

University of Nottingham
UK | CHINA | MALAYSIA

BIOLOGICAL ASSESSMENT - DEVELOPING A NOVEL TOOL TO GUIDE TREATMENT DECISION IN OLDER WOMEN WITH PRIMARY BREAST CANCER

Thesis submitted to the University of Nottingham

for the Degree of Doctor of Philosophy

Ruth Parks

Student ID: 20154914

Submitted: 29th August 2022

Declaration

I hereby declare that the work presented in this thesis is my own work unless specified otherwise and that it has not been submitted to any other university or institution for consideration for any degree.

Ruth M Parks

Ethical approval

The study was conducted at the University of Nottingham, City Hospital Campus and University Park Campus, Nottingham, United Kingdom, and was approved by the Research Ethics Committee (Nottingham Research Ethics Committee 1) under REC reference 19/YH/0293 with the title 'Refining the prognosis of breast cancer and breast cancer precursor lesions for personalised therapy' (Appendix 1).

Dedication

This work is gratefully dedicated to all breast cancer patients who make up the Nottingham cohort of older women with primary breast cancer, without whom this thesis would not be possible.

Abstract

Introduction

The incidence of breast cancer is increasing as our population is ageing and growing. There is a disparity between recommendations for treatment of primary breast cancer and the treatment that older women receive. The majority of existing prognostic tools on the market which can aid in treatment decision making, alongside conventional measures, focus on prognosis after adjuvant therapy following surgery and do not consider the unique needs of the older population at diagnosis of breast cancer.

Objective

To comprehensively evaluate the biology of primary breast cancer in older women which will result in generation of a prognostic algorithm using bioinformatics principles, specific for older women with primary breast cancer to aid them with treatment decision making at diagnosis, primarily when decision between surgery and non-operative therapy is challenging.

Methods

Through a systematic review of the literature and previous work conducted in Nottingham, a panel of 31 biomarkers were deemed to be of potential significance in this cohort (oestrogen receptor (ER), progesterone receptor (PR), androgen receptor (AR), epidermal growth factor receptor (EGFR), human epidermal growth factor receptor (HER) 2, 3 and 4, vascular endothelial growth factor (VEGF), tumour protien p53, breast cancer gene (BRCA) 1 and 2, murine double minute (MDM) 2 and 4, phosphatase and tensin homolog (PTEN), liver kinase (LK) B1, Ki67, B cell lymphoma (BCL) 2, E-cadherin, cluster of differentiation (CD) 44, cytokeratin (CK) 5, 5/6, 7/8, 14, 17, 18 and 19, mucin (MUC) 1, amplified in breast cancer (AIB) 1, solute carrier family 1 member 5 (SLC1A5), SLC3A2, SLC7A5. These markers were

stained by immunohistochemistry (IHC) in all available core needle biopsy (CNB) and surgical excision (SE) samples from a cohort of 1758 women aged ≥70 years treated consecutively in Nottingham for whom long-term follow-up data is available, including some results from previous work by the research team.

Comparison of measurement of biomarkers between CNB and SE specimens was made using concordance analysis. Cluster analysis was performed in all CNB samples (this had previously been done in the SE samples).

Bioinformatics technologies comprising and combining artificial neural network modelling of risk and principal and independent components analysis was performed to produce algorithms to predict the following variables: breast cancer specific survival (BCSS); overall survival; time to metastasis; time to regional recurrence; time to progression of disease.

Results

Comparison of H-score results of biomarkers between CNB and SE samples for 15 biomarkers in the same patient, showed large variation in results, meaning that results could not be directly compared at this stage. The percentage concordance rates between positive/negative cut-offs were above 60% for the markers ER, PR, EGFR, HER2, VEGF, p53, Ki67, BCL2, CK 7/8, MUC1. The markers HER3, HER4, CK 5/6, BRCA1 and LKB1 had a percentage concordance of <45%.

Cluster analysis in CNB samples using a panel of 17 biomarkers, showed three biological clusters of disease, consistent with the literature: luminal A-type tumours, luminal B-type tumours, low-ER luminal (a cluster which appears to be unique to the older population). This is in keeping with previous work by the research team in SE samples and also ER-positive CNB samples. The clusters corresponding to HER2-

positive and triple negative breast cancer were absent, due to differences in how staining is scored in the CNB, compared to SE, samples.

The prognostic models developed for BCSS using the panel of 15 biomarks, showed strong concordance at predicting outcome in both CNB and SE samples ($r = 0.716$ and $r = 0.742$ respectively). The models developed for all other variables were poor at predicting the desired outcome (concordance $r < 0.4$). It was not possible to improve the algorithm further by using only biomarkers which were independently prognostically significant for BCSS.

Conclusions

Older women with breast cancer have distinct biological features which can be utilised to help with treatment decision making. Presently, there are differences between measurement of a biomarker in CNB compared to SE tissue, however, cluster analysis in both tissue types reveals the same clusters within each group. It is feasible to develop a prognostic algorithm specific to older women with breast cancer at diagnosis, with potential useful clinical application. The next step of the project is to refine the algorithm by removing biomarkers with conflicting or low contribution.

Relevant work published during this PhD

Original articles

RXN Lee, MJ Cardoso, KL Cheung, **RM Parks.** Uptake of post-mastectomy immediate breast reconstruction in older women with primary breast cancer – a systematic review of the literature. Br J Surg, 2022 Nov, 109 (11): 1063 – 1072. Impact factor: 11.122. DOI: https://doi.org/10.1093/bjs/znac251

R Parks, KL Cheung. Challenges in Geriatric Oncology – A Surgeon's Perspective. Curr Oncol, 2022 Jan 29, 29 (2), 659 – 674. Impact factor: 3.109. DOI: https://doi.org/10.3390/curroncol29020058

Z Chia, **RM Parks,** KL Cheung. Does breast cancer surgery impact functional status and independence in older patients? A narrative review. Oncol Ther, 2021 Dec, 9 (2): 373 – 383. DOI: https://doi.org/10.1007/s40487-021-00170-4

J Kalvala, **RM Parks,** AR Green, KL Cheung. Concordance between core needle biopsy and surgical excision specimens for Ki-67 in breast cancer – a systematic review of the literature. Histopathology, 2022 Feb, 80 (3): 468 – 484. Impact factor: 7.778. DOI: https://doi.org/10.1111/his.14555

RM Parks, LH Alfarsi, AR Green, KL Cheung. Biology of primary breast cancer in older women beyond routine biomarkers. Breast Cancer, 2021 Sep, 28 (5): 991 – 1001. Impact factor: 3.307. DOI: 10.1007/s12282-021-01266-5

CA Harrison, **RM Parks,** KL Cheung. The impact of breast cancer surgery on functional status in older women – A systematic review of the literature. Eur J Surg

Oncol, 2021 Aug, 47 (8), 1891 – 1899. Impact factor: 4.424.
DOI: 10.1016/j.ejso.2021.04.010

RM Parks, AR Green, KL Cheung. The impact of tumour biology on the management of primary breast cancer in older women – based on a research programme in Nottingham. Ann Breast Surg. 2021 Mar 30, 5: 5. DOI: 10.21037/abs-20-130

RM Parks, MA Albanghali, BM Syed, AR Green, IO Ellis, KL Cheung. Patterns of biomarker expression in patients treated with primary endocrine therapy – a unique insight using core needle biopsy tissue microarray. Breast Cancer Res Treat. 2021 Feb, 185 (3): 647-655. Impact factor: 4.845. DOI: 10.1007/s10549-020-06023-4

S Gordon-Craig, **RM Parks** and KL Cheung. The Potential Use of Tumor-Based Prognostic and Predictive Tools in Older Women with Primary Breast Cancer: A Narrative Review. Oncol Ther, 2020 Dec, 8 (2): 231-250. DOI: 10.1007/s40487-020-00123-3

F Gomes, A Lewis, R Morris, **R Parks,** T Kalsi, G Babic-Illamn et al. The care of geriatric oncology patients in the United Kingdom. Ecancermedicalscience. 2020 Sep 15, 14: 1101. DOI: 10.3332/ecancer.2020.1101

RM Parks, MA Albanghali, BM Syed, AR Green, IO Ellis, KL Cheung. Biology of Oestrogen-Receptor Positive Primary Breast Cancer in Older Women with Utilisation of Core Needle Biopsy Samples and Correlation with Clinical Outcome. Cancers (Basel). 2020 Jul 27, 12 (8): 2067. Impact factor: 6.575.
DOI: 10.3390/cancers12082067

L Alfarsi, R El Ansari, B Masisi, **R Parks,** O Mohammed, I Ellis, E Rakha and A Green. Integrated Analysis of Key Differentially Expressed Genes Identifies DBN1 as a Predictive Marker of Response to Endocrine Therapy in Luminal Breast Cancer. Cancers (Basel). 2020 Jun 12, 12 (6): 1549. Impact factor: 6.575. DOI: 10.3390/cancers12061549f

RM Parks, Y Jauhari, J Morgan, L Wyld and KL Cheung. Special issues for older women with primary breast cancer. Breast Cancer Management. 2020 Jun 10, 9: 3, online ahead of print. DOI: 10.2217/bmt-2020-0022

RM Parks and KL Cheung. Personalising Care in the Older Woman with Primary Breast Cancer. Ann Acad Med Singap. 2019 Nov 15, 48 (11): 370-375. URL: https://www.annals.edu.sg/pdf/48VolNo11Nov2019/V48N11p370.pdf

Book chapters

RM Parks and KL Cheung. 2020. Chapter 3: Clinical Relevance of Biomarkers in Oestrogen Receptor Positive Breast Cancer. In: Estrogen Receptors: Structure, Function and Clinical Aspects. 1st edition. Ed: GG Chen, MCF Tong, CA van Hasselt. Publisher: Nova Science Publishers, New York.

Published abstracts

J Francis, M Saundh, **R Parks,** KL Cheung. Molecular biomarker expression within window of opportunity studies for oestrogen receptor-positive breast cancer. EJSO. May 2022, 48 (5): e224. Association of Breast Surgery annual conference, 16th – 17th May 2022, Liverpool, UK. Poster presentation.

RXN Lee, **R Parks,** MJ Cardoso, KL Cheung. Post-mastectomy immediate breast reconstruction in older women – A systematic review. EJSO. May 2022, 48 (5): e229 – e230. Association of Breast Surgery annual conference, 16th – 17th May 2022, Liverpool, UK. Poster presentation.

R Lee, **RM Parks,** MJ Cardoso, KL Cheung. Factors affecting update of post-mastectomy immediate breast reconstruction in older women – a systematic review. Future Oncology (online ahead of print). 6th Symposium on Primary Breast Cancer in Older Women, 4th March 2022, Nottingham, UK. Oral presentation.

N Tse, **RM Parks,** P Howard, R Zahit, AH Kayani, R Lakshmanan, L Winterbottom, MA Jahan, HM Holmes, KL Cheung. The association between polypharmacy and functional status in older women with primary breast cancer. Future Oncology (online ahead of print). 6th Symposium on Primary Breast Cancer in Older Women, 4th March 2022, Nottingham, UK Poster presentation.

R Parks, P Howard, R Zahit, A Kayani, R Lakshmaman, H Blake, L Winterbottom, M Jahan, KL Cheung. Short-term (6 months) impact of primary breast cancer treatment on functional status and quality of life of older women. J Geriatr Oncol. Dec 2021, 12 (8): suppl 1, S41. International Society of Geriatric Oncology 21st annual meeting, 4th-5th November 2021, virtual. Poster presentation.

N Tse, **RM Parks,** P Howard, R Zahit, A Kayani, R Lakshamanan, L Winterbottom, M Jahan, H Holmes, KL Cheung. The association between medication use and

functional status in older women with primary breast cancer treatment. J Geriatr Oncol. Dec 2021, 12 (8): suppl, S34-35. International Society of Geriatric Oncology 21st annual meeting, 4th-5th November 2021, virtual. Oral presentation.

D Suen, WY Li, **R Parks,** L Wong, A Kwong, KL Cheung. Impact of COVID-19 on Older Women Presenting with Breast Cancer. J Geriatr Oncol. Dec 2021, 12 (8): suppl 1, S8. International Society of Geriatric Oncology 21st annual meeting, 4th-5th November 2021, virtual. Poster presentation.

RXN Lee, **R Parks**, MJ Cardoso, KL Cheung. Uptake of Immediate Breast Reconstruction in Older Women Post Mastectomy – A Systematic Review of the Literature. J Geriatr Oncol. Dec 2021, 12 (8): suppl 1, S2-3. International Society of Geriatric Oncology 21st annual meeting, 4th-5th November 2021, virtual. Oral presentation.

RM Parks, L Alfarsi, AR Green, KL Cheung. The role of amino acid transporters SLC1A5 and SLC3A2 in primary breast cancer in older women. BJS. Oct 2021, Suppl 7: EP.TH.275. Association of Surgeons of Great Britain and Ireland annual conference 2021, 4th-8th May 2021, virtual. Poster presentation.

RM Parks, SJ Johnston, BM Syed, IO Ellis, AR Green, K Keyomarsi, KK Hunt, KL Cheung. Cytoplasmic cyclin E: An exquisite marker of tumor biology in older women with primary breast cancer. J Geriatr Oncol. 2019, 10 (6) Suppl 1: S20. International Society of Geriatric Oncology 19th annual meeting, 15th-16th November 2019, Geneva, Switzerland. Oral presentation.

RM Parks, J Abdi, E Rakha, IO Ellis, AR Green and KL Cheung. The role of the androgen receptor in primary breast cancer in older women. J Geriatr Oncol. 2019, 10 (6), Suppl 1: S48. International Society of Geriatric Oncology 19th annual meeting, 15th-16th November 2019, Geneva, Switzerland. Poster presentation. This presentation won best poster prize – see below.

Related prizes and awards during this PhD

RM Parks. University of Nottingham, Nottingham Breast Cancer Research Centre Early Career Researcher Award. Bursary to attend the International Society of Geriatric Oncology (SIOG) conference in Geneva, Switzerland, 28th-30th October 2022, to present PhD work. This is a local prize. The total value of the award is £1300.

RM Parks. University of Nottingham, Tri Campus Awards 2022 – Dean Moore Award Winner. This is a local prize. The total value of the award is £150.

RM Parks. Developing a prognostic algorithm to assist older women in treatment decision making at diagnosis of early breast cancer. University of Nottingham, Sue Watson Oral Presentation Event, first prize. This is a local prize. The total value of the award is £75.

RM Parks. University of Nottingham, best live presentation in sandpit: Opportunities and Skills Development in Applying for Funding. This is a local prize.

RM Parks. University of Nottingham, Online Conference Fund. Registration fee provided to attend the Association of Surgeons of Great Britain and Ireland Annual Congress (virtual), 4th-8th May 2021. This is a local prize. The total value of the award is £100.

RM Parks. Association of Surgeons in Training, Virtual Event Bursary. Registration fee provided to attend the 17th St. Gallen International Breast Cancer Conference (virtual), 17th-21st March 2021. This is a national prize. The total value of the award is £150.

RM Parks, SJ Johnston, BM Syed, IO Ellis, AR Green, K Keyomarsi, KK Hunt, KL Cheung. Cytoplasmic cyclin E: an exquisite marker of tumour biology in older women with primary breast cancer. University of Nottingham, Cancer Research Network Symposium, 10th-11th November 2020. First prize for short talk (out of 12 short talks). This is a local prize. The total value of the award is £100.

RM Parks, J Abdi, E Rakha, IO Ellis, AR Green and KL Cheung. The role of the androgen receptor in primary breast cancer in older women. International Society of Geriatric Oncology annual conference, 14th-16th November 2019. Best poster award (out of 166 posters). This is an international prize. The award is registration at the 2022 conference, which has a value of 200 euros.

Research fellowships obtained during this thesis

RM Parks. Bio-geriatric assessment – optimising treatment of breast cancer in older women. Royal College of Surgeons of England, Honorary Research Fellowship, July 2020. This was an honorary research fellowship due to prior successful award of fellowship from Nottingham Hospitals Charity (initial value £71,469.65). This is a national award. This award entitled me to attend a Research Methods course of the value of £270.

RM Parks. Bio-geriatric assessment (BGA), a novel tool to optimise treatment of primary breast cancer in older women. Nottingham Hospitals Charity, Oncology Research Fellowship, December 2019. This award has enabled me to be employed by the University of Nottingham as a Clinical Research Fellow for the remainder of my PhD project. The total value of the award is £155,538.03.

Acknowledgements

First and foremost I would like to thank my supervisors Professor Kwok-Leung Cheung and Dr Andrew Green for their guidance, support and counsel during this thesis. I would like to extend my gratitude to Dr Binafsha Syed and Dr Mohammad Albanghali, two excellent former students who worked on this project before me and paved the way for this thesis.

I thank those who collaborated with me on elements of this thesis I could not do alone including Professor Graham Ball for his assistance with bioinformatics analysis, Dr Grazziela Figueredo for her help with cluster analysis, Dr Andrea Venn for her statistics expertise and Ms Alison Ashmore for her support with literature reviewing.

I acknowledge the help and support of all members of the Breast Pathology Research Group at the University of Nottingham, for their inspiration, assistance with laboratory work and analysis, including: Professor Emad Rakha, Professor Ian Ellis, Dr Omar Mohammed, Dr Chitra Joseph, Dr Michael Toss, Mr Christopher Nolan, Ms Holly Nicholls, Dr Lutfi Alfarsi, Ms Brendah Masisi and Dr Yousif Kariri.

I would like to thank my clinical Training Programme Director, now Head of School and mentor Miss Gill Tierney for her support and encouragement to pursue time out of training to complete this thesis.

Finally, I would like to thank my parents Mr Roger Parks and Mrs Norma Parks and all my extended family, friends and supporters for their encouragement, prayers and good wishes.

Table of contents

BIOLOGICAL ASSESSMENT - DEVELOPING A NOVEL TOOL TO GUIDE TREATMENT DECISION IN OLDER WOMEN WITH PRIMARY BREAST CANCER I
Declaration II
Ethical approval III
Dedication IV
Abstract V
Relevant work published during this PhD VIII
Original articles VIII
Book chapters X
Published abstracts X
Related prizes and awards during this PhD XIV
Research fellowships obtained during this thesis XVI
Acknowledgements XVII
Table of contents XVIII
Legend of tables XXVII
Legend of figures XXX
Abbreviations XXXVIII
INTRODUCTION 1
Chapter 1: Current challenges facing older women with primary breast cancer 1
1.1 Overview 2
1.2 Ageing and breast cancer 2
1.2.1 How do we define 'older' women with breast cancer? 2
1.2.2 Breast cancer incidence in older women worldwide 3
1.2.3 Breast cancer incidence in older women in the UK 5
1.3 Treatment of primary breast cancer in older women 6
1.3.1 Primary treatment 6
1.3.1.1 Surgery 6
1.3.1.2 Primary endocrine therapy 7
1.3.1.3 Alternative options for women with ER-negative breast cancer 8
1.3.2 Adjuvant treatment 8
1.3.2.1 Local treatment 9
1.3.2.2 Systemic therapies 9

Endocrine therapy 9
Chemotherapy 10
Trastuzumab 11
1.3.3 What are the challenges with current treatment of primary breast cancer in older women? 11
1.3.3.1 Primary treatment 11
1.3.3.2 Adjuvant treatment 14
1.4 Outcome measures 17
1.5 The potential role of biology in predicting cancer outcomes in primary breast cancer 17
1.5.1 Understanding subtypes of breast cancer 18
1.5.2 The challenges of assessment of biology in primary breast cancer in older women 19
1.6 Background to the research programme in Nottingham 20
1.6.1 Overview of the research programme 20
1.6.2 Work on the biology arm of the research programme to date 21
1.6.2.1 Clinical data 21
1.6.2.2 Biological data 21
Surgical series 21
Core needle biopsy (CNB) series 21
1.6.2.3 Summary of significant findings to date 21
1.6.3 Next steps 22
1.6.3.1 Full characterisation of the biology of breast cancer in older women 22
1.6.3.2 Association between biomarker expression and survival outcome 22
1.6.3.3 Cluster analysis 22
1.6.3.4 Bioinformatics analysis and development of prognostic algorithm 22
1.6.3.5 Validation of prognostic algorithm 23
1.6.4 Hypothesis of this thesis and how the work will contribute to the overall research programme 23
1.7 Summary 24
Chapter 2: A review of the literature on the biology of primary breast cancer in older women beyond routinely measured biomarkers 25
2.1 Introduction 26
2.2 Objectives 26
2.3 Materials and methods 27
2.3.1 Search strategy 27

2.3.2 Data extraction 29
2.3.3 Critical appraisal 29
2.4 Results 29
2.4.1 Summary 29
2.4.2 Reporting standard 31
2.4.3 General characteristics of the studies 34
2.4.4 Significant findings 43
2.4.4.1 Differences between older and younger women 52
2.4.4.2 Relationship of biomarker with survival 52
2.4.4.3 Associations of biomarker with chemotherapy use 52
2.5 Discussion 53
2.5.1 Reporting standard 53
2.5.2 Differences between older and younger women 53
2.5.3 Relationship of biomarker with survival 54
2.5.4 Associations of biomarker with chemotherapy use 56
2.5.5 Limitations of this review 57
2.6 Conclusions and implications for this thesis 58
Chapter 3: Clinical relevance of biomarkers in primary breast cancer in older women 60
3.1 Introduction 61
3.2 Biomarkers – Association with age, predictive and prognostic values, future directions 67
3.2.1 Hormonal receptors 67
3.2.1.1 Oestrogen receptor (ER) 67
3.2.1.2 Progesterone receptor (PR) 70
3.2.1.2 Androgen receptor (AR) 71
3.2.2 Growth factor receptors 73
3.2.2.1 Epidermal growth factor receptor (EGFR) 74
3.2.2.2 Human epidermal growth factor receptor 2 (HER2) 75
3.2.2.3 Human epidermal growth factor receptor 3 (HER3) 77
3.2.2.4. Human epidermal growth factor receptor 4 (HER4) 78
3.2.2.5 Vascular endothelial growth factor (VEGF) 79
3.2.3 Tumour suppressor genes 81
3.2.3.1 Tumour suppressor protein 53 (p53) 81
3.2.3.2 Breast cancer gene (BRCA) 1 and 2 82

3.2.3.3 Murine double minutes (MDM) 2 & 4 84
3.2.3.4 Phosphatase and tensin homolog (PTEN) 85
3.2.3.5 Liver kinase B1 (LKB1) 88
3.2.4 Cell proliferative markers 89
3.2.4.1 Ki67 89
3.2.5 Anti-apoptosis markers 90
3.2.5.1 B cell lymphoma 2 (BCL2) 90
3.2.6 Cell adhesion molecules 92
3.2.6.1 E-cadherin 92
3.2.6.2 Cluster of differentiation (CD) 44 93
3.2.7 Cell differentiation markers 94
3.2.7.1 Cytokeratins (CKs) 94
3.2.8 Others 96
3.2.8.1 Mucin 1 (MUC1) 96
3.2.8.2 Amplified in breast cancer 1 (AIB1) 97
3.2.8.3 Solute carriers 98
3.2.9 Other potential biomarkers 99
3.3 Predictive and prognostic tools using tumour tissue 100
3.3.1 Background to predictive and prognostic tools 100
3.3.2 Evidence base for use of predictive and prognostic tools in older women with primary breast cancer 101
3.3.2.1 Presentation of the evidence base 101
3.3.2.2 Summary of the evidence base 105
3.3.2.3 Current issues with use of tumour profiling tools in older women 106
Age range 106
Setting 107
Tissue sample required 107
3.3.3 Future directions of predictive and prognostic tools in older women 108
3.4 Conclusions and implications for this thesis 109
METHODS 111
Chapter 4: Outline of materials and methods for remainder of thesis 111
4.1 Introduction 112
4.2 Patients 112
4.2.1 Older women series 112
4.2.1.1 Historic database 112

4.2.1.2 Treatment protocols 115
1973 – 1989 115
1990 – 1999 115
4.2.1.3 Follow-up 117
4.2.2 Younger women series 117
4.3 Construction of tissue microarrays (TMAs) 117
4.3.1 Rationale for using TMAs 118
4.3.2 Surgical excision (SE) TMA 118
4.3.3 Core needle biopsy (CNB) TMA 119
4.4 Western blot (WB) 119
4.5 Immunohistochemistry 119
4.5.1 Antibody optimisation 122
4.5.2 Immunohistochemical staining 122
4.6 Evaluation of staining 124
4.6.1 Digitalisation 124
4.6.2 Scoring 124
4.6.3 Validation 125
4.7 Association between biomarker and clinico-pathological features 126
4.8 Cluster analysis 126
4.9 Bioinformatics 126
4.10 Summary of methods and next steps of thesis 127
4.11 Change in methodology and updated thesis plan 127
RESULTS AND DISCUSSION 131
Chapter 5: Immunohistochemistry and H-scoring 131
5.1 Introduction 132
5.1.1 Patient characteristics 134
5. 2 Immunohistochemistry (IHC) 135
5.2.1 Biomarkers measured in SE TMA 136
5.2.1.1 AIB1 136
5.2.1.2 PTEN 137
5.2.1.3 AR 138
5.2.1.4 Solute carrier family 1 member 5 (SLC1A5) 140
5.2.1.5 SLC3A2 141
5.2.1.6 SLC7A5 142
5.2.2 Biomarkers measured in CNB TMAs 143

5.2.2.1 BRCA2 143
5.2.2.2 CD44 144
5.2.2.3 E cadherin 145
5.2.2.4 MDM2 146
5.2.2.5 MDM4 148
5.2.2.6 CK5 149
5.2.2.7 CK14 150
5.2.2.8 CK17 151
5.2.2.9 CK18 152
5.2.2.10 CK19 153
5.2.3 Comparison of IHC staining to previous work and literature 154
5.3 H-scoring of biomarkers stained by IHC for this thesis 156
5.3.1 H-score results 156
5.3.2 Comparison of H scores to the literature 159
5.4 Summary of this chapter and next steps 164
Chapter 6: Core needle biopsy (CNB) versus surgical excision (SE) comparison 166
6.1 Introduction 167
6.2 Comparison of CNB TMA with SE TMA results using basic statistical analysis 167
6.2.1 Methods 167
6.2.2 Results 169
6.2.3 Interpretation of results 172
6.3 Comparison of CNB TMA with SE TMA results using principal components analysis (PCA) 174
6.3.1 Methods 176
6.3.2 Results 176
6.3.3 Interpretation of results 180
6.4 Conclusions of this chapter and next steps 180
Chapter 7: Cluster analysis of all cases in core needle biopsy tissue microarrays (CNB TMAs) 182
7.1 Introduction 183
7.2 Previous cluster analysis 183
7.3 Methods 188
7.3.1 Handling of missing values 189
7.3.2 Validity Indices 190
7.3.3 Characterisation of clusters 190

7.3.4 Cluster validation 190
7.3.5 Identification of key drivers and reconstruction of clusters 191
7.3.6 Breast cancer-specific survival (BCSS) 191
7.4 Results 191
7.4.1 Characterisation of clusters 191
7.4.2 Characterisation of clusters and validation 194
7.4.3 Box-plots demonstrating the biomarker contribution to each cluster 195
7.4.4 Identification of key drivers 197
7.4.5 Association between clusters and BCSS 198
7.5 Explanation of findings 199
7.6 Conclusions of this chapter and next steps 201
Chapter 8: Generation of a prognostic algorithm 203
8.1 Introduction 204
8.2 Methods 204
8.2.1 Software utilised 204
8.2.2 Handling of missing data 205
8.2.2.1 Method for filling missing data 205
8.2.3 Disease outcomes explored 206
8.2.4 Data mining 207
8.2.4.1 Artificial Neural Network 207
8.2.4.2 Architecture 207
8.2.4.3 Formulation of the prognostic algorithm 208
8.2.4.4 Validation within the existing Nottingham dataset 208
8.3 Results 208
8.3.1 Surgical excision (SE) tissue microarrays (TMAs) 209
8.3.1.1 SE TMAs – breast cancer specific survival 209
8.3.1.2 SE TMAs – overall survival 211
8.3.1.3 SE TMAs – time to metastasis 213
8.3.1.4 SE TMAs – time to regional recurrence 215
8.3.1.5 SE TMAs – time to progression of disease 217
8.3.2 Core needle biopsy (CNB) tissue microarrays (TMAs) 219
8.3.2.1 CNB TMAs – breast cancer specific survival 219
8.3.2.2 CNB TMAs – overall survival 221
8.3.2.3 CNB TMAs – time to metastasis 223
8.3.2.4 CNB TMAs – time to regional recurrence 225

8.3.2.5 CNB TMAs – time to progression of disease 227
8.4 Further results 228
8.4.1 Algorithm to predict BCSS in SE TMA BCSS 229
8.4.1 Algorithm to predict BCSS in CNB TMA 238
8.5 Explanation of results 248
8.6 Conclusions of this chapter and next steps 250
Chapter 9: Refining the prognostic algorithm 251
9.1 Introduction 252
9.2 Refining prognostic algorithms to predict BCSS 252
9.2.1 Introduction 252
9.2.2 Methods 252
9.2.2.1 Calculating BCSS in SE TMAs 253
9.2.2.2 Calculating BCSS in CNB TMAs 254
9.2.2.3 Optimised panel of biomarkers 256
9.2.3 Results 257
9.2.3.1 Optimised panel of biomarkers for SE TMAs - BCSS 257
9.2.3.2 Optimised panel of biomarkers for CNB TMAs - BCSS 259
9.2.4 Explanation of results and potential next steps 261
9.2.4.1 Biomarkers of independent prognostic significance for BCSS 262
9.2.4.2 Algorithms generated using optimised panel of biomarkers 264
9.2.4.3 Potential next steps 265
9.3 Conclusions of this chapter and next steps 267
GENERAL DISCUSSION 268
Chapter 10: General discussion 268
10.1 Introduction 269
10.2 Key findings 269
Chapter 1 269
Chapter 2 270
Chapter 3 270
Chapter 4 271
Chapter 5 271
Chapter 6 272
Chapter 7 272
Chapter 8 273
Chapter 9 273

10.3 Limitations 274
10.4 Addition to knowledge and change in practice 276
10.5 Informing future research 277
10.5.1 Next steps of this present work 277
10.5.2 Concurrent work in the biology arm of the research programme 278
10.5.3 Geriatric assessment arm of the research programme 280
10.5.4 Other potential applications of this work 281
10.6 The future of treatment for older women with primary breast cancer 281
10. 7 Conclusions of this thesis 282
References 283
Appendices 339

Legend of tables

Table 1: Results of Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK) assessment of full-text paper included 33

Table 2: Characteristics of the 23 studies included in the systematic review 42

Table 3: Findings of the 23 studies including implications for clinical practice and future directions 51

Table 4: Summary of biomarkers associations with age, efficacy of endocrine therapy and survival..... 66

Table 5: Percentage of patients with oestrogen receptor-positive breast cancer, by age group in a cohort of 18586 breast cancer patients (1) 67

Table 6: A summary of six most common tumour-based predictive and prognostic tools 104

Table 7: Information available in database of older women with primary breast cancer 114

Table 8: Details of biomarkers, antibody clone and retrieval methods used for panel of 31 biomarkers 121

Table 9: Panel of 15 biomarkers with clinical data available 129

Table 10: An indication of which of the panel of 31 biomarkers had been measured prior to this thesis. X indicates marker measured prior to this thesis; - indicates marker has not yet been measured. 134

Table 11: General characteristics of the older women cohort 135

Table 12: The expression of biological markers in the SE TMAs 157

Table 13: The expression of biological markers in the CNB TMAs 159

Table 14: The number of positive/negative samples of biological markers in the SE TMAs based on prior definitions 160

Table 15: The number of positive/negative samples of biological markers in the CNB TMAs based on prior definitions 163

Table 16: Summary of 15 biomarkers for comparative analysis between CNB TMA and SE TMA 168

Table 17: Concordance between CNB TMA and SE TMA based on two categories 'positive' (+ve) or 'negative'(-ve) 171

Table 18: Concordance between CNB TMA and SE TMA based on continuous H-score 171

Table 19: Eigenvalues and percentage variance for each component of principal components analysis 177

Table 20: Number of cases allocated to each cluster in CNB TMAs 195

Table 21: Performance of the algorithm designed to predict BCSS in SE TMAs ... 209

Table 22: Performance of the algorithm designed to predict OS in SE TMAs 211

Table 23: Performance of the algorithm designed to predict time to metastasis in SE TMAs 213

Table 24: Performance of the algorithm designed to predict TRR in SE TMAs 215

Table 25: Performance of the algorithm designed to predict time to disease progression in SE TMAs 217

Table 26: Performance of the algorithm designed to predict BCSS in CNB TMAs 219

Table 27: Performance of the algorithm designed to predict OS in CNB TMAs..... 221

Table 28: Performance of the algorithm designed to predict TTM in CNB TMAs . 223

Table 29: Performance of the algorithm designed to predict TRR in CNB TMAs.. 225

Table 30: Performance of the algorithm designed to predict TTP in CNB TMAs... 227

Table 31: Summary of the association between each biomarker within the prognostic algorithms generated and BCSS 248

Table 32: The optimised panel of biomarkers for each set of TMAs taken from biomarkers which have shown independent significance for BCSS 257

Table 33: Performance of the algorithm designed to predict BCSS in SE TMAs using optimised panel of biomarkers 258

Table 34: Performance of the algorithm designed to predict BCSS in CNB TMAs using optimised panel of biomarkers..... 260

Table 35: Comparison of r-values for prognostic models generated for BCSS 262

Legend of figures

Figure 1: Estimated number of incident cases of breast cancer worldwide in population aged ≥70 years, from 2020 to 2040 (13) - permission to use obtained 12th May 2021 4

Figure 2: Breast cancer average number of new cases per year and age-specific incidence rates per 100,000 population, females, UK, 2014-2016 (2) - permission to use not required 5

Figure 3: Comparison of surgery plus adjuvant endocrine therapy versus PET with overall survival (26)..... 7

Figure 4: Comparison of surgery plus adjuvant endocrine therapy versus PET with local disease control (26) - permission to use obtained 26th May 2020..... 7

Figure 5: Subgroup analysis by age of breast cancer mortality, chemotherapy versus no chemotherapy from the Early Breast Cancer Trialists' Collaborative Group meta-analysis on adjuvant chemotherapy (52) - permission to use obtained 16th May 2021 10

Figure 6: Current understanding of subtypes of breast cancer (100) - permission to use not required 19

