediated signal amplification. *Proteome Sci,* 8**,** 36.

BRENNAN, R. G. & MATTHEWS, B. W. 1989. The helix-turn-helix DNA binding motif. *J Biol Chem,* 264**,** 1903-6.

BROECKX, V., BOONEN, K., PRINGELS, L., SAGAERT, X., PRENEN, H., LANDUYT, B., SCHOOFS, L. & MAES, E. 2016. Comparison of multiple protein extraction buffers for GeLC-MS/MS proteomic analysis of liver and colon formalin-fixed, paraffin-embedded tissues. *Mol Biosyst,* 12**,** 553-65.

BROWN, J., JONES, M. & BENSON, E. A. 1993. Comment on the Nottingham Prognostic Index. *Breast Cancer Res Treat,* 25**,** 283.

BROWN, J. R., DIGIOVANNA, M. P., KILLELEA, B., LANNIN, D. R. & RIMM, D. L. 2014. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. *Lab Invest,* 94**,** 98-106.

BULLWINKEL, J., BARON-LUHR, B., LUDEMANN, A., WOHLENBERG, C., GERDES, J. & SCHOLZEN, T. 2006. Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells. *J Cell Physiol,* 206**,** 624-35.

BYERS, L. A., WANG, J., NILSSON, M. B., FUJIMOTO, J., SAINTIGNY, P., YORDY, J., GIRI, U., PEYTON, M., FAN, Y. H., DIAO, L., MASRORPOUR, F., SHEN, L., LIU, W., DUCHEMANN, B., TUMULA, P., BHARDWAJ, V., WELSH, J., WEBER, S., GLISSON, B. S., KALHOR, N., WISTUBA, II, GIRARD, L., LIPPMAN, S. M., MILLS, G. B., COOMBES, K. R., WEINSTEIN, J. N., MINNA, J. D. & HEYMACH, J. V. 2012. Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel

therapeutic targets including PARP1. *Cancer Discov,* 2**,** 798-811.

CALAO, M., SEKYERE, E. O., CUI, H. J., CHEUNG, B. B., THOMAS, W. D., KEATING, J., CHEN, J. B., RAIF, A., JANKOWSKI, K., DAVIES, N. P., BEKKUM, M. V., CHEN, B., TAN, O., ELLIS, T., NORRIS, M. D., HABER, M., KIM, E. S., SHOHET, J. M., TRAHAIR, T. N., LIU, T., WAINWRIGHT, B. J., DING, H. F. & MARSHALL, G. M. 2013. Direct effects of Bmi1 on p53 protein stability inactivates oncoprotein stress responses in embryonal cancer precursor cells at tumor initiation. *Oncogene,* 32**,** 3616-26.

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.

CAO, L., BOMBARD, J., CINTRON, K., SHEEDY, J., WEETALL, M. L. & DAVIS, T. W. 2011. BMI1 as a novel target for drug discovery in cancer. *J Cell Biochem,* 112**,** 2729-41.

CARDIFF, R. D. 1998. Are the TDLU of the human the same as the LA of mice? *J Mammary Gland Biol Neoplasia,* 3**,** 3-5.

CARDOSO, F., VAN'T VEER, L., RUTGERS, E., LOI, S., MOOK, S. & PICCART-GEBHART, M. J. 2008. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol,* 26**,** 729-35.

CARMECI, C., THOMPSON, D. A., RING, H. Z., FRANCKE, U. & WEIGEL, R. J. 1997. Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics,* 45**,** 607-17.

CARTER, C. L., ALLEN, C. & HENSON, D. E. 1989. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer,* 63**,** 181-7.

CAUSSINUS, E. & GONZALEZ, C. 2005. Induction of tumor growth by altered stem-cell asymmetric division in Drosophila melanogaster. *Nat Genet,* 37**,** 1125-9.

CENCI, T., MARTINI, M., MONTANO, N., D'ALESSANDRIS, Q. G., FALCHETTI, M. L., ANNIBALI, D., SAVINO, M., BIANCHI, F., PIERCONTI, F., NASI, S., PALLINI, R. & LAROCCA, L. M. 2012. Prognostic relevance of c-Myc and BMI1 expression in patients with glioblastoma. *Am J Clin Pathol,* 138**,** 390-6.

CHAN, S. M. & UTZ, P. J. 2005. The challenge of analyzing the proteome in humans with autoimmune diseases. *Ann N Y Acad Sci,* 1062**,** 61-8.

CHANG, X., SUN, Y., HAN, S., ZHU, W., ZHANG, H. & LIAN, S. 2015. MiR-203 inhibits melanoma invasive and proliferative abilities by targeting the polycomb group gene BMI1. *Biochem Biophys Res Commun,* 456**,** 361-6.

CHARAFE-JAUFFRET, E., MONVILLE, F., GINESTIER, C., DONTU, G., BIRNBAUM, D. & WICHA, M. S. 2008. Cancer stem cells in breast: current opinion and future challenges. *Pathobiology,* 75**,** 75-84.

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.

CHEPKO, G. & SMITH, G. H. 1997. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell,* 29**,** 239-53.

CHOI, Y. J., CHOI, Y. L., CHO, E. Y., SHIN, Y. K., SUNG, K. W., HWANG, Y. K., LEE, S. J., KONG, G., LEE, J. E., KIM, J. S., KIM, J. H., YANG, J. H. & NAM, S. J. 2009. Expression of Bmi-1 protein in tumor tissues is associated with favorable prognosis in breast cancer patients. *Breast Cancer Res Treat,* 113**,** 83-93.

CHOO, J. R. & NIELSEN, T. O. 2010. Biomarkers for Basal-like Breast Cancer. *Cancers (Basel),* 2**,** 1040-65.

CHOY, B., BANDLA, S., XIA, Y., TAN, D., PENNATHUR, A., LUKETICH, J. D., GODFREY, T. E., PETERS, J. H., SUN, J. & ZHOU, Z. 2012. Clinicopathologic characteristics of high expression of Bmi-1 in esophageal adenocarcinoma and squamous cell carcinoma. *BMC Gastroenterol,* 12**,** 146.

CHRISTGEN, M., VON AHSEN, S., CHRISTGEN, H., LANGER, F. & KREIPE, H. 2015. The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer. *Hum Pathol,* 46**,** 1341-9.

CLAHSEN, P. C., VAN DE VELDE, C. J., DUVAL, C., PALLUD, C., MANDARD, A. M., DELOBELLE-DEROIDE, A., VAN DEN BROEK, L. & VAN DE VIJVER, M. J. 1999. The utility of mitotic index, oestrogen receptor and Ki-67 measurements in the creation of

novel prognostic indices for node-negative breast cancer. *Eur J Surg Oncol,* 25**,** 356-63.

CLARKE, R. B. 2005. Isolation and characterization of human mammary stem cells. *Cell Prolif,* 38**,** 375-86.

CLEVERS, H. 2005. Stem cells, asymmetric division and cancer. *Nat Genet,* 37**,** 1027-8.

COATES, A. S., WINER, E. P., GOLDHIRSCH, A., GELBER, R. D., GNANT, M., PICCART-GEBHART, M., THURLIMANN, B. & SENN, H. J. 2015. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol,* 26**,** 1533-46.

COHEN, K. J., HANNA, J. S., PRESCOTT, J. E. & DANG, C. V. 1996. Transformation by the Bmi-1 oncoprotein correlates with its subnuclear localization but not its transcriptional suppression activity. *Mol Cell Biol,* 16**,** 5527-35.

COLDITZ, G. A., ROSNER, B. A., CHEN, W. Y., HOLMES, M. D. & HANKINSON, S. E. 2004. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst,* 96**,** 218-28.

COLLEONI, M., VIALE, G., ZAHRIEH, D., PRUNERI, G., GENTILINI, O., VERONESI, P., GELBER, R. D., CURIGLIANO, G., TORRISI, R., LUINI, A., INTRA, M., GALIMBERTI, V., RENNE, G., NOLE, F., PERUZZOTTI, G. & GOLDHIRSCH, A. 2004. Chemotherapy is more effective in patients with breast cancer not expressing steroid hormone receptors: a study of preoperative treatment. *Clin Cancer Res,* 10**,** 6622-8.

COLLINS, L. C., BOTERO, M. L. & SCHNITT, S. J. 2005. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. *Am J Clin Pathol,* 123**,** 16-20.

COLOZZA, M., SIDONI, A. & PICCART-GEBHART, M. 2010. Value of Ki67 in breast cancer: the debate is still open. *Lancet Oncol,* 11**,** 414-5.

CORTES, J., FUMOLEAU, P., BIANCHI, G. V., PETRELLA, T. M., GELMON, K., PIVOT, X., VERMA, S., ALBANELL, J., CONTE, P., LLUCH, A., SALVAGNI, S., SERVENT, V., GIANNI, L., SCALTRITI, M., ROSS, G. A., DIXON, J., SZADO, T. & BASELGA, J. 2012. Pertuzumab monotherapy after trastuzumab-based treatment and subsequent reintroduction of trastuzumab: activity and tolerability in patients with advanced human

epidermal growth factor receptor 2-positive breast cancer. *J Clin Oncol,* 30**,** 1594-600.

CREIGHTON, C. J., FU, X., HENNESSY, B. T., CASA, A. J., ZHANG, Y., GONZALEZ-ANGULO, A. M., LLUCH, A., GRAY, J. W., BROWN, P. H., HILSENBECK, S. G., OSBORNE, C. K., MILLS, G. B., LEE, A. V. & SCHIFF, R. 2010. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res,* 12**,** R40.

CROWE, J. P., HUBAY, C. A., PEARSON, O. H., MARSHALL, J. S., ROSENBLATT, J., MANSOUR, E. G., HERMANN, R. E., JONES, J. C., FLYNN, W. J. & MCGUIRE, W. L. 1982. Estrogen receptor status as a prognostic indicator for stage I breast cancer patients. *Breast Cancer Res Treat,* 2**,** 171-6.

CUI, X., SCHIFF, R., ARPINO, G., OSBORNE, C. K. & LEE, A. V. 2005. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *J Clin Oncol,* 23**,** 7721-35.

CURTIS, C., SHAH, S. P., CHIN, S. F., TURASHVILI, G., RUEDA, O. M., DUNNING, M. J., SPEED, D., LYNCH, A. G., SAMARAJIWA, S., YUAN, Y., GRAF, S., HA, G., HAFFARI, G., BASHASHATI, A., RUSSELL, R., MCKINNEY, S., GROUP, M., LANGEROD, A., GREEN, A., PROVENZANO, E., WISHART, G., PINDER, S., WATSON, P., MARKOWETZ, F., MURPHY, L., ELLIS, I., PURUSHOTHAM, A., BORRESEN-DALE, A. L., BRENTON, J. D., TAVARE, S., CALDAS, C. & APARICIO, S. 2012a. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature,* 486**,** 346-52.

CURTIS, C., SHAH, S. P., CHIN, S. F., TURASHVILI, G., RUEDA, O. M., DUNNING, M. J., SPEED, D., LYNCH, A. G., SAMARAJIWA, S., YUAN, Y., GRAF, S., HA, G., HAFFARI, G., BASHASHATI, A., RUSSELL, R., MCKINNEY, S., LANGEROD, A., GREEN, A., PROVENZANO, E., WISHART, G., PINDER, S., WATSON, P., MARKOWETZ, F., MURPHY, L., ELLIS, I., PURUSHOTHAM, A., BORRESEN-DALE, A. L., BRENTON, J. D., TAVARE, S., CALDAS, C. & APARICIO, S. 2012b. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature,* 486**,** 346-52.

CUYLEN, S., BLAUKOPF, C., POLITI, A. Z., MULLER-REICHERT, T., NEUMANN, B., POSER, I., ELLENBERG, J., HYMAN, A. A. & GERLICH, D. W. 2016. Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature,* 535**,** 308-12.

CUZICK, J., DOWSETT, M., PINEDA, S., WALE, C., SALTER, J., QUINN, E., ZABAGLO, L., MALLON, E., GREEN, A. R., ELLIS, I. O.,

HOWELL, A., BUZDAR, A. U. & FORBES, J. F. 2011. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol,* 29**,** 4273-8.

CUZICK, J., SESTAK, I., BONANNI, B., COSTANTINO, J. P., CUMMINGS, S., DECENSI, A., DOWSETT, M., FORBES, J. F., FORD, L., LACROIX, A. Z., MERSHON, J., MITLAK, B. H., POWLES, T., VERONESI, U., VOGEL, V. & WICKERHAM, D. L. 2013. Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *Lancet,* 381**,** 1827-34.

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.