Figure 7: Selection of studies for review with age cut-off ≥65 years selection 30

Figure 8: The Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) checklist (121)..... 32

Figure 9: Effects of about 5 years of tamoxifen on the 15-year probabilities of recurrence and of breast cancer mortality, for ER-positive disease (45) - permission to use not required 69

Figure 10: Structure of ErbB receptors (200) - permission to reuse not required..... 73

Figure 11: PI3K/Akt/mTOR and Ras/Raf/MEK/ERK signalling pathways (202) - permission to use obtained 2nd February 2020..... 74

Figure 12: A summary of therapeutic targets including HER3, HER2, HER4 and EGFR (222) - permission to use not required..... 79

Figure 13: Overview of angiogenic pathways (235) - permission to use not required80

Figure 14: Mechanism of action of olaparib (251) - permission to use not required. 84

Figure 15: Kaplan-Meier estimates of progression-free survival of patients treated with everolimus plus exemestane versus exemestane alone. CI confidence interval, HR hazard ratio, EVE everolimus, EXE exemestane, PBO placebo (264) - permission to use not required 87

Figure 16: Action of BH3 mimetic drugs (283) - permission to use not required 92

Figure 17: AIB IHC staining attempt 1 in SE TMAs..... 137

Figure 18: AIB IHC staining attempt 2 in SE TMAs..... 137

Figure 19: PTEN IHC staining in SE TMAs..... 138

Figure 20: AR IHC staining in SE TMAs 139

Figure 21: AR IHC staining in CNB TMAs..... 140

Figure 22: SLC1A5 IHC staining in SE TMAs..... 141

Figure 23: SLC3A2 IHC staining in SE TMAs..... 142

Figure 24: SLC7A5 IHC staining in SE TMAs..... 143

Figure 25: BRCA2 IHC staining in CNB TMAs 144

Figure 26: CD44 IHC staining in CNB TMAs..... 145

Figure 27: E cadherin IHC staining in CNB TMAs 146

Figure 28: MDM2 IHC staining in CNB TMAs 147

Figure 29: CK5 IHC staining in CNB TMAs..... 149

Figure 30: CK14 IHC staining in CNB TMAs..... 150

Figure 31: CK17 IHC staining in CNB TMAs..... 151

Figure 32: CK18 IHC staining in CNB TMAs..... 152

Figure 33: CK17 IHC staining in CNB TMAs..... 153

Figure 34: Scatterplots to demonstrate the variance between each component. CNB TMAs cases are represented by red and SE TMAs by black. A – component 1 (y-axis) versus component 2 (x-axis); B – component 2 (y) versus component 3 (x); C - component 3 (y) versus component 4 (x); D - component 4 (y) versus component 5 (x); E - component 5 (y) versus component 6 (x); F - component 6 (y) versus component 7 (x); G - component 7 (y) versus component 8 (x); H - component 8 (y) versus component 9 (x); I - component 9 (y) versus component 10 (x)..... 178

Figure 35: Scatterplots to demonstrate the variance between each component. CNB TMAs cases are represented by red and SE TMAs by black. J - component 10 (y-axis) versus component 11 (x-axis); K - component 11 (y) versus component 12 (x); L - component 12 (y) versus component 13 (x); M - component 13 (y) versus component 14 (x); N – component 14 (y) versus component 15 (x); O - component 15 (0x) versus component 1 (x)..... 179

Figure 36: Biological classes of early operable primary breast cancer in SE TMAs – A) ≥70 years versus B) <70 years women. C) BCSS of older women with early operable primary breast cancer according to the biological clusters (3) – permission to use not required 186

Figure 37: Biological classes of early operable primary breast cancer in CNB TMAs (113) – permission to use not required 187

Figure 38: BCSS of biological cluster in ER-positive CNB TMAs in older women – permission to use not required..... 188

Figure 39: Validity indices for K-means showing the possible number of clusters using six of the most widely used indices in data mining. The red dot indicates the suggested optimal number of clusters for each method. 192

Figure 40: Validity indices for PAM showing the possible number of clusters using six of the most widely used indices in data mining. The red dot indicates the suggested optimal number of clusters for each method. 193

Figure 41: Bi-plots for validity indices showing a pattern of four clusters – three true clusters (red, yellow and purple dots) and one group which was unclustered (blue dots). The contribution of each biomarker is given by the blue lines, demonstrating that ER, PR and MUC1 are important in determining cluster allocation. 194

Figure 42: The contribution of each marker in the individual clusters. 196

Figure 43: BCSS of biological clusters in all CNB TMAs 198

Figure 44: Actual BCSS compared to predicted BCSS in SE TMAs 210

Figure 45: Actual OS compared to predicted OS in SE TMAs..... 212

Figure 46: Actual TTM compared to predicted TTM in SE TMAs 214

Figure 47: Actual time to RR compared to predicted TRR in SE TMAs..... 216

Figure 48: Actual TTP compared to predicted TTP in SE TMAs..... 218

Figure 49: Actual BCSS compared to predicted BCSS in CNB TMAs 220

Figure 50: Actual OS compared to predicted OS in CNB TMAs 222

Figure 51: Actual TTM compared to predicted TTM in CNB TMAs..... 224

Figure 52: Actual TRR compared to predicted TRR in CNB TMAs..... 226

Figure 53: Actual TTP compared to predicted TTP in CNB TMA..... 228

Figure 54: Contribution of each biomarker to model predicting BCSS in SE TMAs 229

Figure 55: Relationship between ER expression and ability to predict BCSS within the model 230

Figure 56: Relationship between PR expression and ability to predict BCSS within the model 231

Figure 57: Relationship between HER2 expression and ability to predict BCSS within the model 231

Figure 58: Relationship between Ki67 expression and ability to predict BCSS within the model 232

Figure 59: Relationship between P53 expression and ability to predict BCSS within the model 232

Figure 60: Relationship between CK 5/6 expression and ability to predict BCSS within the model 233

Figure 61: Relationship between CK 7/8 expression and ability to predict BCSS within the model 233

Figure 62: Relationship between EGFR expression and ability to predict BCSS within the model Within this model, increasing EGFR positivity was associated with a decrease in BCSS..... 234

Figure 63: Relationship between BCL2 expression and ability to predict BCSS within the model Within this model, increasing BCL2 positivity was associated with an increase in BCSS. 234

Figure 64: Relationship between MUC1 expression and ability to predict BCSS within the model 235

Figure 65: Relationship between VEGF expression and ability to predict BCSS within the model 235

Figure 66: Relationship between LKB1 expression and ability to predict BCSS within the model 236

Figure 67: Relationship between BRCA1 expression and ability to predict BCSS within the model 237

Figure 68: Relationship between HER3 expression and ability to predict BCSS within the model 237

Figure 69: Relationship between HER4 expression and ability to predict BCSS within the model 238

Figure 70: Contribution of each biomarker to model predicting BCSS in CNB TMAs 239

Figure 71: Relationship between ER expression and ability to predict BCSS within the model 239

Figure 72: Relationship between PR expression and ability to predict BCSS within the model 240

Figure 73: Relationship between HER2 expression and ability to predict BCSS within the model 240

Figure 74: Relationship between Ki-67 expression and ability to predict BCSS within the model 241

Figure 75: Relationship between p53 expression and ability to predict BCSS within the model 241

Figure 76: Relationship between CK 5/6 expression and ability to predict BCSS within the model 242

Figure 77: Relationship between CK 7/8 expression and ability to predict BCSS within the model 242

Figure 78: Relationship between EGFR expression and ability to predict BCSS within the model 243

Figure 79: Relationship between BCL-2 expression and ability to predict BCSS within the model 243

Figure 80: Relationship between MUC1 expression and ability to predict BCSS within the model 244

Figure 81: Relationship between VEGF expression and ability to predict BCSS within the model 245

Figure 82: Relationship between LKB1 expression and ability to predict BCSS within the model 246

Figure 83: Relationship between BRCA1 expression and ability to predict BCSS within the model 246

Figure 84: Relationship between HER3 expression and ability to predict BCSS within the model 247

Figure 85: Relationship between HER4 expression and ability to predict BCSS within the model 247

Figure 86: Prognostic significance of A) ER, B) PR, C) BCL2, D) MUC1 in early operable primary breast cancer treated by surgery – BCSS (3) - permission to use not required 253

Figure 87: Prognostic significance of A) HER2, B) EGFR, C) Ki67, D) p53 in early operable primary breast cancer treated by surgery – BCSS (3) – permission to use not required 254

Figure 88: Prognostic significance of A) ER, B) PR, C) BCL2, D) VEGF in CNB TMAs in early operable primary breast cancer in older women– BCSS 255

Figure 89: Prognostic significance of A) Ki67, B) p53, C) EGFR, D) MUC1 in CNB TMAs in early operable primary breast cancer in older women– BCSS 256

Figure 90: Actual BCSS compared to predicted BCSS in SE TMAs using optimised panel of biomarkers 259

Figure 91: Actual BCSS compared to predicted BCSS in CNB TMAs using optimised panel of biomarkers 261

Figure 92: BCSS in cases of CNB TMAs with positive VEGF expression, comparison between primary surgery and non-surgical treatment..... 264

Abbreviations

#	study number
+ve	positive
-ve	negative
ABCSG	Austrian Breast and Colorectal Cancer Study Group
ASCO	American Society of Clinical Oncology
AI	aromatase inhibitor
AIB	amplified in breast cancer
ALDH	aldehyde dehydrogenase
ALND	axillary lymph node dissection
AR	androgen receptor
ATAC	Arimidex, Tamoxifen Alone, or in Combination
B2M	beta-2 macroglobulin
BC	breast cancer
BCFI	breast cancer-free interval
BCL	B-cell lymphoma
BCS	breast conserving surgery
BCSS	breast cancer specific survival
BIG	breast international group
BGA	bio-geriatric assessment

BOLERO Breast Cancer Trials of OraL EveROlimus-2

BRCA breast cancer gene

BTAG Bridging the age gap in breast cancer

CA(number) cancer antigen

CAM cell adhesion molecule

c-cyclin E cytoplasmic cyclin E

CD cluster of differentiation

CDK cyclin dependent kinase

CEA carcinoembryonic antigen

CHK1 checkpoint kinase 1

CISH chromogenic in situ hybridization

CK cytokeratin

CNB core needle biopsy

CT chemotherapy

DAB (3,3'-Diaminobenzidine)

DFI disease free interval

DFS disease free survival

DM distant metastasis

DMFS distant metastasis free survival

DRFI disease recurrence free interval

DRFS distant recurrence free survival

DSS disease specific survival

EBCTCG Early Breast Cancer Trialists' Collaborative Group

ECOG Eastern Cooperative Oncology Group

EGFR epidermal growth factor receptor

ER oestrogen receptor

EUSOMA European Society of Cancer Specialists

FDA Food and Drug Administration (US)

FOXP3 forkhead box P3

FFPE formalin-fixed paraffin-embedded

FISH fluorescence in situ hybridization

GA geriatric assessment

Gy gray (unit)

H&E hematoxylin and eosin stain

HER human epidermal growth factor receptor

HERA herceptin adjuvant

HR hormone receptor

ICC intraclass correlation coefficient

IDC invasive ductal carcinoma

IGF insulin growth factor

IL interleukin

IHC immunohistochemistry

IQR interquartile range

IMS industrial methylated spirit

KNN K-nearest neighbour

LK liver kinase

LR local recurrence

LT long-term

LN lymph node

LVEF left ventricular ejection fraction

MAE mean absolute error

MAP2K mitogen-activated protein kinase kinase

MCP monocyte chemoattractant protein

MDM mouse double minute homolog

MDS ADLs minimum data set activities of daily living

MMPs matrix metalloproteases

MoAB monoclonal antibody

MOS Medical Outcomes Study

mTOR mammalian target of rapamycin

MUC mucin

N number

N0 node negative

NA neoadjuvant

NABCOP National Audit of Breast Cancer in Older Patients

NACT neoadjuvant chemotherapy

NAET neoadjuvant endocrine therapy

NCoR nuclear receptor corepressor

NICE National Institute for Health and Care Excellence

NMAE Normalised mean absolute error

NPI Nottingham Prognostic Index

NRMSE Normalised root mean squared error

NSABP National Surgical Adjuvant Breast and Bowel Project

OR odds ratio

OS overall survival

p53 tumour protein p53

PAM Prosigna (tool)

PAM Partition around medoids (cluster analysis)

PARP Poly (ADP-ribose) polymerases

PCA principal component analysis

pCR pathological complete response

PCR polymerase chain reaction

PD programmed cell death protein

PET primary endocrine therapy

PFS progression-free survival

PR progesterone receptor

PRISMA Preferred Reporting Items for Systematic Review and Meta-Analysis

PNAd peripheral node addressin

PTEN phosphatase and tensin homolog

QoL quality of life

r r-value (correlation coefficient)

REMARK Reporting Recommendations for Tumour Marker Prognostic Studies

RCT randomised controlled trial

RIBBON Regimens in Bevacizumab for Breast Oncology

RFS recurrence free survival

RMSE root mean squared error

RR recurrence risk (in relation to cancer)

RR relative risk (in relation to statistics)

RS recurrent score

RT radiotherapy

RT-PCR reverse transcriptase polymerase chain reaction

RUNX runt-related transcription factor

RxPONDER The Clinical Trial for Treatment for Positive Node, Endocrine Responsive Breast Cancer

SE surgical excision

SIOG International Society for Geriatric Oncology

SISH silver in situ hybridization

SLC solute carrier

SLNB sentinel lymph node biopsy

SPARC secreted protein acidic and rich in cysteine

SPSS Statistical Package for Social Sciences

SUMO small ubiquitin-like proteins

T Tumor (in relation to size)

TAILORx Trial Assigning Individualized Options for Treatment

TBS-T tris-buffered saline with tween

TC docetaxel plus cyclophosphamide

TEAM The Tamoxifen Exemestane Adjuvant Multinational trial

TFF1 trefoil factor 1

TIL tumour-infiltrating lymphocyte

TP translocator protein

TTM time to metastasis

TTP time to progression

TRR time to regional recurrence

TIMPs tissue inhibitors

TLS tertiary lymphoid structure

TMA tissue microarray

TNBC triple negative breast cancer

tPA tissue plasminogen activator

TTR time to recurrence

TOP2A topoisomerase 2-alpha

U unknown

uPA urokinase

VEGF vascular endothelial growth factor

VES Vulnerable Elders Survey

vs versus

WB Western blot

INTRODUCTION

Chapter 1: Current challenges facing older women with primary breast cancer

1.1 Overview

Breast cancer risk increases with age and our population is ageing. This chapter outlines why there is an urgent need for greater understanding of the biology of primary breast cancer specifically in older women and the implications this has for alterations in standard care offered to this group.

Firstly, the definition of 'older' women with breast cancer will be explained, followed by establishing the size of the problem worldwide and then specifically in the UK. The current treatment strategies for older women with primary breast cancer will be briefly summarised, followed by a discussion of the areas where current treatment dilemmas most frequently arise. Next, commonly reported cancer outcome measures will be defined and suggestions as to how these might differ in the older patient deliberated. The role of the biology of breast cancer in helping to predict cancer outcome will then be considered. Following this, the chapter will introduce the Nottingham research programme on primary breast cancer in older women and summarise work to date. The next steps in the research programme will then be identified and an outline of the contribution this current thesis will make.

1.2 Ageing and breast cancer

1.2.1 How do we define 'older' women with breast cancer?

Breast cancer is the most common cancer in women worldwide (1). After female gender, age is the highest risk factor for development of breast cancer; risk increasing proportionally with age (2). Existing data suggests that the biology of breast cancer in older women is different compared to their younger counterparts (3) and this may impact treatment decision making, alongside conventional factors, patient preference and discussion with the multidisciplinary healthcare team.

Biological changes occur as a result of the physiological ageing process. An arbitrary cut-off is required for reporting purposes and provision of services to 'older' persons. There is no current international consensus to define 'older' age or 'elderly' (4) as there are vast worldwide differences in mortality, health and socioeconomic status and life expectancy. The Office for National Statistics in the UK currently defines 'elderly' as ≥65 years of age which until recently, was the age of receiving the State Pension (5). Traditionally, the United Nations have defined 'older persons' as those aged 60 or 65 years or over (6).

Although it is recognised that there are differences between chronological and biological age, a linear relationship between them has been identified (7); therefore, chronological age can be used as a surrogate of biological age in most cases. However, the milestone ages of 40 and 70 years have been shown to mark a significant change in the pattern of breast cancer; <40 years appears to be the phase with the most aggressive phenotypes; >70 the less aggressive phase and 40-70 years a transitional phase (8).

Historically, medical investigations and treatments were often limited in the older population; however, with modern practices, change in opinion and increased life expectancy, this is no longer the case. With the above in mind, for the purpose of this thesis, older women will be considered as a concept rather than using a predetermined age cut-off.

1.2.2 Breast cancer incidence in older women worldwide

In 2018, the number of new breast cancers diagnosed worldwide was approximately 2 million (9) with around 25% of cases in patients ≥70 years (10). Due to increase in average age of the world's population, it is expected that the number of older women

living with breast cancer will double by 2040 to nearly 1 million cases in patients ≥70 worldwide per year (11, 12) (Figure 1).

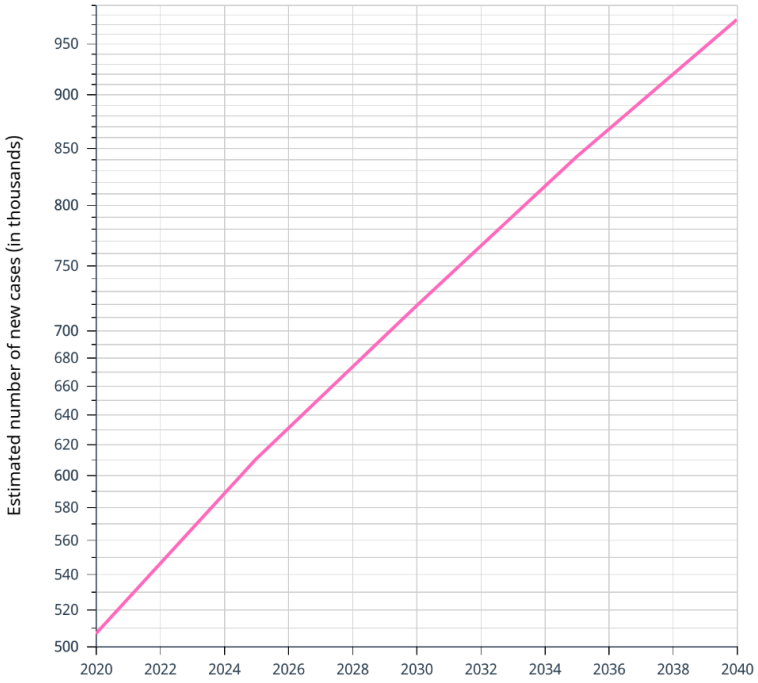


Figure 1: Estimated number of incident cases of breast cancer worldwide in population aged ≥70 years, from 2020 to 2040 (13) - permission to use obtained 12th May 2021

The ageing of the world population is the result of the continued decline in fertility rates and increased life expectancy. This demographic change has resulted in increasing proportions of older persons; older women currently outnumber older men due to the gender gap in longevity (14). In 2019, approximately 9% of the global population were ≥65 years of age (15) and this is expected to double by 2050 (16).

The crisis of our ageing population has been well documented (17, 18); the burden on health service resources in terms of financial and service provision will be unprecedented. Breast cancer in older women will have a significant impact on this.

1.2.3 Breast cancer incidence in older women in the UK

The global burden of breast cancer is mirrored in the UK. Age-specific incidence rates in the UK rise steadily from around age 30-34 years and more steeply from around age 70-74 (2) (Figure 2).

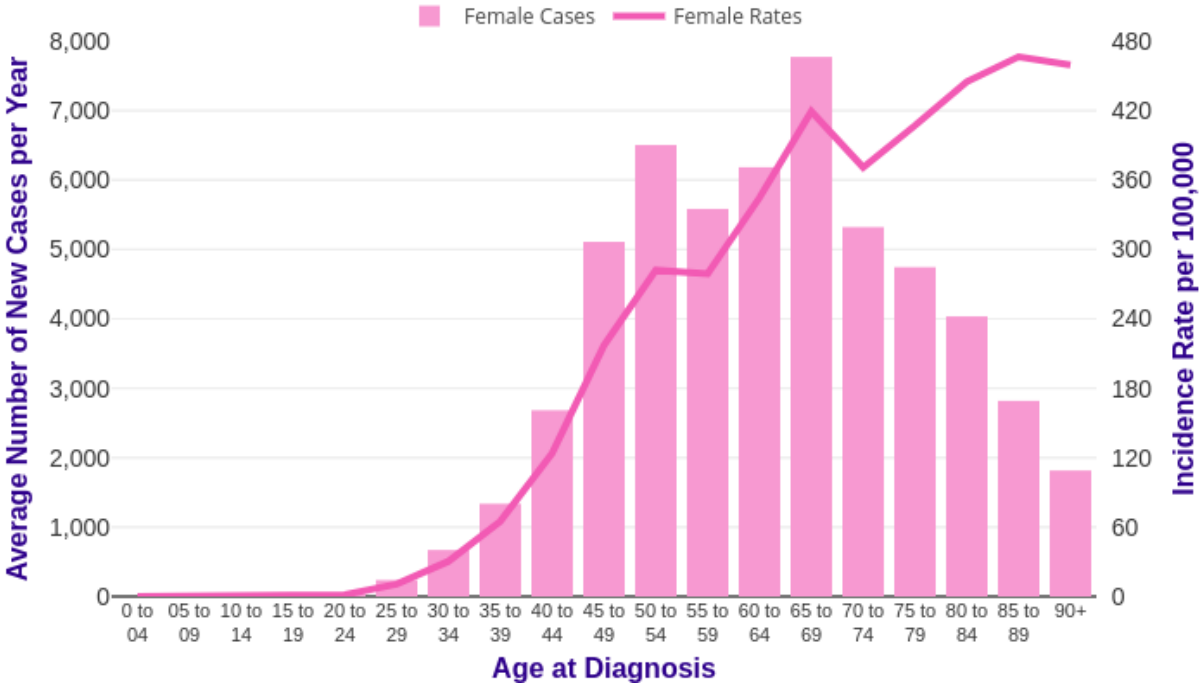


Figure 2: Breast cancer average number of new cases per year and age-specific incidence rates per 100,000 population, females, UK, 2014-2016 (2) - permission to use not required

In 2012, there were around 340,000 women aged 65 and over, out of around 570,000 women of all ages living with breast cancer in the UK. By 2040, this is projected to increase to 1.2 million (out of a total of 1.68 million) – an almost four-fold increase; almost three-quarters of all women living with breast cancer in the UK will be aged over 65 (11, 19).

Despite this stark reality, there remains insufficient evidence specific to older women with breast cancer (20).

1.3 Treatment of primary breast cancer in older women

Regardless of age, recommendations suggest that primary breast cancer is treated surgically with mastectomy or breast conserving surgery (BCS) (21) with consideration of adjuvant treatments (22). The current treatments offered to older women will now be discussed, followed by an outline of the challenges in terms of primary and adjuvant therapies.

1.3.1 Primary treatment

Surgery is the primary treatment of choice for all women irrespective of age (20). For older, more frail, women with oestrogen receptor (ER)-positive early breast cancer, primary endocrine therapy (PET) may be an option. Primary radiotherapy has historically been used in a small group of patients with ER-negative early breast cancer as an alternative to surgery.

1.3.1.1 Surgery

The National Institute for Health and Care Excellence (NICE) in the UK advise surgery as the first-line treatment of primary operable breast cancer, irrespective of age (21). This advice is echoed in European (23), American (24) and international (25) guidelines.

The current standard of care for all patients with a diagnosis of invasive breast cancer and clinically negative axillary nodes is sentinel lymph node biopsy (SLNB). More extensive axillary treatment is beyond the scope of this thesis, which is concerned with early stage, primary breast cancer and will therefore, not be discussed.

1.3.1.2 Primary endocrine therapy

The justification for the use of PET comes from the landmark Cochrane meta-analysis by Hind et al (26), published in 2006, comparing surgery and PET in older women, which did not show a significant difference in favour of either surgery or PET in respect to overall survival (Figure 3), however, progression-free survival (PFS) was in favour of the surgical group (Figure 4).

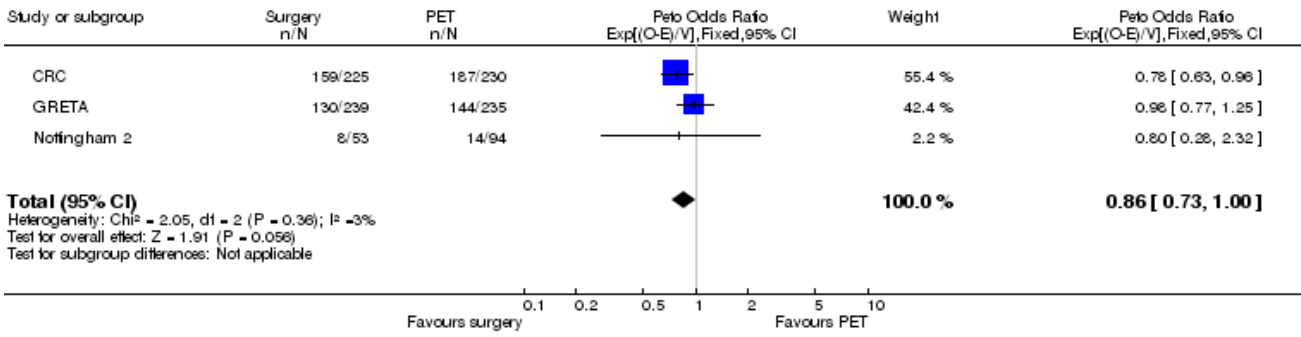


Figure 3: Comparison of surgery plus adjuvant endocrine therapy versus PET with overall survival (26)

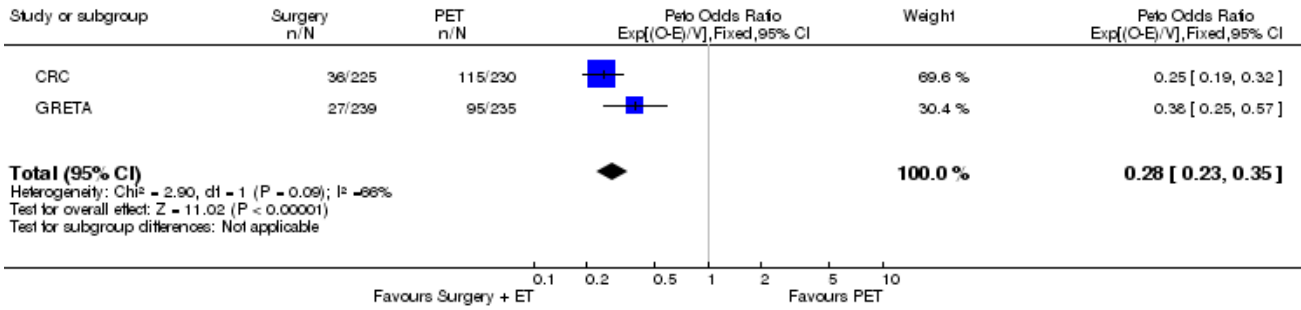


Figure 4: Comparison of surgery plus adjuvant endocrine therapy versus PET with local disease control (26) - permission to use obtained 26th May 2020

More recent work has confirmed these findings. A systematic review of the literature on the subject published in 2014 (27), concluded that data suggests an advantage for surgery over PET in terms of disease control and a likely survival benefit in patients with a predicted life expectancy of ≥5 years.

These studies come with the caveat that patients undergoing primary surgery will also have had other adjuvant therapies, so this is not a direct like-for-like comparison with patients having PET alone. It is expected that patients on PET may recur earlier for this reason.

Tamoxifen was historically the first choice of PET, until aromatase inhibitors (AI) were available (mid 1990s) (28).

1.3.1.3 Alternative options for women with ER-negative breast cancer

There is limited, outdated literature on the use of radiotherapy as primary treatment for ER-negative breast cancer, due to the small number of women this option is available to (29, 30). However, given the size of the older population, there still remain a significant number of women, in absolute terms, who develop ER-negative tumours (31). Primary radiotherapy can present issues in terms of tolerability and side effects. Another option in this group may be chemotherapy; however it would be expected that if the patient could not tolerate surgery, then this would also be true of chemotherapy. Historically, some patients with ER-negative tumours received endocrine therapy regardless of receptor status (32); this is no longer recommended (33).

1.3.2 Adjuvant treatment

Adjuvant treatment can be considered in terms of local and systematic therapies. Current NICE guidelines (34) recommend the use of a prognostic tool (such as Predict (35)) and/or gene profiling (such as Oncotype Dx (36)), to quantify benefit from adjuvant therapy, in numerical terms. However, these instruments tend to be derived from biological data obtained from mainly younger women and have little or no consideration of geriatric issues of the patient. The use of predictive and prognostic tools will be discussed in Chapter 3.

1.3.2.1 Local treatment

Whole-breast radiotherapy is offered to women with invasive breast cancer who have had BCS (21). Partial breast radiotherapy may be considered for women at a low risk of recurrence (aged >50 years, tumours ≤3cm that are node negative (N0), ER-positive, human epidermal growth factor receptor (HER)2-negative and grade 1 or 2) also taking 5 years of adjuvant endocrine therapy. Conversely, a recent trial has found that a shorter duration of radiotherapy is not inferior to the standard regimen (37), thereby being a consideration to partial radiotherapy.

There has been a further move towards omission of radiotherapy in older women having BCS and at very low risk of recurrence (aged ≥65 years, tumours that are T1N0 (size ≤2cm, node negative), ER-positive, human epidermal growth factor receptor (HER)2-negative and grade 1 or 2) who are willing to take endocrine therapy for a minimum of five years (21, 38-42) resulting in a change of clinical guidance in the UK (43) and the US (44).

1.3.2.2 Systemic therapies

This includes endocrine therapy, chemotherapy and targeted therapies such as trastuzumab.

Endocrine therapy

Traditionally adjuvant endocrine therapy with an AI is given for 5 years in post-menopausal women (45), although there is still debate over extension beyond this (46). Specific to the postmenopausal population, the MA.17R randomised trial (47) found improvement in disease free survival (DFS) with 10 years of AI (5 year DFS 95% for patients receiving AI versus 91% for placebo). However, these findings were not reproduced in two other trials (48, 49). The Association of Breast Surgery (50) and

American Society of Clinical Oncology (51) suggest that extended endocrine therapy up to 10 years should be considered, taking into account patient preference, tolerability of therapy, relapse risk and comorbidity.

Chemotherapy

NICE recommends chemotherapy should be offered to patients with high risk of recurrence (33). The recommended regimen should contain both a taxane and an anthracycline.

The 2011 landmark Early Breast Cancer Trialists' Collaborative Group (EBCTCG) meta-analysis (52) included more than 100 chemotherapy trials of approximately 100,000 women of all ages. The study found that combination adjuvant chemotherapy overall reduced annual risk of recurrence by almost 25% and risk of death by around 14%, compared to single agent chemotherapy, but only up to 70 years of age (Figure 5). This is likely related to biological characteristics of breast cancer being more favourable to respond to chemotherapy in younger women, for example, younger women are more likely to have ER-negative tumours (53).

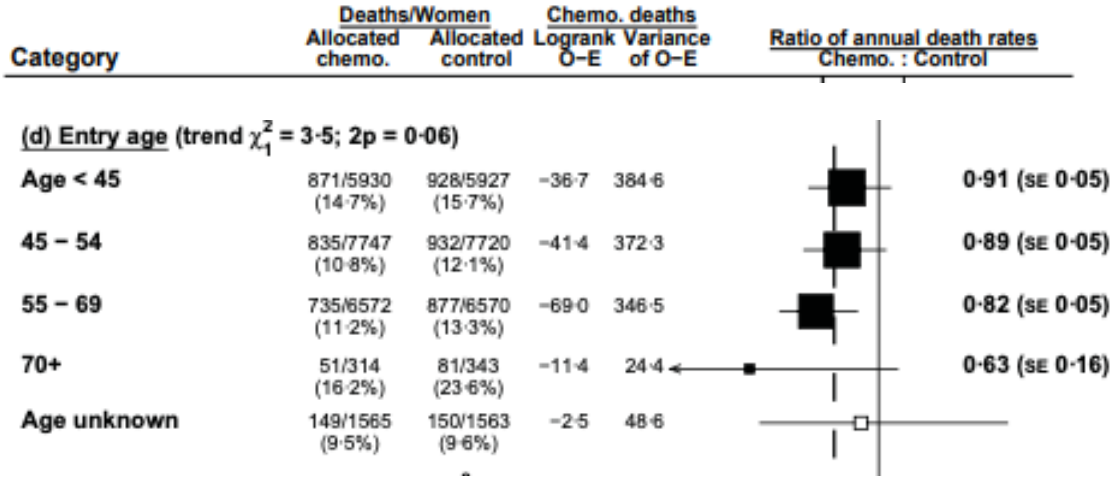


Figure 5: Subgroup analysis by age of breast cancer mortality, chemotherapy versus no chemotherapy from the Early Breast Cancer Trialists' Collaborative Group meta-analysis on adjuvant chemotherapy (52) - permission to use obtained 16th May 2021

Trastuzumab

The most utilised targeted therapy is the anti-HER2 agent trastuzumab (Herceptin). Adjuvant trastuzumab should be offered for patients with a HER2-positive invasive breast cancer ≥1cm in size, given at 3-week intervals for 1 year, in combination with surgery, chemotherapy and radiotherapy as appropriate. Trastuzumab should be used with caution in patients with a history of cardiac disease (21). Older women remain under-represented in clinical trials involving anti-HER2 therapies but there is evidence to suggest that trastuzumab is well tolerated in a select group of older women with appropriate cardiac monitoring (54, 55).

1.3.3 What are the challenges with current treatment of primary breast cancer in older women?

There is a lack of research focussing on older women with primary breast cancer. This is due to a variety of reasons including exclusion of older adults in clinical trials, which tend to recruit fitter, hence younger adults, lack of stratification that includes the older age group and minimal participation leading to under-representation (56). In consequence, older women are being treated in the same way as their younger counterparts, with an often-unproven assumption their disease has the same biological characteristics and behaviour and older patients have the same treatment goals compared to younger patients.

1.3.3.1 Primary treatment

Within the realm of primary treatment, the most significant challenge facing older women is in deciding who should have surgery or non-operative therapy.
The Bridging The Age Gap in breast cancer (BTAG) study included over 3400

women ≥70 years of age and published in 2021 (57), found that PET was associated with an inferior overall survival, however not breast cancer specific survival (BCSS) up to around 4 – 5 years of follow-up. This suggests that individualised decision making regarding PET versus surgery may be appropriate when predicted life expectancy is <5 years.

As shown in the 2020 Annual Report of the National Audit of Breast Cancer in Older Patients (NABCOP) (58), up to 25% of women aged ≥70 years do not receive primary surgery. The latest International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA) guidelines (20) state that older women (≥ 70 years) should be offered the same surgery as younger women and PET should only be offered to patients with a life expectancy of <5 years despite optimisation of medical conditions (20).

Breast cancer surgery carries with it a much lower morbidity compared to colorectal cancer surgery or other visceral surgeries, so there are few patients who cannot undergo breast surgery in general terms (59). Despite this, up to 25% of women ≥70 years of age still do not receive primary surgery (58).

In terms of type of surgery offered to older women, studies have shown that they are equally interested in discussing BCS in preference to mastectomy (60, 61) as well as reconstruction techniques (62), however the literature suggests that they are less likely to receive these treatments (61, 63, 64) compared to their younger counterparts. Reasons for this are likely to be multifactorial and related to patient preferences, consideration of quality of life (QoL) and physical fitness. One reason may be concerns over decline in health following surgery. However, there is limited evidence in the literature to support or deny the claim that functional status and independence decline after surgery (65). There is some evidence in the literature to suggest that

more extensive surgery, may cause not only a physical detriment to the patient, but also a decrease in QoL (57, 66-69). A recent systematic review of the literature by Nottingham (70) concluded that level of physical activity was negatively impacted by breast cancer surgery and this was compounded by extent of surgical procedure (more extensive axillary surgery, reconstructive procedures, mastectomy compared to breast conserving surgery). However, evidence for the impact of breast cancer surgery on QoL, fatigue and cognition, is conflicting.

Another area of debate in older women undergoing primary surgery, is management of the axilla. De-escalation of axillary surgery in recent years has arisen based on results from a number of key clinical trials (71) which has seen a move towards omission of axillary lymph node dissection (ALND) in selected patients. More recent trials have been launched to assess omission of SLNB.

The Society of Surgical Oncology's Choosing Wisely guidelines (72) first released in 2016, recommend against SLNB in patients with node-negative, hormone receptor-positive breast cancer aged ≥70 years, who will be treated with adjuvant hormonal therapy. This has been supported by retrospective (73) and prospective (74) studies which have shown no increased rated of locoregional recurrence or BCSS with omission of SLNB, however, other studies have been disputed this (75).

The Ontario Health and American Society of Clinical Oncology (ASCO) updated guideline published in 2021, suggests that SLNB may be omitted in women ≥70 years if they have low risk disease that is hormone receptor-positive, HER2-negative and they will be treated with adjuvant hormonal therapy. (76). Axillary staging should still be considered if results may impact radiation or systemic therapy decisions.

1.3.3.2 Adjuvant treatment

No adjuvant treatment is risk-free and risks versus benefits of all potential treatments should be discussed with an older patient as part of the shared decision making progress. Clearly, with use of multidrug therapy, risk of adverse events increases.

The biggest treatment dilemma in older women concerning adjuvant treatment is whether or not adjuvant chemotherapy should be given. Overall, chemotherapy benefit in older women remains controversial with the majority of high-quality randomised controlled trials to evaluate systemic chemotherapy, excluding or poorly recruiting older patients (77)**.** Older patients tolerate chemotherapy poorly compared to their younger counterparts because of progressive reduction of organ function and comorbidities related to age (78).

A systematic review on the topic conducted by the SIOG team, included prospective studies evaluating chemotherapy in older women adults with hormone receptor (HR), HER2-negative early breast cancer (79). They concluded that chemotherapy survival benefits are unclear and gene expression profiling much less used in this population.

Recently published results from the BTAG trial (80) demonstrate that chemotherapy was associated with reduced risk of metastatic recurrence in older women, but only in patients with ER-negative cancers. They noted that QOL impacts were significant but did recover. Furthermore, recently presented results from the ASTER phase III trial in which patients ≥70 years in ER-positive, HER2-negative, high genomic grade breast cancer were randomised to receive adjuvant endocrine therapy +/- chemotherapy, did not find a statistically significant survival benefit with the addition of chemotherapy (81).

A possible solution is to use single agent chemotherapy. Clinical trials of chemotherapy in older women are few. The Cancer and Leukaemia Group B 49907 trial recruited 633

women aged ≥65 years with early breast cancer and randomly assigned them to either standard adjuvant chemotherapy or capecitabine (single agent oral chemotherapy) (82). At 10-year follow-up recurrence-free survival (RFS) remains superior in the standard adjuvant chemotherapy group compared to capecitabine, however, there were many deaths from non-breast cancer causes, diluting the benefits of adjuvant treatment (82, 83).

Another solution in older patients who have HER2-positive tumours may be to use trastuzumab monotherapy and withhold chemotherapy. The National Surgical Adjuvant Study of Breast Cancer-07 (RESPECT) trial was a phase III randomised trial in 275 patients ≥70 years randomised to receive either trastuzumab or trastuzumab plus chemotherapy (84). Noninferiority for trastuzumab monotherapy was not seen and therefore, the authors concluded monotherapy can be considered a potential option in selected older patients (85). This could have significant implications for patients with HER2-positive tumours; as they would only have to face toxicity of a single treatment.

Although generally, the side effects from adjuvant endocrine and radiotherapy are understood to be less detrimental than those of chemotherapy, they are still extremely significant in the group who suffer them.

The majority of older women with early breast cancer will receive an AI as adjuvant therapy. Patient discontinuation of endocrine therapy within 5 years of surgery is reported in 38-50% of women (22, 86, 87). In cases of serious adverse events, management depends on several factors including risk of recurrence, duration of therapy and QoL. Options include stopping therapy or switching to another medication. If considering stopping adjuvant therapy, risk of recurrence must be discussed with the patient. Alternatively, the clinician and patient may decide that side

effects can be managed. For example, joint pain may be treated successfully with non-steroidal anti-inflammatories (88) with input from a rheumatologist, allowing continuation of an AI in high-risk patients.

Despite advances in radiation techniques, the toxicities of radiotherapy in older patients still occur and most commonly include damage to the skin, fatigue and cardiac toxicity (40). De-escalating the side effects of radiotherapy, such as by hypofractionated radiation, selective delivery of a boost dose to the lumpectomy cavity and the introduction of accelerated partial breast irradiation, including brachytherapy, are all potential options (89). In addition, older people may not want the additional burden of having to travel to the hospital daily, to receive treatment. The recent FAST-Forward (37) phase III trial randomly allocated 4096 patients with invasive breast cancer (pT1-3, pN0-1, M0) after BCS or mastectomy to either 40 gray (Gy) in 15 fractions (over three weeks), 27 Gy in five fractions (over one week) or 26 Gy in five fractions (over 1 week) to the whole breast or chest wall. Receipt of the 26 Gy in five fractions regime was non-inferior to the standard 40 Gy in 15 fractions for local tumour control regime A one-week schedule of radiotherapy after primary surgery has potential major benefits in the older population in terms of convenience, cost for patients and health care services.

In summary, treatment decision making regarding primary and adjuvant treatment options in older women with primary breast cancer is complex and hampered by the lack of inclusion of older adults in clinical trials. Hand-in-hand with treatment decision making are patient preferences and an older woman's expectations of their breast cancer treatment.

1.4 Outcome measures

The goals of breast cancer treatment and in fact, geriatric oncology as a whole, may be different in some older adults compared to their younger counterparts and this opinion in shared by the majority of surgical oncologists (90). For example, whilst the focus of treatment in younger women with breast cancer may be curative and to prolong life, older women may be more concerned with maintaining independence and QoL (91). Currently oncological success of breast cancer treatment is based on measures of survival (92). These measures essentially put a numerical value on a patient's estimated life expectancy. The number of chronic health conditions increases with age; 30% of adults aged <65 years report at least two chronic conditions compared to 65% of those aged ≥65 years (93) - older women are much more likely to die from other comorbidities than of breast cancer (94). Therefore, a measure of overall survival (OS) or local recurrence (LR) for example, may not be appropriate in this patient population. Breast cancer-specific-survival (BCSS) identifies deaths due directly to breast cancer over a given period of time and may be more useful in this population to consider whether a specific treatment will be worthwhile. It is also likely to be a good surrogate marker of the biology of the disease which is the focus of this thesis.

1.5 The potential role of biology in predicting cancer outcomes in primary breast cancer

Given the challenges highlighted above, it becomes important to see if treatment decisions could be guided by tumour biology, available at diagnosis i.e. based on core needle biopsy (CNB) sampling, which the majority of patients will have.

This is important in several contexts. Firstly, as discussed previously, in helping to decide which older women should have surgery or PET. Although surgery is still the recommended primary treatment irrespective of age, a significant proportion of older women will not have surgery and the proportion of older women with breast cancer is increasing; CNB may be the only tissue sample obtainable in these patients.
Secondly, in the context of neoadjuvant therapy. Currently CNB specimens are routinely used to guide neoadjuvant treatment in all breast cancer patients. Specifically in older women, alongside geriatric assessment, it may be possible to select a population who would benefit from neoadjuvant use of chemotherapy, in a similar way to younger patients, as well as neoadjuvant endocrine therapy, which has become a significant treatment strategy during the recent COVID-19 pandemic (95, 96).
Comprehensive CNB assessment is also of value in locally advanced and metastatic disease, which is beyond the scope of this present thesis, however, systemic therapies tend to be the initial treatment of choice.
Tumour biology can be assessed measuring 'biomarkers'; a characteristic that can be objectively evaluated as an indicator of a normal biological or pathogenic process or pharmacological response to a therapeutic intervention (97).

1.5.1 Understanding subtypes of breast cancer

Traditionally, there are four main molecular subtypes of breast cancer, each with a differing response to treatment pattern and prognosis (Figure 6), these are luminal A, luminal B, HER2-positive and triple negative (98, 99) breast cancer (TNBC). The molecular subtypes are commonly represented by their immunohistochemistry (IHC) surrogates. Luminal A breast cancer is ER-positive and/or progesterone receptor (PR)-positive, HER2-negative with low levels of the protein Ki67. Luminal B breast cancer

is ER-positive and/or PR-positive and either HER2-positive or negative with high levels of Ki67. HER2-positive breast cancer is ER and PR-negative and HER2-positive. TNBC is negative for ER, PR and HER2.

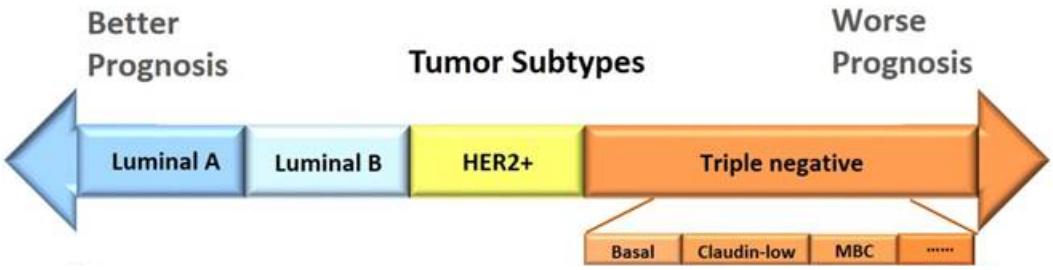


Figure 6: Current understanding of subtypes of breast cancer (100) - permission to use not required

It is now understood that breast cancer is a heterogeneous complex of diseases, with a spectrum of many subtypes with distinct biology (101-103) that leads to differences in response patterns to various treatment modalities. The ages of 40 and 70 years, appear to be significant milestones in marking a change in the pattern of breast cancer biology (8). Therefore, traditional treatment plans based on routinely measured biomarkers and our current understanding of disease subtypes (Figure 6) may no longer be adequate (3, 104).

1.5.2 The challenges of assessment of biology in primary breast cancer in older women

The correlation between ER positivity and age is well documented; ER positivity increases with age with the highest proportion of ER-positive cases seen in patients >65-70 years of age (105-108). This association is fundamental in the development of PET and adjuvant therapies in older women, however, we are now beginning to understand that even within a cohort of 'ER-positive' patients, patients have different outcomes, leading to the school of thought that there are 'subgroups' within this category (109).

Other favourable characteristics noted in older women with primary breast cancer are lower frequency of HER2 positivity, lower frequency of tumour protein 53 (p53) mutations and overexpression of B-cell lymphoma (BCL)2 protein (110). Generally, breast cancers in older women appear to be more slow-growing and display less aggressive biological features. These features may help to guide treatment decision making where a patient is considering surgery versus PET and when deliberating adjuvant therapies.

Although, ER, PR and HER2 are now routinely measured (and often Ki67 as well), numerous other 'biomarkers' such as those discussed above, may have a potential role in breast cancer and are currently only considered in a research setting. Many of these have been explored by the Nottingham breast pathology research group, primarily in younger women (111).

1.6 Background to the research programme in Nottingham

1.6.1 Overview of the research programme

The Nottingham team has a dedicated research programme focused on primary operable breast cancer in older women. The centre describes a large consecutive series of older women with primary breast cancer, for whom long-term follow-up data is available (3, 112). There is also a younger cohort of primary breast cancer patients, available for comparative analysis (111). There are currently two arms of the research programme in older women, one focussing on profiling the unique biology of primary breast cancer in this series and the second arm, focussing on the application of geriatric assessment (GA). This current thesis is contributing to the biology arm of the programme.

1.6.2 Work on the biology arm of the research programme to date

1.6.2.1 Clinical data

The cohort consists of 1758 women with early operable breast cancer aged ≥70 years. Long-term follow-up data (up to 37 years) and survival outcomes for the whole cohort has been collected and described (112) for all patients, where available.

1.6.2.2 Biological data

Surgical series

From the whole series, 813 patients underwent primary surgery. It has been possible to construct tissue microarrays (TMAs) from 575 cases using the surgical excision (SE) samples where a panel of 25 biomarkers has been assessed (3).

Core needle biopsy (CNB) series

It has been possible to construct TMAs from CNB samples from 693 cases in the overall cohort. A panel of 17 biomarkers has been measured in the CNB TMAs in patients who had ER-positive breast cancer (113).

1.6.2.3 Summary of significant findings to date

Cluster analysis in both the SE and CNB TMAs identified a novel biological cluster of disease, which is distinct from standard clusters seen in younger women (3, 113). The unique cluster, termed 'low ER luminal' had high expression of luminal cytokeratin (CK)s, mucin (MUC)1 and HER3 compared to the other conventional clusters and had divergent BCSS compared to conventional clusters.

Similar findings suggesting a differing biology (with differing clinical outcomes) according to age, have also been found in terms of histological type, as well as in HER2-positive (114) and TNBC (115) and in response to treatment (116).

Both the work in the SE and CNB TMAs have identified potential biomarkers which may be of use in predicting response to therapy (surgery, PET, adjuvant endocrine therapy) and survival in older women with primary breast cancer.

The following section will outline the next steps of the research programme.

1.6.3 Next steps

1.6.3.1 Full characterisation of the biology of breast cancer in older women

In addition to the biomarkers already stained in both the SE and CNB TMAs, it is important to identify any additional biomarkers that may be relevant to the study population. This may be due to identification of new biomarkers since construction of the initial TMAs or because of greater understanding of how existing biomarkers function in relation to breast cancer.

1.6.3.2 Association between biomarker expression and survival outcome

Expression of each biomarker will be correlated with survival outcomes including OS, BCSS and rates of recurrence to determine which biomarkers can independently add prognostic value.

1.6.3.3 Cluster analysis

When staining by IHC of the full panel of biomarkers has been complete, cluster analysis will be performed to investigate whether the provisional findings of a unique biological cluster in older women remain, or whether this only applies to a particular subgroup of patients.

1.6.3.4 Bioinformatics analysis and development of prognostic algorithm

Alongside conventional statistical analysis, bioinformatic principles and technologies will be applied to the dataset. This will enable detection of the most significant

combination of biomarkers, in relation to survival. This information will be used to form a prognostic algorithm. The idea is that when a new diagnosis of primary breast cancer is made in an older woman, their breast cancer biopsy specimen would be tested for the biomarkers that form the prognostic algorithm and predicted survival outcomes generated based on an individual patient's unique cancer biology.

1.6.3.5 Validation of prognostic algorithm

After creation of the prognostic algorithm as described, testing in unseen data sets for validation would be required. This would be using a similar model to that used by existing tools on the market, for example, Oncotype Dx (117).

Initial validation would be retrospective in existing populations of older women who have previously been recruited to randomised controlled trials (RCTs) to receive different treatment outcomes and for whom biological samples are available. To validate patients with ER-positive disease, trials which have randomised patients to receive different treatments (i.e. PET or surgery), for example, those contributing to the landmark Cochrane review on the subject (26) could be utilised.

1.6.4 Hypothesis of this thesis and how the work will contribute to the overall research programme

The hypothesis of this thesis is that the biology of primary breast cancer in older women can be utilised to generate a prognostic algorithm specific to this population for use at diagnosis.

This thesis will include a review of the current literature to identify gaps in the research relating to the biology of primary breast cancer in older women and secondly, to identify biomarkers relevant to this study. Following this, an outline of the laboratory methods will follow. A complete panel of biomarkers will then be measured using IHC. The next stage of this thesis will involve analysis of the

measured biomarkers using conventional statistics as well as cluster analysis and bioinformatics.

At the conclusion of this thesis, it is anticipated that a prognostic algorithm to estimate survival and/or recurrence in older women with primary breast cancer, will have been formulated.

1.7 Summary

Despite most primary breast cancer diagnoses being made in older women, robust and reliable evidence leading to treatment guidelines specific to this cohort are lacking. As such, older women may potentially receive suboptimal treatments leading to poor outcome.

The main challenges currently facing this cohort of patients include decision of primary treatment, which may be the decision between surgery and PET for patients with ER-positive tumours and in deciding who should receive adjuvant systematic therapy, primarily chemotherapy.

This chapter has outlined emerging work to suggest that breast cancer in older women displays different biological features compared to younger women. To quantify this and identify deficits in the research base of older women with primary breast cancer, the next step will be to conduct a systematic review of the literature to attempt to characterise the biology of primary operable invasive breast cancer in older women to date.

Chapter 2: A review of the literature on the biology of primary breast cancer in older women beyond routinely measured biomarkers

2.1 Introduction

Emerging work suggests that breast cancer in older women displays different biological features compared to younger women, as outlined in Chapter 1. To quantity this and identify deficits in the research base of older women with primary breast cancer, a systematic review of the literature was performed to characterise the biology of primary operable invasive breast cancer in older women to date.

This chapter will start by outlining the objectives of the systematic literature review before describing methods used. Following this, a critical appraisal of the relevant articles will be performed alongside assessment of the quality of evidence. The discussion will summarise the work that has been done to date to characterise primary operable breast cancer in older women and identify the clinical relevance of this. Finally, the chapter will suggest future directions for work in this field.

Part of the work described in this chapter has been published:

RM Parks, LH Alfarsi, AR Green, KL Cheung. Biology of primary breast cancer in older women beyond routine biomarkers. Breast Cancer, 2021 Sep, 28 (5): 991 – 1001. DOI: 10.1007/s12282-021-01266-5

2.2 Objectives

The biology of breast cancer is heterogeneous and there are numerous biomarkers which may have potential predictive and prognostic significance. This is extremely important in older women, who may opt for less aggressive therapy and need accurate predictors of life expectancy and response to treatment. This work systematically reviews the literature on biological assessment outside of standard biomarkers (defined as ER, PR, HER2 and Ki67) in women ≥65 years with primary operable

invasive breast cancer, to determine which additional biomarkers are relevant in this cohort.

This was achieved through the following objectives:

1) Systematic review of the literature to extract relevant information on measurement of biomarkers in primary operable invasive breast cancer in older women
2) Appraisal of the quality of evidence available
3) Identify biological markers outside of those routinely measured in clinical practice and highlight differences in older compared to younger women
4) Identify future research required to complete an overall profile of primary operable breast cancer in older women.

2.3 Materials and methods

A review protocol was developed based on the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement (118) and using the support of a Senior Research Librarian from the University of Nottingham.

2.3.1 Search strategy

The Medline and Embase databases were searched using the Ovid interface on the 1st July 2020. The biology of breast cancer has become more understood in the last decade therefore, the search was limited to this time frame. Studies were restricted to those published in English language with full-text available. Additional articles were identified by checking selected conference abstracts for updated full-text papers describing the same work.

Titles and abstracts were searched using the following terms: (breast cancer) AND (older OR elderly OR geriatric OR aged) AND (primary OR operable OR non-metastatic OR surgery) AND (biology OR biomarker(s) OR marker(s) or subtype). Duplicate publications were excluded.

Articles were screened by two independent researchers in two stages: screening of titles and abstracts followed by the retrieval and screening of full-text articles (second reviewer Dr Lutfi Alfarsi). Discrepancies were resolved by discussion.

Inclusion criteria were as follows:

- Female participants only
- Studies focused on women ≥65 years of age, or included women of all age groups with clear representation of older women
- Where studies included women of all age groups, must include stratification of results according to age, with a category of patients ≥65 years
- Includes primary operable invasive breast cancer
- Measurement of biomarker outside of standard assay (standard assay defined as ER, PR), HER2 and Ki67; luminal A, luminal B, basal or triple negative, HER2-enriched, or synonymous descriptions) on breast tissue sample.

Studies were excluded if:

- Biology or breast cancer not discussed
- Restricted access to study report/data/full-text
- Review article, letter to the editor, editorial report, case report

- Cell culture or animal study
- Conference abstract where the same data was later presented as a full-text article.

2.3.2 Data extraction

All data was extracted directly from the study text. When available, the following variables were extracted: year of publication; study published as full paper or conference abstract; country; definition of older women; number of women ≥65 years with primary invasive breast cancer included; study design; type of patient included; aim of study; whether or not there was comparison with younger patients; biomarkers measured; summary of how biomarker(s) were measured; findings specific to older women; implications for clinical practice; future research directions.

2.3.3 Critical appraisal

Methodological quality of full-text papers was assessed using the Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK) criteria (119) and in accordance with the PRISMA statement (120). Since the included studies were extremely varied in methodology and context, meta-analysis was not performed.

2.4 Results

2.4.1 Summary

Details of the literature search and study selection are shown in Figure 7. Twenty-three studies met the inclusion criteria and were included in this review. These studies included a total of 7508 women aged ≥65 years with primary operable invasive breast cancer.

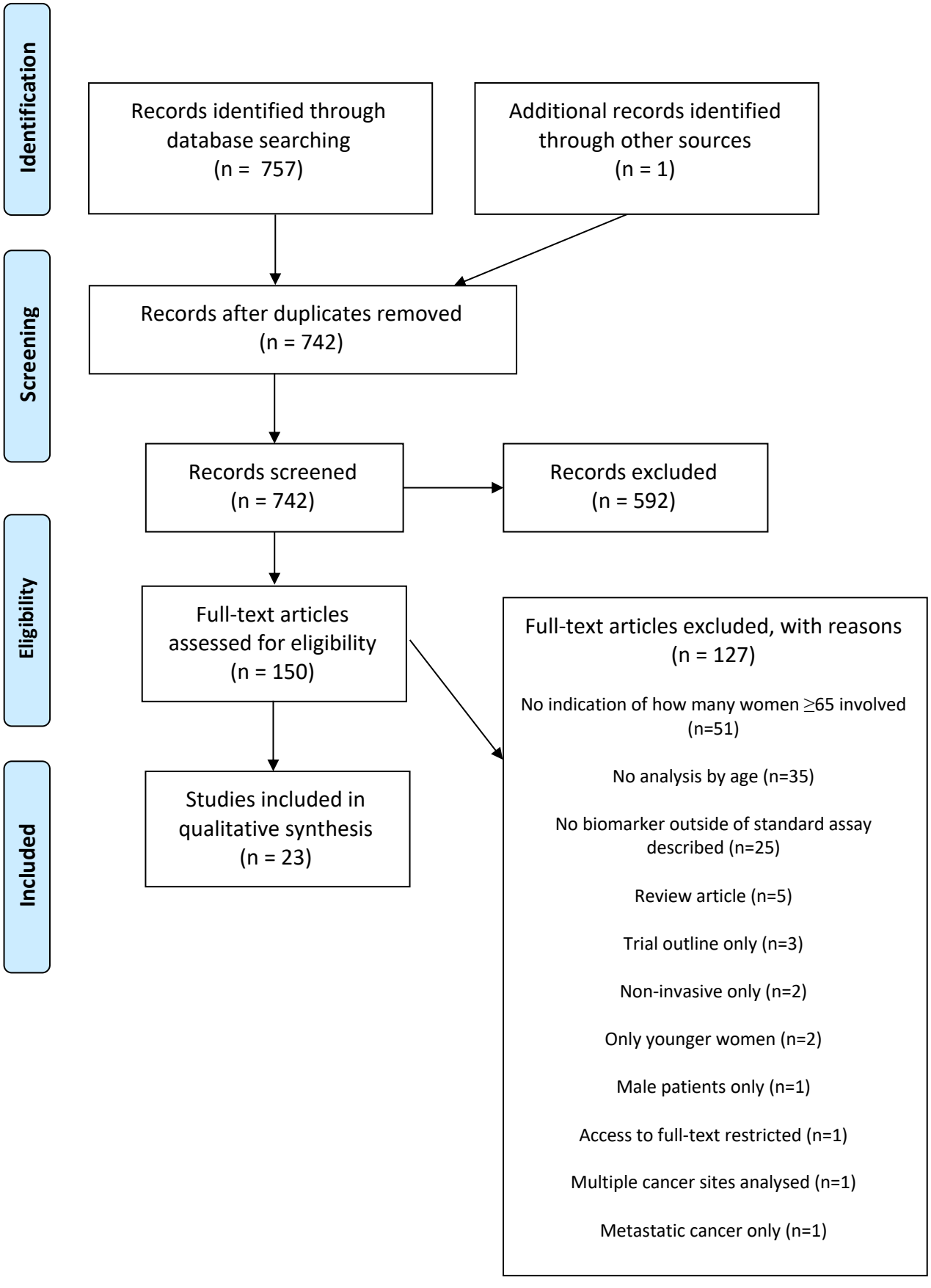


Figure 7: Selection of studies for review with age cut-off ≥65 years selection

2.4.2 Reporting standard

The REMARK checklist is given below for reference (Figure 8), taken from an explanatory article about the REMARK tool (121).

The results of REMARK assessment of full-text papers are given in Table 1.

All papers achieved at least 75% of the REMARK criteria for reporting of a study including a biomarker. The most frequently missed criteria were number 9 in relation to rationale of sample size and 16 and 17 in the reporting of multivariate analyses and confidence intervals.

INTRODUCTION	
1	State the marker examined, the study objectives, and any pre-specified hypotheses.
MATERIALS AND METHODS	
Patients	
2	Describe the characteristics (for example, disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.
3	Describe treatments received and how chosen (for example, randomized or rule-based).
Specimen characteristics	
4	Describe type of biological material used (including control samples) and methods of preservation and storage.
Assay methods	
5	Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.
Study design	
6	State the method of case selection, including whether prospective or retrospective and whether stratification or matching (for example, by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7	Precisely define all clinical endpoints examined.
8	List all candidate variables initially examined or considered for inclusion in models.
9	Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.
Statistical analysis methods	
10	Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11	Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.
RESULTS	
Data	
12	Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the number of patients and the number of events.
13	Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.
Analysis and presentation	
14	Show the relation of the marker to standard prognostic variables.
15	Present univariable analyses showing the relation between the marker and outcome, with the estimated effect (for example, hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.
16	For key multivariable analyses, report estimated effects (for example, hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
17	Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18	If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.
DISCUSSION	
19	Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.
20	Discuss implications for future research and clinical value.

Note: we have changed 'univariate' to 'univariable' in item 15 for consistency with 'multivariable'.
doi:10.1371/journal.pmed.1001216.t001

Figure 8: The Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) checklist (121)

#	Author	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
1	Lu G-W	Y	Y	Y	Y	X	Y	X	Y	X	Y	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	16
2	Johnston S	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	18
5	Di Benedetto A	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	X	Y	Y	Y	Y	Y	Y	Y	Y	Y	18
8	Syed BM	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	X	X	Y	Y	Y	17
11	Extermann M	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	X	Y	X	X	Y	Y	Y	16
12	Brouwers B	Y	Y	Y	Y	Y	X	Y	Y	X	Y	X	Y	Y	X	Y	X	Y	Y	Y	Y	15
13	Engels EC	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	19
15	Brouwers B	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	X	Y	Y	Y	Y	X	Y	Y	Y	Y	17
17	Garimella V	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	X	Y	Y	Y	Y	18
18	Syed BM	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	X	X	Y	X	Y	16
19	Syed BM	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	X	X	Y	Y	Y	17
20	Syed BM	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	X	X	Y	Y	Y	17
21	Mieog JSD	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	20
23	Lumachi F	Y	Y	Y	Y	Y	Y	Y	Y	X	Y	Y	Y	Y	Y	Y	X	X	Y	X	X	15

Y = yes, criteria met, X = did not meet criteria

Table 1: Results of Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK) assessment of full-text paper included

2.4.3 General characteristics of the studies

Characteristics of the included 23 studies are presented in Table 2.

The papers were published from 2014 to the present time. Sixteen of the studies were conducted in Europe (3, 114, 115, 122-134), four in the USA (135-138) and three in Asia (139-141).

The number of eligible patients included per study varied enormously from 11 (135) to 1698 (126). Twelve studies (3, 114, 115, 122, 123, 125, 127, 131-135) defined older women as ≥70 years and nine (126, 128-130, 136, 137, 139, 140) ≥65 years. One study each used 75 years (141) and 77 years (124) as a definition of 'older'. Only four studies considered patients who did not have surgery (124, 127, 133, 135).

The range of biomarkers covered was varied. The most frequently measured biomarker outside of standard assay (as defined in the methodology) was p53 measured in seven studies (3, 114, 115, 127, 133, 138, 141), followed by BCL2 measured in five studies (3, 114, 115, 127, 133) and PIK3CA (126, 138, 140), interleukin (IL)6 (122, 123, 136) and insulin growth factor (IGF)1 (122, 123, 136) measured in three studies each.