DANG, Z., XU, W. H., LU, P., WU, N., LIU, J., RUAN, B., ZHOU, L., SONG, W. J. & DOU, K. F. 2014. MicroRNA-135a inhibits cell proliferation by targeting Bmi1 in pancreatic ductal adenocarcinoma. *Int J Biol Sci,* 10**,** 733-45.

DAVIES, C., GODWIN, J., GRAY, R., CLARKE, M., CUTTER, D., DARBY, S., MCGALE, P., PAN, H. C., TAYLOR, C., WANG, Y. C., DOWSETT, M., INGLE, J. & PETO, R. 2011. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet,* 378**,** 771-84.

DAWSON, S. J., RUEDA, O. M., APARICIO, S. & CALDAS, C. 2013. A new genome-driven integrated classification of breast cancer and its implications. *Embo j,* 32**,** 617-28.

DE AZAMBUJA, E., CARDOSO, F., DE CASTRO, G., JR., COLOZZA, M., MANO, M. S., DURBECQ, V., SOTIRIOU, C., LARSIMONT, D., PICCART-GEBHART, M. J. & PAESMANS, M. 2007. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer,* 96**,** 1504-13.

DE KOK, J. B., ROELOFS, R. W., GIESENDORF, B. A., PENNINGS, J. L., WAAS, E. T., FEUTH, T., SWINKELS, D. W. & SPAN, P. N.

2005. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest,* 85**,** 154-9.

DENKERT, C., LOIBL, S., MULLER, B. M., EIDTMANN, H., SCHMITT, W. D., EIERMANN, W., GERBER, B., TESCH, H., HILFRICH, J., HUOBER, J., FEHM, T., BARINOFF, J., JACKISCH, C., PRINZLER, J., RUDIGER, T., ERBSTOSSER, E., BLOHMER, J. U., BUDCZIES, J., MEHTA, K. M. & VON MINCKWITZ, G. 2013. Ki67 levels as predictive and prognostic parameters in pretherapeutic breast cancer core biopsies: a translational investigation in the neoadjuvant GeparTrio trial. *Ann Oncol,* 24**,** 2786-93.

DENT, R., TRUDEAU, M., PRITCHARD, K. I., HANNA, W. M., KAHN, H. K., SAWKA, C. A., LICKLEY, L. A., RAWLINSON, E., SUN, P. & NAROD, S. A. 2007. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res,* 13**,** 4429-34.

DEOME, K. B., FAULKIN, L. J., JR., BERN, H. A. & BLAIR, P. B. 1959. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res,* 19**,** 515-20.

DERIBE, Y. L., PAWSON, T. & DIKIC, I. 2010. Post-translational modifications in signal integration. *Nat Struct Mol Biol,* 17**,** 666-72.

DEYARMIN, B., KANE, J. L., VALENTE, A. L., VAN LAAR, R., GALLAGHER, C., SHRIVER, C. D. & ELLSWORTH, R. E. 2013. Effect of ASCO/CAP guidelines for determining ER status on molecular subtype. *Ann Surg Oncol,* 20**,** 87-93.

DIAB, S. G., CLARK, G. M., OSBORNE, C. K., LIBBY, A., ALLRED, D. C. & ELLEDGE, R. M. 1999. Tumor characteristics and clinical outcome of tubular and mucinous breast carcinomas. *J Clin Oncol,* 17**,** 1442-8.

DIMRI, G. P., MARTINEZ, J. L., JACOBS, J. J., KEBLUSEK, P., ITAHANA, K., VAN LOHUIZEN, M., CAMPISI, J., WAZER, D. E. & BAND, V. 2002. The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. *Cancer Res,* 62**,** 4736-45.

DOBRESCU, A., CHANG, M., KIRTANI, V., TURI, G. K., HENNAWY, R. & HINDENBURG, A. A. 2011. Study of Estrogen Receptor and Progesterone Receptor Expression in Breast Ductal Carcinoma In Situ by Immunohistochemical Staining in ER/PgR-Negative Invasive Breast Cancer. *ISRN Oncol,* 2011**,** 673790.

DOMAGALA, W., MARKIEWSKI, M., HAREZGA, B., DUKOWICZ, A. & OSBORN, M. 1996. Prognostic significance of tumor cell proliferation rate as determined by the MIB-1 antibody in breast carcinoma: its relationship with vimentin and p53 protein. *Clin Cancer Res,* 2**,** 147-54.

DONTU, G., JACKSON, K. W., MCNICHOLAS, E., KAWAMURA, M. J., ABDALLAH, W. M. & WICHA, M. S. 2004. Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. *Breast Cancer Res,* 6**,** R605-15.

DOWSETT, M., CUZICK, J., WALE, C., HOWELL, T., HOUGHTON, J. & BAUM, M. 2005. Retrospective analysis of time to recurrence in the ATAC trial according to hormone receptor status: an hypothesis-generating study. *J Clin Oncol,* 23**,** 7512-7.

DOWSETT, M., NIELSEN, T. O., A'HERN, R., BARTLETT, J., COOMBES, R. C., CUZICK, J., ELLIS, M., HENRY, N. L., HUGH, J. C., LIVELY, T., MCSHANE, L., PAIK, S., PENAULT-LLORCA, F., PRUDKIN, L., REGAN, M., SALTER, J., SOTIRIOU, C., SMITH, I. E., VIALE, G., ZUJEWSKI, J. A. & HAYES, D. F. 2011. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst,* 103**,** 1656-64.

DOWSETT, M., SESTAK, I., LOPEZ-KNOWLES, E., SIDHU, K., DUNBIER, A. K., COWENS, J. W., FERREE, S., STORHOFF, J., SCHAPER, C. & CUZICK, J. 2013. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol,* 31**,** 2783-90.

DU MANOIR, S., GUILLAUD, P., CAMUS, E., SEIGNEURIN, D. & BRUGAL, G. 1991. Ki-67 labeling in postmitotic cells defines different Ki-67 pathways within the 2c compartment. *Cytometry,* 12**,** 455-63.

DUBEY, A. K., GUPTA, U. & JAIN, S. 2015. Breast cancer statistics and prediction methodology: a systematic review and analysis. *Asian Pac J Cancer Prev,* 16**,** 4237-45.

DUFFY, M. J. 2006. Estrogen receptors: role in breast cancer. *Crit Rev Clin Lab Sci,* 43**,** 325-47.

DUMONTET, C., KRAJEWSKA, M., TREILLEUX, I., MACKEY, J. R., MARTIN, M., RUPIN, M., LAFANECHERE, L. & REED, J. C. 2010. BCIRG 001 molecular analysis: prognostic factors in node-positive breast cancer patients receiving adjuvant chemotherapy. *Clin Cancer Res,* 16**,** 3988-97.

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.

EKHOLM, M., BEGLERBEGOVIC, S., GRABAU, D., LOVGREN, K., MALMSTROM, P., HARTMAN, L. & FERNO, M. 2014. Immunohistochemical assessment of Ki67 with antibodies SP6 and MIB1 in primary breast cancer: a comparison of prognostic value and reproducibility. *Histopathology,* 65**,** 252-60.

ELLEDGE, R. M., GREEN, S., PUGH, R., ALLRED, D. C., CLARK, G. M., HILL, J., RAVDIN, P., MARTINO, S. & OSBORNE, C. K. 2000. Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. *Int J Cancer,* 89**,** 111-7.

ELLIS, I., CARDER, P., HALES, S., LEE, A. H., PINDER, S., RAKHA, E., STEPHENSON, T., AL-SAM, S., DEB, R., HANBY, A., LIEBMANN, R., PROVENZANO, E., ROWLANDS, D., WELLS, C., ANDERSON, N., GIRLING, A., IBRAHIM, M., MALLON, E. & QUINN, C. 2016. *Pathology reporting of breast disease in surgical excision specimens*

incorporating the dataset for histological reporting of breast cancer.

ELLIS, M. J. & PEROU, C. M. 2013. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov,* 3**,** 27-34.

ELLIS, M. J., TAO, Y., LUO, J., A'HERN, R., EVANS, D. B., BHATNAGAR, A. S., CHAUDRI ROSS, H. A., VON KAMEKE, A., MILLER, W. R., SMITH, I., EIERMANN, W. & DOWSETT, M. 2008. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst,* 100**,** 1380-8.

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.

ELSTON, C. W., ELLIS, I. O. & PINDER, S. E. 1999. Pathological prognostic factors in breast cancer. *Crit Rev Oncol Hematol,* 31**,** 209-23.

ENDL, E. & GERDES, J. 2000. Posttranslational modifications of the KI-67 protein coincide with two major checkpoints during mitosis. *J Cell Physiol,* 182**,** 371-80.

ERTEL, A. 2010. Bimodal gene expression and biomarker discovery. *Cancer Inform,* 9**,** 11-4.

ESPINA, V., MEHTA, A. I., WINTERS, M. E., CALVERT, V., WULFKUHLE, J., PETRICOIN, E. F., 3RD & LIOTTA, L. A. 2003. Protein microarrays: molecular profiling technologies for clinical specimens. *Proteomics,* 3**,** 2091-100.

EYSTER, K. M. 2016. The Estrogen Receptors: An Overview from Different Perspectives. *Methods Mol Biol,* 1366**,** 1-10.

FANG, X., CAI, Y., LIU, J., WANG, Z., WU, Q., ZHANG, Z., YANG, C. J., YUAN, L. & OUYANG, G. 2011. Twist2 contributes to breast cancer progression by promoting an epithelial-mesenchymal transition and cancer stem-like cell self-renewal. *Oncogene,* 30**,** 4707-20.

FARATIAN, D., MUNRO, A., TWELVES, C. & BARTLETT, J. M. 2009. Membranous and cytoplasmic staining of Ki67 is associated with HER2 and ER status in invasive breast carcinoma. *Histopathology,* 54**,** 254-7.

FEDCHENKO, N. & REIFENRATH, J. 2014. Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn Pathol,* 9**,** 221.

FERGUSON, D. J. 1985. Intraepithelial lymphocytes and macrophages in the normal breast. *Virchows Arch A Pathol Anat Histopathol,* 407**,** 369-78.

FISHER, B., REDMOND, C., FISHER, E. R. & CAPLAN, R. 1988. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncol,* 6**,** 1076-87.

FOULKES, W. D. 2012. Size surprise? Tumour size, nodal status, and outcome after breast cancer. *Curr Oncol,* 19**,** 241-3.

FOWLER, C. B., WAYBRIGHT, T. J., VEENSTRA, T. D., O'LEARY, T. J. & MASON, J. T. 2012. Pressure-assisted protein extraction: a novel method for recovering proteins from archival tissue for proteomic analysis. *J Proteome Res,* 11**,** 2602-8.

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.

GASPARINI, G., POZZA, F. & HARRIS, A. L. 1993. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. *J Natl Cancer Inst,* 85**,** 1206-19.

GERDES, J., LEMKE, H., BAISCH, H., WACKER, H. H., SCHWAB, U. & STEIN, H. 1984. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J Immunol,* 133**,** 1710-5.

GERDES, J., SCHWAB, U., LEMKE, H. & STEIN, H. 1983. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int J Cancer,* 31**,** 13-20.

GEYER, F. C., LOPEZ-GARCIA, M. A., LAMBROS, M. B. & REIS-FILHO, J. S. 2009. Genetic characterization of breast cancer and implications for clinical management. *J Cell Mol Med,* 13**,** 4090-103.

GIANGRANDE, P. H. & MCDONNELL, D. P. 1999. The A and B isoforms of the human progesterone receptor: two functionally different transcription factors encoded by a single gene. *Recent Prog Horm Res,* 54**,** 291-313; discussion 313-4.

GIL-SANCHIS, C., CERVELLO, I., KHURANA, S., FAUS, A., VERFAILLIE, C. & SIMON, C. 2015. Contribution of different bone marrow-derived cell types in endometrial regeneration using an irradiated murine model. *Fertil Steril,* 103**,** 1596-605.e1.

GINESTIER, C., CHARAFE-JAUFFRET, E., BERTUCCI, F., EISINGER, F., GENEIX, J., BECHLIAN, D., CONTE, N., ADELAIDE, J., TOIRON, Y., NGUYEN, C., VIENS, P., MOZZICONACCI, M. J., HOULGATTE, R., BIRNBAUM, D. & JACQUEMIER, J. 2002. Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol,* 161**,** 1223-33.

GINJALA, V., NACERDDINE, K., KULKARNI, A., OZA, J., HILL, S. J., YAO, M., CITTERIO, E., VAN LOHUIZEN, M. & GANESAN, S. 2011. BMI1 is recruited to DNA breaks and contributes to DNA damage-induced H2A ubiquitination and repair. *Mol Cell Biol,* 31**,** 1972-82.

GLOYESKE, N. C., DABBS, D. J. & BHARGAVA, R. 2014. Low ER+ breast cancer: Is this a distinct group? *Am J Clin Pathol,* 141**,** 697-701.

GOETZ, M. P., SUMAN, V. J., INGLE, J. N., NIBBE, A. M., VISSCHER, D. W., REYNOLDS, C. A., LINGLE, W. L., ERLANDER, M., MA, X. J., SGROI, D. C., PEREZ, E. A. & COUCH, F. J. 2006. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clin Cancer Res,* 12**,** 2080-7.

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.

GOLDHIRSCH, A., WINER, E. P., COATES, A. S., GELBER, R. D., PICCART-GEBHART, M., THURLIMANN, B. & SENN, H. J. 2013. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol,* 24**,** 2206-23.

GOLDHIRSCH, A., WOOD, W. C., COATES, A. S., GELBER, R. D., THURLIMANN, B. & SENN, H. J. 2011. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol,* 22**,** 1736-47.

GOLDSTEIN, N. S., FERKOWICZ, M., ODISH, E., MANI, A. & HASTAH, F. 2003. Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *Am J Clin Pathol,* 120**,** 86-92.

GRANTZDORFFER, I., YUMLU, S., GIOEVA, Z., VON WASIELEWSKI, R., EBERT, M. P. & ROCKEN, C. 2010. Comparison of different tissue sampling methods for protein extraction from formalin-fixed and paraffin-embedded tissue specimens. *Exp Mol Pathol,* 88**,** 190-6.

GREEN, A. R., POWE, D. G., RAKHA, E. A., SORIA, D., LEMETRE, C., NOLAN, C. C., BARROS, F. F., MACMILLAN, R. D., GARIBALDI, J. M., BALL, G. R. & ELLIS, I. O. 2013. Identification of key clinical phenotypes of breast cancer using a reduced panel of protein biomarkers. *Br J Cancer,* 109**,** 1886-94.

GREEN, A. R., SORIA, D., STEPHEN, J., POWE, D. G., NOLAN, C. C., KUNKLER, I., THOMAS, J., KERR, G. R., JACK, W., CAMERON, D., PIPER, T., BALL, G. R., GARIBALDI, J. M., RAKHA, E. A., BARTLETT, J. M. & ELLIS, I. O. 2016. Nottingham Prognostic Index Plus: Validation of a clinical decision making tool in breast cancer in an independent series. *J Pathol Clin Res,* 2**,** 32-40.

GREEN, A. R., YOUNG, P., KRIVINSKAS, S., RAKHA, E. A., CLAIRE PAISH, E., POWE, D. G. & ELLIS, I. O. 2009. The expression of ERalpha, ERbeta and PR in lobular carcinoma in situ of the breast determined using laser microdissection and real-time PCR. *Histopathology,* 54**,** 419-27.

GREEN, S., KUMAR, V., KRUST, A., WALTER, P. & CHAMBON, P. 1986. Structural and functional domains of the estrogen receptor. *Cold Spring Harb Symp Quant Biol,* 51 Pt 2**,** 751-8.

GUO, B. H., FENG, Y., ZHANG, R., XU, L. H., LI, M. Z., KUNG, H. F., SONG, L. B. & ZENG, M. S. 2011. Bmi-1 promotes invasion and metastasis, and its elevated expression is correlated with an advanced stage of breast cancer. *Mol Cancer,* 10**,** 10.

GWAK, J. M., KIM, M., KIM, H. J., JANG, M. H. & PARK, S. Y. 2017. Expression of embryonal stem cell transcription factors in breast cancer: Oct4 as an indicator for poor clinical outcome and tamoxifen resistance. *Oncotarget,* 8**,** 36305-36318.

GYGI, S. P., ROCHON, Y., FRANZA, B. R. & AEBERSOLD, R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol,* 19**,** 1720-30.

HAJAGE, D., DE RYCKE, Y., BOLLET, M., SAVIGNONI, A., CALY, M., PIERGA, J. Y., HORLINGS, H. M., VAN DE VIJVER, M. J., VINCENT-SALOMON, A., SIGAL-ZAFRANI, B., SENECHAL, C., ASSELAIN, B., SASTRE, X. & REYAL, F. 2011. External validation of Adjuvant! Online breast cancer prognosis tool. Prioritising recommendations for improvement. *PLoS One,* 6**,** e27446.

HAMMOND, M. E., HAYES, D. F., DOWSETT, M., ALLRED, D. C., HAGERTY, K. L., BADVE, S., FITZGIBBONS, P. L., FRANCIS, G., GOLDSTEIN, N. S., HAYES, M., HICKS, D. G., LESTER, S., LOVE, R., MANGU, P. B., MCSHANE, L., MILLER, K., OSBORNE, C. K., PAIK, S., PERLMUTTER, J., RHODES, A., SASANO, H., SCHWARTZ, J. N., SWEEP, F. C., TAUBE, S., TORLAKOVIC, E. E., VALENSTEIN, P., VIALE, G., VISSCHER, D., WHEELER, T., WILLIAMS, R. B., WITTLIFF, J. L. & WOLFF, A. C. 2010a. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol,* 28**,** 2784-95.

HAMMOND, M. E., HAYES, D. F., WOLFF, A. C., MANGU, P. B. & TEMIN, S. 2010b. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Oncol Pract,* 6**,** 195-7.

HARRISON, H., FARNIE, G., HOWELL, S. J., ROCK, R. E., STYLIANOU, S., BRENNAN, K. R., BUNDRED, N. J. & CLARKE, R. B. 2010. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer Res,* 70**,** 709-18.

HARVEY, J. M., CLARK, G. M., OSBORNE, C. K. & ALLRED, D. C. 1999. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol,* 17**,** 1474-81.

HATTORI, H. 2002. Sclerosing haemangioma of the lung is positive for MIB-1 in cell membrane and cytoplasmic staining pattern. *Histopathology,* 40**,** 291-3.

HAUPT, Y., ALEXANDER, W. S., BARRI, G., KLINKEN, S. P. & ADAMS, J. M. 1991. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell,* 65**,** 753-63.

HAYBITTLE, J. L., BLAMEY, R. W., ELSTON, C. W., JOHNSON, J., DOYLE, P. J., CAMPBELL, F. C., NICHOLSON, R. I. & GRIFFITHS, K. 1982. A prognostic index in primary breast cancer. *Br J Cancer,* 45**,** 361-6.

HAYES, D. F., TROCK, B. & HARRIS, A. L. 1998. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res Treat,* 52**,** 305-19.

HEFTI, M. M., HU, R., KNOBLAUCH, N. W., COLLINS, L. C., HAIBE-KAINS, B., TAMIMI, R. M. & BECK, A. H. 2013. Estrogen receptor negative/progesterone receptor positive breast cancer is not a reproducible subtype. *Breast Cancer Res,* 15**,** R68.

HENNESSY, B. T., GONZALEZ-ANGULO, A. M., STEMKE-HALE, K., GILCREASE, M. Z., KRISHNAMURTHY, S., LEE, J. S., FRIDLYAND, J., SAHIN, A., AGARWAL, R., JOY, C., LIU, W., STIVERS, D., BAGGERLY, K., CAREY, M., LLUCH, A., MONTEAGUDO, C., HE, X., WEIGMAN, V., FAN, C., PALAZZO, J., HORTOBAGYI, G. N., NOLDEN, L. K., WANG, N. J., VALERO, V., GRAY, J. W., PEROU, C. M. & MILLS, G. B. 2009. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res,* 69**,** 4116-24.

HENNESSY, B. T., LU, Y., GONZALEZ-ANGULO, A. M., CAREY, M. S., MYHRE, S., JU, Z., DAVIES, M. A., LIU, W., COOMBES, K., MERIC-BERNSTAM, F., BEDROSIAN, I., MCGAHREN, M., AGARWAL, R., ZHANG, F., OVERGAARD, J., ALSNER, J., NEVE, R. M., KUO, W. L., GRAY, J. W., BORRESEN-DALE, A. L. & MILLS, G. B. 2010. A Technical Assessment of the Utility of

Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clin Proteomics,* 6**,** 129-51.

HENSON, D. E., RIES, L., FREEDMAN, L. S. & CARRIAGA, M. 1991. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index. *Cancer,* 68**,** 2142-9.

HERSCHKOWITZ, J. I., SIMIN, K., WEIGMAN, V. J., MIKAELIAN, I., USARY, J., HU, Z., RASMUSSEN, K. E., JONES, L. P., ASSEFNIA, S., CHANDRASEKHARAN, S., BACKLUND, M. G., YIN, Y., KHRAMTSOV, A. I., BASTEIN, R., QUACKENBUSH, J., GLAZER, R. I., BROWN, P. H., GREEN, J. E., KOPELOVICH, L., FURTH, P. A., PALAZZO, J. P., OLOPADE, O. I., BERNARD, P. S., CHURCHILL, G. A., VAN DYKE, T. & PEROU, C. M. 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol,* 8**,** R76.

HESS, K. R., ANDERSON, K., SYMMANS, W. F., VALERO, V., IBRAHIM, N., MEJIA, J. A., BOOSER, D., THERIAULT, R. L., BUZDAR, A. U., DEMPSEY, P. J., ROUZIER, R., SNEIGE, N., ROSS, J. S., VIDAURRE, T., GOMEZ, H. L., HORTOBAGYI, G. N. & PUSZTAI, L. 2006. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol,* 24**,** 4236-44.

HIROKAWA, M. & CARNEY, J. A. 2000. Cell membrane and cytoplasmic staining for MIB-1 in hyalinizing trabecular adenoma of the thyroid gland. *Am J Surg Pathol,* 24**,** 575-8.

HODI, Z., CHAKRABARTI, J., LEE, A. H., RONAN, J. E., ELSTON, C. W., CHEUNG, K. L., ROBERTSON, J. F. & ELLIS, I. O. 2007. The reliability of assessment of oestrogen receptor expression on needle core biopsy specimens of invasive carcinomas of the breast. *J Clin Pathol,* 60**,** 299-302.

HOFMANN, K. & BUCHER, P. 1995. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem Sci,* 20**,** 347-9.

HONMA, N., HORII, R., IWASE, T., SAJI, S., YOUNES, M., ITO, Y. & AKIYAMA, F. 2015. Ki-67 evaluation at the hottest spot predicts clinical outcome of patients with hormone receptor-positive/HER2-negative breast cancer treated with adjuvant tamoxifen monotherapy. *Breast Cancer,* 22**,** 71-8.

HOUGHTON, J., STOICOV, C., NOMURA, S., ROGERS, A. B., CARLSON, J., LI, H., CAI, X., FOX, J. G., GOLDENRING, J. R. & WANG, T. C. 2004. Gastric cancer originating from bone marrow-derived cells. *Science,* 306**,** 1568-71.

HOWARD, B. A. & GUSTERSON, B. A. 2000. Human breast development. *J Mammary Gland Biol Neoplasia,* 5**,** 119-37.

HOWELL, A., CUZICK, J., BAUM, M., BUZDAR, A., DOWSETT, M., FORBES, J. F., HOCTIN-BOES, G., HOUGHTON, J., LOCKER, G. Y. & TOBIAS, J. S. 2005. Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet,* 365**,** 60-2.

HU, 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.

HUGH, J., HANSON, J., CHEANG, M. C., NIELSEN, T. O., PEROU, C. M., DUMONTET, C., REED, J., KRAJEWSKA, M., TREILLEUX, I., RUPIN, M., MAGHERINI, E., MACKEY, J., MARTIN, M. & VOGEL, C. 2009. Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *J Clin Oncol,* 27**,** 1168-76.

IBARRA-DRENDALL, C., TROCH, M. M., BARRY, W. T., BROADWATER, G., PETRICOIN, E. F., 3RD, WULFKUHLE, J., LIOTTA, L. A., LEM, S., BAKER, J. C., JR., FORD, A. C., WILKE, L. G., ZALLES, C., KUDERER, N. M., HOFFMAN, A. W., SHIVRAJ, M., MEHTA, P., WILLIAMS, J., TOLBERT, N., LEE, L. W., PILIE, P. G., YU, D. & SEEWALDT, V. L. 2012. Pilot and feasibility study: prospective proteomic profiling of mammary epithelial cells from high-risk women provides evidence of activation of pro-survival pathways. *Breast Cancer Res Treat,* 132**,** 487-98.

IGNATIADIS, M. & SOTIRIOU, C. 2008. Understanding the molecular basis of histologic grade. *Pathobiology,* 75**,** 104-11.