Biomarker expression was based on tissue samples in 15 studies; 11 on SE specimens, three in CNB specimens, one using both SE and CNB and the type of sample was unknown in one study. A further seven studies used peripheral blood samples (122, 123, 128, 129, 137, 139, 141). The method was not explicitly stated in one study, however, assumed to be tissue sample based on measurement of other markers in the study (141)

#	Author and date	FP or CA	Country	Age	N	Type of study (P = prospective; R = retrospective)	Treatment	Aim	YC	Biomarkers analysed	Method
1	Lu G-W June 2020 (139)	FP	China	≥65	152	P single-cohort	Any patient undergoing Mx	Investigate association between translocator protein and postoperative delirium	Y	Translocator protein	Blood, ELISA
2	Johnston S Mar 2020 (132)	FP	UK	≥70	517	R multi-cohort	Any surgical	Identify the role of c-cyclin E against a large panel of already measured biomarkers	N	C-cyclin E	SE TMA, IHC

3	Selenica P Feb 2020 (138)	CA	USA	≥65	510	R multi-cohort	Unknown	Describe the spectrum of somatic genetic alterations and mutational signatures in breast cancer according to age	Y	PIK3CA, TP53, CDH1, AKT1, MAP2K4, RUNX1, BRCA1, BRCA2, PALB2, RAD51C, RAD51D	Targeted and whole-exome sequencing
4	Parks RM Nov 2019 (134)	CA	UK	≥70	509	R single-cohort	Any surgical	Investigate the role of AR as a prognostic marker alongside a panel of measured biomarkers	N	AR	SE TMA, IHC
5	Di Benedetto A Apr 2019 (124)	FP	Italy	>77	27	P single-cohort	ER/PR +ve treated with NAET	Exploration of the prognostic role of DNA damage and repair biomarkers	Y	pATR, ATM, H2AX	CNB and SE, IHC

6	Okumura Y, Mar 2019 (141)	CA	Japan	≥75	342	R analysis	Any surgical	Examination of clinicopathological and prognostic factors for older patients	Y	p53	Unknown
7	Aft R, Feb 2019 (135)	CA	USA	>70	11	P single-arm phase II	Stage I/II, ER+ve, Ki67<30 or RS <18	Correlate response to NAET at 6 months with Oncotype DX RS	N	Oncotype Dx RS	CNB, Oncotype Dx
8	Ma D, Feb 2019 (140)	CA	China	≥65	69	Single-cohort observational	TNBC	Compare genomic and transcriptome features of young and older TNBC patients	Y	AR, PIK3CA, KMT2C, KMT2D, MES, TYPMS, HRD, BRCA1, BRCA2 FAN1, KDH4A	Gene set enrichment analysis
9	Syed BM, Jan 2019 (131)	FP	UK	≥70	407	R single-cohort analysis	Any surgical	Investigate the role of LKB1 in clinical outcome alongside a panel of measured biomarkers	N	LKB1	SE TMA, IHC

10	Cheung KL, May 2018 (133)	CA	UK	≥70	639	R single-cohort	ER+ve	Investigate tumour biology in older women with ER+ve breast cancer, based on CNB	N	p53, EGFR, HER3, HER4, BCL2, CK 5/6, CK 7/8, MUC1, VEGF, PTEN, AIB1, LBK1, BRCA1	CNB TMA, IHC
11	Exterman n M, Jan 2017 (136)	FP	USA	≥65	56	P single-cohort	ER+ve, 1-2 years post-surgery, on AI	To evaluate physical performance of breast cancer survivors with and without adjuvant chemotherapy	Y	IL-6, d-dimer, albumin, IGF1, IGFBP3, TNF-alpha, vitamin D	Blood, ELISA
12	Brouwers B, May 2016 (123)	FP	Belgium	≥70	109	P multi-cohort	Any surgical	Assessment of the effect of chemotherapy on biological and clinical ageing biomarker	N	Leucocyte telomere length, IL6, RANTES, MCP-1, IGF1, IL10	Blood, qPCR and ELISA
13	Engels EC, Mar 2016	FP	The Netherlan ds	≥65	1698	R population-based	Surgical patients in FOCUS	Evaluate the prognostic value of HER2	N	PIK3CA	SE TMA, IHC

	(126)							overexpression and PIK3CA mutations			
14	Kozlova L, Sep 2015 (128)	CA	Russia	>65	176	Unknown	T2-3, N0-1, M0	Analyse components of tissue fibrinolytic system in tissues of various types of breast cancer	Y	plasminogen, plasmin, plasminogen activators (uPA, tPA) and PAI-I	Blood, ELISA
15	Brouwers B, May 2015 (122)	FP	Belgium	≥70	162	R single-cohort	All patients with GA before Rx	Explore the relationship between biomarkers with age and clinical frailty	Y	leukocyte telomere length, IL6, RANTES, MCP-1, IGF1	Blood, ELISA
16	Dubot C, May 2015 (125)	CA	France and Belgium	>70	932	P multi-cohort phase III	ER+ve, HER2-ve	Compare the impact of CT + HT to HT alone on survival according to genomic grade	N	Genomic grade	SE, RT-PCR
17	Garimella V, Aug 2014	FP	UK	>70	45	R Single-cohort	ER/PR+ve treated	To identify possible predictive biomarkers of response to PET	N	p53, EGFR, BCL2	CNB, IHC

	(127)					case-series	with letrozole				
18	Syed BM, Jul 2014 (115)	FP	UK	≥70	127	R single-cohort	Surgical and TNBC	Analyse the biology and clinical outcome of TNBCs in older compared to younger women	Y	p53, CK5, CK5/6, CK7/8, CK14, CK17, CK18, CK19, BCL2, E-cadherin, MUC 1, EGFR	SE TMA, IHC
19	Syed BM, Apr 2014 (114)	FP	UK	≥70	45	R Single-cohort	Surgical and HER2+ve	Analyse the biology and clinical outcome of HER2+ve in older compared to younger women	Y	p53, BCL2, MUC1, E-cadherin, basal and luminal cytokeratins	SE TMA, IHC
20	Syed BM, Mar 2013 (3)	FP	UK	≥70	575	R single-cohort	Any surgical	Analyse tumour biology using TMAs and partitional clustering analysis	Y	p53, BCL2, MUC1, E-cadherin, basal and luminal cytokeratins	SE TMA, IHC

21	Mieog JSD, Jan 2012 (130)	FP	The Netherlands	>65	170	R Single-cohort	Any surgical	Compare the expression and prognostic effect of the breast cancer stem cell marker ALDH1	Y	ALDH1	SE TMA, IHC
22	Muss HB May 2011 (137)	CA	USA	≥65	73	P observational	Any surgical	Determine the effects of breast cancer therapy on a molecular marker of ageing	N	p16INK41	Blood, magnetic bead sorting
23	Lumachi F, Nov 2010 (129)	CA	Italy	>65	157	R analysis	PT1-2 having surgery	Analyse whether pre-operative levels of CEA and CA 15-3 correlate with ER, PR and MIB-1	Y	CEA, CA 15-3, MIB-1	Blood, ELISA

Abbreviations: AI - aromatase inhibitor; ALDH - aldehyde dehydrogenase; AR - androgen receptor; BCL- B-cell lymphoma ; BCSS - breast cancer specific survival; CA – conference abstract; CA(number) – cancer antigen; c-cyclin E – cytoplasmic cyclin E; CEA – carcinoembryonic antigen; CK – cytokeratin; CNB – core needle biopsy; EGFR- epidermal growth factor receptor; ER - oestrogen receptor; FP – full paper; GA - geriatric assessment; HER- human epidermal growth factor receptor; IGF- insulin growth factor; IHC – immunohistochemistry; IL – interleukin; LK – liver kinase; MAP2K- mitogen-activated protein kinase kinase; MCP- monocyte chemoattractant protein; MUC – mucin; NAET – neoadjuvant endocrine therapy; MX – mastectomy; NCoR- nuclear receptor corepressor; p53 - tumour protein p53; PET- primary endocrine therapy; RS – recurrence score; RUNX - runt-related transcription factor; TMA – tissue microarray; TNBC - triple negative breast cancer; TP – translocator protein; tPA – tissue plasminogen activator; uPA - urokinase

Table 2: Characteristics of the 23 studies included in the systematic review

2.4.4 Significant findings

Findings of the studies, implications for clinical practice and future directions are given in Table 3. P-values, correlation coefficients or relative risks are given where provided.

The following themes emerged: differences between older and younger women; relationship of biomarker with survival; association of biomarker with chemotherapy use.

#	Author	Findings specific to older women	Findings specific to survival	Implications for clinical practice	Suggested future directions
1	Lu G-W (139)	Postoperative serum translocator protein >0.385ng/mL increases risk of delirium (OR=26.5, CI 10.0 – 70.9, p<0.001)	Greater TP level associated with increased stage (r=0.442, p<0.001) and grade (r=0.280, p<0.001)	TP can be used to detect post-operative delirium in breast cancer patients undergoing mastectomy	Needs verifying in larger cohort
2	Johnston S (132)	59% of cohort had +ve C-cyclin E expression	C-cyclin E was the only biological factor independently predictive of BCSS (hazard ratio =6.23, CI 1.93-20.14, p=0.002)	Patient who are c-cyclin E negative are unlikely to die from breast cancer	Suggest verification of findings in prospective setting, irrespective of initial treatment
3	Selenica P (138)	Lower frequency of mutations in TP53, AKT1, MAP2K4 and high frequency in RUNX1 and NCOR1	-	As expected, breast cancers in older patients have higher incidence of ageing signature	None suggested

4	Parks RM (134)	No difference between AR expression and age	AR positivity associated with lower grade ($p<0.001$), lower mitotic frequency ($p<0.001$), ER positivity ($p<0.001$) and better BCSS ($p=0.009$)	AR has potential prognostic use in determining BCSS	Further work needed to investigate AR as therapeutic target
5	Di Benedetto A (124)	No difference between expression of pATR, ATM, H2AX and age	H2aX/pATM associated with ki-67 reduction ($p=0.011$), no association with pathological response	DNA damage and repair biomarkers may inform decision making regarding use of NAET	Multilevel analysis integrating genomic and transcriptomic profiling
6	Okumura Y (141)	p53 overexpression associated with recurrence on univariate analysis	On multivariate analysis only high grade and Ki67 significant for recurrence	Treatment strategies should be tailored for older women	None suggested

7	Aft R (135)	Nine patients tolerate NAET for at least 6 months	At 6 months, 71% had partial response, 25% had stable disease	Oncotype Dx may identify which patients can remain on endocrine therapy alone	Recruitment to the trial is ongoing, aim to recruit 50 women
8	Ma D (140)	46% of older women with TNBC were luminal AR subtype	Older women also had more frequent PIK3CA and KMT2C/KMT2D mutations	Should reconsider the benefit of neoadjuvant chemotherapy in this group	These molecular features may be potential therapeutic targets
9	Syed BM (131)	Positive LKB1 expression seen in 78.1% of patients	LKB1 is associated with better BCSS in patients receiving adjuvant endocrine therapy (93% versus 74%, p=0.03)	LKB1 may be used as a biomarker in predicting survival among patients receiving adjuvant endocrine therapy	Further studies needed to compare expression of LKB1 in younger patients
10	Cheung KL (133)	Within clusters identified, subgroups (luminal A and low ER luminal) had a high chance of benefit from PET	MUC1 identified as prognostic for BCSS, BCL2 and LKB1 were prognostic for TTP.	Identified subgroups had high chance of benefit from PET with comparable BCSS to surgery	Analysis of large panel of biomarkers in ER-ve cohort

11	Extermann M (136)	Vitamin D higher in group who did not have chemotherapy (42.4 +/- 21.9 nmol/L vs 55.8 +/-23.6 nmol/L, p=0.041)	Older women who received chemotherapy reported lower physical function but had similar physical performance to women who did not	Increasing Vitamin D levels by supplementation in patients receiving chemotherapy, should be considered	Better understanding of effect of inflammatory cytokines and physical function in older breast cancer survivors on AI
12	Brouwers B (123)	Chemotherapy does not impact on biomarker expression	Chemotherapy is unlikely to amplify clinical ageing or induce frailty at 1 year and does not increase mortality	Older women should not be denied chemotherapy based on age alone	None suggested
13	Engels EC (126)	Among patients with PIK3CA status available, 30% had a mutation; PIK3CA mutation not associated with HER2 status	PIK3CA mutations had no impact on overall survival (p=0.298)	Older women with HER2+ve cancer may benefit from additional targeted anti-HER2 treatment, regardless of age	Establish an effective anti-HER2 regimen with minimal toxicity for the older population

14	Kozlova L (128)	Luminal B tumours show maximum release of plasminogen and plasminogen activator uPA-act and minimum levels of tPA-act and plasmin.	In luminal B tumours, degradation of extracellular matrix components can be assisted by UPA-act	None suggested	None suggested
15	Brouwers B (122)	IL-6 associated with chronological age ($r=0/272$, $p<0.0001$) and measures of frailty on GA; telomore length, IGF1 ($r=-0.529$, $p<0.001$) and MCP1 ($r=0.412$, $p<0.0001$) only correlated with age	No association between biomarkers and standard clinicopathological features	IL-6 is a frailty biomarker that could partly replace geriatric assessment	Prospective assessment of these biomarkers and in predicting outcome from different treatments

16	Dubot C (125)	Of cases with genomic grade available, 40% low risk, 20% equivocal and 39% high risk	The proportion of high-risk patients is like the general breast cancer population	This multi-centre trial is feasible and may help to better selected older patients for treatments	Recruitment to the trial is ongoing
17	Garimella V (127)	BCL was expressed in 100% of patients, EGFR in 28% and p53 in 11%	Expression of p53 is associated with poor progression free survival (median survival 20 vs 54 months, p=0.03)	p53 may be used as a screening tool to select patients who would benefit from primary letrozole therapy	Larger studies evaluating the role of p53 as a marker of response to PET are needed
18	Syed BM (115)	Older patients had lower levels of Ki67 and CK 7/8 and higher levels of BCL2 and CK18 compared to younger (p<0.05)	Positive EGFR station associated with better BCSS (p=0.01)	Despite not having received adjuvant chemotherapy, the older series had clinical outcomes similar to the younger patients	Other biomarkers in addition to ER/PR and HER2 may influence prognosis and needs exploring
19	Syed BM (114)	Older patients had tumours with higher levels of BCL2 and lower	No significant different in BCSS of the two age groups	Despite not having received adjuvant chemotherapy, the older	Exploration of the role of trastuzumab as

		levels of Ki67, compared to younger ($p<0.05$)		series had clinical outcome similar to the younger patients	monotherapy and combined with endocrine therapy
20	Syed BM (3)	Older women had higher expression of BCL2, MUC 1, BRCA1, BRCA2 and lower expression of p53, EGFR, CK17	A novel cluster (lower ER luminal) was identified, unique to the older cohort with different survival outcomes	Measurement of these biomarkers may impact treatment decisions in this cohort	The findings of a novel cluster warrant further investigation in a larger series
21	Mieog JSD (130)	Less likely to express ALDH1 with increasing age ($p<0.01$)	ALDH1 expression not associated with any clinical outcome in older women	With increasing age, the role of tumour stem cells becomes more limited	Assessment of ALDH1A3 expression may correlate better with ALDH1 activity
22	Muss HB (137)	In the oldest patients, chemotherapy use was associated with increased expression of p16INK41	Expression of p16INK41was not associated with radiotherapy or endocrine therapy use	Chemotherapy use is associated with more than one decade increase in age as measured by a molecular marker	Accrual to the study continues and will include toxicity assessment

23	Lumachi F (129)	There was an inverse correlation between ER and MIB-1 expression ($p<0.05$)	Associations between CA15-3 in younger patients only	Only MIB-1 index should be considered an effective parameter for assessing tumour proliferation	None suggested

Abbreviations: AI - aromatase inhibitor; ALDH - aldehyde dehydrogenase; AR - androgen receptor; BCL- B-cell lymphoma ; BCSS - breast cancer specific survival; CA – cancer antigen; c-cyclin E – cytoplasmic cyclin E; EGFR- epidermal growth factor receptor; ER - oestrogen receptor; GA - geriatric assessment; HER- human epidermal growth factor receptor; IGF- insulin growth factor; IL – interleukin; LK – liver kinase; MAP2K- mitogen-activated protein kinase kinase; MCP- monocyte chemoattractant protein; MUC – mucin; NAET – neoadjuvant endocrine therapy; NCoR- nuclear receptor corepressor; p53 - tumour protein p53; OR – odds ratio, PET- primary endocrine therapy; r – correlation coefficient, RUNX - runt-related transcription factor; TNBC - triple negative breast cancer; TP – translocator protein; tPA – tissue plasminogen activator; uPA - urokinase

Table 3: Findings of the 23 studies including implications for clinical practice and future directions

2.4.4.1 Differences between older and younger women

There was a lower frequency of mutations in p53, AKT1, mitogen-activated protein kinase (MAP2K)4 and higher frequency of runt-related transcription factor (RUNX)1, nuclear receptor corepressor (NCOR)1 (138), PIK3CA, KMT2C and KMT2D mutations (140) in older women.

Older patients had tumours with lower levels of cytokeratin (CK) 7/8 (115), p53, epidermal growth factor receptor (EGFR) (3) and aldehyde dehydrogenase (ALDH)1 (130). They had tumours with higher levels of BCL2 (3, 114, 115), MUC1, breast cancer gene (BRCA)1, BRCA2 (3) and CK18 (115).

2.4.4.2 Relationship of biomarker with survival

The following markers were associated with improved BCSS in patients treated by surgery: negative expression of cytoplasmic cyclin E (p=0.002) (132), positive expression of androgen receptor (AR) (p=0.009)(134), positive EGFR expression (p=0.01) (115).

Liver kinase (LK) B1 was associated with better BCSS in patients receiving adjuvant endocrine therapy (p=0.03) (131). MUC1 was associated with BCSS in patients with ER-positive tumours, irrespective of treatment (133).

BCL2, LKB1 and p53 (127) were associated with time to disease progression (133) in ER-positive tumours. PIK3CA mutation did not impact on overall survival (p=0.298) (142).

2.4.4.3 Associations of biomarker with chemotherapy use

Chemotherapy use did not amplify biomarkers of ageing (123) but was associated with increased expression of p16INK41(137) and lower levels of Vitamin D (136).

2.5 Discussion

Most studies in this review add to the growing body of evidence that breast cancer in older women is less aggressive when compared to younger women and therefore, treatments offered should be tailored, instead of blanket treatment policies. From now on, studies in this review are referred to by their Study number (#) from Table 2, for the rest of the chapter.

2.5.1 Reporting standard

All papers achieved at least 75% of the REMARK criteria which is an acceptable rate. Many studies did not present information related to selection of sample size.

The main problem with research of this kind is recruitment of older women to these studies. Recruitment of older women to clinical studies is lacking due to bias by the clinician or patient, concerns regarding side effects of treatment and competing comorbidities. Therefore, a prescribed sample size may not be an appropriate goal in studies of older women, but simply to recruit as many participants as possible.

2.5.2 Differences between older and younger women

The most frequently reported biomarkers in this area were p53 and BCL2.

The work presented by study #3 (138) found lower rates of p53 mutations in tumours in older women. This is contrary to a large International Agency for Research on Cancer study which found that the likelihood of a p53 mutation increased with age (143). Studies linking p53 mutation to incidence of breast cancer are inconclusive (144), however, it appears that the numbers of breast cancers related to p53 mutation in older women is much lower. Around 25% of breast cancers in older women are

due to p53 mutations (143). Younger women with mutations of p53 have an 85% risk of breast cancer by the age of 60 years (145).

BCL2 was found in higher proportions in breast cancers of older women in studies #18 #19 and #20 (3, 114, 115, 146). BCL2 regulates cell death by either inhibiting or inducing apoptosis, therefore positivity is associated with improved outcomes. This agrees with findings from a large biological characterisation study in 2003 by Daidone et al (147) of over 14,000 primary breast cancers.

2.5.3 Relationship of biomarker with survival

The markers cytoplasmic cyclin E (132), AR (134) and EGFR (115) were associated with improved BCSS in three separate studies (study #2, #4 and #18).

Interestingly, in study #2 (132), cytoplasmic cyclin E was the only independent biomarker of BCSS alongside a large panel of other biomarkers including EGFR. There is a growing body of evidence to support its oncogenic role in various cancer types (148, 149). Johnston et al (132), showed that at ten years, 42% of older patients with cytoplasmic cyclin E-positive tumours had died of breast cancer versus 8% of negative cases ($p<0.0005$). They concluded that older women with cytoplasmic cyclin E-negative tumours are unlikely to ever die of breast cancer.

Although study #4 (134) did not find any association between AR and age, AR positivity was associated with improved BCSS. There is contradictory information in the literature regarding the role of AR in breast cancer prognosis. In the Breast International Group (BIG) trial 1-98 (150), the value of AR expression in 3021 postmenopausal ER-positive breast cancer patients was evaluated. AR expression was not associated with prognosis in any arms of the trial. Specifically, in TNBC, AR has been shown to be suggestive of outcome (151, 152) in some studies, but this has been

refuted in a large international study (153). A review on the subject by Venema et al (154), suggests that the prognostic value of AR remains uncertain due to the use of various antibodies and methods of assessment.

Specifically in older women with TNBC in study #18 (115), EGFR was more common in younger compared to older women. High expression of EGFR is known to correlate with poor clinical outcome in breast cancer (155). There are now EGFR inhibitors on the market developed for use in other cancer types including lung and pancreatic cancer (22). For breast cancer however, EGFR inhibitors have produced variable response rates in TNBC both as monotherapy and in combination with chemotherapy (156, 157).

LKB1, a tumour suppressor gene, did not show any significant influence on survival in the overall population undergoing surgery in study #9, however, in those receiving adjuvant endocrine therapy for ER-positive tumours, five year BCSS was better for those with positive expression (93% versus 74%, p=0.03). Other studies to date examining LKB1 are mainly cell line studies (158, 159). LKB1 could be a potential therapeutic target.

High expression of MUC1 was associated with worse BCSS in patients with ER-positive tumours, irrespective of treatment (133). This concurs with evidence in younger women, that high levels are associated with poor clinical outcome (160, 161).

PIK3CA is mutated in a wide variety of tumour types (162). Results from a recent phase III clinical trial (163) have shown that treatment with alpelisib (an oral PI3K inhibitor) improved PFS compared with placebo in patients with PIK3CA-mutated advanced breast cancer. A systematic review and meta-analysis on PIK3CA mutations and response to neoadjuvant treatment in early breast cancer, by Fan et al (164),

included 20 studies with 4392 patients. In the present review, study #13 (126) found that PIK3CA mutations had no significant prognostic effect in older breast cancer patients.

Study #7 (135) presented evidence for the use of Oncotype Dx, a genomic tool to identify patients who may benefit from primary endocrine therapy, rather than surgery. In their study cohort, patients with stage 1 or 2 ER-positive breast cancer, of grade 1-2, with Ki67<30% or recurrence score (RS) <18, were safely treated with endocrine therapy for six months, without developing progressive disease. However, a major flaw of this study was that it reported recruitment of only 11 patients, of whom only nine were able to tolerate endocrine therapy. Furthermore the duration of six months follow-up is relatively short, especially given the expected slow rate of progression of ER-positive breast cancer. Therefore, it is difficult to draw any firm conclusions from this study at present, but will certainly be of interest after accrual of further patients. Currently, Oncotype Dx is only routinely used to predict benefit of adjuvant chemotherapy. However, it has recently been extensively tested and validated in over 900000 CNB specimens (165). This confirms its potential use in the primary and neoadjuvant settings.

2.5.4 Associations of biomarker with chemotherapy use

In study #12 (123), older women after surgery were assigned to either chemotherapy or a control group. The biomarkers that represent ageing including IL-10, tumour necrosis factor-alpha, IGF1 and monocyte chemoattractant protein (MCP)1 suggested a minor biological effect in the group receiving chemotherapy, but this was not clinically relevant. Therefore, the authors suggested that older breast cancer patients should not be denied chemotherapy based on age. Contrary to the above study,

chemotherapy use was associated with increased expression of a molecular marker of ageing, p16INK41, in the study by Muss et al (137). There is little else in the literature to add to the evidence for this.

Study #11 (136) did not find any differences in markers of inflammation, oxidation or frailty amongst patients receiving chemotherapy. Lower levels of Vitamin D were reported in patients who were treated by chemotherapy, however there was little impact on measures of functional outcome at two years post-diagnosis. The authors suggested that supplementation of Vitamin D levels by exercise and/or oral supplements may be useful in these patients.

2.5.5 Limitations of this review

Comparability of studies is limited due to such a varied selection of studies with different methodology and aims. Studies which have included the largest number of patients, are retrospective in nature. One of the main hurdles of biological studies is availability of tissue samples and long-term follow-up data, which is often available in retrospective datasets. With updated ethical and legal guidance regarding handling of human tissue and consent, new prospective studies of this kind could face difficulty.

The majority of studies examining patient tissue used SE specimens. This excludes the important group of older women who have non-surgical treatment and creates potential bias towards ER-negative tumours (who are not suitable for PET). Therefore, results may not be necessarily applicable to all older women with breast cancer.

A number of the biomarkers detected were measured in blood samples (compared to tissue based samples). The exclusion of blood samples was not made as part of this review as it was felt important to identify all biomarkers relevant to breast cancer in older women, irrespective of source of measurement first and to later determine the

most appropriate method to measure these biomarkers. A future avenue to explore by the research team may be specifically biomarkers measured in blood samples. As the majority of biomarkers measured were done so in tissue samples, these will be focused on for the remainder of this thesis.

2.6 Conclusions and implications for this thesis

This chapter has presented initial work investigating the importance of biomarkers outside those currently measured in clinical practice, specific to the context of older women with primary operable breast cancer, however, the range of biomarkers studied and various methodologies used makes it difficult to form robust conclusions.

The most frequently reported biomarkers measured in this review include: p53, BCL2, cyclin E, AR, EGFR, LKB1 and MUC1 which have all been measured in tissue samples (compared to blood samples). These markers are relevant to the present thesis which will use the same methodology.

Rates of p53 appear to decrease in breast cancer with age, which is the inverse finding with BCL2. Expression of AR is associated with improved BCSS in older women, contrary to cytoplasmic cyclin-E, MUC1 and EGFR expression, which are associated with poor BCSS. LKB1 may have prognostic implications in patients receiving adjuvant therapy only. However, there is not enough evidence to suggest that other biomarkers identified in this review, are of no significance.

Further measurement of these biomarkers using standardised methodology in larger numbers of older women will enable firm conclusions to be drawn about their predictive and prognostic value. Biomarker assessment on a similar scale in CNB is urgently needed to address the same issues in those patients having non-operative

therapies as outlined in section 1.5 of this thesis. This present thesis will help to address this.

The identified biomarkers can be largely classified into the following categories based on their functions: hormonal receptors, such as ER and PR; growth factor receptors, including HER family and vascular endothelial growth factor (VEGF); tumour suppressor genes, such as p53, BRCA and mouse double minute homolog (MDM)s; cell proliferative markers, for example Ki67; anti-apoptosis markers, for example BCL2; cell adhesion molecules, including E-cadherin; cell differentiation markers, including CKs.

Following on from this review, the next step of this thesis was to examine the current literature to understand the role of individual biomarkers in primary operable breast cancer in older women as identified in this chapter and to assess whether further investigation with IHC within the remit of this thesis, is feasible.

Chapter 3: Clinical relevance of biomarkers in primary breast cancer in older women

3.1 Introduction

A number of relevant biomarkers have been identified in Chapter 2. The next step of this thesis was to understand the precise role of these biomarkers as well as others that have been identified from previous work (111, 166-171). This will generate the panel of biomarkers that will be measured by IHC in the following stages of this thesis.

A summary of all biomarkers discussed in this chapter is given in Table 4. For each biomarker, the frequency of its occurrence in breast cancer will be given if known. The current understanding of the relationship of the biomarker with age, as well as the predictive (how the cancer will respond to a certain therapy) and prognostic (what the biomarker can tell us about clinical outcome) use of that biomarker will then be detailed.

This chapter will group biomarkers according to their function, as explained in Chapter 2: hormonal receptors, growth factors receptors, tumour suppressor genes, proliferative markers, anti-apoptosis markers, cell adhesion molecules, cell differentiation markers and others.

In the latter part of the chapter, the evidence base to support the use of current tumour-based tools in older women with primary breast cancer will be presented. Finally, the concept of assessment specific to older women with primary breast cancer, will be introduced. The studies presented have used IHC to measure biomarkers on patient tissue, unless stated.

Part of the work in this chapter has been published as a book chapter:
RM Parks and KL Cheung. Chapter 3: Clinical Relevance of Biomarkers in Oestrogen Receptor Positive Breast Cancer. In: Estrogen Receptors: Structure,

Functions and Clinical Aspects. 2020; 1st Edition. Editor: GC Chen, MCF Tong, CA van Hasselt; Publisher: Nova Science Publishers, New York, USA.

Receptor	**Association with age**	**Association with efficacy of endocrine/ targeted therapy**	**Association with survival**
Hormonal receptors			
Oestrogen receptor (ER)	Increases with age	Increases	Positivity confers better outcome
Progesterone receptor (PR)	No association confirmed	Increases	Negativity linked to poor outcome
Androgen receptor (AR)	Conflicting evidence	Nil associated with use of endocrine therapy	Possibly linked with good prognosis in certain groups
Growth factor receptors			
Epidermal growth factor receptor (EGFR)	Positivity more likely in younger age	Associated with therapy resistance	Poor clinical outcome in terms of survival
Human epidermal growth factor receptor (HER2)	Decreases with age	Established use of trastuzumab	Positivity confers poor outcome

HER3	Possible increase with age	Linked to resistance	Conflicting evidence, possible poor outcomes
HER4	Possible increase with age	Expression may be linked to resistance	Expression associated with good prognosis
Vascular endothelial growth factor (VEGF)	Conflicting evidence	Anti-VEGF therapy established	Positivity associated with lower survival
Tumour suppressor genes			
P53	Chance of mutation increases with age	Linked to therapy resistance	High expression associated with lower survival
Breast cancer gene (BRCA)	Possible loss of expression with age	BRCA 1 associated with resistance	Conflicting results
Mouse double minute (MDM) homolog	Possible increase with age	Conflicting evidence	Possibly associated with poor prognosis
Phosphatase and tensin (PTEN) homolog	Unknown	Downregulation associated with poor response to tamoxifen	PTEN loss predicts worse survival

Liver kinase (LK)B1	Possible increase with age	Not enough evidence	Conflicting evidence
Cell proliferative markers			
Ki67	Lower expression in older women	High expression is an independent predictor of pathological complete response (pCR)	Positivity associated with worse survival
Anti-apoptosis markers			
B-cell lymphomas (BCL)2	Conflicting evidence	Better response to tamoxifen	High expression associated with better survival
Cell adhesion molecules			
E-cadherin	Higher expression in older women	Possibly downregulated in tamoxifen resistance	High expression associated with better survival
Cluster of differentiation (CD)44	Unknown	Expression associated to response to endocrine therapy in vitro	Conflicting evidence
Cell differentiation markers			
Cytokeratins (CKs)	Higher expression of CK5/6, CK5, CK18 in the older	Unknown – little evidence at present	Basal cytokeratins linked to poor prognosis

	population and lower expression of CK 7/8		
Others			
Mucin (MUC)1	Unknown	Associated with resistance	High levels associated with poor outcome
Amplified in breast cancer (AIB)1	Unknown	Associated with good response	High levels associated with poor outcome
Solute carriers	Unknown	Association with failure of endocrine therapy	High levels associated with poor outcome

Table 4: Summary of biomarkers associations with age, efficacy of endocrine therapy and survival.

3.2 Biomarkers – Association with age, predictive and prognostic values, future directions

3.2.1 Hormonal receptors

If present and activated in breast cancer cells, hormonal receptors will stimulate growth of that cell.

3.2.1.1 Oestrogen receptor (ER)

The ERs are a group of nuclear receptor proteins, of which ERα is the key member, which are activated by the hormone oestrogen. Around 70% of breast cancers overall are ER-positive (172).

Association with age

ER positivity increases with age with the highest proportion of ER-positive cases in patients >65 years of age (105) (106) (Table 5).

Age group (years)	**Oestrogen-receptor positive breast cancer (%)**
35-39	73.4
40-44	78.0
45-49	76.3
50-54	80.9
55-59	81.6
60-64	83.6
65-69	84.4
70-74	84.3
Overall	80.5

Table 5: Percentage of patients with oestrogen receptor-positive breast cancer, by age group in a cohort of 18586 breast cancer patients (1)

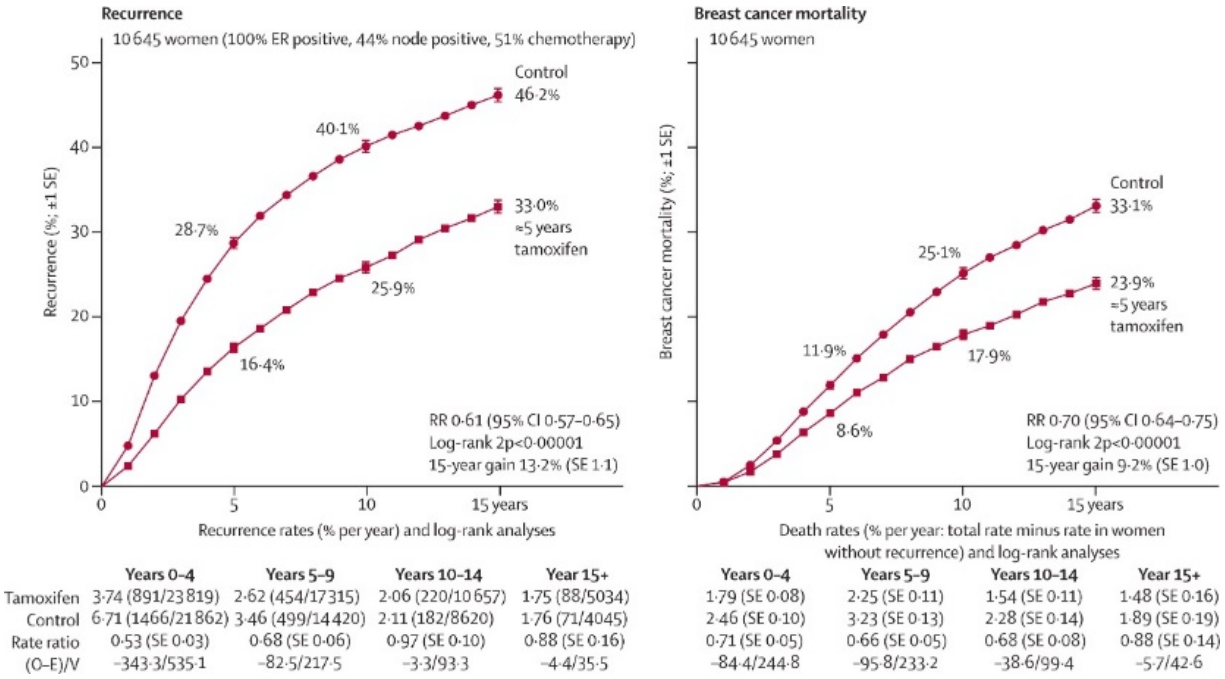
Predictive value

Endocrine agents targeting the ER are effective treatment for breast cancer. Tamoxifen and raloxifene are approved by the Food and Drug Administration (FDA) for chemoprevention in high risk women (173). In the International Breast Cancer Prevention Trial (174), 13388 women deemed to be at an increased risk of breast cancer received tamoxifen or placebo. Tamoxifen reduced the risk of invasive breast cancer by 49% overall, however, there was an increase in side effects (compared to placebo).

Neoadjuvant endocrine therapy (NAET) has been used in postmenopausal women with ER-positive breast cancer when chemotherapy is not suitable, either due to tolerability or patient choice, with the aim to improve operability whilst simultaneously treating presumed micrometastases. However, it is noted that the pCR of NAET is less than with chemotherapy (175), longer durations of treatment may be required (176) and there is greater disease progression when compared to chemotherapy (177).

Primary endocrine therapy has been shown to be equivalent in terms of survival outcome compared to surgery (26) and is frequently used in the older population with multiple comorbidity or who decline surgery (178).

Adjuvant endocrine therapy is given for at least 5 years after surgery. The EBCTCG meta-analysis published in 2011 (45) showed a 30% reduction in breast cancer mortality (absolute reduction in mortality of 9%) maintained for 15 years, following five years of adjuvant tamoxifen (Figure 9); tamoxifen is more effective when compared to control at reducing breast cancer mortality, in ER-positive breast cancer with a RR of 0.7 (CI 0.64 – 0.75).



	Years 0–4	Years 5–9	Years 10–14	Year 15+
Tamoxifen	3·74 (891/23 819)	2·62 (454/17 315)	2·06 (220/10 657)	1·75 (88/5034)
Control	6·71 (1466/21 862)	3·46 (499/14 420)	2·11 (182/8620)	1·76 (71/4045)
Rate ratio	0·53 (SE 0·03)	0·68 (SE 0·06)	0·97 (SE 0·10)	0·88 (SE 0·16)
(O–E)/V	−343·3/535·1	−82·5/217·5	−3·3/93·3	−4·4/35·5

Years 0–4	Years 5–9	Years 10–14	Year 15+
1·79 (SE 0·08)	2·25 (SE 0·11)	1·54 (SE 0·11)	1·48 (SE 0·16)
2·46 (SE 0·10)	3·23 (SE 0·13)	2·28 (SE 0·14)	1·89 (SE 0·19)
0·71 (SE 0·05)	0·66 (SE 0·05)	0·68 (SE 0·08)	0·88 (SE 0·14)
−84·4/244·8	−95·8/233·2	−38·6/99·4	−5·7/42·6

Figure 9: Effects of about 5 years of tamoxifen on the 15-year probabilities of recurrence and of breast cancer mortality, for ER-positive disease (45) - permission to use not required

Prognostic value

Above the age of 40 years, positive ER status is associated with improved BCSS and OS (179). Patients with ER-positive tumours generally have lower risk of mortality compared to those with ER-negative tumours (180), although this is not uniform across all groups. In BRCA1/2 carriers, there is evidence that ER-positivity confers a negative outcome (181).