IKEDA, K., MONDEN, T., KANOH, T., TSUJIE, M., IZAWA, H., HABA, A., OHNISHI, T., SEKIMOTO, M., TOMITA, N., SHIOZAKI, H. & MONDEN, M. 1998. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J Histochem Cytochem,* 46**,** 397-403.

INGOLF, J. B., RUSSALINA, M., SIMONA, M., JULIA, R., GILDA, S., BOHLE, R. M., ANDREA, H., ERICH, S. & DANIEL, H. 2014. Can ki-67 play a role in prediction of breast cancer patients' response to neoadjuvant chemotherapy? *Biomed Res Int,* 2014**,** 628217.

IOANNIDIS, J. P., ALLISON, D. B., BALL, C. A., COULIBALY, I., CUI, X., CULHANE, A. C., FALCHI, M., FURLANELLO, C., GAME, L., JURMAN, G., MANGION, J., MEHTA, T., NITZBERG, M., PAGE, G. P., PETRETTO, E. & VAN NOORT, V. 2009. Repeatability of published microarray gene expression analyses. *Nat Genet,* 41**,** 149-55.

ITAHANA, K., ZOU, Y., ITAHANA, Y., MARTINEZ, J. L., BEAUSEJOUR, C., JACOBS, J. J., VAN LOHUIZEN, M., BAND, V., CAMPISI, J. & DIMRI, G. P. 2003. Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. *Mol Cell Biol,* 23**,** 389-401.

IWAMOTO, T., BOOSER, D., VALERO, V., MURRAY, J. L., KOENIG, K., ESTEVA, F. J., UENO, N. T., ZHANG, J., SHI, W., QI, Y., MATSUOKA, J., YANG, E. J., HORTOBAGYI, G. N., HATZIS, C., SYMMANS, W. F. & PUSZTAI, L. 2012. Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. *J Clin Oncol,* 30**,** 729-34.

JACOBS, J. J., KIEBOOM, K., MARINO, S., DEPINHO, R. A. & VAN LOHUIZEN, M. 1999. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature,* 397**,** 164-8.

JALAVA, P., KUOPIO, T., JUNTTI-PATINEN, L., KOTKANSALO, T., KRONQVIST, P. & COLLAN, Y. 2006. Ki67 immunohistochemistry: a valuable marker in prognostication but with a risk of misclassification: proliferation subgroups formed based on Ki67 immunoreactivity and standardized mitotic index. *Histopathology,* 48**,** 674-82.

JENSEN, E. V. & JACOBSON, H. I. 1962. Basic guides to the mechanism of estrogen action. . *Recent Prog Horm Res,,* 18**,** 387.

JENSEN, E. V. & JORDAN, V. C. 2003. The estrogen receptor: a model for molecular medicine. *Clin Cancer Res,* 9**,** 1980-9.

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., MA, X. J., LI, H., SALUNGA, R., KESTY, N. C., ERLANDER, M. G., SGROI, D. C., HOLMLUND, B., SKOOG, L., FORNANDER, T., NORDENSKJOLD, B. & STAL, O. 2011. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer,* 104**,** 1762-9.

JIANG, X., JIANG, X., FENG, S., TIAN, R., YE, M. & ZOU, H. 2007. Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J Proteome Res,* 6**,** 1038-47.

JOHNSON, M. 2010. *Anatomy and physiology of the breast. In: Jatoi I,*

Kaufmann M, eds. Management of Breast Diseases., Springer-Verlag Berlin Heidelberg.

KAMEL, D., BRADY, B., TABCHY, A., MILLS, G. B. & HENNESSY, B. 2012. Proteomic classification of breast cancer. *Curr Drug Targets,* 13**,** 1495-509.

KAMETAKA, A., TAKAGI, M., HAYAKAWA, T., HARAGUCHI, T., HIRAOKA, Y. & YONEDA, Y. 2002. Interaction of the chromatin compaction-inducing domain (LR domain) of Ki-67 antigen with HP1 proteins. *Genes Cells,* 7**,** 1231-42.

KARAMBOULAS, C. & AILLES, L. 2013. Developmental signaling pathways in cancer stem cells of solid tumors. *Biochim Biophys Acta,* 1830**,** 2481-95.

KASTNER, P., KRUST, A., TURCOTTE, B., STROPP, U., TORA, L., GRONEMEYER, H. & CHAMBON, P. 1990. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *Embo j,* 9**,** 1603-14.

KATOH, Y. & KATOH, M. 2008. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review). *Int J Mol Med,* 22**,** 271-5.

KENNECKE, H., YERUSHALMI, R., WOODS, R., CHEANG, M. C., VODUC, D., SPEERS, C. H., NIELSEN, T. O. & GELMON, K. 2010. Metastatic behavior of breast cancer subtypes. *J Clin Oncol,* 28**,** 3271-7.

KHOSHNOUD, M. R., LOFDAHL, B., FOHLIN, H., FORNANDER, T., STAL, O., SKOOG, L., BERGH, J. & NORDENSKJOLD, B. 2011. Immunohistochemistry compared to cytosol assays for determination of estrogen receptor and prediction of the long-term effect of adjuvant tamoxifen. *Breast Cancer Res Treat,* 126**,** 421-30.

KIM, J. H., YOON, S. Y., JEONG, S. H., KIM, S. Y., MOON, S. K., JOO, J. H., LEE, Y., CHOE, I. S. & KIM, J. W. 2004. Overexpression of Bmi-1 oncoprotein correlates with axillary lymph node metastases in invasive ductal breast cancer. *Breast,* 13**,** 383-8.

KOBAYASHI, T., IWAYA, K., MORIYA, T., YAMASAKI, T., TSUDA, H., YAMAMOTO, J. & MATSUBARA, O. 2013. A simple immunohistochemical panel comprising 2 conventional markers, Ki67 and p53, is a powerful tool for predicting patient outcome in luminal-type breast cancer. *BMC Clin Pathol,* 13**,** 5.

KOBIERZYCKI, C., PULA, B., WOJNAR, A., PODHORSKA-OKOLOW, M. & DZIEGIEL, P. 2012. Tissue microarray technique in evaluation of proliferative activity in invasive ductal breast cancer. *Anticancer Res,* 32**,** 773-7.

KUIPER, G. G., ENMARK, E., PELTO-HUIKKO, M., NILSSON, S. & GUSTAFSSON, J. A. 1996. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A,* 93**,** 5925-30.

LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-5.

LANDIS, J. R. & KOCH, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics,* 33**,** 159-74.

LAURINAVICIUS, A., PLANCOULAINE, B., LAURINAVICIENE, A., HERLIN, P., MESKAUSKAS, R., BALTRUSAITYTE, I., BESUSPARIS, J., DASEVICIUS, D., ELIE, N., IQBAL, Y. & BOR, C. 2014. A methodology to ensure and improve accuracy of Ki67 labelling index estimation by automated digital image analysis in breast cancer tissue. *Breast Cancer Res,* 16**,** R35.

LEHMANN, B. D., BAUER, J. A., CHEN, X., SANDERS, M. E., CHAKRAVARTHY, A. B., SHYR, Y. & PIETENPOL, J. A. 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest,* 121**,** 2750-67.

LEHR, H. A., HANSEN, D. A., KUSSICK, S., LI, M., HWANG, H., KRUMMENAUER, F., TROUET, S. & GOWN, A. M. 1999. Assessment of proliferative activity in breast cancer: MIB-1 immunohistochemistry versus mitotic figure count. *Hum Pathol,* 30**,** 1314-20.

LEMOUND, J., STUCKI-KOCH, A., STOETZER, M., KOKEMULLER, H., GELLRICH, N. C., KREIPE, H. & HUSSEIN, K. 2015. Aberrant expression of caspase 14 in salivary gland carcinomas. *J Oral Pathol Med,* 44**,** 444-8.

LEONHARDT, S. A., BOONYARATANAKORNKIT, V. & EDWARDS, D. P. 2003. Progesterone receptor transcription and non-transcription signaling mechanisms. *Steroids,* 68**,** 761-70.

LESSARD, J. & SAUVAGEAU, G. 2003. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature,* 423**,** 255-60.

LEVENSON, V. V. 2007. Biomarkers for early detection of breast cancer: what, when, and where? *Biochim Biophys Acta,* 1770**,** 847-56.

LEWIS, F., JACKSON, P., LANE, S., COAST, G. & HANBY, A. M. 2004. Testing for HER2 in breast cancer. *Histopathology,* 45**,** 207-17.

LI, Z., CAO, R., WANG, M., MYERS, M. P., ZHANG, Y. & XU, R. M. 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J Biol Chem,* 281**,** 20643-9.

LIN, C. M., JASWAL, J., VANDENBERG, T., TUCK, A. & BRACKSTONE, M. 2013. Weakly hormone receptor-positive breast cancer and use of adjuvant hormonal therapy. *Curr Oncol,* 20**,** e612-3.

LIN, M. X., WEN, Z. F., FENG, Z. Y. & HE, D. 2008. [Expression and significance of Bmi-1 and Ki67 in colorectal carcinoma tissues]. *Ai Zheng,* 27**,** 1321-6.

LIU, J., CAMPEN, A., HUANG, S., PENG, S. B., YE, X., PALAKAL, M., DUNKER, A. K., XIA, Y. & LI, S. 2008a. Identification of a gene signature in cell cycle pathway for breast cancer prognosis using gene expression profiling data. *BMC Med Genomics,* 1**,** 39.

LIU, J. H., SONG, L. B., ZHANG, X., GUO, B. H., FENG, Y., LI, X. X., LIAO, W. T., ZENG, M. S. & HUANG, K. H. 2008b. Bmi-1 expression predicts prognosis for patients with gastric carcinoma. *J Surg Oncol,* 97**,** 267-72.

LIU, R., WANG, X., CHEN, G. Y., DALERBA, P., GURNEY, A., HOEY, T., SHERLOCK, G., LEWICKI, J., SHEDDEN, K. & CLARKE, M. F. 2007. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med,* 356**,** 217-26.

LIU, S., DONTU, G., MANTLE, I. D., PATEL, S., AHN, N. S., JACKSON, K. W., SURI, P. & WICHA, M. S. 2006. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res,* 66**,** 6063-71.

LIU, Y., LIU, F., YU, H., ZHAO, X., SASHIDA, G., DEBLASIO, A., HARR, M., SHE, Q. B., CHEN, Z., LIN, H. K., DI GIANDOMENICO, S., ELF, S. E., YANG, Y., MIYATA, Y., HUANG, G., MENENDEZ, S., MELLINGHOFF, I. K., ROSEN, N., PANDOLFI, P. P., HEDVAT, C. V. & NIMER, S. D. 2012. Akt phosphorylates the transcriptional repressor bmi1 to block its effects on the tumor-suppressing ink4a-arf locus. *Sci Signal,* 5**,** ra77.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LOEBKE, C., SUELTMANN, H., SCHMIDT, C., HENJES, F., WIEMANN, S., POUSTKA, A. & KORF, U. 2007. Infrared-based protein detection arrays for quantitative proteomics. *Proteomics,* 7**,** 558-64.

LUEBKER, S. A., WOJTKIEWICZ, M. & KOEPSELL, S. A. 2015. Two methods for proteomic analysis of formalin-fixed, paraffin embedded tissue result in differential protein identification, data quality, and cost. *Proteomics,* 15**,** 3744-53.

LUNDIN, J., LUNDIN, M., HOLLI, K., KATAJA, V., ELOMAA, L., PYLKKANEN, L., TURPEENNIEMI-HUJANEN, T. & JOENSUU, H. 2001. Omission of histologic grading from clinical decision making may result in overuse of adjuvant therapies in breast cancer: results from a nationwide study. *J Clin Oncol,* 19**,** 28-36.

LUPORSI, E., ANDRE, F., SPYRATOS, F., MARTIN, P. M., JACQUEMIER, J., PENAULT-LLORCA, F., TUBIANA-MATHIEU, N., SIGAL-ZAFRANI, B., ARNOULD, L., GOMPEL, A., EGELE, C., POULET, B., CLOUGH, K. B., CROUET, H., FOURQUET, A., LEFRANC, J. P., MATHELIN, C., ROUYER, N., SERIN, D., SPIELMANN, M., HAUGH, M., CHENARD, M. P., BRAIN, E., DE CREMOUX, P. & BELLOCQ, J. P. 2012. Ki-67: level of evidence and methodological considerations for its role in the clinical management of breast cancer: analytical and critical review. *Breast Cancer Res Treat,* 132**,** 895-915.

LUSA, L., MCSHANE, L. M., REID, J. F., DE CECCO, L., AMBROGI, F., BIGANZOLI, E., GARIBOLDI, M. & PIEROTTI, M. A. 2007. Challenges in projecting clustering results across gene expression-profiling datasets. *J Natl Cancer Inst,* 99**,** 1715-23.

MA, W., XIAO, G. G., MAO, J., LU, Y., SONG, B., WANG, L., FAN, S., FAN, P., HOU, Z., LI, J., YU, X., WANG, B., WANG, H., WANG, H., XU, F., LI, Y., LIU, Q. & LI, L. 2015. Dysregulation of the miR-34a-SIRT1 axis inhibits breast cancer stemness. *Oncotarget,* 6**,** 10432-44.

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., 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.

MACCALLUM, D. E. & HALL, P. A. 1999. Biochemical characterization of pKi67 with the identification of a mitotic-specific form associated with hyperphosphorylation and altered DNA binding. *Exp Cell Res,* 252**,** 186-98.

MANI, S. A., GUO, W., LIAO, M. J., EATON, E. N., AYYANAN, A., ZHOU, A. Y., BROOKS, M., REINHARD, F., ZHANG, C. C., SHIPITSIN, M., CAMPBELL, L. L., POLYAK, K., BRISKEN, C., YANG, J. & WEINBERG, R. A. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell,* 133**,** 704-15.

MARTINET, W., ABBELOOS, V., VAN ACKER, N., DE MEYER, G. R., HERMAN, A. G. & KOCKX, M. M. 2004. Western blot analysis of a limited number of cells: a valuable adjunct to proteome analysis of paraffin wax-embedded, alcohol-fixed tissue after laser capture microdissection. *J Pathol,* 202**,** 382-8.

MAS, A., CERVELLO, I., FERNANDEZ-ALVAREZ, A., FAUS, A., DIAZ, A., BURGUES, O., CASADO, M. & SIMON, C. 2015. Overexpression of the truncated form of High Mobility Group A proteins (HMGA2) in human myometrial cells induces leiomyoma-like tissue formation. *Mol Hum Reprod,* 21**,** 330-8.

MCCARTY, K. S., JR., MILLER, L. S., COX, E. B., KONRATH, J. & MCCARTY, K. S., SR. 1985. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med,* 109**,** 716-21.

MERCHANT, A. A. & MATSUI, W. 2010. Targeting Hedgehog--a cancer stem cell pathway. *Clin Cancer Res,* 16**,** 3130-40.

MIKAMI, Y., UENO, T., YOSHIMURA, K., TSUDA, H., KUROSUMI, M., MASUDA, S., HORII, R., TOI, M. & SASANO, H. 2013. Interobserver concordance of Ki67 labeling index in breast cancer: Japan Breast Cancer Research Group Ki67 ring study. *Cancer Sci,* 104**,** 1539-43.

MOLOFSKY, A. V., HE, S., BYDON, M., MORRISON, S. J. & PARDAL, R. 2005. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. *Genes Dev,* 19**,** 1432-7.

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,* 10**,** 1070-6.

MOORHEAD, G. B., TRINKLE-MULCAHY, L., NIMICK, M., DE WEVER, V., CAMPBELL, D. G., GOURLAY, R., LAM, Y. W. & LAMOND, A. I. 2008. Displacement affinity chromatography of protein phosphatase one (PP1) complexes. *BMC Biochem,* 9**,** 28.

MOREL, A. P., LIEVRE, M., THOMAS, C., HINKAL, G., ANSIEAU, S. & PUISIEUX, A. 2008. Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One,* 3**,** e2888.

MOSSELMAN, S., POLMAN, J. & DIJKEMA, R. 1996. ER beta: identification and characterization of a novel human estrogen receptor. *FEBS Lett,* 392**,** 49-53.

MUFTAH, A. A., ALESKANDARANY, M., SONBUL, S. N., NOLAN, C. C., DIEZ RODRIGUEZ, M., CALDAS, C., ELLIS, I. O., GREEN, A. R.

& RAKHA, E. A. 2016. Further Evidence to Support Bimodality of Oestrogen Receptor Expression in Breast Cancer. *Histopathology*.

MUFTAH, A. A., ALESKANDARANY, M. A., AL-KAABI, M. M., SONBUL, S. N., DIEZ-RODRIGUEZ, M., NOLAN, C. C., CALDAS, C., ELLIS, I. O., RAKHA, E. A. & GREEN, A. R. 2017. Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections. *Breast Cancer Res Treat*.

MUKAKA, M. M. 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J,* 24**,** 69-71.

MULAC-JERICEVIC, B., LYDON, J. P., DEMAYO, F. J. & CONNEELY, O. M. 2003. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc Natl Acad Sci U S A,* 100**,** 9744-9.

MYLONA, E., GIANNOPOULOU, I., FASOMYTAKIS, E., NOMIKOS, A., MAGKOU, C., BAKARAKOS, P. & NAKOPOULOU, L. 2008. The clinicopathologic and prognostic significance of CD44+/CD24(-/low) and CD44-/CD24+ tumor cells in invasive breast carcinomas. *Hum Pathol,* 39**,** 1096-102.

NACERDDINE, K., BEAUDRY, J. B., GINJALA, V., WESTERMAN, B., MATTIROLI, F., SONG, J. Y., VAN DER POEL, H., PONZ, O. B., PRITCHARD, C., CORNELISSEN-STEIJGER, P., ZEVENHOVEN, J., TANGER, E., SIXMA, T. K., GANESAN, S. & VAN LOHUIZEN, M. 2012. Akt-mediated phosphorylation of Bmi1 modulates its oncogenic potential, E3 ligase activity, and DNA damage repair activity in mouse prostate cancer. *J Clin Invest,* 122**,** 1920-32.

NADERI, A., TESCHENDORFF, A. E., BARBOSA-MORAIS, N. L., PINDER, S. E., GREEN, A. R., POWE, D. G., ROBERTSON, J. F., APARICIO, S., ELLIS, I. O., BRENTON, J. D. & CALDAS, C. 2007. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene,* 26**,** 1507-16.

NADJI, M., GOMEZ-FERNANDEZ, C., GANJEI-AZAR, P. & MORALES, A. R. 2005. Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers. *Am J Clin Pathol,* 123**,** 21-7.

NARDULLI, A. M., GREENE, G. L., O'MALLEY, B. W. & KATZENELLENBOGEN, B. S. 1988. Regulation of progesterone receptor messenger ribonucleic acid and protein levels in MCF-7 cells by estradiol: analysis of estrogen's effect on progesterone receptor synthesis and degradation. *Endocrinology,* 122**,** 935-44.

NEGM, O. H., MUFTAH, A. A., ALESKANDARANY, M. A., HAMED, M. R., AHMAD, D. A., NOLAN, C. C., DIEZ-RODRIGUEZ, M., TIGHE, P. J., ELLIS, I. O., RAKHA, E. A. & GREEN, A. R. 2016. Clinical utility of reverse phase protein array for molecular classification of breast cancer. *Breast Cancer Res Treat,* 155**,** 25-35.

NETWORK, C. G. A. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70.

NGUYEN, L. V., VANNER, R., DIRKS, P. & EAVES, C. J. 2012. Cancer stem cells: an evolving concept. *Nat Rev Cancer,* 12**,** 133-43.

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.

NIELSEN, T. O., PARKER, J. S., LEUNG, S., VODUC, D., EBBERT, M., VICKERY, T., DAVIES, S. R., SNIDER, J., STIJLEMAN, I. J., REED, J., CHEANG, M. C., MARDIS, E. R., PEROU, C. M., BERNARD, P. S. & ELLIS, M. J. 2010. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res,* 16**,** 5222-32.

NISHIZUKA, S., CHARBONEAU, L., YOUNG, L., MAJOR, S., REINHOLD, W. C., WALTHAM, M., KOUROS-MEHR, H., BUSSEY, K. J., LEE, J. K., ESPINA, V., MUNSON, P. J., PETRICOIN, E., 3RD, LIOTTA, L. A. & WEINSTEIN, J. N. 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A,* 100**,** 14229-34.

NUSSE, R. & VARMUS, H. E. 1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell,* 31**,** 99-109.

O'ROURKE, J. P. & NESS, S. A. 2008. Alternative RNA splicing produces multiple forms of c-Myb with unique transcriptional activities. *Mol Cell Biol,* 28**,** 2091-101.

OFFERSEN, B. V., SORENSEN, F. B., KNOOP, A. & OVERGAARD, J. 2003. The prognostic relevance of estimates of proliferative activity in early breast cancer. *Histopathology,* 43**,** 573-82.

OGAWA, S., INOUE, S., WATANABE, T., HIROI, H., ORIMO, A., HOSOI, T., OUCHI, Y. & MURAMATSU, M. 1998. The complete primary structure of human estrogen receptor beta (hER beta)

and its heterodimerization with ER alpha in vivo and in vitro. *Biochem Biophys Res Commun,* 243**,** 122-6.

OGLE, B. M., CASCALHO, M. & PLATT, J. L. 2005. Biological implications of cell fusion. *Nat Rev Mol Cell Biol,* 6**,** 567-75.

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.

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.

PALMER-TOY, D. E., KRASTINS, B., SARRACINO, D. A., NADOL, J. B., JR. & MERCHANT, S. N. 2005. Efficient method for the proteomic analysis of fixed and embedded tissues. *J Proteome Res,* 4**,** 2404-11.

PARK, I. K., MORRISON, S. J. & CLARKE, M. F. 2004. Bmi1, stem cells, and senescence regulation. *J Clin Invest,* 113**,** 175-9.

PARK, I. K., QIAN, D., KIEL, M., BECKER, M. W., PIHALJA, M., WEISSMAN, I. L., MORRISON, S. J. & CLARKE, M. F. 2003. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature,* 423**,** 302-5.

PARKER, J. S., MULLINS, M., CHEANG, M. C., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X., HU, Z., QUACKENBUSH, J. F., STIJLEMAN, I. J., PALAZZO, J., MARRON, J. S., NOBEL, A. B., MARDIS, E., NIELSEN, T. O., ELLIS, M. J., PEROU, C. M. & BERNARD, P. S. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol,* 27**,** 1160-7.

PATHMANATHAN, N., BALLEINE, R. L., JAYASINGHE, U. W., BILINSKI, K. L., PROVAN, P. J., BYTH, K., BILOUS, A. M., SALISBURY, E. L. & BOYAGES, J. 2014. The prognostic value of Ki67 in

systemically untreated patients with node-negative breast cancer. *J Clin Pathol,* 67**,** 222-8.

PAWELEK, J. M. 2000. Tumour cell hybridization and metastasis revisited. *Melanoma Res,* 10**,** 507-14.

PAWELETZ, C. P., CHARBONEAU, L., BICHSEL, V. E., SIMONE, N. L., CHEN, T., GILLESPIE, J. W., EMMERT-BUCK, M. R., ROTH, M. J., PETRICOIN, I. E. & LIOTTA, L. A. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene,* 20**,** 1981-9.

PENAULT-LLORCA, F., ANDRE, F., SAGAN, C., LACROIX-TRIKI, M., DENOUX, Y., VERRIELE, V., JACQUEMIER, J., BARANZELLI, M. C., BIBEAU, F., ANTOINE, M., LAGARDE, N., MARTIN, A. L., ASSELAIN, B. & ROCHE, H. 2009. Ki67 expression and docetaxel efficacy in patients with estrogen receptor-positive breast cancer. *J Clin Oncol,* 27**,** 2809-15.

PEPINSKY, R. B., RAYHORN, P., DAY, E. S., DERGAY, A., WILLIAMS, K. P., GALDES, A., TAYLOR, F. R., BORIACK-SJODIN, P. A. & GARBER, E. A. 2000. Mapping sonic hedgehog-receptor interactions by steric interference. *J Biol Chem,* 275**,** 10995-1001.

PEPPERCORN, J., PEROU, C. M. & CAREY, L. A. 2008. Molecular subtypes in breast cancer evaluation and management: divide and conquer. *Cancer Invest,* 26**,** 1-10.

PEREIRA, H., PINDER, S. E., SIBBERING, D. M., GALEA, M. H., ELSTON, C. W., BLAMEY, R. W., ROBERTSON, J. F. & ELLIS, I. O. 1995. Pathological prognostic factors in breast cancer. IV: Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. *Histopathology,* 27**,** 219-26.

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.

PETZ, L. N., ZIEGLER, Y. S., SCHULTZ, J. R., KIM, H., KEMPER, J. K. & NARDULLI, A. M. 2004. Differential regulation of the human progesterone receptor gene through an estrogen response element half site and Sp1 sites. *J Steroid Biochem Mol Biol,* 88**,** 113-22.