Future directions

Endocrine therapies are well established as outlined above. Current research involves combinations with newer targeted agents, for example, cyclin-dependent kinase (CDK) 4/6 inhibitors. This combination is currently utilised in the setting of advanced disease (182) and now approved in the adjuvant setting, following the publication of

the monarchE trial in 2022 (183). Further clinical trials using CDK 4/6 inhibitors in neoadjuvant (184) and early breast cancer (185) settings are underway.

3.2.1.2 Progesterone receptor (PR)

PR is a nuclear receptor, activated by the hormone progesterone. Between 65-75% of breast cancers are PR-positive (186). Most PR-positive tumours are also ER-positive, so its significance is often downplayed. Progesterone receptor serves as an indicator of a functionally intact nuclear ER pathway because adequate levels of oestrogen and nuclear ER are required to transcribe PR; therefore, it may help predict which patients will respond to hormonal therapy (187).

Association with age

There is no confirmed relationship between expression of PR and age (3, 106, 188). Biological characterisation of 14007 primary breast cancer cases in a single series (147) has suggested that PR expression may increase with age, but this was not statistically significant.

Predictive value

In the adjuvant setting, PR expression is usually associated with greater benefit from endocrine therapy (147, 189). One study which found conflicting results was the Arimidex, Tamoxifen Alone, or in Combination (ATAC) trial (190). A total of 9366 postmenopausal patients with primary breast cancer had significantly improved disease-free survival (DFS) when treated with adjuvant anastrozole compared to tamoxifen. The time to recurrence (TTR) was longer for patients who were ER-positive and PR-positive or negative, however the OS benefit was greater in the PR-negative subgroup. The significance of this retrospective subgroup analysis remains controversial.

Prognostic value

A lack of PR positivity has been linked to poor survival regardless of age and independent of ER status (191) and increases the risk of relapse and death from breast cancer (192).

In a recent study by Van Asten et al (193) in 9647 patients with primary operable ER-positive, HER2-negative breast cancer, absent PR expression predicted worse outcome in terms of disease free recurrence interval (DFRI).

In a study performed by Albanghali et al (167) of 536 older women with ER-positive primary breast cancer, PR expression was associated with better BCSS.

Future directions

Both the ATAC trial (190) and the study by Van Asten et al (193) have shown results which require validating in larger prospective cohorts.

3.2.1.2 Androgen receptor (AR)

The AR is a nuclear hormone transcription factor, activated by the binding of androgen to its ligand binding domain (194). AR is expressed in 60-80% of breast cancers and is correlated with ER-positivity
(150).

Association with age

In a retrospective IHC analysis of 509 breast cancer samples in women of all ages, there was no significant difference in AR positivity in older (≥70 years) compared to younger women (134). Tokunaga et al (195) analysed AR expression in 250 breast cancers and found positive AR expression was associated with increased age in ER-positive cases only.

Predictive value

The predictive and prognostic value of AR in 3021 postmenopausal patients with ER-positive tumours recruited to the BIG 1-98 trial (150) was evaluated. The BIG 1-98 trial was a four-armed phase III trial comparing five years of tamoxifen or letrozole monotherapy or sequences of treatment with one drug and then the other. There was no heterogeneity of treatment effect across the continuum of AR expression.

Witzel et al (196) measured AR status in a total of 493 patients undergoing neoadjuvant chemotherapy (NACT) and found that overall, high AR mRNA levels were associated with lower pCR rates but better DFS and OS. However, Chintamani et al (197) found that higher AR mRNA expression levels before NACT were associated with a good response to therapy.

Prognostic value

In the BIG 1-98 study (150), AR expression was not associated with breast cancer-free interval (BCFI).

A multi-institutional international study by Bhattarai et al (153) evaluated AR in resected primary TNBCs from six international centres comprising a total of 1407 patients. AR-positivity was associated with good prognosis in patients from the US and Nigeria, poor prognosis in patients from Norway, Ireland and India and neutral in UK patients.

Future directions

AR has shown promise as a good prognostic marker, however, the exact significance of this is not yet confirmed. Consensus guidelines on how to measure AR, optimal cut-off points and differences in expression in different subtypes of breast cancer is required (198). AR targeted therapies have only been approved for treatment of prostate cancer, however, is an area of active investigation for breast cancer (199).

3.2.2 Growth factor receptors

Breast cancer cells require activated growth factor receptors to proliferate, invade and metastasise. Growth factor receptors transduce extracellular signals through activation of intracellular messengers. The human epidermal growth factor (HER) family is one of the most extensively studied and consists of four members: EGFR, HER2, HER3 and HER4. The receptors are structurally related and require dimerisation (Figure 10) activation of RAS-MAPK and PI3K-AKT-mTOR downstream signalling pathways (200, 201) (Figure 11).

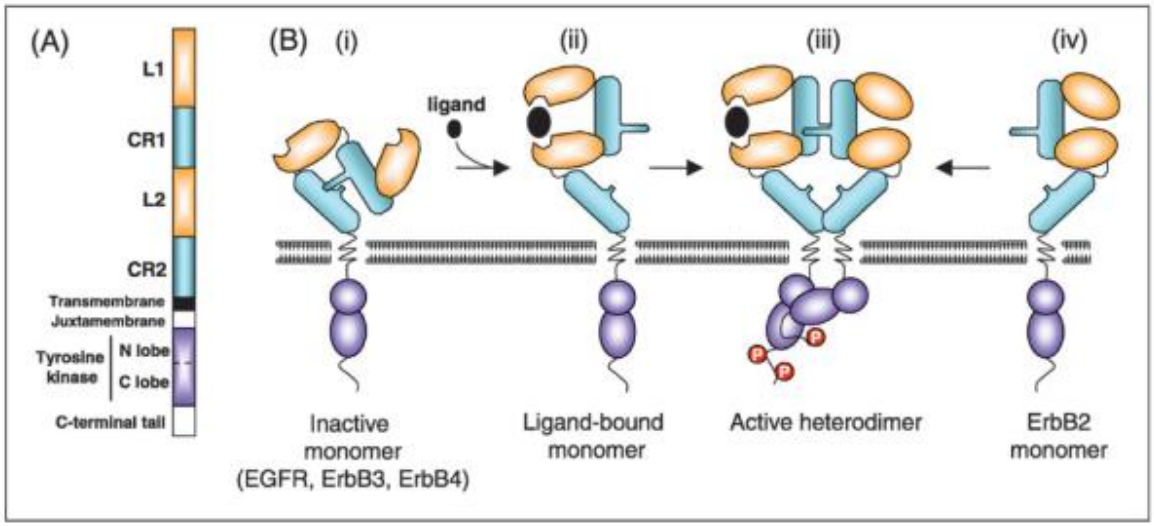


Figure 10: Structure of ErbB receptors (200) - permission to reuse not required

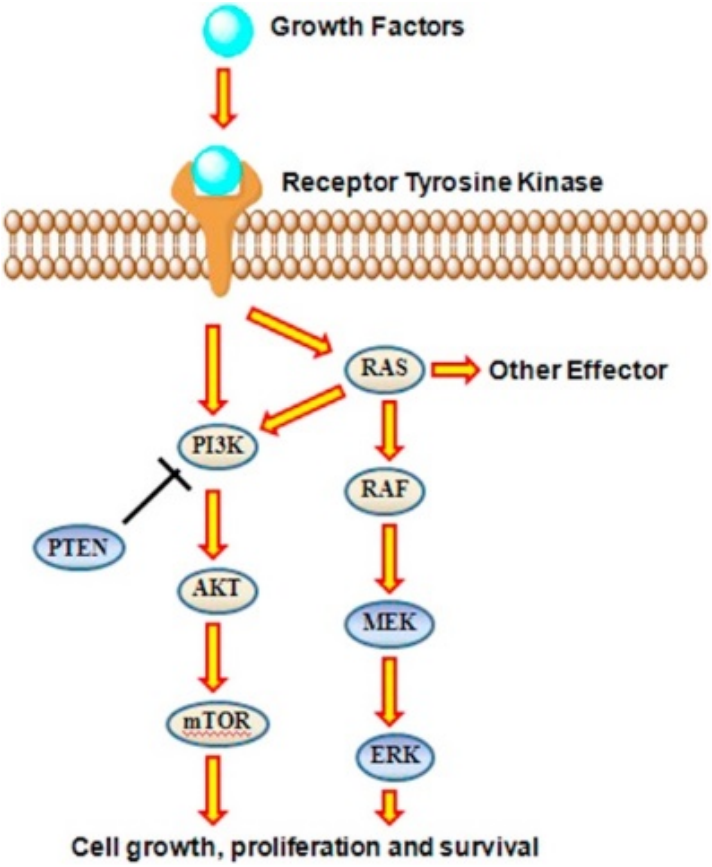


Figure 11: PI3K/Akt/mTOR and Ras/Raf/MEK/ERK signalling pathways (202) - permission to use obtained 2nd February 2020

3.2.2.1 Epidermal growth factor receptor (EGFR)

In an American database of 2567 breast cancer patients on whom EGFR expression was measured, 18% were EGFR positive (155).

Association with age

In the same study, EGFR expression was more common in breast tumours in younger (<50 years) compared to older women (155).

Predictive value

Expression of EGFR is associated with lower HR levels, higher tumour proliferation and HER2 overexpression (155); which are correlated with higher risk of relapse in

patients receiving adjuvant treatment. This has been observed in other studies where high levels of EGFR have contributed to acquired endocrine resistance with tamoxifen (203).

Prognostic value

High expression of EGFR correlates with poor clinical outcome in breast cancer (108, 204, 205). In a study by Tsutsui et al (204) expression of EGFR was analysed in 241 patients with recurrent breast cancer. Patients with positive EGFR expression had worse post-relapse survival than those with negative expression.

Future directions

There are two types of EGFR inhibitors currently on the market; these include tyrosine kinase inhibitors for lung and pancreatic cancer and monoclonal antibodies for colorectal and head and neck squamous cell carcinoma. For breast cancer however, clinical trials of EGFR inhibitors as monotherapy and in combination with chemotherapy, have not been as successful in universally producing good response rates. Only a handful of trials have noted a good response to EGFR inhibitors in a small number of patients (156, 206). A phase II trial performed by Johnston et al (206) tested lapatinib, an oral inhibitor of EGFR and HER2 in 45 patients. A clinical response was seen in 50% of HER2-positive patients. Going forward, it may be necessary to stratify patients to enhance the efficacy of EGFR inhibitors and consider combination therapies (157).

3.2.2.2 Human epidermal growth factor receptor 2 (HER2)

Approximately 15% of breast cancers have positive expression of HER2 (207, 208). There is an inverse relationship between HER2 and ER positivity, leading to approximately 10% of ER-positive tumours being HER2-positive and around 50% of HER2-positive tumours being ER-positive (209).

Association with age

There is lower expression of HER2 with increasing age (3, 207, 210); HER2-positive tumours are less common over the age of 70 years (210).

Predictive value

Clinical trials have shown that trastuzumab (Herceptin) significantly improves OS in women with HER2-positive breast cancer and this has been confirmed with long-term follow-up data from the HERA (HERceptin Adjuvant) trial (211); one year of Herceptin significantly reduced the risk of death (hazard ratio 0.74, CI 0.64 – 0.86). NICE guidelines recommend one year of adjuvant trastuzumab for HER2-positive invasive breast tumours ≥1cm, in combination with chemotherapy (21).

Retrospective studies have suggested that HER2 overexpression may have a predictive role for response to chemotherapy and endocrine therapy, but results are conflicting (212).

Other HER2 directed therapies approved for clinical use include the monoclonal antibody, pertuzumab; the small molecule kinase inhibitors lapatinib (213) and neratinib (214); the toxin-carrying antibodies, trastuzumab emtansine (215) and trastuzumab deruxtecan (216).

Prognostic value

HER2 overexpression, as a consequence of amplifications of its gene, confers an aggressive phenotype and poor patient outcome (217). Overexpression is associated with increased tumour aggressiveness, increased rates of recurrence and increased mortality in patients with node-positive disease (212).

Future directions

Pharmacological resistance to trastuzumab and other HER2 directed therapies is recognised. It is hypothesised that HER2-positive breast cancer can be targeted by

immunotherapeutic interventions and compounds are largely currently in the clinical trial setting (201). One example is immune checkpoint inhibitors, which include cytotoxic T-lymphocyte associated protein-4 inhibitors, anti-programmed cell death protein (PD)-1 and anti PD-L1 antibodies which are currently in phase I and II trials for early and advanced HER2-positive breast cancer and have recently been approved for locally advanced or metastatic TNBC (218). HER2 specific peptide-based vaccines have been developed to treat cancers by enhancing the antitumour immune response. One example is Nelipepimut-S (E75) which is currently under investigation in phase II and III clinical trials (201).

3.2.2.3 Human epidermal growth factor receptor 3 (HER3)

Co-expression of HER2 and HER3 is common in breast cancer (219). The rate of overexpression of HER3 in breast cancer is widely varied according to the literature and ranges from 30-75% (220), due to the capricious nature of antibodies used to measure this biomarker.

Association with age

High expression has been noted in older women (≥70 years) with primary breast cancer, however, no direct comparison to younger patients has been made (3).

Predictive value

HER3 has been linked to resistance to tamoxifen, due to the ability of a breast cancer cell to bypass a response to normal endocrine therapies; this is more likely with co-expression of HER2 and HER3 (221). Although there are currently no specific therapeutic targets to HER3, several studies indicate that activation of HER3 is a major cause of failure of anti-HER2 or anti-ER based therapies. Overexpression of HER3 has been shown to predict resistance to trastuzumab and lapatinib (222).

Prognostic value

The prognostic value of HER3 is unknown. Studies have shown conflicting evidence and suggest that overexpression of HER3 may be associated with worse survival or of no prognostic use (223). In a study of 177 primary breast cancers, low HER3 expression predicted recurrence in HER2 amplified breast cancer (223) and more aggressive clinicopathological features.

Future directions

There are a number of HER3-targeted biologic therapies in preclinical phases of study in breast, colorectal, lung and ovarian cancer. One example is Seribantumab (MM-121) (224), a monoclonal antibody that inhibits HER3 activation, which is in phase II testing for preoperative TNBC and HR-positive, HER2-negative breast cancer in combination with chemotherapy. However, concerns have been raised regarding intolerable side effects of HER3-targeted therapies, resulting in discontinuation of a previous study (221, 224).

3.2.2.4. Human epidermal growth factor receptor 4 (HER4)

HER4 mediates both anti-proliferative and pro-apoptotic activity; its function is not yet fully understood (225, 226) therefore, its overall expression in breast cancer remains unknown.

Association with age

As with HER3, high expression has been noted in older women (≥70 years) with primary breast cancer, but no direct comparison to younger patients has been made (3).

Predictive value

Recent studies have shown that expression of HER4 improves OS of breast cancer

patients with ER-positive tumours, raising the possibility that HER4 influences response to endocrine therapy (227). This has been shown in mouse models where HER4 suppression has been associated with tamoxifen resistance (228).

Prognostic value

Expression of HER4 is associated with better BCSS in early breast cancer and correlates with ER-positivity and lower tumour grade (229-231).

Future directions

There are few clinical trials to investigate anti-HER4 therapies, compared to anti-HER3 therapies. Anti-HER4 antibodies, such as clone P6-1, have been shown to cause growth inhibition of breast cancer cells in vitro (222).

Figure 12 summarises therapeutic targets against the HER family, currently under investigation.

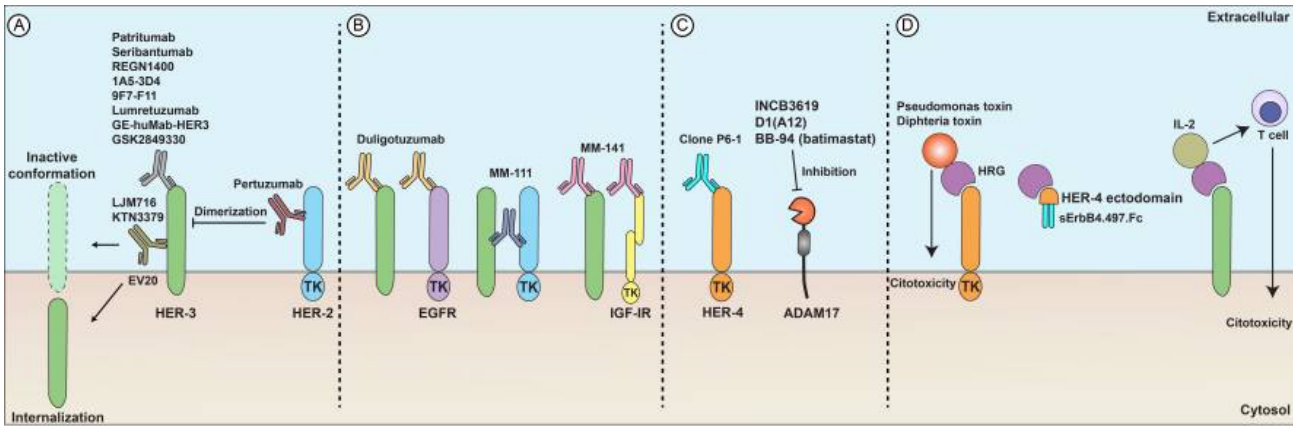


Figure 12: A summary of therapeutic targets including HER3, HER2, HER4 and EGFR (222) - permission to use not required

3.2.2.5 Vascular endothelial growth factor (VEGF)

VEGF is a signalling protein expressed on vascular endothelial cells as well as carcinoma cells. Isoform VEGF-A has been shown to have an essential role in angiogenesis and tumour blood supply which is a major element in tumour growth and

metastasis (232, 233) (Figure 13) and is overexpressed in approximately 72-98% of breast cancers (234).

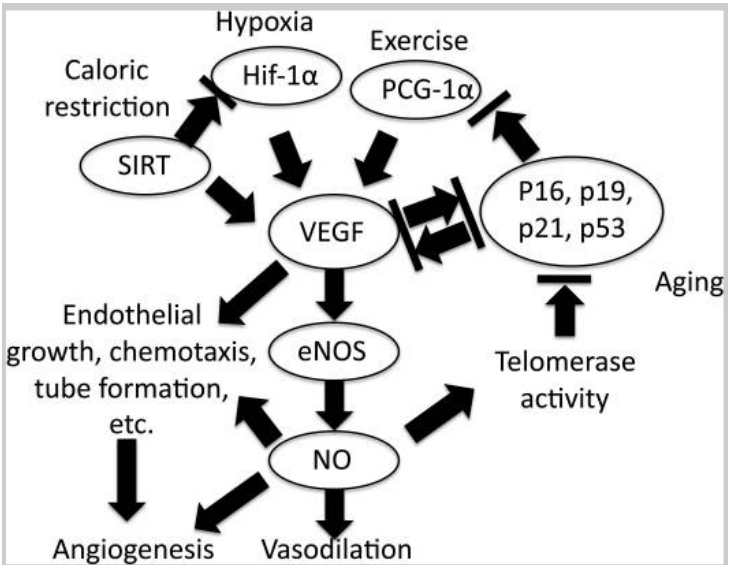


Figure 13: Overview of angiogenic pathways (235) - permission to use not required

Association with age

Advancing age is associated with a decline in VEGF in normal tissue (235). Conversely, in a study of 575 surgical specimens from older women with primary breast cancer by Syed et al (3), high levels were found.

Predictive value

High VEGF expression has been linked to failure of tamoxifen therapy (236). Foekens et al (236) measured levels of VEGF in primary breast tumours from 845 patients who developed recurrence during follow-up. Patients with higher or intermediate levels of VEGF showed a poor rate of response to first-line tamoxifen therapy compared to those with low VEGF levels.

The benefits of anti-VEGF therapy, for example, bevacizumab (a monoclonal antibody) when combined with standard chemotherapy for first-line treatment of metastatic HER2-negative breast cancer have been confirmed in the phase III

RIBBON-1 trial (Regimens in Bevacizumab for Breast Oncology) (237). The combination of bevacizumab with chemotherapy improved clinical benefit in terms of increased PFS. Bevacizumab is approved for use in the US.

Prognostic value

In the same study by Foekens et al (236), higher levels of VEGF were associated with shorter PFS and OS. This has been confirmed in other breast cancer studies (234, 238).

Future directions

Following RIBBON-1 (237), further clinical trials of bevacizumab are ongoing. The current focus is on identifying subgroups of patients, based on expression of other biomarkers, who may benefit from this therapy (239).

3.2.3 Tumour suppressor genes

These are a group of genes prone to genetic change, which result in tumourigenesis, usually by inhibiting cell proliferation and tumour development. In many tumours, these genes are lost or inactivated, which results in removal of negative regulation of cell growth; tumour cells abnormally grow. TP53 mutations are the most commonly acquired mutations in cancer and are estimated to occur in 50% of all cancers (240), however, a comprehensive meta-analysis revealed that only approximately 20% of all breast cancer cases express mutant p53 (241).

3.2.3.1 Tumour suppressor protein 53 (p53)

The TP53 gene provides instructions to produce tumour protein p53. The normal function of this protein is to regulate cell division. A mutation of the TP53 gene results in the ability of damaged cells to induce apoptosis (242), resulting in cell death.

Association with age

The likelihood of a TP53 mutation increases with age. A study based on the

GLOBOCAN and International Agency for Research on Cancer databases showed that p53 mutations accounted for approximately 25% of the ageing-related rise in breast cancer worldwide (143).

Predictive value

The impact of p53 status on antihormone treatments has been investigated in two retrospective series, which provided evidence that p53 mutation may affect tamoxifen response (243).

Yamashita et al (244) found that p53 protein accumulation was associated with greater likelihood of resistance to endocrine therapy (244) in metastatic breast cancer, as well as reduced post-relapse survival, compared to cases where p53 expression was low.

Prognostic value

Low expression of mutant p53 has been linked to better DFS and OS in breast cancer (245). These findings have been found to be similar across all age groups (3, 107).

Future directions

Retrospective studies have investigated a potential prognostic and therapeutic predictive role for mutant p53 in breast cancer, however with conflicting results (246). Several compounds are being investigated which can reactivate the mutant p53 protein and convert it to a conforming type. Some of these compounds (e.g. PRIMA-1) exhibit anticancer activity in preclinical models of breast cancer (246).

3.2.3.2 Breast cancer gene (BRCA) 1 and 2

BRCA1 and 2 are genes that produce tumour suppressor proteins. Specific inherited mutations in BRCA1 and BRCA2 increase the risk of breast, ovarian and pancreatic cancers. BRCA mutations are the most common genetically inherited mutations which increase the risk of breast cancer; around 5% of breast cancers are caused by a BRCA mutation (172).

Associations with age

Estimates from pooled data of 8139 patients, of whom 500 had a germline mutation in BRCA1 or BRCA2, give an average cumulative risk of BRCA1 mutation carriers by age 70 years of 65% and 45% for BRCA2. The risk of developing breast cancer with a BRCA mutation increases with age up to 70 years and then plateaus (247). There is a loss of proteomic expression of BRCA1 and BRCA2 in older women (3) compared to their younger counterparts.

Predictive value

There is some laboratory and clinical data to suggest that BRCA1 associated breast cancer may be resistant to tamoxifen (248). A retrospective study, comparing outcomes in early sage BRCA mutation and sporadic breast cancer treated with endocrine therapy, notes a lower OS in the BRCA carrier group (249).

Prognostic value

In an analysis of 1940 breast cancer cases (168), altered BRCA1 protein was associated with shorter disease free interval (DFI), as well as development of recurrence. Van den Broek et al (250) examined the evidence for breast cancer prognosis of BRCA1/BRCA2 mutation carriers. They concluded that current evidence did not support reduced breast cancer survival in mutation carriers. This is contradicted in a recent retrospective cohort study by Vocka et al (181) who analysed the effect of clinicopathological features on prognosis in BRCA1/2 carriers. They found that mutation carriers with ER-positive disease were high-risk patients in terms of risk of breast cancer death and recurrence.

Future directions

Poly (ADP-ribose) polymerases (PARPs) are important enzymes in DNA damage repair mechanisms, which tumours with BRCA1/2 mutations rely on (Figure 14). The

PARP inhibitor, olaparib, has been approved for use in HER2-negative metastatic breast cancer with an inherited BRCA1/2 mutation. The phase III OlympiAD trial has shown a PFS benefit of olaparib compared to single-agent chemotherapy in this setting, but no OS benefit (251). Further findings of this trial have shown the benefit of olaparib as adjuvant therapy in early breast cancer, but this is not yet NICE approved in the UK (252). There are currently over 30 actively recruiting trials involving olaparib to determine the exact benefit as well as comparison with other immunotherapy agents.

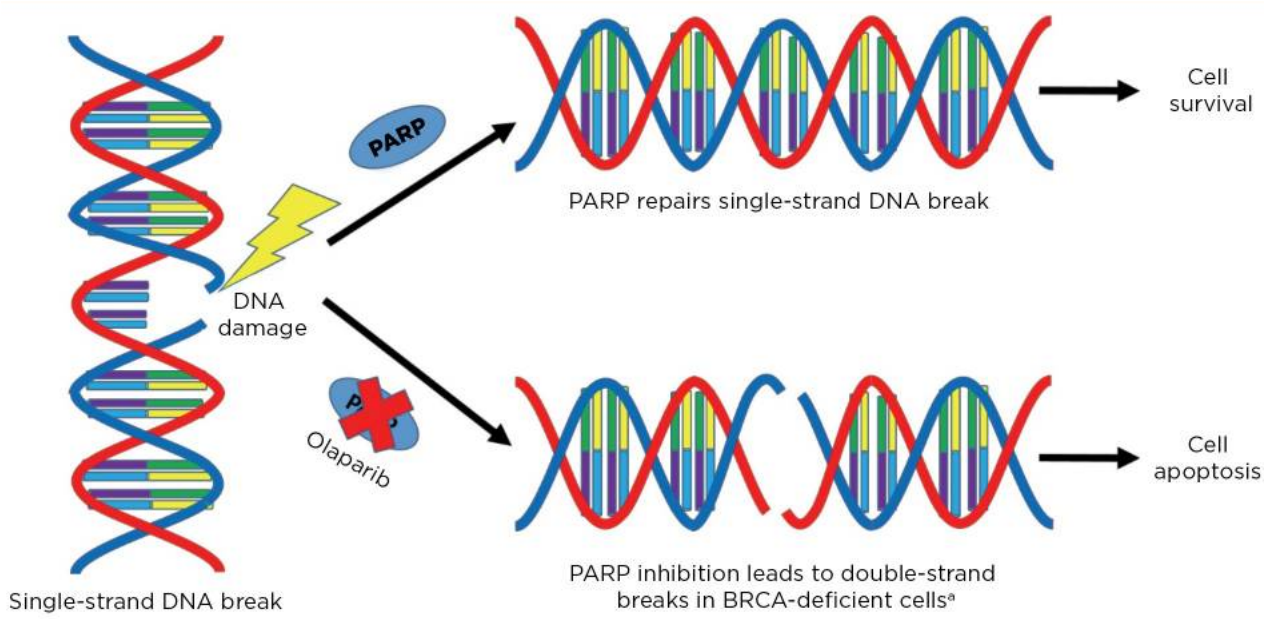


Figure 14: Mechanism of action of olaparib (251) - permission to use not required

3.2.3.3 Murine double minutes (MDM) 2 & 4

Murine double minutes genes 2 and 4, inhibit the function of p53 and therefore, maintain cell cycle regulation. Downregulation of these genes indicates mutant p53 (253).

Association with age

Most studies examining MDM in breast cancer are limited to younger populations. In

the previously referenced study by Syed et al (3) of 575 primary breast cancer patients aged ≥70 years, expression of MDM2 and MDM4 was higher in this population compared to a younger cohort.

Predictive value

Genotyping of 823 breast cancer cases by Kulkarni et al (253) suggested that a polymorphism in the MDM4 gene was associated with ER-negative breast cancer only and had no role in ER-positive breast cancer. This is contradictory to the work by Swetzig et al (254) who found that expression of both MDM2 and MDM 4 was associated with ER-positive breast cancers. They went on to show by in-vitro models that ER regulates MDM4 expression and therefore, these effects could be blocked by relevant endocrine therapies (254).

Prognostic value

A review by Haupt et al (255) of the literature concerning MDM2 and MDM4 expression in breast cancer, suggests that high expression is associated with poor prognosis.

Future directions

Although the role of MDM2 and MDM4 in breast cancer has been studied in the literature, there is little evidence on how this relates to response to therapy or prognosis.

3.2.3.4 Phosphatase and tensin homolog (PTEN)

PTEN is a negative regulator of the PI3K/ALT signalling pathway (Figure 11) which is associated with tumourigenesis in multiple cancers including breast. Less than 5% of breast cancers occur because of a PTEN mutation, however, around 25% have significantly low levels of PTEN (256).

Association with age

Association of PTEN with age is not clearly understood. Dean et al (257) studied 101 breast cancers and found that PTEN loss was associated with younger age at the onset of breast cancer. Shoman et al (258) found no association between PTEN expression and age (range 28-93 years) in 100 patients.

Predictive value

PTEN expression is postulated to have an effect on acquired resistance to tamoxifen. Shoman et al (258) found an association between downregulation of PTEN expression in patients with ER-positive tumours and failure to respond to adjuvant tamoxifen.

Prognostic value

Reduced PTEN expression has been associated with more advanced disease (258) and PTEN gene deletion has direct association with HER2 amplification and poorer prognosis in HER2-positive breast cancer (259).

A study of 49 primary breast cancer patients treated with adjuvant tamoxifen by Tanic et al (260) found that the loss of PTEN expression was associated with shorter DFS, BCSS and OS, compared to high expression.

A meta-analysis by Li et al (261) based on 27 studies involving 10231 breast cancer patients showed that PTEN loss was associated with reduced DFS and OS, larger tumour size, lymph node metastasis and triple-negative phenotype.

Future directions

Each stage of the downstream signalling pathway following on from PTEN consists of potential therapeutic targets (Figure 11). Mammalian target of rapamycin (mTOR) is a serine/threonine protein kinase, downstream of AKT. Once activated it is responsible for limiting the proliferative signals transmitted by upstream effectors (262).

Everolimis is a rapamycin analogue that targets mTOR and was the first drug to be approved for the treatment of HR-positive, HER2-negative metastatic breast cancer, in post-menopausal women, who have progressed on first-line therapy with an aromatase inhibitor (AI) (263). The Breast Cancer Trials of OraL EveROlimus (BOLERO)2 trial (264) demonstrated that the addition of everolimus to exemestane (a steroidal AI) markedly prolonged PFS in HR positive, advanced breast cancer with disease recurrence following prior non-steroidal AIs (Figure 15).

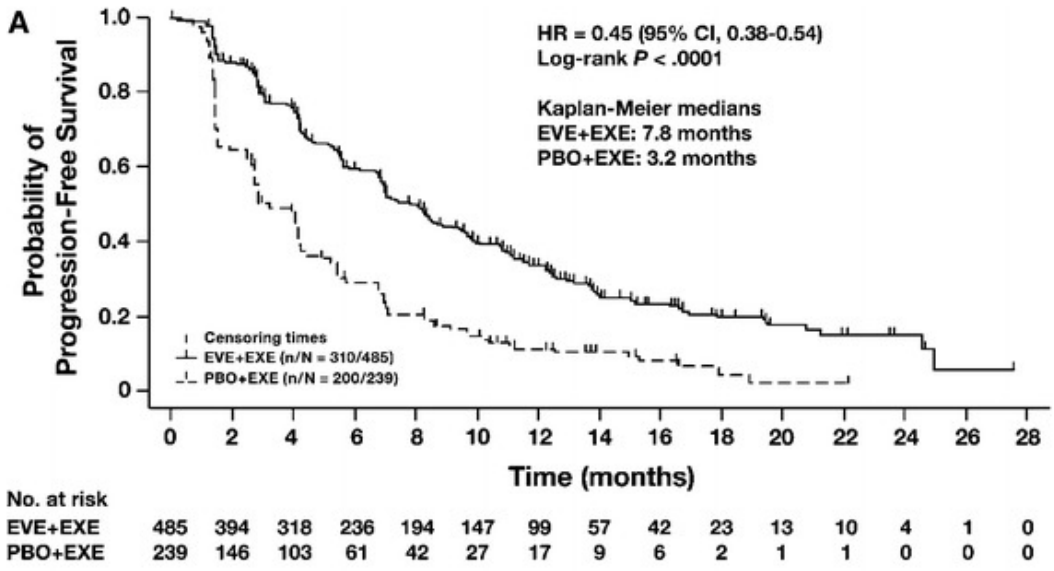


Figure 15: Kaplan-Meier estimates of progression-free survival of patients treated with everolimus plus exemestane versus exemestane alone. CI confidence interval, HR hazard ratio, EVE everolimus, EXE exemestane, PBO placebo (264) - permission to use not required

An in vitro study by Owusu-Brackett et al (265) tested TAK228 (an mTOR inhibitor). TAK228 enhanced the efficacy of eribulin (a chemotherapy agent) in TNBC, but further investigation is required to determine patient selection for this combination therapy.

3.2.3.5 Liver kinase B1 (LKB1)

LKB1 is a serine-threonine kinase described in the development of Peutz-Jeghers syndrome where around 45% of patients develop breast cancer in their lifetime (131), among other cancers.

Association with age

High expression of LKB1 has been noted in women aged >70 years compared to younger patients (3, 131).

Predictive value

There is little in the present literature about the predictive value of LKB1 in breast cancer. In a study of older breast cancer patients undergoing surgery (131), LKB1 expression was associated with better survival outcome among patients receiving adjuvant endocrine therapy.

Prognostic value

In the same study (131), within the older age group, LKB1 expression was associated with poor prognostic factors including high tumour grade and high expression of Ki67 and HER2.

Contrary to this, Chen et al (266) found that high LKB1 expression was predictive of better survival in patients with HER2-positive tumours.

Future directions

Future studies are required to investigate and compare the expression of LKB1 in the younger compared to older patient population and delineate the precise role of LKB1 as a therapeutic target. Metformin has been shown to selectively inhibit proliferation of LKB1 positive cancers including breast cancer (267), which needs further evaluation.

3.2.4 Cell proliferative markers

3.2.4.1 Ki67

Ki67 is a protein that is associated with cell proliferation. It is difficult to determine the overall expression of Ki67 in breast cancer due to the lack of consensus over the method of detection and definition of 'high' expression (268, 269).

Association with age

In a cohort of 462 patients with stage 1-3 breast cancer (270), younger age (≤40 years) was associated with high expression of Ki67. In a cohort of 575 older women (≥70 years), low expression of Ki67 was seen (3). These findings suggest a less proliferative disease in older women.

Predictive value

Measurement of Ki67 has been utilised to monitor the tumour proliferation index in neoadjuvant settings, particularly for chemotherapy treatment. A high Ki67 proliferation index is an independent predictive factor for pCR after NACT (271). The Immediate Preoperative Anastrozole, Tamoxifen, or Combined With Tamoxifen (IMPACT) trial was a randomised trial evaluating neoadjuvant anastrozole or tamoxifen or both before surgery in 330 postmenopausal women with ER-positive breast cancer (272). A secondary objective of the study was to evaluate changes in Ki67 expression after two weeks of treatment. Higher expression of Ki67 after two weeks of endocrine therapy was associated with shorter RFS, larger tumour size at baseline and lower ER level.

Prognostic value

A meta-analysis by Petrelli et al (273) exploring the prognostic value of Ki67 in early breast cancer analysed data from 64,196 patients in 41 studies. Ki67 was an

independent prognostic indicator in terms of OS and expression of >25% was associated with greater risk of death compared with lower expression rates.

A meta-analysis by de Azambuja et al (274) identified 12,155 patients from 46 studies where Ki67 was measured. Ki67 positivity was associated with higher rate of relapse and worse survival in patients with early breast cancer.

Future directions

Despite endorsements by several international guidelines, measurement of Ki67 is yet to gain widespread application as a prognostic or predictive marker, which is largely due to wide variation in methodology of detection and lack of standardisation in this area (269).

Recent studies in renal cancer have focused on Ki67 targeted therapy but this is not translated to breast cancer (275); therapies targeting Ki67 in the setting of renal carcinoma are in the preclinical stages.

3.2.5 Anti-apoptosis markers

3.2.5.1 B cell lymphoma 2 (BCL2)

BCL2 gene is the founding member of the BCL2 family of regulator proteins that regulate cell death by either inhibiting or inducing apoptosis. In a retrospective study of 605 cases of breast cancer, 54% showed BCL2 expression (276).

Association with age

In the same study (276), BCL2 expression was associated with young age (<50 years). High expression of BCL2 was observed in a cohort of women ≥70 years with primary breast cancer treated by surgery (3), compared to younger patients (<70 years) and specifically in older women with TNBC (277) (≥70 years). This is in agreement with the biological characterisation of 14,007 primary breast cancers by Daidone et al

(147), who showed an increase in BCL2 expression with age (≥65 years compared to <65 years).

Predictive value

High BCL2 has been linked to better clinical response to tamoxifen in ER-positive metastatic breast cancer (278).

Tamoxifen induces apoptosis in breast cancer cells. BCL2 genes interfere with apoptosis in various ways. A study by Zhang et al (279) investigated the competing effects of tamoxifen and BCL2 gene products and found that tamoxifen induced down-regulation of BCL2, thus inducing apoptosis.

Prognostic value

A meta-analysis by Callagy et al (280) reviewed 5892 cases of female breast cancer; high expression of BCL2 equated to better OS and DFS. This has been reproduced in other studies (281, 282).

Specifically in older women, BCL2 has been shown to correlate with better DFS and BCSS (3, 147).

Future directions

Therapeutic compounds have been developed to attempt to disrupt the action of BCL2, known as BH3-mimetics (Figure 16). These BCL2 inhibitors restrict tumour growth in xenograft models (283) and there are a number currently under further investigation in the clinical trial setting for example, Navitoclax (284).

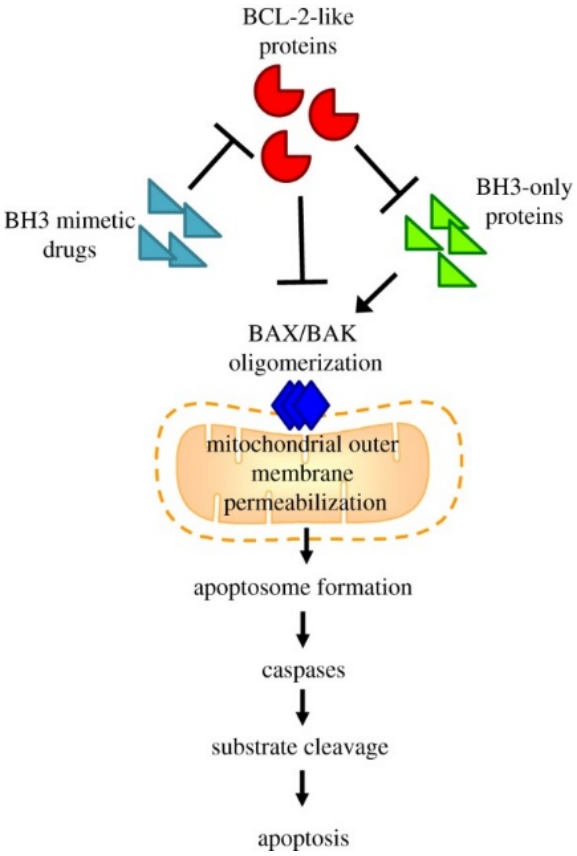


Figure 16: Action of BH3 mimetic drugs (283) - permission to use not required

3.2.6 Cell adhesion molecules

3.2.6.1 E-cadherin

E-cadherin is a type of cell adhesion molecule (CAM) expressed in most normal epithelial tissue. Selective loss of E-cadherin can promote invasion in human carcinomas and has been demonstrated in some breast cancers, in particular of the ER-positive lobular type (285).

Association with age

On examination of 450 cases of invasive breast cancer by Yang et al (286), age >60 years was associated with the positive expression of E-cadherin. This is replicated in a study of 575 women ≥70 years, when compared to a younger cohort (3).

Predictive value

Arthur et al (287) examined the biological characteristics of 14 postmenopausal breast cancer patients. Although the pre-treatment profiles of invasive lobular carcinomas compared to invasive ductal carcinomas, showed different expression of E-cadherin signalling, this was not a marker of response to neoadjuvant endocrine therapy.
Early in-vitro work has shown E-cadherin downregulation in tamoxifen resistant breast cancer cells (288).

Prognostic value

A study of 54 breast cancer patients by Younis et al (289) found a significant correlation between E-cadherin expression and node-negative cases. Furthermore, loss of E-cadherin was associated with advanced stages of breast cancer.
This agrees with the study by Yang et al (286) who found that high expression of E-cadherin was associated with better overall survival in cases of TNBC.

Future directions

E-cadherin appears to be a marker of poor prognosis in breast cancer. The role of E-cadherin in predicting response to treatment is currently unclear.
A number of monoclonal antibodies to a type of E-cadherin (VE-cadherin) that is mainly expressed in endothelial cells, where it plays a role in vascular integrity and therefore angiogenesis, have been developed (290). These largely remain in the pre-clinical phases, especially in the setting of breast cancer (291).

3.2.6.2 Cluster of differentiation (CD) 44

CD44 is a glycoprotein that promotes tumorigenesis, as well as maintaining the function of cancer stem cells in tumour regeneration following treatment (292).

Association with age

There is little in the literature regarding association of CD44 expression with age in the setting of breast cancer, however, differences in expression according to age are seen in thyroid carcinoma (293) and hepatocellular carcinoma (294).

Predictive value

Cell line studies have suggested that breast cancer cells with high expression of CD44 have a better survival response to endocrine therapy, than those with low expression (295). There is little in the literature to support this beyond in-vitro studies.

Prognostic value

A study of 140 primary breast cancer patients with stage 1-3 disease showed that high CD44 expression was an independent factor for decreased PFS and OS (296). These findings are consistent with a study of 145 patients with TNBC where expression of CD44 was also associated with poor prognosis (297). However, this is contrary to evidence presented by Liu et al (298) and Honeth et al (299).

Future directions

The exact role and function of CD44 needs to be more clearly understood before it can be used as a therapeutic target.

3.2.7 Cell differentiation markers

3.2.7.1 Cytokeratins (CKs)

Breast cancer originates from basal (myoepithelial) or luminal (glandular) tissue that express specific cytokeratins. In normal breast tissue, CK5 and CK14 are expressed in myoepithelial cells and CK7, CK8, CK18, CK19, in ductal epithelium (300).

Association with age

High expression of CK5, CK18 and low expression of CK7/8 has been seen in older populations of primary breast cancer patients (3), compared to younger women (<70

years). In a cohort of patients with TNBC, CK7/8 positivity was lower and CK18 higher, compared to a comparative younger series (115). A meta-analysis to determine the clinicopathological significance of CK18 in breast cancer, confirmed higher expression in the older population ≥50 years (301).

Predictive value

There is minimal evidence in the literature to support the predictive values of cytokeratins, which encompass a large heterogeneous group of markers.

An experimental study by Bozionellou et al (302) has shown that chemotherapy resistant CK19 mRNA-positive cells in the peripheral blood and bone marrow can be effectively targeted by trastuzumab administration.

Prognostic value

A study of 611 breast cancers by van de Rijn et al (303) found that expression of CK17 and CK5/6 in tumour cells was associated with poor clinical outcome. Multivariate analysis showed that in node-negative cancer, expression of CK17 and CK5/6 was a prognostic factor independent of size and grade.

A literature review by Haupt et al (304) on basal-like breast carcinomas confirmed that expression of basal CKs was an independent prognostic factor in lymph node negative cases, but in lymph node positive cases the relationship was unclear. Overall, the evidence suggests that basal-like breast carcinomas are associated with worse clinical outcome.

Future directions

Cytokeratin expression is associated with response to PET in a small cohort of older women with primary operable breast cancer (167) and these findings need validating in a larger cohort.

3.2.8 Others

3.2.8.1 Mucin 1 (MUC1)

MUC1 is an epithelial cell surface protein that is aberrantly overexpressed in over 90% of breast cancer as well as other cancers (305).

Association with age

There is little in the current literature to describe the association of the expression of MUC1 with age in breast cancer. A study in older compared to younger women having surgery did not show any statistically significant difference in expression (3).

Predictive value

In ER-positive patients, increased expression of MUC1 is associated with tamoxifen resistance (306).

Prognostic value

Breast cancers which exhibit increased and mislocalised expression of MUC1 are more likely to metastasise (306).

High levels of MUC1 are associated with poor clinical outcome (160, 161, 307). Jing et al (160) assessed expression levels of MUC1 in breast cancer and normal tissue from the Oncomine database. Abnormally high levels of MUC1 was associated with poor prognosis, with reduced BCSS.

An analysis of 171 cases of invasive breast cancer by McGuckin et al (161) showed patients whose tumours had high expression of MUC1 had significantly poorer DFS and OS, when compared to those with lower expression.

Rakha et al (169) demonstrated that that MUC1 overexpression was associated with lower tumour grade, smaller size and absence of metastasis, however, noted that aberrant cytoplasmic and membranous localisation of MUC1 was associated with poor

outcomes, compared with apical localisation, which is the normal physiological site of expression.

Future directions

MUC1 targeted therapies have been in development for the last 30 years, however effective benefit in clinical trials has not yet been achieved (308). A monoclonal antibody GGSK-1/30 targeting human tumour-associated MUC1 has shown promising initial results in human breast tissue analysis (305).

3.2.8.2 Amplified in breast cancer 1 (AIB1)

AIB1 is a member of the steroid receptor coactivator family and mediates the activities of nuclear receptors including ER and PR. AIB1 is overexpressed in 30-60% of breast cancers (309).

Association with age

AIB1 expression was measured in 185 breast cancer surgical resection specimens; there was no difference noted between expression and age (≤50 years compared to >50 years) (310).

Predictive value

In postmenopausal women with primary breast cancer, high AIB1 has been implicated as a marker of good response to tamoxifen treatment as adjuvant therapy. Furthermore, it was a prognostic marker of decreased RFS in patients who did not receive adjuvant systemic therapy (311).

Osborne et al (312) measured AIB1 and HER2 levels in 316 breast cancer patients. In patients who received tamoxifen, high AIB1 expression was associated with worse DFS, indicative of tamoxifen resistance. Patients whose tumours expressed AIB1 and HER2 had worse outcomes with tamoxifen therapy than other combinations. These two studies are conflicting in their findings.

Prognostic value

The BIG 1-98 trial randomised 1396 Danish patients to adjuvant tamoxifen and/or AIs (313). AIB1 expression correlated to expression of HER2 and high grade tumours as well as poor DFS and OS.

In a study of 185 breast cancers by Lee at al (310), high AIB1 expression was associated with lower DFS and OS. These findings were most significant in patients with ER-negative cancers.

Future directions

Steroid receptor coactivators such as AIB1 lack a high affinity ligand binding domain and therefore are considered difficult drug targets (314).

An AIB1 small molecule inhibitor, bufalin, has been shown to effectively inhibit tumour growth in TNBC in a mouse model (315), suggesting a potential target for TNBC treatment. Bufalin and other small molecule inhibitors are still in pre-clinical phases of investigation (314).

3.2.8.3 Solute carriers

Glutamine is the second primary metabolite to fuel cancer cell proliferation after glucose. The solute carrier family 1 member 5 (SLC1A5) and SLC7A5 are the key cellular glutamine transporters which are upregulated in a variety of cancers with some demonstrating glutamine dependence (316). SLC7A5 requires association with SLC3A2 for its expression (171).

Association with age

There is no evidence in the current literature to specifically associate expression of glutamine transporters with age.

Predictive value

Co-expression of SLC7A5 and SLC3A2 is associated with poor outcome in patients

receiving endocrine therapy (317). SLC1A5 has been associated with a high risk of relapse after endocrine treatment (318).

Prognostic value

The combined high expression of SLC1A5, SLC7A5 and SLC3A2 has been associated with shorter BCSS in ER-positive breast cancer (171).

Expression of SLC7A5 has been associated with larger tumour size and higher grade and with poor outcome in highly proliferative ER-positive, HER2-positive subtypes of breast cancer (319).

Future directions

Glutamine transporters could present exciting new therapeutic targets to halt cancer cell growth and endocrine therapy resistance. Inhibitors of SLC7A5 and SLC1A5 are in the laboratory testing phase, with one inhibitor of SLC7A5 in phase I clinical trials (171). The consequences of blocking the action of these solute carriers is not yet clear. Association of glutamine transporters with age is important to ascertain. It can be hypothesised that because breast cancers in older women are generally less proliferative than their younger counterparts, they may be less reliant on glutamine.

3.2.9 Other potential biomarkers

It is not possible to discuss every biomarker relevant to breast cancer in this chapter, of which are there many hundreds. The biomarkers discussed are known to be relevant in breast cancer and are under further investigation in this setting.

Some more recent biomarkers which have been discovered to have a prognostic role in breast cancer and involvement in acquired endocrine resistance include zinc transporters (320), cathepsin (321), cell division regulators (such as CDC20) (322) and other members of the nuclear receptor superfamily (such as RXRG) (323).

With such a plethora of literature of copious numbers of biomarkers, it can be daunting to understand how to use this information in clinical practice. A number of predictive and prognostic tools, in the form of computer models, have been developed and are currently on the market. These will now be discussed.

3.3 Predictive and prognostic tools using tumour tissue

3.3.1 Background to predictive and prognostic tools

A move is being made towards personalised cancer treatment, where tumour biology of an individual patient is examined, to give unique predictive and prognostic information. This is even more important in the setting of older women, to direct the healthcare team and the patient to decide on the most appropriate treatment plan. Prognostic tests inform us about a possible cancer outcome (recurrence of the disease, likelihood of death from cancer) independent of treatment, over a specified time-period (324). Predictive tests provide information on response to a specific treatment (325).

Some tools to help inform the prognosis and response to therapy do exist. The Nottingham Prognostic Index (NPI) (326) was the first tool of its kind to assess a combination of factors, including histological grade which reflects tumour biology, together with size of tumour and nodal status (time-dependent factors), to inform prognosis following surgery. Assessment that is more comprehensive, for example Adjuvant! Online (327), and Predict (35), which use more clinico-pathological features, have been developed, but recruitment of older women in their conception is lacking and the aims of these tools are not focused to the treatment dilemmas of the older population. Furthermore, these tools do not assess biology beyond grade and status of ER, PR and HER2 (+/- Ki67). To provide information specific to an individual patient, unique material from the patient's tumour is required, rather than

using standard clinicopathological variables such as tumour grade and size. Assessment of patient tissue can be protein or genomic based.

3.3.2 Evidence base for use of predictive and prognostic tools in older women with primary breast cancer

3.3.2.1 Presentation of the evidence base

Six of the most well-known tumour based tools currently available include: MammaPrint, Oncotype DX, Mammostrat, Prosigna (PAM 50), EndoPredict and IHC4. A summary of the clinical use of each tool as well as the validation of the tools, with specific interest in validation in older women is given in Table 6. Part of this work has been published as a narrative review:

S Gordon-Craig, **RM Parks** and KL Cheung. The Potential Use of Tumour-Based Prognostic and Predictive Tools in Older Women with Primary Breast Cancer: A Narrative Review. Oncology and Therapy. Dec 2020, 8 (2): 231 – 250.

A summary of the findings are now presented.

Tool	Recommended source and nature of sample	Central laboratory required	Setting	Prognostic value and validation	Predictive value and validation	Validation of tool in older women
MammaPrint	SE; Fresh/frozen or FFPE	Yes	Adjuvant treatment following surgery	Early stage breast cancer, tumour size ≤5cm, irrespective of HR status, up to 3 positive nodes; Netherlands Cancer Institute, 295 patients, <53 years of age (no age range given) (328)	No	MINDACT phase III RCT, 6693 patients, median age 55 years (range 23-71), assessing prognosis of low genomic, high clinical risk, without chemotherapy (329, 330)
Oncotype DX	SE; FFPE	Yes	Adjuvant treatment following surgery	Early stage breast cancer, ER-positive, HER2-negative, up to 3 positive nodes; NSABP B-14: 668 patients in total- 301(45%) were aged >60 years (no age range given) (331)	Early stage breast cancer, ER-positive, HER2-negative, up to 3 positive nodes; NSABP B-20: 651 patients 196 (30%) >60 years (no age range given) (331)	RxPONDER phase III RCT, of 5018 women, 66.8% postmenopausal, 11.6% ≥70 years (range 18 – 87)
Mammostrat	SE; FFPE	No	Adjuvant treatment following surgery	Early stage breast cancer, ER-positive, LN negative;	No	4598 TEAM patients, ER-positive post-menopausal patients, RCT, accurately predicted DRFS

				Cleveland Clinic Foundation, 299 patients, 75% >50 years (no age range given) (332); British Columbia Cancer Agency: 344 patients (no age range given) (332)		median age 64 years (range 34-96) (333)
Prosigna (PAM 50)	SE; FFPE	No	Adjuvant treatment following surgery	Post-menopausal, early breast cancer, ER-positive, HER2-negative, up to 3 positive nodes; 1478 post-menopausal women with ER-positive early breast cancer, ABCSG-8 group, Median age: 64 years (range 41.4-80) (334)	No	1478 post-menopausal women with ER-positive early breast cancer, ABCSG-8 group, Median age: 64 years (range 41.4-80) (334)
EndoPredict	SE; FFPE	Central lab or local pathology	Adjuvant treatment following surgery	Early breast cancer, ER-positive, HER2-negative, up to 3 positive nodes; 2 cohorts of 1702 samples acquired from patients enrolled into 2 randomised trials at the ABCSG; ABCSG-6 median age 65.6 years (range 43.7 - 80.7; ABCSG-8 median age for	No	GEICAM 9906, phase III, RCT 555 patients randomised to receive either FEC or FEC-P 250 (45%) of the patients were <50 305 (55%) were ≥50 (no age range given), Confirmed independent prognostic marker of recurrence at 10 years (336)

				this group was 63.4 years (range 41.5 - 80.5) (335)		
IHC4	SE; FFPE	No	Adjuvant treatment following surgery	Post-menopausal, early stage breast cancer, ER-positive, up to 3 positive nodes, 5 years of adjuvant endocrine therapy; 786 patients in Nottingham Median age: 55 (range 48-63) (337)	No	TEAM trial: 2919 samples post-menopausal patients with early PR/ER positive, invasive breast cancer who had completed primary, endocrine therapy confirmed IHC4 ability to predict residual risk (no age range given) (338)

Table 6: A summary of six most common tumour-based predictive and prognostic tools

Abbreviations: ABCSG – Austrian Breast and Colorectal Cancer Study Group; ER – oestrogen receptor; DRFS – distant recurrence free survival; FEC-P – fluorouracil, epirubicin, cyclophosphamide and paclitaxel; FFPE - formalin-fixed paraffin-embedded; HER – human epidermal growth factor receptor; HR – hormone receptor; LN – lymph node; NSABP – National Surgical Adjuvant Breast and Bowel Project; PR – progesterone receptor; RCT – randomised controlled trial; RxPONDER - The Clinical Trial for Treatment for Positive Node, Endocrine Responsive Breast Cancer ; SE – surgical excision; TEAM – The Tamoxifen Exemestane Adjuvant Multinational trial

3.3.2.2 Summary of the evidence base

Oncotype DX stands out as the only tool validated for both predictive and prognostic utility. It is the most used frequently tool in the USA and has the largest body of evidence in terms of trials conducted and research undertaken. The large-scale Trial Assigning Individualized Options for Treatment (TAILORx) study (339) was promising in its recruitment of older women but in the intermediate risk group the study focused on, only 4.3% of the women were over 70 years old. The Clinical Trial for Treatment for Positive Node, Endocrine Responsive Breast Cancer (RxPONDER) (340) included women with ER-positive, HER2-negative breast cancer with positive lymph nodes and a RS of ≤25 and had slightly better success with recruitment of older women; they included 11.6% women ≥70 years. Oncotype Dx has been recently extensively tested and validated in over 900000 CNB specimens (165). This confirms its potential use in the primary and neoadjuvant settings.

As the oldest tool, MammaPrint also has bountiful evidence and scope for use as a predictive as well as prognostic tool. However, in the most recent trial of its efficacy, only patients up to the age of 70 were recruited (330). It was also validated on younger women of <53 years (341).

Many of the key trials concerning Mammostrat only provide age information by categories of <50 or >50 years. There is overall less evidence for Mammostrat. The TEAM trial (333) (also concerning IHC4) includes patients up to the age of 96, but further breakdown by age has not been provided, which makes the results difficult to analyse in this way.

PAM 50 is specifically recommended for postmenopausal women only, which implies it

may be preferential to an older group. In addition, unlike the other tests its classification by subtype may pose an advantage in terms of tumour heterogeneity by age. However, since its validation there has not been further contribution to the evidence for its use in older women.

EndoPredict has evidence to support its use in patients up to the age of 80 years but like PAM 50, has not been included in trials concerning older women since its inception. The GEICAM 9906 trial (336) includes 55% who are over 50 years of age, but further breakdown is not provided.

The newest tool, IHC4 has also been validated in the TEAM trial but as previously mentioned, information about age is lacking. Interestingly, IHC4 stands out as having the largest evidence base in the literature of CNB, which makes it potentially exciting for use in older women.

3.3.2.3 Current issues with use of tumour profiling tools in older women

A recent systematic review by the SIOG team in older women with early luminal breast cancer, concluded that there was limited evidence to show that gene expression profiling (primarily Oncotype Dx) aided treatment decisions in this population and use was offered less frequently compared to younger patients (79).

Age range

From this review, it is clear that overall there is a lack of evidence specific to older women, which is echoed in clinical trials of new therapies and research in general (56). This is partly due to lack of recruitment of older women and where older women are

recruited, numbers are few. This may be due to failure of inclusion by the researchers or lack of willingness to participate by the patient due to competing comorbidities, concern of impact on treatment or quality of life.

Where older women are included, often analysis by age has not been performed, or a breakdown of results by age has not been given. This makes it difficult to interpret the results in this specific cohort.

Setting

Current tools available are focused mainly on ER-positive tumours and considering adjuvant treatment decision-making, especially chemotherapy. Although the majority of breast cancers in older women are ER-positive, there is an important group of patients who are ER-negative. Especially if the patient is unfit for or declines surgery, it is important to understand what potential alternative treatments are available to them, as PET is not recommended.

Tissue sample required

The tumour profiling tools have been developed on SE specimens rather than CNB samples and have not been comprehensively validated in this source, (342, 343), with the exception of Oncotype Dx; guidance for all the current tools suggest SE as the most reliable form of tissue to generate a result.

Using CNB presents a degree of uncertainty and some studies have shown discordance between different markers particularly PR when it comes to CNB (344). The evidence is unclear and although all of the tools have been tested in CNB to some extent, in general most studies of this kind involve small sample sizes. This is an important area for future

research.

3.3.3 Future directions of predictive and prognostic tools in older women

The evidence for use of the tools discussed in this chapter specific to older women with primary breast cancer is lacking. These tools are most often derived from and validated in cohorts of patients consisting entirely of younger women or including a small number of older women. The tools are mainly focused on predicting benefit of adjuvant therapy after surgery, therefore, cannot provide any useful information in the cohort of older women who have PET.

This is extremely concerning giving our ageing population; the number of older women living with breast cancer worldwide will double by 2040 to nearly 1 million cases in patients ≥70 per year (11, 12). Despite recognition of the importance and potential benefit of these tools, they have not been adequately tested or modified for the older population; there needs to be greater recognition and focus specifically on the older population when it comes to validating existing tools and in the development of new tumour based tools. A lack of information means that categories such as post-menopausal and >50 years of age are falsely attributed to the evidence base for older women. The biology of breast cancer in older women has been shown to have different biological features compared to their younger counterparts with a significant change in biology noted at >70 years (8); in the era of personalised treatments, the unique tumour biology of this cohort can no longer be ignored.

Measurement of biomarkers on CNB samples is a process that could greatly help in the setting of primary disease in older women. Small studies have already shown the

possibility of using existing tools in CNB specimens; however, these findings require validation in large prospective studies.

3.4 Conclusions and implications for this thesis

There is a role for extensive biomarker research, as described in this chapter. Biomarkers have significance in detection, treatment and measurement of recurrence risk of breast cancer. Current tools are not largely validated for use in older women (≥70 years) and are primarily concerning adjuvant treatment. In the future, optimal personalised treatment for the individual patient with breast cancer based on unique biological characteristics of their tumour is achievable.

ER, PR and HER2 are biomarkers routinely measured by IHC on CNB taken at diagnosis for breast cancer. Many other biomarkers are associated with predicting prognosis and response to therapy in breast cancer but are not currently measured outside of a research setting. Differences in expression of biomarkers between younger and older women has been confirmed for ER, HER2, P53 and cytokeratins. The mechanism of endocrine resistance is complex, and biomarkers associated with this are HER3, HER4, EGFR, P53, BRCA1, PTEN, BCL2, MUC1, AIB1 and amino acid transporters. The biomarkers ER, PR, HER2, HER4, EGFR, VEGF, p53, PTEN, Ki67, BCL2, E-cadherin cytokeratins, MUC1 and AIB1 have an association with survival. There are numerous other biomarkers, such as MDM2, MDM4 and CD44 for which there is emerging evidence to suggest they have a role in breast cancer, but the exact nature of this role is unclear at present. Once there is a better understanding of the complex roles and interactions of biomarkers in breast cancer, there is potential for prediction tools to be developed and employed at all stages of the treatment pathway.

A further biomarker of great interest is cytoplasmic cyclin E (132), which is currently only performed in a limited number of research laboratories worldwide (345). Due to the logistical and financial issues associated with this, this was not pursued as part of the present thesis.

The next stage of this thesis will be to analyse a panel of 31 biomarkers discussed in this chapter, on all available tissue specimens (both CNB and SE) in the Nottingham cohort of older women with primary breast cancer. The biomarkers to be studied are: ER, PR, AR, EGFR, HER2, HER3, HER4, VEGF, P53, BRCA1, BRCA2, MDM2, MDM4, PTEN, LKB1, Ki67, BCL2, E-Cadherin, CD44, CK5, CK 5/6, CK 7/8, CK14, CK17, CK18, CK19, Muc1, AIB1, SLC1A5, SLC3A2, SLC7A5.

METHODS

Chapter 4: Outline of materials and methods for remainder of thesis

4.1 Introduction

A description of the general materials and methods utilised in this study are included in this chapter. A detailed description of methods will be given where relevant, in later chapters. This PhD project is part of work within the Nottingham Breast Cancer Research Centre, which has appeared numerous times in this thesis so far. The project is based on an existing cohort of 1758 older women with primary operable breast cancer, which will be described in detail below.

Following the above work in Chapter 2 and 3, a panel of 31 biomarkers examined by IHC in this cohort of older women with primary breast cancer, has been selected (section 3.4).

4.2 Patients

4.2.1 Older women series

4.2.1.1 Historic database

This cohort consists of 1758 patients who consecutively attended a dedicated clinic for older women with primary breast cancer, in Nottingham from 1973 to 2010. Patients were female, aged ≥70 years, diagnosed with invasive early operable primary breast cancer (≤5cm tumour size) with no evidence of metastatic disease at diagnosis. This database has been extensively described in the literature (3, 112).

The database was collated retrospectively at the University of Nottingham, City Hospital Campus and University Park Campus, Nottingham, United Kingdom, and was approved by the Research Ethics Committee (Nottingham Research Ethics Committee 1) under REC reference 19/YH/0293 with the title 'Refining the prognosis of breast cancer and

breast cancer precursor lesions for personalised therapy' (Appendix 1); the protocol is given in Appendix 2. All data was collected from patient electronic or case notes and therefore the reliability of the data is dependent on what was input by others previously. An outline of the information collated in the database is given in Table 7.

In summary, the series consists of: clinical data collected at diagnosis, pathological features of the disease identified using diagnostic CNB, additional pathological data obtained from patients treated by surgery, outcome data until last follow-up or death.

Clinical information	**Additional information from surgical excision (if had surgery)**
Age	Pathological size
Date of death	Histological type
Cause of death	Grade
Axillary nodal status	
Grade of tumour	
Clinical size of tumour	
Histological type	
Primary treatment	
Adjuvant hormone therapy	
Response at six months	
Progression status and time to progression	
Further treatment/response/progression status/time to progression	
Local recurrence status and time to recurrence	
Regional recurrence status and time	
Metastatic status and time to metastasis	
Date of review	
Survival in months	
Radiotherapy status	

Table 7: Information available in database of older women with primary breast cancer

4.2.1.2 Treatment protocols

The pattern of management over the period is given below.

1973 – 1989

All patients were offered surgery. Those who declined or were unfit for surgery received tamoxifen. ER status was not routinely assessed and there were no standard criteria for adjuvant systemic therapy.

1990 – 1999

Measurement of ER status as part of standard clinical practice was introduced, initially by ligand binding, then immunoassays and later IHC. All patients with a confirmed diagnosis of ER-positive invasive cancer had the option of surgery or PET. Those with ER-negative tumours were offered surgery. Patients undergoing surgery had the options of mastectomy or BCS if appropriate. Axillary surgery was considered when there were clinically palpable lymph nodes only. Patients who underwent BCS were offered either adjuvant endocrine therapy +/- adjuvant radiotherapy or no further treatment based on clinical judgement, tumour grade, margin status, nodal status if available and degree of ER-positivity.

Tamoxifen was the initial choice of PET. Patients with contraindications to tamoxifen were offered a non-steroidal AI such as anastrozole (available after 1995).

If patients progressed on first line PET, they were considered for surgery, further endocrine therapy or primary radiotherapy. Patients with ER-negative tumours who declined or were unfit for surgery, were offered primary radiotherapy.

2000 – 2010

The treatment protocol became more structured and all cases were discussed at a multi-

disciplinary team meeting, before patients were seen in the clinic. Individual treatment recommendation was based on a joint assessment by a surgeon, oncologist and breast cancer nurses in the clinic. The following describes the general recommendation.

All patients with a confirmed diagnosis of breast cancer regardless of ER status were recommended to have surgery in the form of mastectomy or BCS as appropriate. Axillary surgery became standard procedure. At the advent of the decade this was by four-node sampling. The introduction of blue-dye sentinel node biopsy for a clinically negative axilla, or level 3 axillary clearance for node positive cases proven preoperatively, came later in the decade. Intact breast irradiation was normally given following BCS. Post-mastectomy chest wall irradiation was given if the tumour had two out of the following features: grade 3 disease, positive nodal status, vascular invasion.

Addition of adjuvant endocrine therapy for ER-positive cancers was largely based on the Nottingham Prognostic Index Score (NPI). For those with a Nottingham Prognostic Index (NPI) score of ≥3.4 adjuvant endocrine therapy was recommended. Those with NPI 3.0-3.4 were given the option of adjuvant endocrine therapy. Those with NPI <3.0 did not receive any systemic therapy.

Adjuvant chemotherapy +/- trastuzumab, depending on HER2 status, was considered in fit, 'younger' patients with moderate/poor prognostic tumours, especially if they were ER-negative.

Patients with ER-positive tumours who refused or were deemed unfit for surgery were given PET as described above and those with ER-negative tumours offered primary radiotherapy.

4.2.1.3 Follow-up

All patients were followed up in the Primary Breast Cancer Clinic for Older Women. Patients treated surgically were followed up on a yearly basis or upon request. Patients receiving non-operative therapy were followed up six weeks after commencement of therapy, then after three months and then six months. After this they continued yearly follow-up. The last follow-up status was collated in 2012 either by case note or electronic note review, as part of a previous PhD by Dr Binafsha Syed.

4.2.2 Younger women series

A cohort of younger patients (<70 years) is available for comparison (111). This database consists of a consecutive series of 1944 cases of primary operable invasive breast cancer from patients presenting from 1986-1998 at the same institution and entered into the Nottingham Tenovus Primary Breast Carcinoma Series. Data collection and biological analysis of breast cancer in younger women were performed independent of this current research programme. All patients were managed in the same unit following the same clinical guidelines at any given time.

4.3 Construction of tissue microarrays (TMAs)

From the literature review in Chapter 2, it is apparent that construction of TMAs allows for the most cost-effective and efficient method for examining a large panel of biomarkers. All SE and CNB samples from the older women series, where available, were considered for potential construction of TMA. All blocks with enough tumour tissue to construct a TMA were selected. All TMAs in the older women series were constructed prior to this current thesis and are briefly described here.

During the period of my PhD thesis, no new TMA blocks have been made by my research

group, however, I have been shown the process of how this is done and familiarised myself with the equipment used. I have had the opportunity to discuss the process with several research technicians and experienced researchers in this field.