PICHON, M. F., BROET, P., MAGDELENAT, H., DELARUE, J. C., SPYRATOS, F., BASUYAU, J. P., SAEZ, S., RALLET, A., COURRIERE, P., MILLON, R. & ASSELAIN, B. 1996. Prognostic value of steroid receptors after long-term follow-up of 2257 operable breast cancers. *Br J Cancer,* 73**,** 1545-51.

PIETERSEN, A. M., HORLINGS, H. M., HAUPTMANN, M., LANGEROD, A., AJOUAOU, A., CORNELISSEN-STEIJGER, P., WESSELS, L. F., JONKERS, J., VAN DE VIJVER, M. J. & VAN LOHUIZEN, M. 2008. EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. *Breast Cancer Res,* 10**,** R109.

PINTO, A. E., AREIA, F., PEREIRA, T., CARDOSO, P., APARICIO, M., SILVA, G. L., FERREIRA, M. C. & ANDRE, S. 2013. Clinical relevance of the reappraisal of negative hormone receptor expression in breast cancer. *Springerplus,* 2**,** 375.

POLLEY, M. Y., FREIDLIN, B., KORN, E. L., CONLEY, B. A., ABRAMS, J. S. & MCSHANE, L. M. 2013a. Statistical and practical considerations for clinical evaluation of predictive biomarkers. *J Natl Cancer Inst,* 105**,** 1677-83.

POLLEY, M. Y., LEUNG, S. C., MCSHANE, L. M., GAO, D., HUGH, J. C., MASTROPASQUA, M. G., VIALE, G., ZABAGLO, L. A., PENAULT-LLORCA, F., BARTLETT, J. M., GOWN, A. M., SYMMANS, W. F., PIPER, T., MEHL, E., ENOS, R. A., HAYES, D. F., DOWSETT, M. & NIELSEN, T. O. 2013b. An international Ki67 reproducibility study. *J Natl Cancer Inst,* 105**,** 1897-906.

PONTI, D., COSTA, A., ZAFFARONI, N., PRATESI, G., PETRANGOLINI, G., CORADINI, D., PILOTTI, S., PIEROTTI, M. A. & DAIDONE, M. G. 2005. Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res,* 65**,** 5506-11.

POURTEIMOOR, V., MOHAMMADI-YEGANEH, S. & PARYAN, M. 2016. Breast cancer classification and prognostication through diverse systems along with recent emerging findings in this respect; the dawn of new perspectives in the clinical applications. *Tumour Biol,* 37**,** 14479-14499.

PRAT, A., PARKER, J. S., KARGINOVA, O., FAN, C., LIVASY, C., HERSCHKOWITZ, J. I., HE, X. & PEROU, C. M. 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res,* 12**,** R68.

PRAT, A. & PEROU, C. M. 2011. Deconstructing the molecular portraits of breast cancer. *Mol Oncol,* 5**,** 5-23.

PRENZEL, N., FISCHER, O. M., STREIT, S., HART, S. & ULLRICH, A. 2001. The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr Relat Cancer,* 8**,** 11-31.

PRINCE, M. E., SIVANANDAN, R., KACZOROWSKI, A., WOLF, G. T., KAPLAN, M. J., DALERBA, P., WEISSMAN, I. L., CLARKE, M. F. & AILLES, L. E. 2007. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci U S A,* 104**,** 973-8.

RAAPHORST, F. M. 2005. Deregulated expression of Polycomb-group oncogenes in human malignant lymphomas and epithelial tumors. *Hum Mol Genet,* 14 Spec No 1**,** R93-r100.

RAILO, M., LUNDIN, J., HAGLUND, C., VON SMITTEN, K., VON BOGUSLAWSKY, K. & NORDLING, S. 1997. Ki-67, p53, Er-receptors, ploidy and S-phase as prognostic factors in T1 node negative breast cancer. *Acta Oncol,* 36**,** 369-74.

RAKHA, E. A., EL-SAYED, M. E., GREEN, A. R., PAISH, E. C., POWE, D. G., GEE, J., NICHOLSON, R. I., LEE, A. H., ROBERTSON, J. F. & ELLIS, I. O. 2007. Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. *J Clin Oncol,* 25**,** 4772-8.

RAKHA, E. A., EL-SAYED, M. E., LEE, A. H., ELSTON, C. W., GRAINGE, M. J., HODI, Z., BLAMEY, R. W. & ELLIS, I. O. 2008. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol,* 26**,** 3153-8.

RAKHA, E. A. & ELLIS, I. O. 2011. Modern classification of breast cancer: should we stick with morphology or convert to molecular profile characteristics. *Adv Anat Pathol,* 18**,** 255-67.

RAKHA, E. A., ELSHEIKH, S. E., ALESKANDARANY, M. A., HABASHI, H. O., GREEN, A. R., POWE, D. G., EL-SAYED, M. E., BENHASOUNA, A., BRUNET, J. S., AKSLEN, L. A., EVANS, A. J., BLAMEY, R., REIS-FILHO, J. S., FOULKES, W. D. & ELLIS, I. O. 2009. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clin Cancer Res,* 15**,** 2302-10.

RAKHA, E. A., LEE, A. H., EVANS, A. J., MENON, S., ASSAD, N. Y., HODI, Z., MACMILLAN, D., BLAMEY, R. W. & ELLIS, I. O. 2010a. Tubular carcinoma of the breast: further evidence to support its excellent prognosis. *J Clin Oncol,* 28**,** 99-104.

RAKHA, E. A., LEE, A. H., ROBERTS, J., VILLENA SALINAS, N. M., HODI, Z., ELLIS, I. O. & REIS-FILHO, J. S. 2012. Low-estrogen receptor-positive breast cancer: the impact of tissue sampling,

choice of antibody, and molecular subtyping. *J Clin Oncol,* 30**,** 2929-30; author reply 2931.

RAKHA, E. A., REIS-FILHO, J. S., BAEHNER, F., DABBS, D. J., DECKER, T., EUSEBI, V., FOX, S. B., ICHIHARA, S., JACQUEMIER, J., LAKHANI, S. R., PALACIOS, J., RICHARDSON, A. L., SCHNITT, S. J., SCHMITT, F. C., TAN, P. H., TSE, G. M., BADVE, S. & ELLIS, I. O. 2010b. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res,* 12**,** 207.

RAKHA, E. A., SORIA, D., GREEN, A. R., LEMETRE, C., POWE, D. G., NOLAN, C. C., GARIBALDI, J. M., BALL, G. & ELLIS, I. O. 2014. Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer. *Br J Cancer,* 110**,** 1688-97.

RAPKIEWICZ, A., ESPINA, V., ZUJEWSKI, J. A., LEBOWITZ, P. F., FILIE, A., WULFKUHLE, J., CAMPHAUSEN, K., PETRICOIN, E. F., 3RD, LIOTTA, L. A. & ABATI, A. 2007. The needle in the haystack: application of breast fine-needle aspirate samples to quantitative protein microarray technology. *Cancer,* 111**,** 173-84.

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.

READ, L. D., SNIDER, C. E., MILLER, J. S., GREENE, G. L. & KATZENELLENBOGEN, B. S. 1988. Ligand-modulated regulation of progesterone receptor messenger ribonucleic acid and protein in human breast cancer cell lines. *Mol Endocrinol,* 2**,** 263-71.

RECHSTEINER, M. & ROGERS, S. W. 1996. PEST sequences and regulation by proteolysis. *Trends Biochem Sci,* 21**,** 267-71.

RICHER, J. K., JACOBSEN, B. M., MANNING, N. G., ABEL, M. G., WOLF, D. M. & HORWITZ, K. B. 2002. Differential gene regulation by the two progesterone receptor isoforms in human breast cancer cells. *J Biol Chem,* 277**,** 5209-18.

RING, A. E., SMITH, I. E., ASHLEY, S., FULFORD, L. G. & LAKHANI, S. R. 2004. Oestrogen receptor status, pathological complete response and prognosis in patients receiving neoadjuvant chemotherapy for early breast cancer. *Br J Cancer,* 91**,** 2012-7.

RING, B. Z., SEITZ, R. S., BECK, R., SHASTEEN, W. J., TARR, S. M., CHEANG, M. C., YODER, B. J., BUDD, G. T., NIELSEN, T. O., HICKS, D. G., ESTOPINAL, N. C. & ROSS, D. T. 2006. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol,* 24**,** 3039-47.

RONDON-LAGOS, M., VILLEGAS, V. E., RANGEL, N., SANCHEZ, M. C. & ZAPHIROPOULOS, P. G. 2016. Tamoxifen Resistance: Emerging Molecular Targets. *Int J Mol Sci,* 17.

ROSS, D. T., KIM, C. Y., TANG, G., BOHN, O. L., BECK, R. A., RING, B. Z., SEITZ, R. S., PAIK, S., COSTANTINO, J. P. & WOLMARK, N. 2008. Chemosensitivity and stratification by a five monoclonal antibody immunohistochemistry test in the NSABP B14 and B20 trials. *Clin Cancer Res,* 14**,** 6602-9.

ROSSI, L., LAAS, E., MALLON, P., VINCENT-SALOMON, A., GUINEBRETIERE, J. M., LEREBOURS, F., ROUZIER, R., PIERGA, J. Y. & REYAL, F. 2015. Prognostic impact of discrepant Ki67 and mitotic index on hormone receptor-positive, HER2-negative breast carcinoma. *Br J Cancer,* 113**,** 996-1002.

ROSSI, S., LAURINO, L., FURLANETTO, A., CHINELLATO, S., ORVIETO, E., CANAL, F., FACCHETTI, F. & DEI TOS, A. P. 2005. Rabbit monoclonal antibodies: a comparative study between a novel category of immunoreagents and the corresponding mouse monoclonal antibodies. *Am J Clin Pathol,* 124**,** 295-302.

RUFF, M., GANGLOFF, M., WURTZ, J. M. & MORAS, D. 2000. Estrogen receptor transcription and transactivation: Structure-function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res,* 2**,** 353-9.

RUIZ, C., SEIBT, S., AL KURAYA, K., SIRAJ, A. K., MIRLACHER, M., SCHRAML, P., MAURER, R., SPICHTIN, H., TORHORST, J., POPOVSKA, S., SIMON, R. & SAUTER, G. 2006. Tissue microarrays for comparing molecular features with proliferation activity in breast cancer. *Int J Cancer,* 118**,** 2190-4.

SABATIER, R., FINETTI, P., GUILLE, A., ADELAIDE, J., CHAFFANET, M., VIENS, P., BIRNBAUM, D. & BERTUCCI, F. 2014. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Mol Cancer,* 13**,** 228.

SALISBURY, J. 2009. How to count and measure features on slides. *Diagnostic Histopathology,* 15**,** 213-219.

SAPINO, A., MARCHIO, C., SENETTA, R., CASTELLANO, I., MACRI, L., CASSONI, P., GHISOLFI, G., CERRATO, M., D'AMBROSIO, E. & BUSSOLATI, G. 2006. Routine assessment of prognostic factors

in breast cancer using a multicore tissue microarray procedure. *Virchows Arch,* 449**,** 288-96.

SARTORIUS, C. A., MELVILLE, M. Y., HOVLAND, A. R., TUNG, L., TAKIMOTO, G. S. & HORWITZ, K. B. 1994. A third transactivation function (AF3) of human progesterone receptors located in the unique N-terminal segment of the B-isoform. *Mol Endocrinol,* 8**,** 1347-60.

SASAKI, K., MURAKAMI, T., KAWASAKI, M. & TAKAHASHI, M. 1987. The cell cycle associated change of the Ki-67 reactive nuclear antigen expression. *J Cell Physiol,* 133**,** 579-84.

SAUVAGEAU, M. & SAUVAGEAU, G. 2010. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell,* 7**,** 299-313.

SCHLUTER, C., DUCHROW, M., WOHLENBERG, C., BECKER, M. H., KEY, G., FLAD, H. D. & GERDES, J. 1993. The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. *J Cell Biol,* 123**,** 513-22.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

SCHNITT, S. J. 2006. Estrogen receptor testing of breast cancer in current clinical practice: what's the question? *J Clin Oncol,* 24**,** 1797-9.

SCHOLZEN, T., ENDL, E., WOHLENBERG, C., VAN DER SAR, S., COWELL, I. G., GERDES, J. & SINGH, P. B. 2002. The Ki-67 protein interacts with members of the heterochromatin protein 1 (HP1) family: a potential role in the regulation of higher-order chromatin structure. *J Pathol,* 196**,** 135-44.