From the whole series it has been possible to construct SE TMAs using 575 patient samples and CNB TMAs using 693 patient samples. Ability to construct TMA was based upon availability of specimen (some specimens dating back to the 1970s could not be identified) and availability of tumour tissue within each specimen. Some of the surgical specimens had been overused (for other research studies) and tumour tissue was depleted. Due to the smaller comparative size of CNB (compared to SE), many CNB samples were unsuitable for TMA construction.

4.3.1 Rationale for using TMAs

The TMA technique is a validated and reliable method of protein analysis for a large number of tumours on a single glass slide (346). It is economical in terms of time and money and provides a standardised environment for multiple tumour cores positioned on one slide. Furthermore, it amplifies the utilisation of a small amount of tumour tissue, as tissue resources of this kind are often scarce (166).

4.3.2 Surgical excision (SE) TMA

Tissue microarrays of formalin-fixed paraffin-embedded tumour (FFPE) sections were constructed as described (347). Briefly, 0.6mm diameter cores of the representative part of the tumour blocks were implanted in the TMA blocks using Beecher's manual tissue microarrayer (MP06 Beecher Instruments Inc., Sun Prairie, WI, USA). It was possible to construct four TMAs from each tumour sample. These TMAs were constructed by Dr Binafsha Syed (348).

4.3.3 Core needle biopsy (CNB) TMA

These TMAs were constructed using a novel technique developed within the Nottingham group (167, 349). Briefly, based on tumour tissue availability, haematoxylin and eosin (H&E) slides and paraffin CNB blocks were reviewed and multiple areas of 4mm of tumour were marked from each case, retrieved with skin biopsy punch and implanted in Agarose-paraffin TMA blocks. It was possible to construct only one or two TMAs from each tumour sample (due to the small size of the sample). These TMAs were constructed by Dr Mohammad Albanghali (167).

4.4 Western blot (WB)

Western blot (WB) was used to validate specificity of primary antibodies used to investigate biomarker expression. The procedure was undertaken to detect specific bands of antigen. This was carried out by several members from the research group prior to the current thesis and confirmed the specificity of the primary antibodies. For new antibodies which had not been used previously by the research group, WB was not undertaken due to limited resources and time pressures when this arose during the thesis (due to the COVID-19 pandemic). The data sheet and corresponding references provided by the manufacturers were taken as a guide. This will need to be confirmed by the research team at a later date.

4.5 Immunohistochemistry

From the review of biomarkers in the current literature in Chapter 3, a panel of 31 biomarkers have been selected for analysis. The biomarkers, antibody clones and retrieval methods are given in Table 8. The following section describes the immunohistochemistry process which was conducted by myself during this thesis.

Biomarker	Antibody clone	Dilution	Incubation time	Antigen retrieval method
ER	RM-9101-SP1 / Neomarkers	1:100	45 minutes	Citrate buffer
PR	PgR 636 / Dako	1:200	45 minutes	Citrate buffer
AR	M3562 / Dako	1:50	30 minutes	Citrate buffer
EGFR	31G7 / Zymed	1:30	60 minutes	Proteinase K
HER2	Herceptest Kit / Dako	Pre-diluted	30 minutes	Citrate buffer
HER3	MS-201-P1ABX / Neomarker	1:300	60 minutes	Citrate buffer
HER4	RB-9045 / Thermo Scientific	1:50	60 minutes	Citrate buffer
VEGF	SP28-Ab27620 / Abcam	Pre-diluted	30 minutes	Citrate buffer
P53	DO-7 / Novocastra	1:100	45 minutes	Citrate buffer
BRCA1	MS110 mouse monoclonal / Calbiochem	1:100	40 minutes	Citrate buffer
BRCA2	PC146 rabbit polyclonal / Calbiochem	1:100	Overnight	Citrate buffer
MDM2	NCL-114317 / Novocastra	1:200	24 hours	Citrate buffer
MDM4	00108 / Bethyl Laboratories	1:100	1 hour	Citrate buffer
PTEN	M3627 / Dako	1:25	1 hour	Waterbath
LKB1	Ley37D/G6-ab15095 / Abcam	1:100	30 minutes	Citrate buffer
Ki67	M7240 / Dako	1:100	45 minutes	Citrate buffer
BCL2	M9887 / Dako	1:100	45 minutes	Citrate buffer
E-Cadherin	M3612 / Dako	1:100	45 minutes	Citrate buffer

CD44	DF1485 / Dako	1:50	45 minutes	Citrate buffer
CK5	M26 / Thermo Scientific	1:100	45 minutes	Citrate buffer
CK5/6	M7237 / Dako	1:100	45 minutes	EDTA
CK7/8	34779 / BioSciences	1:100	45 minutes	Citrate buffer
CK14	LL002 / Vector Laboratories	1:100	45 minutes	Citrate buffer
CK17	E3 / Vector Laboratories	1:100	45 minutes	Citrate buffer
CK18	M7010/ Dako	1:100	45 minutes	Citrate buffer
CK19	CM242A / Biocare Medical	1:100	45 minutes	Citrate buffer
Muc1	NCL-Muc1 / Novocastra	1:300	45 minutes	Citrate buffer
AIB1	011105 / BD Transduction Laboratories	1:75	Overnight	Citrate Buffer
SLC1A5	HPA035240 / Sigma	1:200	30 minutes	Citrate Buffer
SLC3A2	HPA017980 / Sigma	1:1000	1 hour	Citrate Buffer
SLC7A5	EPR17573 / Abcam	1:50	Overnight	Citrate Buffer

Table 8: Details of biomarkers, antibody clone and retrieval methods used for panel of 31 biomarkers

Abbreviations: AIB – amplified in breast cancer; AR – androgen receptor; BCL – B-cell lymphoma; BRCA – breast cancer gene; CD – cluster of differentiation; CK – cytokeratin; ER – oestrogen receptor; EGFR – epidermal growth factor receptor; HER – human epidermal growth factor receptor; LK – liver kinase; MDM - mouse double minute homolog, MUC – mucin; PR – progesterone receptor; PTEN – phosphatase and tensin homolog; SLC – solute carrier; VEGF – vascular endothelial growth factor

4.5.1 Antibody optimisation

Optimisation of all antibodies was performed on breast TMA sections prior to IHC in order to achieve specific staining with minimal background. Optimisation involved using different antibody concentrations and incubation times based on the manufacturer's recommendation.

Positive and negative controls using breast TMAs were used in each experiment. Negative controls with omission of the primary antibody were used to evaluate non-specific binding of the secondary antibody. Tissues with known staining patterns were used as a positive control. Where positive control tissues were not available, rabbit polyclonal β2-microglobulin (B2M) (1:2000, Dako) was used. B2M is a protein that is reliably found on the surface of almost all human cells, including mammary tissue (350).

4.5.2 Immunohistochemical staining

IHC staining was performed on 4μm TMA sections using the Novolink Polymer Detection System (Leica Biosystems) as per the standard operating procedure of the Nottingham research group. This system utilises a novel controlled polymerisation technology, which resolves the problem of non-specific staining seen with traditional streptavidin/biotin detection systems due to endogenous blocks.

1. Labelled tissue slides were heated at 60°C for 10 minutes before cooling for 5 minutes at room temperature. This facilitates strong adhesion of tumour cores to the slides.

2. Using either an autostainer (Leica XL, Leica Microsystems) or by manual technique, slides were dewaxed and rehydrated using the following steps: 1. Xylene bath - 5 minutes; 2. Xylene bath - 5 minutes; 3. Alcohol (industrial methylated spirit (IMS)) bath

- 2 minutes; 4. Alcohol (IMS) bath - 2 minutes; 5. Alcohol (IMS) bath - 2 minutes; 6. Water bath - 5 minutes.

3. Antigen retrieval was performed using the appropriate buffer (Table 8) and heated as recommended for 20 minutes on full power using a microwave (Whirlpool JT359 1000W) to achieve a temperature of 100°C.

4. Slides were cooled for at least 5 minutes using running tap water. Slides were then loaded into a flatbed. A hydrophobic barrier pen was used around the tissue section to form a hydrophobic barrier.

5. Next, slides were washed with Tris Buffered Saline containing 0.1% Tween 20 (TBS-T) at pH 7.6. Peroxidase block was applied to the slides for 5 minutes to block endogenous peroxidase activity, followed by 3 x 5 minute washes with TBS-T.

6. Protein block was applied for 5 minutes to block non-specific binding, followed by 3 x 5 minute washes with TBS-T.

7. This was followed by adding the primary antibody for each target (Table 8), optimally diluted in Leica antibody diluent (AR9352, Leica Biosystems). The slides were incubated for the required time (Table 8) before they were washed with TBS-T for 3 x 5 minutes.

8. Next, post-primary solution was applied for 30 minutes followed by another 3 x 5 minute washes with TBS-T.

9. Novolink polymer was applied for 30 minutes before another 3 x 5 minute washes with TBS-T.

10. 3,3'Diaminobenzidine (DAB) solution was applied for 5 minutes before repeating 3 x 5 minute TBS-T washes.

11. Finally, haematoxylin was applied for 6 minutes and then the slides were placed in a water bath.

12. Using either an autostainer (Leica XL, Leica Microsystems) or by manual technique, slides were dehydrated and cleaned using the following steps: 1. Alcohol (IMS) bath – 2 minutes; 2. Alcohol (IMS) bath – 2 minutes; 3. Alcohol (IMS) bath - 2 minutes; 4. Alcohol (IMS) bath - 2 minutes; 5. Xylene bath - 5 minutes; 6. Xylene bath - 5 minutes.

13. Finally, slides were mounted with a glass cover slip by application of Distyrene Plasticizer Xylene (Leica Microsystems) and left to dry at room temperature for a minimum of 24 hours.

4.6 Evaluation of staining

4.6.1 Digitalisation

Stained TMA sections were digitalised using high-resolution digital images (Nanozoomer; Hamamatsu Photonics), at 20x magnification and viewed using Xplore (Philips) software.

4.6.2 Scoring

Evaluation of TMA cores was based on a semi-quantitative assessment of invasive tumour cells using McCarty's immunohistochemical scoring (H-score; range 0-300) (351).

The intensity of tumour cell staining was described as follows:

0 = no staining; 1 = weak staining; 2 = moderate staining; 3 = strong staining.

The percentage of each category was estimated and the H-score was calculated by multiplying intensity and percentage of staining. Cores were only assessed if invasive

tumour burden was >15%. All cases were scored without foreknowledge to patient clinical outcomes or outcome data.

A special scoring system was used to score expression of HER2. This involved scoring the staining of the membrane as follows:

0 = No membrane staining or incomplete membrane staining in >10% of invasive tumour cells

1+ = faint/barely perceptible or weak incomplete membrane staining in >10% of tumour cells

2+ = a weak to moderate complete membrane staining is observed in >10% of tumour cells OR strong complete membrane staining in ≤10% of tumour cells

3+ = a strong (intense and uniform) complete membrane staining is observed in >10% of the invasive tumour cells.

H-score has been used in this present thesis to allow comparison to previous work by the research group. Although all methods of manual scoring are subjective measurement, they have been shown to be comparable to automated methods (352) and can be bolstered by validation.

4.6.3 Validation

The tumour cores were scored three times and the average of the scores taken as the final score. For inter-observer concordance, at least 10% of the cores were randomly scored by a second reviewer and correlation coefficient statistics performed for all markers.

R-values were interpreted as follows using published guidance (353):

r 0.7 – 1 = strong correlation

r 0.5 – 0.7 = moderate correlation

r <0.4 = weak or no correlation.

A correlation coefficient of ≥0.7 between two reviewers was accepted.

Validation was performed by other members of the research team: Dr Lutfi Alfarsi, Ms Brendah Masisi and Dr Yousif Kariri.

4.7 Association between biomarker and clinico-pathological features

The Statistical Package for Social Sciences (SPSS, version 25.0, Chicago, IL, USA) was used for data collection and analysis, as previously described (166). The X-tile Bio-informative software was used to define cut-offs (354). Chi-squared test was use for comparisons of biomarker expression between groups. Survival analysis was performed using Kaplan-Meier methods with application of log-rank. The Cox-regression model was used for multivariate analysis. A p-value of <0.05 was considered significant.

4.8 Cluster analysis

Biological patterns were characterised by partitional clustering methods as described (355), using R (R, version 4.1.2, a free software environment) clustering software. Cluster analysis will be described in more detail in Chapter 7. Cluster analysis was performed with the assistance of Dr Grazziela Figueredo.

4.9 Bioinformatics

Bioinformatics technologies used in this thesis comprise and combine artificial neural network (ANN) modelling of risk and principal and independent components analysis (widely used for population modelling). Regarding survival parameters, models relating to risk using Cox modelling were used. Cross-validation of markers and models using Cox proportional hazards modelling and random forest based decision tree models were

used (356). Bioinformatics methodology will be described in more detail in Chapter 8. Bioinformatics analysis was conducted with the assistance of Professor Graham Ball.

4.10 Summary of methods and next steps of thesis

The methodology of the next steps of this thesis can be summarised as follows:

1) IHC staining of panel of 31 biomarkers in all available CNB and SE specimens in the older women cohort
2) H-scoring of IHC staining
3) Comparison of biomarker staining to standard clinico-pathological features
4) Biological cluster analysis of cohort
5) Using bioinformatics principles to explore the feasibility of development of a prognostic algorithm in this cohort.

4.11 Change in methodology and updated thesis plan

Unfortunately, due to issues outside of my control, a change in the methodology following that described above and overall plan of this thesis was required.

During the course of this thesis the Nottingham University Hospitals (NUH) Biobank (where tissue samples and patient data required for this project originated), were instructed by the Human Tissue Authority to suspend the service and manually audit all samples in the collection. This had an immediate impact on all tissue based research at NUH and the University of Nottingham (UoN); this is a much wider issue between the two organisations.

In February 2020 I was asked to return the database relating to the older women series described in this thesis, for the data to be re-anonymised and to make a new application for access to this dataset when the new standard operating procedure for this, was in

place. As a result of this, I have been unable to link any new IHC data generated during this thesis, to the clinical database.

A Joint Data Access committee has since been established between NUH and UoN and a new data access request was submitted on 21st April 2021. Unfortunately the process for managing and transferring of data by the committee has not been finalised at the time of writing (and not by the time of submitting amendments to this thesis) and therefore, linking of the results from IHC of 16 biomarkers stained as part of this thesis, to the correlating clinical data, has not been possible. The escalation of this issue was made by my lead supervisor to the School including the Dean.

This is a major issue for NUH and UoN to resolve, which has halted all research relating to patient tissue samples currently at both institutions (not just in the field of breast cancer).

A short time after returning the database to NUH/UoN, I became aware this was going to be an ongoing issue and I escalated this initially on a local level within my laboratory research group and my supervisors. This was then escalated to the Pathology team at NUh and the Nottingham Breast Cancer Research Centre whose members from several research groups were also impacted. Following this, my main supervisor escalated the issue to the Dean of the School of Medicine. I was kept informed at all stages.

On 4th August 2021, after consultation with my supervisors and the Dean of the School of Medicine, we made the decision to amend the thesis plan as there was still no access to the required clinical data at this time. I was directly involved in the decision as to when this should occur and also how the thesis plan would be changed.

Therefore, for the rest of this thesis I will refer to clinical data that was made available to me prior to February 2020. This includes data previously described in Chapter 1 from work conducted by two former PhD students which has achieved measurement of 25 biomarkers in all available SE TMAs (3) and 17 biomarkers in all available CNB TMAs (357). From these two pieces of work, a total of 15 biomarkers have been measured in both the SE and CNB TMAs (Table 9).

Biomarker		
ER	PR	HER2
HER3	HER4	EGFR
VEGF	P53	BRCA1
LKB1	Ki67	BCL2
CK5/6	CK7/8	Muc1

Table 9: Panel of 15 biomarkers with clinical data available

Abbreviations: BCL – B-cell lymphoma; BRCA – breast cancer gene; CK – cytokeratin; ER – oestrogen receptor; EGFR – epidermal growth factor receptor; HER – human epidermal growth factor receptor; LK – liver kinase; MUC – mucin; PR – progesterone receptor; VEGF – vascular endothelial growth factor

Association of these 15 biomarkers to standard clinico-pathological features of the dataset, as well as cluster analysis, has been performed and published for all available SE samples (3) and for ER-positive CNB TMA samples (113).

The next steps of this thesis will therefore now be as follows:

1) IHC staining to complete a panel of 31 biomarkers in all available CNB and SE specimens in the older women cohort (this includes staining of all markers

performed in this thesis, for which a link to clinical data was not available – see section 4.11)

2) H-scoring of IHC staining (31 biomarkers – Table 8) (this includes staining of all markers performed in this thesis, for which a link to clinical data was not available – see section 4.11)
3) Comparing biomarker expression in CNB and SE samples to determine if the samples are comparable (15 biomarkers – Table 9)
4) Biological cluster analysis of panel of 17 biomarkers measured in all available CNB TMA samples (irrespective of ER status – Table 10)
5) Using bioinformatics principles to explore the feasibility of development of a prognostic algorithm in this cohort (using 15 biomarkers – Table 9)

Chapter 5 of these thesis will address points 1) and 2); chapter 6 will address point 4); chapter 7 will address point 4); chapters 8 and 9 will explore point 5).

RESULTS AND DISCUSSION

Chapter 5:

Immunohistochemistry

and H-scoring

5.1 Introduction

The next step in this work was to complete IHC staining of the panel of 31 biomarkers and H scoring of these.

This chapter will firstly give a brief overview of the patient characteristics from the database which have previously been published (112), to give context to the results. Next, the results of IHC staining for the remaining biomarkers in the SE TMAs and then the CNB TMAs, conducted as part of this thesis, will be presented. Example images are presented at 5x magnification. Following this, the results of H-scoring for these biomarkers will be given and a comparison made to previous work in the same research group.

As previously stated, all TMAs were constructed prior to commencement of this present thesis. An indication as to which of the panel of 31 biomarkers has been measured prior to this thesis is given in Table 10.

Biomarker	**SE TMA**	**CNB TMA**
ER	X	X
PR	X	X
AR	-	-
EGFR	X	X
HER2	X	X
HER3	X	X
HER4	X	X
VEGF	X	X

P53	X	X
BRCA1	X	X
BRCA2	X	-
MDM2	X	-
MDM4	X	-
PTEN	-	X
LKB1	X	X
Ki67	X	X
BCL2	X	X
E-cadherin	X	-
CD44	X	-
CK5	X	-
CK5/6	X	X
CK7/8	X	X
CK14	X	-
CK17	X	-
CK18	X	-
CK19	X	-
Muc1	X	X
AIB1	-	X
SLC1A5	-	-
SLC7A5	-	-
SLC3A2	-	-

Table 10: An indication of which of the panel of 31 biomarkers had been measured prior to this thesis. X indicates marker measured prior to this thesis; - indicates marker has not yet been measured.

Abbreviations: AIB – amplified in breast cancer; AR – androgen receptor; BCL – B-cell lymphoma; BRCA – breast cancer gene; CD – cluster of differentiation; CK – cytokeratin; CNB – core needle biopsy; ER – oestrogen receptor; EGFR – epidermal growth factor receptor; HER – human epidermal growth factor receptor; LK – liver kinase; MDM - mouse double minute homolog, MUC – mucin; PR – progesterone receptor; PTEN – phosphatase and tensin homolog; SE – surgical excision; SLC – solute carrier; TMA – tissue microarray; VEGF – vascular endothelial growth factor

After receiving training in the methodology and practice of performing IHC (described in Chapter 4), I went on to complete IHC for the remaining 17 biomarkers (six in the SE TMAs and eleven in the CNB TMAs) outstanding in Table 10.

5.1.1 Patient characteristics

Table 11 shows the general characteristics of the older women cohort (N=1758) as described previously (112).

Characteristic	N=1758
Age	Median 77 years (range 70 – 99)
ER status	N = 1380: Positive = 1128 (81.4%); Negative = 252 (18.3%)
5-year BCSS	Decade (average rate per annum %) 1973 - 1999 = 81%; 2000 - 2010 = 91%
5-year OS	Decade (average rate per annum %) 1973 - 1999 = 56%; 2000 - 2010 = 71%
Treatment	Surgery 866 (49.1%); PET 855 (48.8%); Primary radiotherapy 22 (1.2%); No treatment 15 (0.9%)

Table 11: General characteristics of the older women cohort

5. 2 Immunohistochemistry (IHC)

The majority of the biomarkers in Table 10 have been used extensively previously by the research team in Nottingham. Therefore, where possible the same antibody clones and manufacturers were used for this thesis. If the antibody clone used was no longer available from the same manufacturer, then the same clone from another manufacturer in the UK was used. If the antibody was no longer made in the UK, then a new antibody was selected taking into account the following:

- Researching current diagnostic clones in use in the UK
- Investigating publications referring to these clones, particularly in journals of high impact
- Prioritising mouse as the chosen host
- Prioritising monoclonal clonality

5.2.1 Biomarkers measured in SE TMA

The IHC results for the remaining biomarkers from Table 10, measured in the SE TMAs will now be presented.

5.2.1.1 AIB1

Prior to this thesis, AIB1 has been stained by IHC in CNB TMAs from older women with breast cancer by the Nottingham research group.

Optimisation

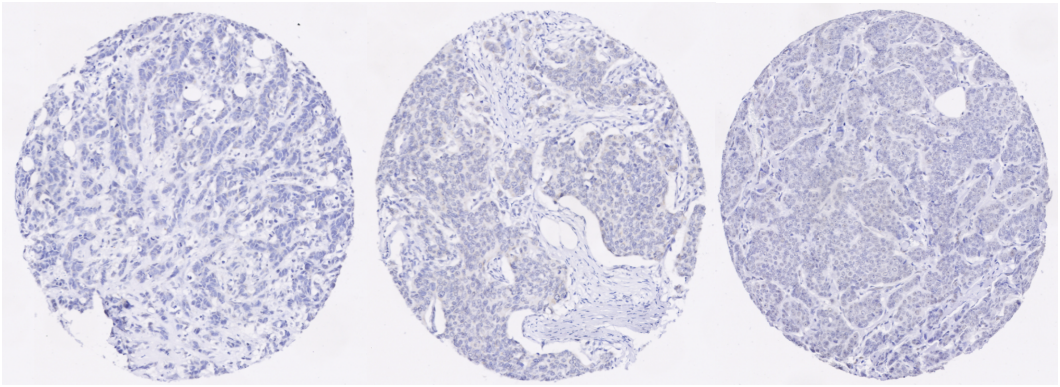
The same antibody product code from the same manufacturer was available (Table 8). Previous optimisation suggested a dilution of 1:75, using retrieval conditions of citrate buffer at pH 6.0 with overnight incubation.

Optimisation in SE TMAs from older women was performed using the following dilutions: 1:25, 1:50, 1:75, 1:100. 1:200.

A dilution of 1:75 using the above conditions was agreed as the optimum staining method.

Staining

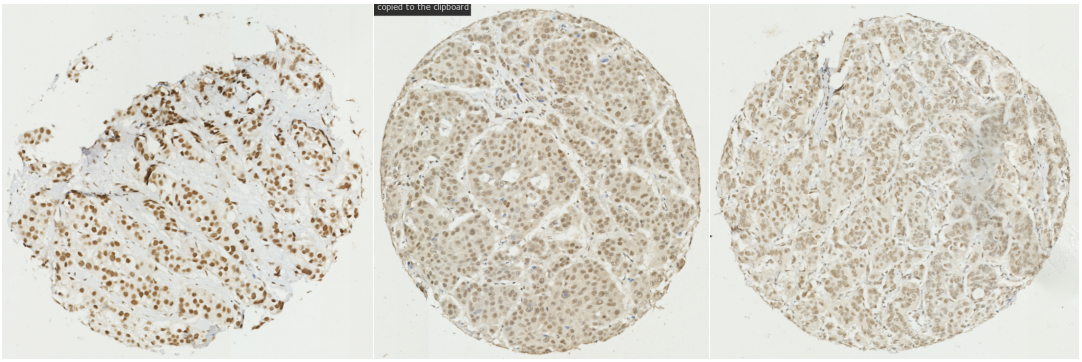
Staining was attempted using the above conditions. On the first attempt there did not seem to be any staining in any of the sections. Please see Figure 17 for representative samples.



Strong staining Moderate staining Mild staining

Figure 17: AIB IHC staining attempt 1 in SE TMAs

The second attempt at IHC for AIB1 using the above conditions was successful with differing degrees of nuclear staining apparent (Figure 18) as expected.



Strong staining Moderate staining Mild staining

Figure 18: AIB IHC staining attempt 2 in SE TMAs

5.2.1.2 PTEN

PTEN has been previously stained by IHC in CNB TMAs from older women with breast cancer in the research group.

Optimisation

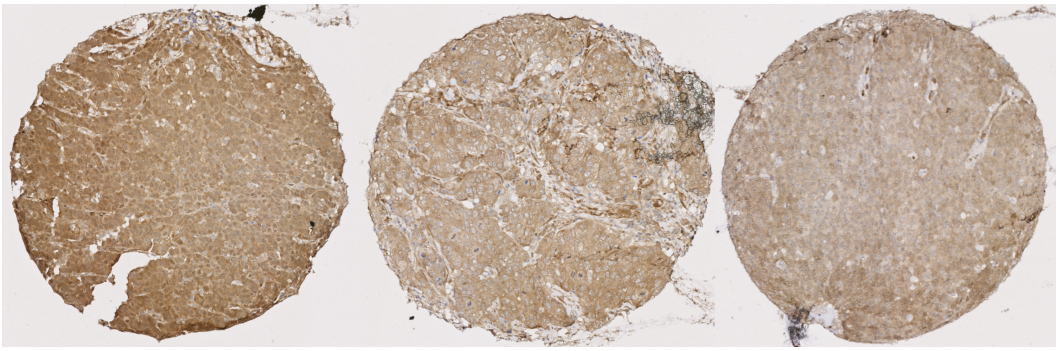
The same antibody product code from the same manufacturer was available. Previous optimisation suggested a dilution of 1:25, using retrieval conditions of water bath at pH 9.0 with an incubation time of one hour.

Optimisation was checked in SE TMAs from older women with breast cancer using the following dilutions: 1:20, 1:25, 1:30, 1:50.

A dilution of 1:25 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining (Figure 19); however, a significant loss of cores was noted in all sections.



Strong staining Moderate staining Mild staining

Figure 19: PTEN IHC staining in SE TMAs

5.2.1.3 AR

AR has been previously stained by IHC in SE TMAs from older women with breast cancer by the research group, however, the stained sections and scores could not be located at the time of this thesis and therefore, this was repeated.

Optimisation

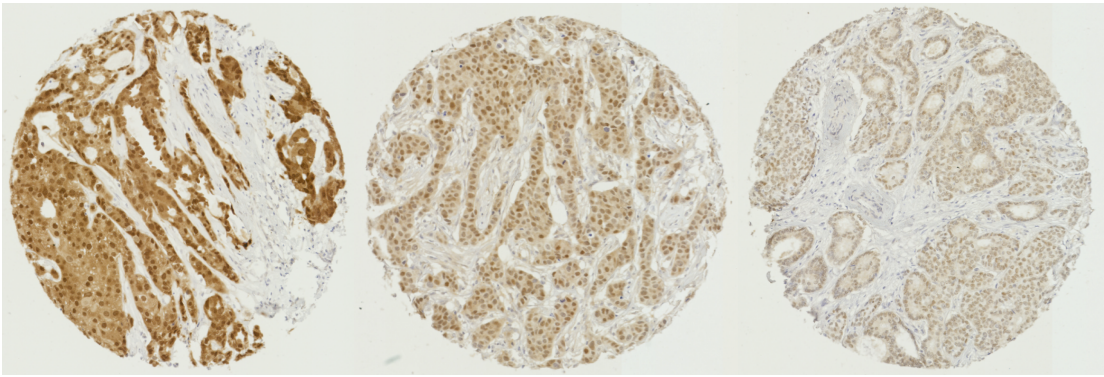
The same antibody product code from the same manufacturer was available. Previous optimisation suggested a dilution of 1:50, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of 30 minutes.

Optimisation was checked in SE TMAs and CNB TMAs from older women with breast cancer using the following dilutions: 1:25, 1:50, 1:75.

A dilution of 1:50 using the above conditions was agreed as the optimum staining method.

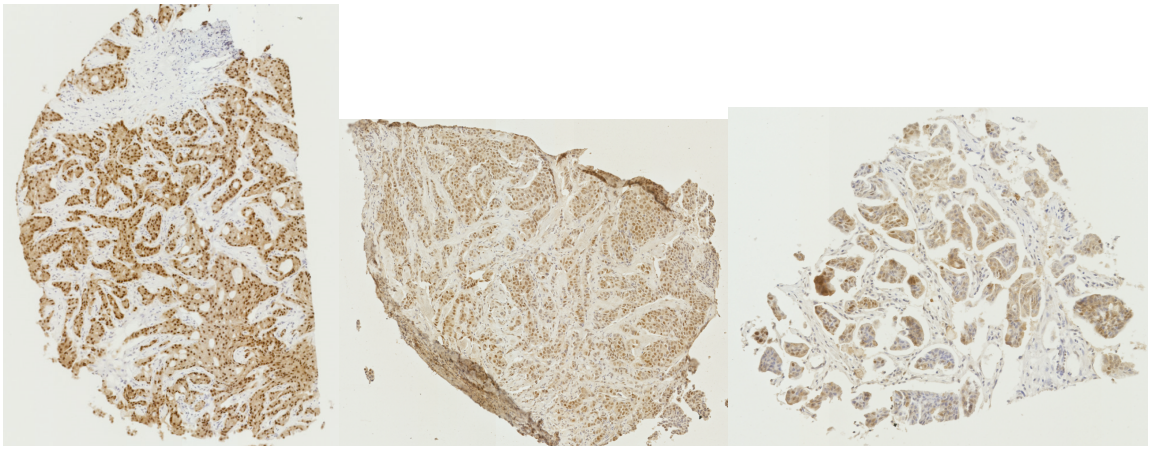
Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of nuclear staining in both SE TMAs (Figure 20) and CNB TMAs (Figure 21).



Strong staining Moderate staining Mild staining

Figure 20: AR IHC staining in SE TMAs



Strong staining Moderate staining Mild staining

Figure 21: AR IHC staining in CNB TMAs

5.2.1.4 Solute carrier family 1 member 5 (SLC1A5)

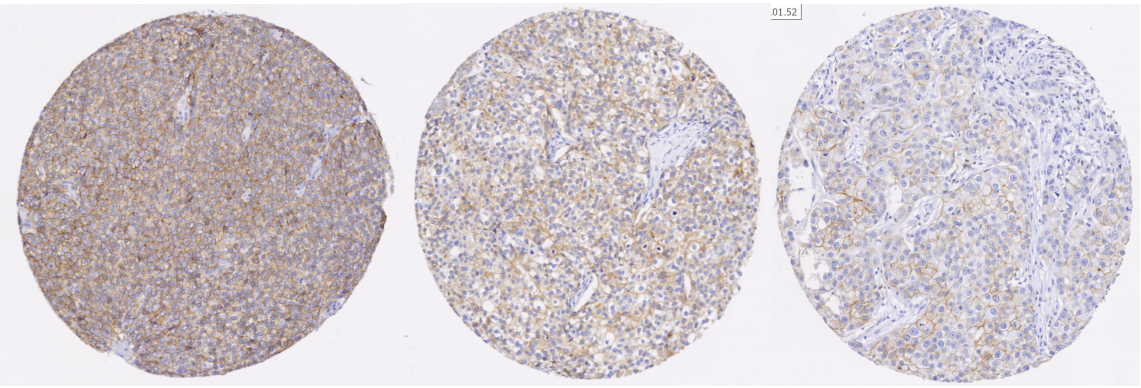
SLC1A5 has been previously stained by IHC in SE TMAs from younger women with breast cancer, by the research group.

Optimisation

The same antibody product code from the same manufacturer was available. Recent optimisation suggested a dilution of 1:200, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of 30 minutes; therefore re-optimisation was not performed. A dilution of 1:200 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of membranous staining in SE TMAs (Figure 22).



Strong staining Moderate staining Mild staining

Figure 22: SLC1A5 IHC staining in SE TMAs

5.2.1.5 SLC3A2

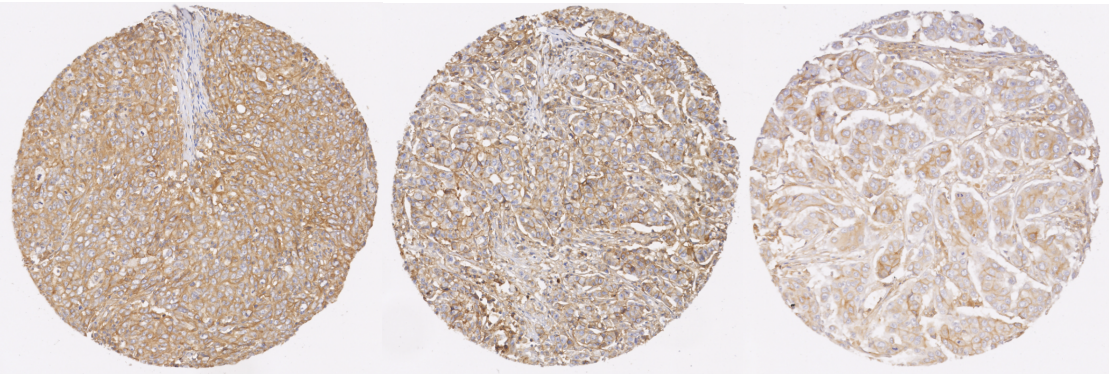
SLC3A2 has been previously stained by IHC in SE TMAs from younger women with breast cancer.

Optimisation

The same antibody product code from the same manufacturer was available. Recent optimisation suggested a dilution of 1:1000, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of one hour; therefore re-optimisation was not performed. A dilution of 1:1000 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of membranous staining in SE TMAs (Figure 23).



Strong staining Moderate staining Mild staining

Figure 23: SLC3A2 IHC staining in SE TMAs

5.2.1.6 SLC7A5

SLC7A5 has been previously stained by IHC in SE TMAs from younger women with breast cancer.