SCHOLZEN, T. & GERDES, J. 2000. The Ki-67 protein: from the known and the unknown. *J Cell Physiol,* 182**,** 311-22.

SCHULTZ, J. R., PETZ, L. N. & NARDULLI, A. M. 2003. Estrogen receptor alpha and Sp1 regulate progesterone receptor gene expression. *Mol Cell Endocrinol,* 201**,** 165-75.

SCICCHITANO, M. S., DALMAS, D. A., BOYCE, R. W., THOMAS, H. C. & FRAZIER, K. S. 2009. Protein extraction of formalin-fixed, paraffin-embedded tissue enables robust proteomic profiles by mass spectrometry. *J Histochem Cytochem,* 57**,** 849-60.

SHACKLETON, M., VAILLANT, F., SIMPSON, K. J., STINGL, J., SMYTH, G. K., ASSELIN-LABAT, M. L., WU, L., LINDEMAN, G. J. & VISVADER, J. E. 2006. Generation of a functional mammary gland from a single stem cell. *Nature,* 439**,** 84-8.

SHAIKHIBRAHIM, Z., OFFERMANN, A., BRAUN, M., MENON, R., SYRING, I., NOWAK, M., HALBACH, R., VOGEL, W., RUIZ, C., ZELLWEGER, T., RENTSCH, C. A., SVENSSON, M., ANDREN, O., BUBENDORF, L., BISKUP, S., DUENSING, S., KIRFEL, J. & PERNER, S. 2014. MED12 overexpression is a frequent event in castration-resistant prostate cancer. *Endocr Relat Cancer,* 21**,** 663-75.

SHAIKHIBRAHIM, Z., OFFERMANN, A., HALBACH, R., VOGEL, W., BRAUN, M., KRISTIANSEN, G., BOOTZ, F., WENZEL, J., MIKUT, R., LENGERKE, C., REISCHL, M., SCHROCK, A. & PERNER, S. 2015. Clinical and molecular implications of MED15 in head and neck squamous cell carcinoma. *Am J Pathol,* 185**,** 1114-22.

SHALABI LA, S. Z., KASASBEH B 2006. Data Mining: A Preprocessing Engine. *J Comput Sci,* 2**,** 735-739.

SHAO, Y., GENG, Y., GU, W., NING, Z., JIANG, J. & PEI, H. 2014. Prognostic role of high Bmi-1 expression in Asian and Caucasian patients with solid tumors: a meta-analysis. *Biomed Pharmacother,* 68**,** 969-77.

SHARPLESS, N. E. 2004. Ink4a/Arf links senescence and aging. *Exp Gerontol,* 39**,** 1751-9.

SHEFFIELD, B. S., KOS, Z., ASLEH-ABURAYA, K., WANG, X. Q., LEUNG, S., GAO, D., WON, J., CHOW, C., RACHAMADUGU, R., STIJLEMAN, I., WOLBER, R., GILKS, C. B., MYLES, N., THOMSON, T., HAYES, M. M., BERNARD, P. S., NIELSEN, T. O. & CHIA, S. K. 2016. Molecular subtype profiling of invasive breast cancers weakly positive for estrogen receptor. *Breast Cancer Res Treat,* 155**,** 483-90.

SHEMER, G. & PODBILEWICZ, B. 2003. The story of cell fusion: big lessons from little worms. *Bioessays,* 25**,** 672-82.

SHERGILL, I. S., SHERGILL, N. K., ARYA, M. & PATEL, H. R. 2004. Tissue microarrays: a current medical research tool. *Curr Med Res Opin,* 20**,** 707-12.

SHERI, A., SMITH, I. E., JOHNSTON, S. R., A'HERN, R., NERURKAR, A., JONES, R. L., HILLS, M., DETRE, S., PINDER, S. E., SYMMANS, W. F. & DOWSETT, M. 2015. Residual proliferative cancer burden to predict long-term outcome following neoadjuvant chemotherapy. *Ann Oncol,* 26**,** 75-80.

SHERR, C. J. 2001. The INK4a/ARF network in tumour suppression. *Nat Rev Mol Cell Biol,* 2**,** 731-7.

SHI, S. R., TAYLOR, C. R., FOWLER, C. B. & MASON, J. T. 2013. Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. *Proteomics Clin Appl,* 7**,** 264-72.

SIDDIQUE, H. R. & SALEEM, M. 2012. Role of BMI1, a stem cell factor, in cancer recurrence and chemoresistance: preclinical and clinical evidences. *Stem Cells,* 30**,** 372-8.

SIEGEL, R. L., MILLER, K. D. & JEMAL, A. 2016. Cancer statistics, 2016. *CA Cancer J Clin,* 66**,** 7-30.

SINGLETARY, S. E. & CONNOLLY, J. L. 2006. Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual. *CA Cancer J Clin,* 56**,** 37-47; quiz 50-1.

SINN, H. P. & KREIPE, H. 2013. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel),* 8**,** 149-54.

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.

SLAMON, D. J., GODOLPHIN, W., JONES, L. A., HOLT, J. A., WONG, S. G., KEITH, D. E., LEVIN, W. J., STUART, S. G., UDOVE, J., ULLRICH, A. & ET AL. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science,* 244**,** 707-12.

SMITH, G. H. & MEDINA, D. 1988. A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci,* 90 (Pt 1)**,** 173-83.

SOBECKI, M., MROUJ, K., CAMASSES, A., PARISIS, N., NICOLAS, E., LLERES, D., GERBE, F., PRIETO, S., KRASINSKA, L., DAVID, A., EGUREN, M., BIRLING, M. C., URBACH, S., HEM, S., DEJARDIN, J., MALUMBRES, M., JAY, P., DULIC, V., LAFONTAINE, D., FEIL, R. & FISHER, D. 2016. The cell proliferation antigen Ki-67 organises heterochromatin. *Elife,* 5**,** e13722.

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.

SORIA, D., GARIBALDI, J. M., AMBROGI, F., GREEN, A. R., POWE, D., RAKHA, E., MACMILLAN, R. D., BLAMEY, R. W., BALL, G., LISBOA, P. J., ETCHELLS, T. A., BORACCHI, P., BIGANZOLI, E. & ELLIS, I. O. 2010. A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients. *Comput Biol Med,* 40**,** 318-30.

SORIA, D., GARIBALDI, J. M., GREEN, A. R., POWE, D. G., NOLAN, C. C., LEMETRE, C., BALL, G. R. & ELLIS, I. O. 2013. A quantifier-based fuzzy classification system for breast cancer patients. *Artif Intell Med,* 58**,** 175-84.

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., LONNING, P. E. & 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.

SPYRATOS, F., FERRERO-POUS, M., TRASSARD, M., HACENE, K., PHILLIPS, E., TUBIANA-HULIN, M. & LE DOUSSAL, V. 2002. Correlation between MIB-1 and other proliferation markers: clinical implications of the MIB-1 cutoff value. *Cancer,* 94**,** 2151-9.

STAAF, J., RINGNER, M., VALLON-CHRISTERSSON, J., JONSSON, G., BENDAHL, P. O., HOLM, K., ARASON, A., GUNNARSSON, H., HEGARDT, C., AGNARSSON, B. A., LUTS, L., GRABAU, D., FERNO, M., MALMSTROM, P. O., JOHANNSSON, O. T., LOMAN, N., BARKARDOTTIR, R. B. & BORG, A. 2010. Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome. *J Clin Oncol,* 28**,** 1813-20.

STUART-HARRIS, R., CALDAS, C., PINDER, S. E. & PHAROAH, P. 2008. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast,* 17**,** 323-34.

SUGIHARA, H., ISHIMOTO, T., WATANABE, M., SAWAYAMA, H., IWATSUKI, M., BABA, Y., KOMOHARA, Y., TAKEYA, M. & BABA, H. 2013. Identification of miR-30e* regulation of Bmi1 expression mediated by tumor-associated macrophages in gastrointestinal cancer. *PLoS One,* 8**,** e81839.

SUN, S., YU, J. P., CHEN, F., ZHAO, T. J., FANG, X. H., LI, Y. Q. & SUI, S. F. 2008. TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in Arabidopsis. *J Biol Chem,* 283**,** 6261-71.

SUNDQUIST, M., THORSTENSON, S., BRUDIN, L. & NORDENSKJOLD, B. 1999. Applying the Nottingham Prognostic Index to a Swedish breast cancer population. South East Swedish Breast Cancer Study Group. *Breast Cancer Res Treat,* 53**,** 1-8.

SYMMANS, W. F., PEINTINGER, F., HATZIS, C., RAJAN, R., KUERER, H., VALERO, V., ASSAD, L., PONIECKA, A., HENNESSY, B., GREEN, M., BUZDAR, A. U., SINGLETARY, S. E., HORTOBAGYI, G. N. & PUSZTAI, L. 2007. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol,* 25**,** 4414-22.

TAKAGI, M., MATSUOKA, Y., KURIHARA, T. & YONEDA, Y. 1999. Chmadrin: a novel Ki-67 antigen-related perichromosomal protein possibly implicated in higher order chromatin structure. *J Cell Sci,* 112 (Pt 15)**,** 2463-72.

TAN, K. P., YANG, M. & ITO, S. 2007. Activation of nuclear factor (erythroid-2 like) factor 2 by toxic bile acids provokes adaptive defense responses to enhance cell survival at the emergence of oxidative stress. *Mol Pharmacol,* 72**,** 1380-90.

TANCA, A., PAGNOZZI, D., FALCHI, G., BIOSA, G., ROCCA, S., FODDAI, G., UZZAU, S. & ADDIS, M. F. 2011. Impact of fixation

time on GeLC-MS/MS proteomic profiling of formalin-fixed, paraffin-embedded tissues. *J Proteomics,* 74**,** 1015-21.

TASHIRO, T., HIROKAWA, M., HARADA, H., YOKOYAMA, S. & SANO, T. 2002. Cell membrane expression of MIB-1 in salivary gland pleomorphic adenoma. *Histopathology,* 41**,** 559-61.

TEMPLIN, M. F., STOLL, D., SCHRENK, M., TRAUB, P. C., VOHRINGER, C. F. & JOOS, T. O. 2002. Protein microarray technology. *Trends Biotechnol,* 20**,** 160-6.

TESCHENDORFF, A. E., NADERI, A., BARBOSA-MORAIS, N. L., PINDER, S. E., ELLIS, I. O., APARICIO, S., BRENTON, J. D. & CALDAS, C. 2006. A consensus prognostic gene expression classifier for ER positive breast cancer. *Genome Biol,* 7**,** R101.

THOR, A. D., LIU, S., MOORE, D. H., 2ND & EDGERTON, S. M. 1999. Comparison of mitotic index, in vitro bromodeoxyuridine labeling, and MIB-1 assays to quantitate proliferation in breast cancer. *J Clin Oncol,* 17**,** 470-7.

TONG, Y. Q., LIU, B., ZHENG, H. Y., HE, Y. J., GU, J., LI, F. & LI, Y. 2012. Overexpression of BMI-1 is associated with poor prognosis in cervical cancer. *Asia Pac J Clin Oncol,* 8**,** e55-62.

TRIHIA, H., MURRAY, S., PRICE, K., GELBER, R. D., GOLOUH, R., GOLDHIRSCH, A., COATES, A. S., COLLINS, J., CASTIGLIONE-GERTSCH, M. & GUSTERSON, B. A. 2003. Ki-67 expression in breast carcinoma: its association with grading systems, clinical parameters, and other prognostic factors--a surrogate marker? *Cancer,* 97**,** 1321-31.

TYERS, M. & MANN, M. 2003. From genomics to proteomics. *Nature,* 422**,** 193-7.

TZANKOV, A., WENT, P., ZIMPFER, A. & DIRNHOFER, S. 2005. Tissue microarray technology: principles, pitfalls and perspectives--lessons learned from hematological malignancies. *Exp Gerontol,* 40**,** 737-44.

UMEMURA, S., ITOH, J., ITOH, H., SERIZAWA, A., SAITO, Y., SUZUKI, Y., TOKUDA, Y., TAJIMA, T. & OSAMURA, R. Y. 2004. Immunohistochemical evaluation of hormone receptors in breast cancer: which scoring system is suitable for highly sensitive procedures? *Appl Immunohistochem Mol Morphol,* 12**,** 8-13.

URRUTICOECHEA, A., SMITH, I. E. & DOWSETT, M. 2005. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol,* 23**,** 7212-20.

VALK-LINGBEEK, M. E., BRUGGEMAN, S. W. & VAN LOHUIZEN, M. 2004. Stem cells and cancer; the polycomb connection. *Cell,* 118**,** 409-18.

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 DIEST, P. J., VAN DER WALL, E. & BAAK, J. P. A. 2004. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology,* 57**,** 675-681.

VAN LOHUIZEN, M., VERBEEK, S., SCHEIJEN, B., WIENTJENS, E., VAN DER GULDEN, H. & BERNS, A. 1991. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell,* 65**,** 737-52.

VERHEIJEN, R., KUIJPERS, H. J., SCHLINGEMANN, R. O., BOEHMER, A. L., VAN DRIEL, R., BRAKENHOFF, G. J. & RAMAEKERS, F. C. 1989a. Ki-67 detects a nuclear matrix-associated proliferation-related antigen. I. Intracellular localization during interphase. *J Cell Sci,* 92 (Pt 1)**,** 123-30.

VERHEIJEN, R., KUIJPERS, H. J., VAN DRIEL, R., BECK, J. L., VAN DIERENDONCK, J. H., BRAKENHOFF, G. J. & RAMAEKERS, F. C. 1989b. Ki-67 detects a nuclear matrix-associated proliferation-related antigen. II. Localization in mitotic cells and association with chromosomes. *J Cell Sci,* 92 (Pt 4)**,** 531-40.

VIALE, G. 2012. The current state of breast cancer classification. *Ann Oncol,* 23 Suppl 10**,** x207-10.

VIALE, G., GIOBBIE-HURDER, A., REGAN, M. M., COATES, A. S., MASTROPASQUA, M. G., DELL'ORTO, P., MAIORANO, E., MACGROGAN, G., BRAYE, S. G., OHLSCHLEGEL, C., NEVEN, P., OROSZ, Z., OLSZEWSKI, W. P., KNOX, F., THURLIMANN, B., PRICE, K. N., CASTIGLIONE-GERTSCH, M., GELBER, R. D., GUSTERSON, B. A. & GOLDHIRSCH, A. 2008. Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. *J Clin Oncol,* 26**,** 5569-75.

VILLADSEN, R., FRIDRIKSDOTTIR, A. J., RONNOV-JESSEN, L., GUDJONSSON, T., RANK, F., LABARGE, M. A., BISSELL, M. J. & PETERSEN, O. W. 2007. Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol,* 177**,** 87-101.

VON MINCKWITZ, G., UNTCH, M., NUESCH, E., LOIBL, S., KAUFMANN, M., KUMMEL, S., FASCHING, P. A., EIERMANN, W., BLOHMER, J. U., COSTA, S. D., MEHTA, K., HILFRICH, J., JACKISCH, C., GERBER, B., DU BOIS, A., HUOBER, J., HANUSCH, C., KONECNY, G., FETT, W., STICKELER, E., HARBECK, N., MULLER, V. & JUNI, P. 2011. Impact of treatment characteristics on response of different breast cancer phenotypes: pooled analysis of the German neo-adjuvant chemotherapy trials. *Breast Cancer Res Treat,* 125**,** 145-56.

VRTACNIK, P., OSTANEK, B., MENCEJ-BEDRAC, S. & MARC, J. 2014. The many faces of estrogen signaling. *Biochem Med (Zagreb),* 24**,** 329-42.

WALTER, P., GREEN, S., GREENE, G., KRUST, A., BORNERT, J. M., JELTSCH, J. M., STAUB, A., JENSEN, E., SCRACE, G., WATERFIELD, M. & ET AL. 1985. Cloning of the human estrogen receptor cDNA. *Proc Natl Acad Sci U S A,* 82**,** 7889-93.

WANG, F., FLANAGAN, J., SU, N., WANG, L. C., BUI, S., NIELSON, A., WU, X., VO, H. T., MA, X. J. & LUO, Y. 2012a. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn,* 14**,** 22-9.

WANG, H., LIU, H., LI, X., ZHAO, J., ZHANG, H., MAO, J., ZOU, Y., ZHANG, H., ZHANG, S., HOU, W., HOU, L., MCNUTT, M. A. & ZHANG, B. 2014. Estrogen receptor alpha-coupled Bmi1 regulation pathway in breast cancer and its clinical implications. *BMC Cancer,* 14**,** 122.

WANG, J., WEN, S., SYMMANS, W. F., PUSZTAI, L. & COOMBES, K. R. 2009. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform,* 7**,** 199-216.

WANG, Y., ZHE, H., DING, Z., GAO, P., ZHANG, N. & LI, G. 2012b. Cancer stem cell marker Bmi-1 expression is associated with basal-like phenotype and poor survival in breast cancer. *World J Surg,* 36**,** 1189-94.

WARTH, A., FINK, L., FISSELER-ECKHOFF, A., JONIGK, D., KELLER, M., OTT, G., RIEKER, R. J., SINN, P., SODER, S., SOLTERMANN, A., WILLENBROCK, K. & WEICHERT, W. 2013. Interobserver agreement of proliferation index (Ki-67) outperforms mitotic count in pulmonary carcinoids. *Virchows Arch,* 462**,** 507-13.

WEI, L. L., KRETT, N. L., FRANCIS, M. D., GORDON, D. F., WOOD, W. M., O'MALLEY, B. W. & HORWITZ, K. B. 1988. Multiple human progesterone receptor messenger ribonucleic acids and their autoregulation by progestin agonists and antagonists in breast cancer cells. *Mol Endocrinol,* 2**,** 62-72.

WEIGELT, B., BAEHNER, F. L. & REIS-FILHO, J. S. 2010a. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol,* 220**,** 263-80.

WEIGELT, B., GEYER, F. C. & REIS-FILHO, J. S. 2010b. Histological types of breast cancer: how special are they? *Mol Oncol,* 4**,** 192-208.

WEIGELT, B., MACKAY, A., A'HERN, R., NATRAJAN, R., TAN, D. S., DOWSETT, M., ASHWORTH, A. & REIS-FILHO, J. S. 2010c. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol,* 11**,** 339-49.

WEIGELT, B. & REIS-FILHO, J. S. 2009. Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol,* 6**,** 718-30.

WEISS, R. B., WOOLF, S. H., DEMAKOS, E., HOLLAND, J. F., BERRY, D. A., FALKSON, G., CIRRINCIONE, C. T., ROBBINS, A., BOTHUN, S., HENDERSON, I. C. & NORTON, L. 2003. Natural history of more than 20 years of node-positive primary breast carcinoma treated with cyclophosphamide, methotrexate, and fluorouracil-based adjuvant chemotherapy: a study by the Cancer and Leukemia Group B. *J Clin Oncol,* 21**,** 1825-35.

WERNER, M., CHOTT, A., FABIANO, A. & BATTIFORA, H. 2000. Effect of formalin tissue fixation and processing on immunohistochemistry. *Am J Surg Pathol,* 24**,** 1016-9.

WILLIAMS, J. M. & DANIEL, C. W. 1983. Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol,* 97**,** 274-90.

WISHART, G. C., AZZATO, E. M., GREENBERG, D. C., RASHBASS, J., KEARINS, O., LAWRENCE, G., CALDAS, C. & PHAROAH, P. D. 2010. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res,* 12**,** R1.

WISHART, G. C., BAJDIK, C. D., DICKS, E., PROVENZANO, E., SCHMIDT, M. K., SHERMAN, M., GREENBERG, D. C., GREEN, A. R., GELMON, K. A., KOSMA, V. M., OLSON, J. E., BECKMANN, M. W., WINQVIST, R., CROSS, S. S., SEVERI, G., HUNTSMAN,

D., PYLKAS, K., ELLIS, I., NIELSEN, T. O., GILES, G., BLOMQVIST, C., FASCHING, P. A., COUCH, F. J., RAKHA, E., FOULKES, W. D., BLOWS, F. M., BEGIN, L. R., VAN'T VEER, L. J., SOUTHEY, M., NEVANLINNA, H., MANNERMAA, A., COX, A., CHEANG, M., BAGLIETTO, L., CALDAS, C., GARCIA-CLOSAS, M. & PHAROAH, P. D. 2012. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer,* 107**,** 800-7.

WISHART, G. C., RAKHA, E., GREEN, A., ELLIS, I., ALI, H. R., PROVENZANO, E., BLOWS, F. M., CALDAS, C. & PHAROAH, P. D. 2014. Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. *BMC Cancer,* 14**,** 908.

WODARZ, A. & GONZALEZ, C. 2006. Connecting cancer to the asymmetric division of stem cells. *Cell,* 124**,** 1121-3.

WOLFF, C., SCHOTT, C., MALINOWSKY, K., BERG, D. & BECKER, K. F. 2011a. Producing reverse phase protein microarrays from formalin-fixed tissues. *Methods Mol Biol,* 785**,** 123-40.

WOLFF, C., SCHOTT, C., PORSCHEWSKI, P., REISCHAUER, B. & BECKER, K. F. 2011b. Successful protein extraction from over-fixed and long-term stored formalin-fixed tissues. *PLoS One,* 6**,** e16353.

WOOD, B., JUNCKERSTORFF, R., STERRETT, G., FROST, F., HARVEY, J. & ROBBINS, P. 2007. A comparison of immunohistochemical staining for oestrogen receptor, progesterone receptor and HER-2 in breast core biopsies and subsequent excisions. *Pathology,* 39**,** 391-5.

WOODWARD, W. A., CHEN, M. S., BEHBOD, F. & ROSEN, J. M. 2005. On mammary stem cells. *J Cell Sci,* 118**,** 3585-94.

WRANGE, O., CARLSTEDT-DUKE, J. & GUSTAFSSON, J. A. 1979. Purification of the glucocorticoid receptor from rat liver cytosol. *J Biol Chem,* 254**,** 9284-90.

WU, K. J. 2011. Direct activation of Bmi1 by Twist1: implications in cancer stemness, epithelial-mesenchymal transition, and clinical significance. *Chang Gung Med J,* 34**,** 229-38.

WULFKUHLE, J. D., SPEER, R., PIEROBON, M., LAIRD, J., ESPINA, V., DENG, J., MAMMANO, E., YANG, S. X., SWAIN, S. M., NITTI, D., ESSERMAN, L. J., BELLUCO, C., LIOTTA, L. A. & PETRICOIN, E. F., 3RD 2008. Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy. *J Proteome Res,* 7**,** 1508-17.

YADAV, A. K., SAHASRABUDDHE, A. A., DIMRI, M., BOMMI, P. V., SAINGER, R. & DIMRI, G. P. 2010. Deletion analysis of BMI1 oncoprotein identifies its negative regulatory domain. *Mol Cancer,* 9**,** 158.

YANG, M. H., HSU, D. S., WANG, H. W., WANG, H. J., LAN, H. Y., YANG, W. H., HUANG, C. H., KAO, S. Y., TZENG, C. H., TAI, S. K., CHANG, S. Y., LEE, O. K. & WU, K. J. 2010. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat Cell Biol,* 12**,** 982-92.

YERUSHALMI, R., WOODS, R., RAVDIN, P. M., HAYES, M. M. & GELMON, K. A. 2010. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol,* 11**,** 174-83.

YI, M., HUO, L., KOENIG, K. B., MITTENDORF, E. A., MERIC-BERNSTAM, F., KUERER, H. M., BEDROSIAN, I., BUZDAR, A. U., SYMMANS, W. F., CROW, J. R., BENDER, M., SHAH, R. R., HORTOBAGYI, G. N. & HUNT, K. K. 2014. Which threshold for ER positivity? a retrospective study based on 9639 patients. *Ann Oncol,* 25**,** 1004-11.

YIN, J., ZHENG, G., JIA, X., ZHANG, Z., ZHANG, W., SONG, Y., XIONG, Y. & HE, Z. 2013. A Bmi1-miRNAs cross-talk modulates chemotherapy response to 5-fluorouracil in breast cancer cells. *PLoS One,* 8**,** e73268.

YOSHIKAWA, R., TSUJIMURA, T., TAO, L., KAMIKONYA, N. & FUJIWARA, Y. 2012. The oncoprotein and stem cell renewal factor BMI1 associates with poor clinical outcome in oesophageal cancer patients undergoing preoperative chemoradiotherapy. *BMC Cancer,* 12**,** 461.

ZHANG, B. N., CAO, X. C., CHEN, J. Y., CHEN, J., FU, L., HU, X. C., JIANG, Z. F., LI, H. Y., LIAO, N., LIU, D. G., TAO, O., SHAO, Z. M., SUN, Q., WANG, S., WANG, Y. S., XU, B. H. & ZHANG, J. 2012. Guidelines on the diagnosis and treatment of breast cancer (2011 edition). *Gland Surg,* 1**,** 39-61.