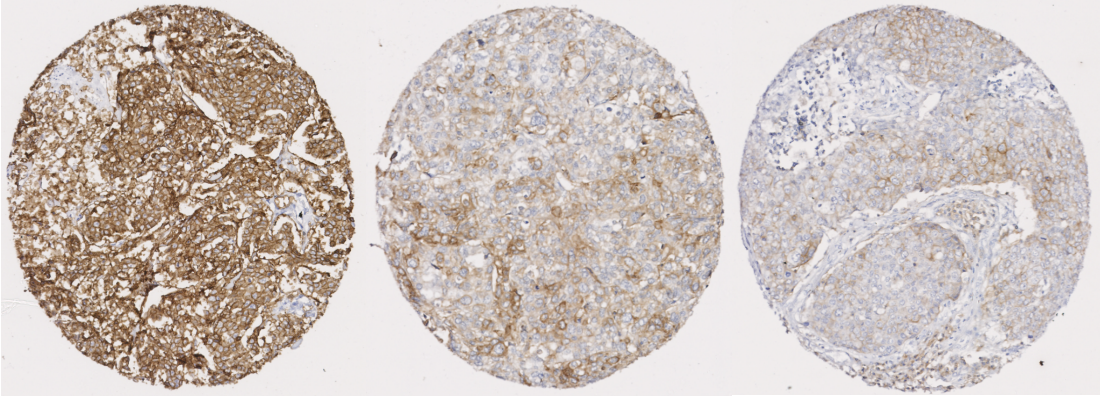
Optimisation

The same antibody product code from the same manufacturer was available. Recent optimisation suggested a dilution of 1:50, using retrieval conditions of citrate bath at pH 6.0 with overnight incubation time; therefore re-optimisation was not performed.

A dilution of 1:50 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of membranous staining in SE TMAs (Figure 24).



Strong staining Moderate staining Mild staining

Figure 24: SLC7A5 IHC staining in SE TMAs

5.2.2 Biomarkers measured in CNB TMAs

The IHC results for the remaining biomarkers from Table 10 measured in the CNB TMAs will now be presented. Staining of AR in CNB TMAs has been described already for ease (Figure 21).

5.2.2.1 BRCA2

BRCA2 has been previously stained by IHC in SE TMAs from older women with breast cancer by the research group.

Optimisation

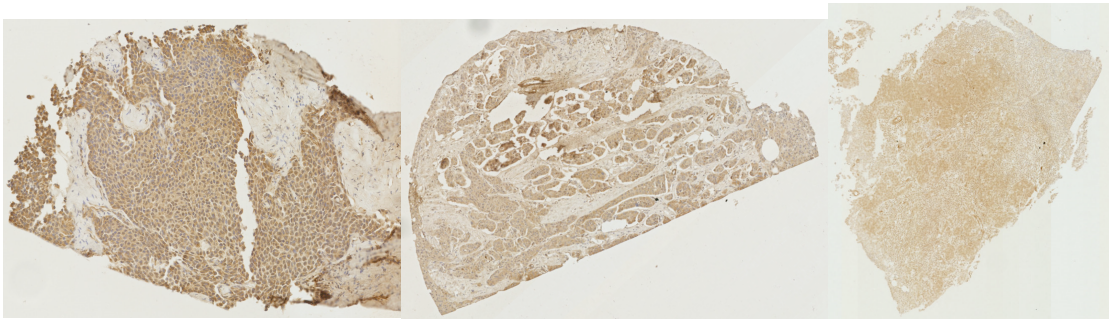
The antibody clone used for staining BRCA2 was no longer available from the same manufacturer or any other manufacturer in the UK therefore, a new antibody was selected.

The datasheet provided with the antibody suggested a dilution of between 1:50 and 1:200. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions, using citrate buffer and incubation time of one hour: 1:50, 1:100. 1:150, 1:200. Staining was very poor, therefore, repeat optimisation using the following

dilutions and overnight incubation was performed: 1:25, 1:50, 1:100. This provided a good staining result, but it was difficult to differentiate between mild and moderate staining. Therefore, a third round of optimisation was performed using the following dilutions and overnight incubation: 1:150, 1:200. Following this, a dilution of 1:200 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 25).



Strong staining Moderate staining Mild staining

Figure 25: BRCA2 IHC staining in CNB TMAs

5.2.2.2 CD44

CD44 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

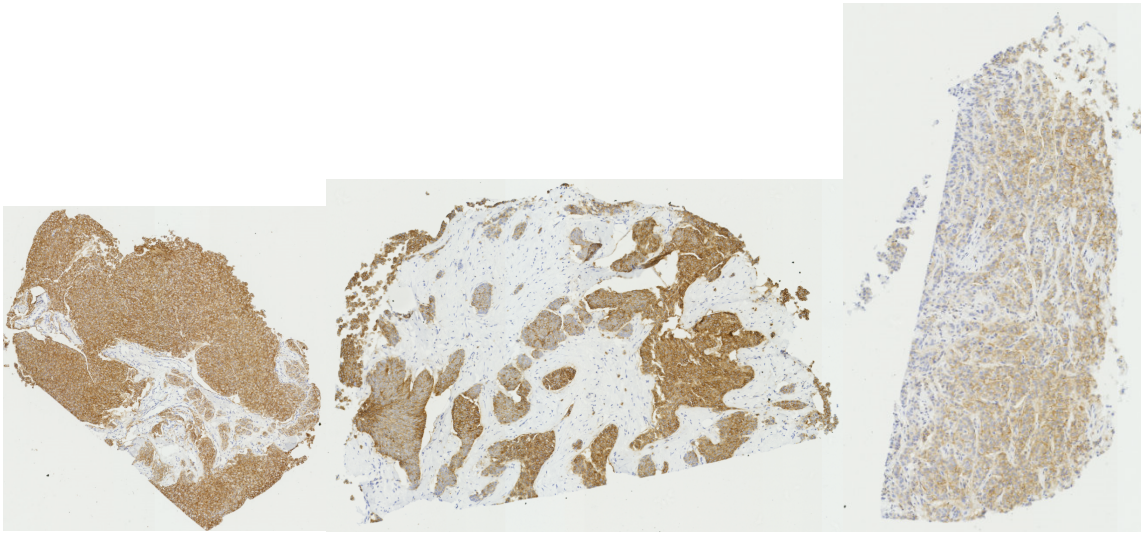
The same antibody product code from the same manufacturer was available. Previous optimisation suggested a dilution of 1:50, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of 45 minutes.

Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:25, 1:40, 1:50, 1:60.

A dilution of 1:50 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of membranous staining in CNB TMAs (Figure 26).



Strong staining Moderate staining Mild staining

Figure 26: CD44 IHC staining in CNB TMAs

5.2.2.3 E cadherin

E cadherin has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

The same antibody product code from the same manufacturer was available. Previous

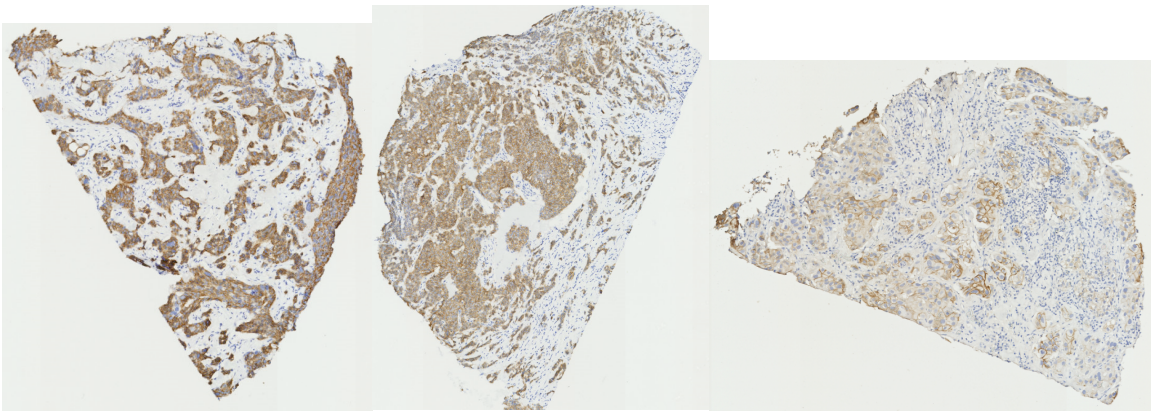
optimisation suggested a dilution of 1:100, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of 45 minutes.

Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:50, 1:75, 1:100, 1:125.

A dilution of 1:100 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of membranous staining in CNB TMAs (Figure 27).



Strong staining Moderate staining Mild staining

Figure 27: E cadherin IHC staining in CNB TMAs

5.2.2.4 MDM2

MDM 2 was previously stained by IHC in SE TMAs from older women with breast cancer.

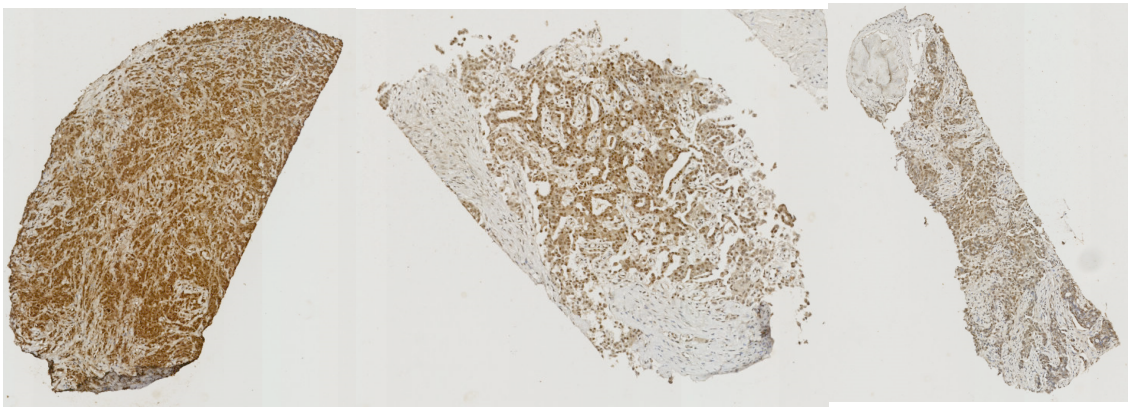
Optimisation

The previous antibody used was no longer available for purchase either by the same or any other company, therefore a different antibody was used.

Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:25, 1:50, 1:75, 1:100, 1:125 initially using EDTA buffer with primary antibody incubation time of 60 minutes. After EDTA buffer wash, the majority of cores had disappeared from the specimen and therefore the exercise was repeated with citrate buffer which resulted in good staining. EDTA buffer was initially used as this was described in the referenced literature given in the manufacturer's datasheet. A dilution of 1:125 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 28).



Strong staining Moderate staining Mild staining

Figure 28: MDM2 IHC staining in CNB TMAs

5.2.2.5 MDM4

MDM 4 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

The previous antibody used was no longer available for purchase either by the same or any other company, therefore a different antibody was used.

Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:50, 1:100, 1:150, 1:200 using citrate buffer. Incubation time of primary antibody was 60 minutes.

A dilution of 1:200 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was unsuccessful with generally poor staining, likely due to error in methodology. Previous work in the research team (166) of MDM4 and MDM2 staining in SE TMAs in older women, found that 100% of the samples were positive for MDM2 and 99.5% of the samples for MDM4. As the majority of samples were positive for both MDM2 and MDM4 expression, it was hypothesised that the findings in this present study would be similar and therefore, there was little to be gained by repeating staining for MDM4 expression at the expense of using more TMA samples.

5.2.2.6 CK5

CK5 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

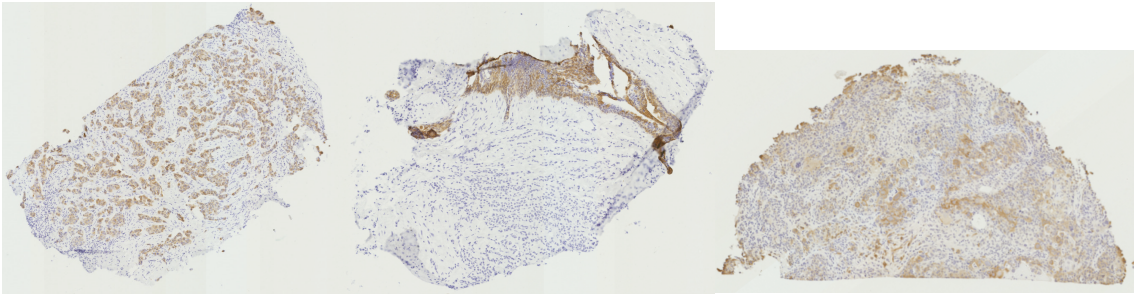
The antibody clone used for this staining was no longer available from the same manufacturer or any other manufacturer in the UK, therefore a new antibody was purchased.

Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:25, 1:50, 1:75, 1:100, 1:125 in a citrate buffer with primary antibody time of 45 minutes.

A dilution of 1:150 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 29).



Strong staining Moderate staining Mild staining

Figure 29: CK5 IHC staining in CNB TMAs

5.2.2.7 CK14

CK14 has been previously stained by IHC in SE TMAs from older women with breast cancer.

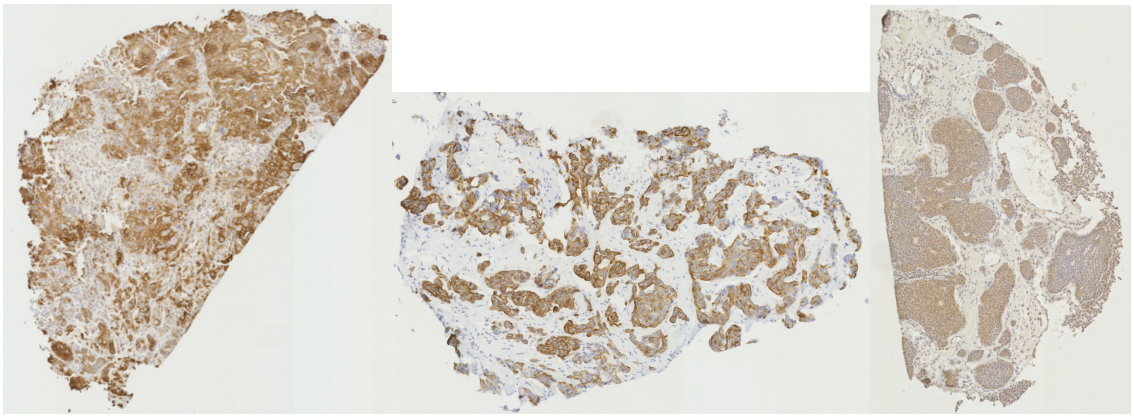
Optimisation

The antibody clone used for this staining was no longer available from the same manufacturer or any manufacturer in the UK, therefore a new antibody was purchased. Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:20, 1:40, 1:60 in a citrate buffer with primary antibody time of 30 minutes.

A dilution of 1:40 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 30).



Strong staining Moderate staining Mild staining

Figure 30: CK14 IHC staining in CNB TMAs

5.2.2.8 CK17

CK17 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

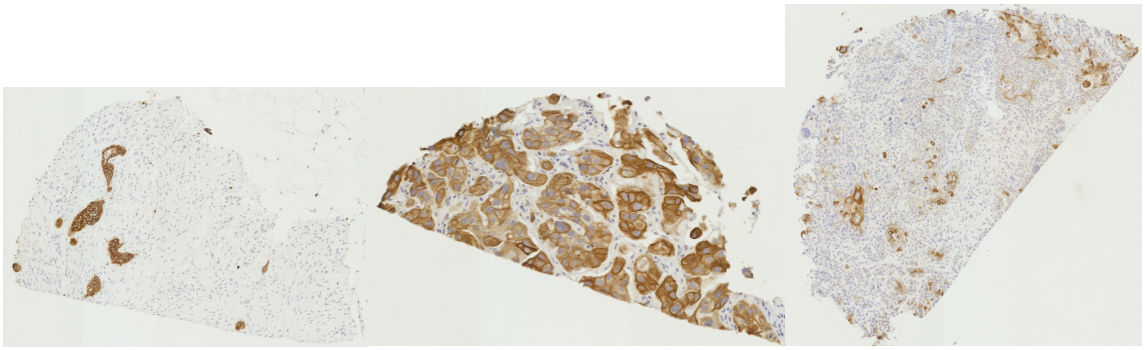
The antibody clone used for this staining was no longer available from the same manufacturer or any other manufacturer in the UK, therefore a new antibody was purchased.

Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:50, 1:75, 1:100, 1:150 in citrate buffer with primary antibody time of 60 minutes.

A dilution of 1:75 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 31).



Strong staining Moderate staining Mild staining

Figure 31: CK17 IHC staining in CNB TMAs

5.2.2.9 CK18

CK18 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

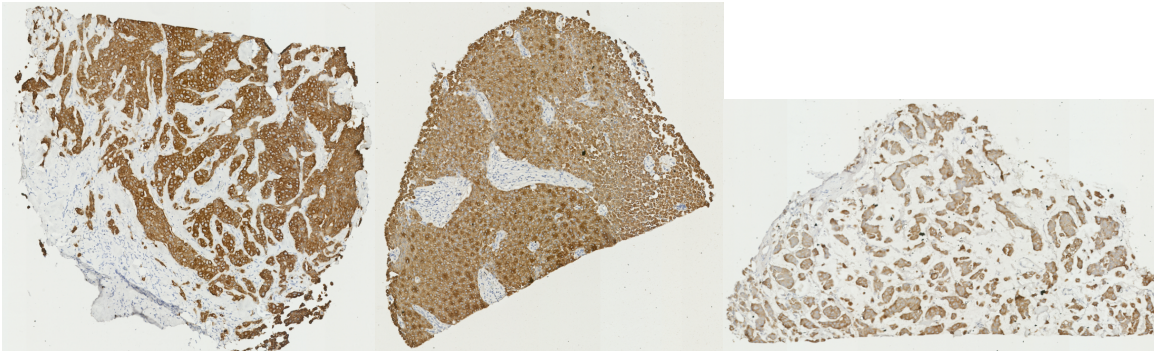
The same antibody product code from the same manufacturer was available. Previous optimisation suggested a dilution of 1:100, using retrieval conditions of citrate bath at pH 6.0 with an incubation time of 45 minutes.

Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1: 25, 1:50, 1:75, 1:100, 1:125. Staining was very strong at all dilutions with difficulty differentiating between high and low expression, therefore repeat optimisation was performed using the following dilutions: 1:125, 1:150, 1:200.

A dilution of 1:150 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 32).



Strong staining Moderate staining Mild staining

Figure 32: CK18 IHC staining in CNB TMAs

5.2.2.10 CK19

CK19 has been previously stained by IHC in SE TMAs from older women with breast cancer.

Optimisation

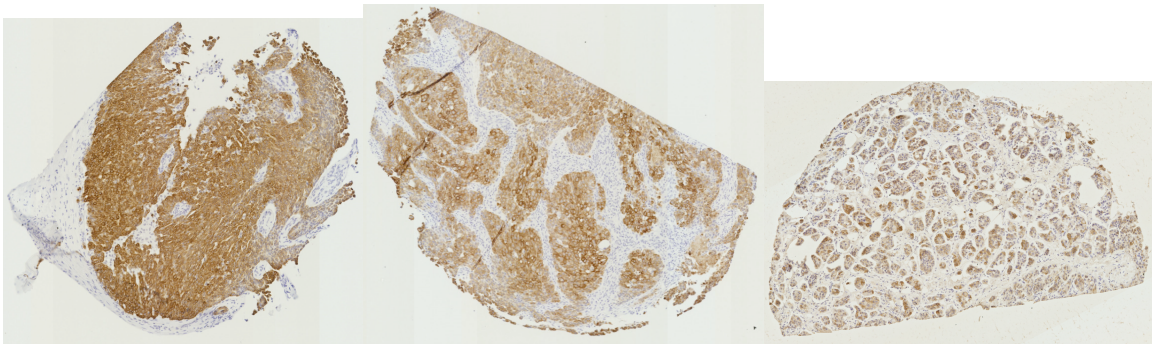
The antibody clone used for this staining was no longer available from the same manufacturer or any other manufacturer in the UK, therefore a new antibody was purchased.

Optimisation and staining conditions were determined using the manufacturer's datasheet as a guide and current literature. Optimisation was checked in CNB TMAs from older women with breast cancer using the following dilutions: 1:50, 1:75, 1:100 in a citrate buffer with primary antibody time of 30 minutes.

A dilution of 1:100 using the above conditions was agreed as the optimum staining method.

Staining

Staining was attempted using the above conditions. The first attempt was successful with a good variation of cytoplasmic staining in CNB TMAs (Figure 33).



Strong staining Moderate staining Mild staining

Figure 33: CK17 IHC staining in CNB TMAs

5.2.3 Comparison of IHC staining to previous work and literature

The majority of biomarkers in the panel have not been extensively measured in primary breast cancer specimens in older women, outside of my research group.

Overall, IHC staining conducted as part of this thesis was successful with the exception of MDM4 which is likely to have been due to human error. An excessive loss of cores was noted during the staining of PTEN in all sections of SE TMA. Of note, this was the only biomarker for which antigen retrieval required a water bath (instead of citrate bath); it is possible that an excess of cores were lost during this stage, however this had not been noted when previously stained by the research team in CNB TMAs (113) and the same standard operating procedure was used.

Previous measurement of a panel of biomarkers in the same cohort of SE TMAs achieved scores for between 401 – 575 cases for each biomarker measured (166). This contrasts with between 184 – 415 cases in the present study.

Previous measurement of a panel of biomarkers in the same cohort of CNB TMAs achieved scores for between 400 – 536 cases for each marker (167). This contrasts with between 323 – 648 cases in the present study. This is a larger range than anticipated, but on average produces a similar number of cases for each biomarker.

There are a number of potential reasons for these discrepancies. The SE TMAs were constructed between 2009 – 2012 and the CNB TMAs between 2013 – 2016. It is possible that some of the TMAs have degraded over time. Repetitive new sectioning of the TMA block is likely to result in loss of tissue material, due to facing the block in the microtome. Furthermore, the uneven thickness of the different samples constituting the TMA block makes it more sensitive to trimming than a block with only a few tissue

pieces (358). The optimal technique would be to section as many samples as possible from the block at once; however, preservation of sectioned slides for a longer period of time is challenging in itself (359).

There is limited evidence in the literature to support construction of TMAs from CNB; the main use of CNB is for diagnostic purposes and as such various methodologies have been described in the literature (349). The method used in this present study is unique (360), therefore, there is no current data on how the CNB TMA samples will fair over time in terms of maintaining integrity and cell structure.

In the present study IHC has been performed by a different researcher than those from the previous studies described (166, 167), in a different laboratory. Although the same methodology was followed, there may be discrepancies in measurements due to different instruments used and also variation between researchers in both techniques for IHC and H-scoring (361).

All antibodies previously used by the research team had reached their expiry date and therefore, new antibodies were purchased. As far as possible, the same clone was purchased from the same manufacturer (Table 8), however, this was not possible for all biomarkers in the panel. Therefore, there may be some variation for those markers which used different clones (CK5, CK14, CK17, CK 19, BRCA2). Furthermore, even for the biomarkers with the same clones, there may be inter-batch variation, especially several years apart (362, 363).

5.3 H-scoring of biomarkers stained by IHC for this thesis

H-scoring was performed as outlined in Chapter 4 for all the markers stained above in section 5.2. All biomarkers were scored three times by the primary reviewer and by a second independent reviewer. Inter- and intra-rater concordance was checked.

5.3.1 H-score results

Table 12 shows the range of H-score expression of biological markers measured in the SE TMAs and Table 13 the biomarkers measured in the CNB TMAs.

For each marker, the total number of cores possible to score is given (Total cases) as well as the mean, median, range and interquartile range (IQR) of H-scores. The intraclass correlation (ICC) coefficient is given to show variation between scores for the same biomarker by the primary reviewer (RP, intra-rater) and also to demonstrate variation between the primary and second reviewer (inter-rater). An ICC >0.7 is considered a strong correlation and was deemed acceptable.

Bio-marker	Total cases	Mean H-score	Median H-score	Range H-score	IQR H-score	Intra-rater ICC	Inter-rater ICC
AR	278	159	201	0 - 300	2 - 263	0.98 (CI 0.98 – 0.99)	0.81 (CI 0.73 – 0.86)
PTEN	184	101	83	0 - 293	13 - 178	0.98 (CI 0.97 – 0.98)	0.78 (CI 0.45 – 0.91)
AIB1	206	70	7	0 - 297	0 - 138	0.96 (CI 0.95 – 0.97)	0.68 (CI 0.53 – 0.78)
SLC1A5	415	20	1	0 - 212	0 - 22	0.96 (CI 0.95 – 0.97)	0.87 (CI 0.82 – 0.93)
SLC3A2	355	59	0	0 – 300	0 - 89	0.96 (CI 0.95 – 0.97)	0.86 (CI 0.79 – 0.90)
SLC7A5	351	16	0	0 - 295	0 - 1	0.95 (CI 0.94 – 0.96)	0.75 (CI 0.34 – 0.88)

Table 12: The expression of biological markers in the SE TMAs

Abbreviations: AIB – amplified in breast cancer; AR – androgen receptor; PTEN – phosphatase and tensin homolog; SLC – solute carrier.

Bio-marker	**Total cases**	**Mean H-score**	**Median H-score**	**Range H-score**	**IQR H-score**	**Intra-rater ICC**	**Inter-rater ICC**
AR	523	201	240	0 - 300	144 - 286	0.98 (CI 0.97 – 0.98)	0.81 (CI 0.73 – 0.85)
BRCA2	523	38	0	0 - 293	0 - 46	0.98 (CI 0.95 – 0.96)	0.86 (CI 0.80 – 0.91)
MDM2	478	173	190	0 - 300	116 - 233	0.89 (CI 0.87 – 0.90)	0.90 (CI 0.83 – 0.94)
E-cadherin	592	142	150	0 - 300	4 - 244	0.96 (CI 0.96 – 0.97)	0.82 (CI 0.71 – 0.89)
CD44	644	33	0	0 - 300	0 – 25	0.97 (CI 0.95 – 0.96)	0.91 (CI 0.87 – 0.94)
CK5	561	4	0	0 – 297	0 - 0	0.97 (CI 0.96 – 0.97)	0.89 (CI 0.84 – 0.92)
CK14	424	5	0	0 – 300	0 – 0	0.99 (CI 0.99 – 0.99)	0.82 (CI 0.71 – 0.88)
CK17	494	5	0	0 – 300	0 – 0	0.98 (CI 0.98 – 0.99)	0.79 (CI 0.68 – 0.86)
CK18	648	114	93	0 - 300	0 – 214	0.97 (CI 0.96 – 0.97)	0.71 (CI 0.67 – 0.84)
CK19	548	167	198	0 – 300	38 – 272	0.98 (CI 0.97 – 0.98)	0.92 (CI 0.87 – 0.95)

Table 13: The expression of biological markers in the CNB TMAs

Abbreviations: AR – androgen receptor; BRCA – breast cancer gene; CK – cytokeratin; MDM - mouse double minute homolog.

5.3.2 Comparison of H scores to the literature

For the reasons outlined in section 4.11 of this thesis, it was not possible to correlate the H-scores for the biomarkers measured in this present thesis, to the associated clinical data. As high/low or positive/negative cut-offs are generated with a specific clinical feature in mind (such as BCSS), the cut-offs are unique to each biomarker for both SE and CNB TMAs.

Table 14 and 15 show the number of cases of positive/negatives using values taken from the measurement of these markers by the Nottingham research group in younger women or in older women but in the converse specimen type (SE or CNB).

Biomarker	Cohort, definition of cut-off originates	H-score definition of 'positive' expression	% of +ve cases suggested	% of +ve cases in present work	% of -ve cases suggested	% of -ve cases in present work
AIB1	Older women, CNB (167)	>80	46%	34%	54%	66%
PTEN	Older women, CNB (167)	>180	15%	25%	85%	75%
AR	Older women, SE (364)*	>40	59%	70%	41%	30%
SLC1A5	Younger women, SE (171)	>40	63%	17%	37%	83%
SLC3A2	Younger women, SE (171)	>15	50%	36%	50%	64%
SLC7A5	Younger women, SE (171)	>15	50%	17%	50%	83%

Table 14: The number of positive/negative samples of biological markers in the SE TMAs based on prior definitions

*Staining of this marker was repeated as results of previous staining could not be found.

Using the definitions given in Table 14, the number of cases positive/negative for AIB1, PTEN and AR are similar to previous work by the Nottingham research group and within 12% difference. The number of cases positive/negative for SLCs, particularly SLC1A5

and SLC7A5 were markedly different, with a preponderance towards negativity for all three SLCs in this current thesis, compared to previous work (171). The previous measurement of SLCs by El-Ansari et al (171) was conducted in primary breast cancer samples in younger women (<70 years alone). As breast cancer in older women is generally likely to be more slow-growing and less proliferative compared with their younger counterparts, one hypothesis for the difference between El-Ansari's work and the present thesis is that breast cancer in older women may therefore be less reliant on SLCs for growth, contributing to the higher rate of negative scores overall. There is presently, nil else in the literature to support or deny this finding.

Biomarker (reference for definition of 'positive')	**Cohort, definition of cut-off taken from**	**H-score definition of 'positive' expression**	**% of +ve cases suggested**	**% of +ve cases in present work**	**% of -ve cases suggested**	**% of -ve cases in present work**
AR	Older women, SE (364)	>40	59%	85%	41%	15%
BRCA2	Older women, SE (166)	>0	78.5%	36%	21.5%	64%
CD44	Older women, SE (166)	>0	21.8%	46%	78.2%	54%
E-cadherin	Older women, SE (166)	>50	61.7%	69%	38.3%	31%
MDM2	Older women, SE (166)	>0	100%	95%	0%	5%
CK5	Older women, SE (166)	>0	29.9%	9%	70.1%	91%
CK14	Older women, SE (166)	>0	24%	11%	76%	89%
CK17	Older women, SE (166)	>0	20%	18%	80%	82%

CK18	Older women, SE (166)	>0	96.2%	67%	3.8%	33%
CK19	Older women, SE (166)	>0	95.5%	80%	4.5%	20%

Table 15: The number of positive/negative samples of biological markers in the CNB TMAs based on prior definitions

All of the definitions used to differentiate positive/negative staining in the CNB TMAs (Table 15) were taken from work in SE TMAs. These markers have not been extensively explored in CNB samples or in large populations of older women in the literature; it is difficult to make direct comparisons. Therefore, variation can be expected. Variation between SE and CNB TMAs will be explored in greater detail in Chapter 6. However, the overall trends appear to be the same for all biomarkers (preponderance towards positive or negative for each marker), with the exception of AR and BRCA2 which demonstrate the greatest differences. In the present work, 85% of all cases were positive for the expression of AR, compared to only 59% in the previous work and only 36% of cases were positive for expression of BRCA2, compared to 79% in previous work.

There have been conflicting results in the literature on definition of AR positivity as well as its prognostic significance, with cut-offs of ≥1% and ≥10% often being explored simultaneously in the same studies (153, 365) as well as some studies using a different definition of AR positivity in the ER-positive subtype of breast cancer (366). Wide variation in AR positivity depending on ethnicity has been noted (153). Androgen receptor positivity in large cohort studies published in the literature range between 19 - 77% (150, 367-369) but this seems highly dependent on breast cancer subtype, with a

lack of AR seen in TNBC (367) and increased AR seen in ER-positive subtypes (150). Therefore, the increased AR-positivity seen in the present study, is likely due to high numbers of corresponding patients with ER-positive breast cancer, given that this is a cohort of older women. The SE TMA cohort is likely to be biased towards older women with ER-negative breast cancer, who will not have the option of PET as an alternative treatment.

The BRCA2 antibody used in the present thesis was different to the antibody previously used by my research group, as this antibody was no longer manufactured in the UK. Furthermore, the BRCA antibody chosen was polyclonal and from a rabbit host as this was the most suitable antibody found specific for the target, available for research use in the UK. This antibody has been developed and validated by the Human Protein Atlas project for use in IHC (370). Unfortunately, there is little in the published literature yet to support use of this antibody in breast cancer, particularly in the setting of IHC; the predominance of research in BRCA2 is exploring mutations at the genomic level. The different antibody used in the present study, does probably therefore explain the difference in H-score expression seen.

5.4 Summary of this chapter and next steps

A panel of 31 biomarkers have been measured by IHC on all available SE TMAs and CNB TMAs in the cohort of older women with primary breast cancer and a H-score assigned. The similarities and differences in measurements of H-scores in this present study, compared to previous studies by the research team have been explored and the results are generally comparable. Where results are not comparable, this has been explained and is largely due to differences between SE TMA and CNB TMA samples and

differences between younger and older women with breast cancer.

The next step is to formally compare measurement of biomarkers between SE and CNB specimens in the present study. All patients will have CNB at diagnosis of breast cancer, irrespective of treatment decision, therefore, the hypothesis is that measurement of a panel of biomarkers in CNB samples is at least equivalent, if not more accurate, to measurement of the same panel in SE specimens. Where differences between CNB and SE measurements arise, it is important to be able to quantify this difference in order to predict SE result, based on CNB sample alone.

Chapter 6: Core needle biopsy (CNB) versus surgical excision (SE) comparison

6.1 Introduction

From the whole series of constructed TMAs (575 SE TMAs and 693 CNB TMAs) there are 202 cases which have both SE and CNB TMA with follow-up data.

This chapter will compare the measurement of biomarkers on CNB TMAs directly with SE TMAs initially by using basic statistical analysis measures of concordance and then using principal components analysis, in an attempt to potentially predict SE TMA staining result, based on CNB TMA result alone (which every patient will have at diagnosis). An interpretation of these findings and resulting impact on this thesis plan will be given after each analysis.

6.2 Comparison of CNB TMA with SE TMA results using basic statistical analysis

6.2.1 Methods

A total of 15 biomarkers from the overall panel have been measured in both the SE TMA and CNB TMA samples for 202 cases. The markers are: ER, PR, EGFR, HER2, HER3, HER4, VEGF, p53, BRCA1, LKB1, Ki67, BCL2, CK5/6, CK 7/8, MUC1. As well as the continuous H score value for each biomarker (except for HER2 for which a different method was used, see section 4.6.2), expression of biomarker was divided into two groups: 'positive' and 'negative' for BCSS using pre-defined cut-offs, which were determined using X-tile software (section 4.7).

Table 16 presents the cellular localisation and cut-off values for each biomarker taken from previous work in SE (166) and CNB TMAs (167).

Biomarker	Cellular localisation	H-score definition of 'positive' expression for SE TMA	H-score definition of 'positive' expression for CNB TMA
ER	Nucleus	>20	>50
PR	Nucleus	>10	>10
EGFR	Membranous	>0	>0
HER2	Membranous	3+*	3+*
HER3	Membranous	>0	>20
HER4	Membranous	>0	>20
VEGF	Cytoplasm	>0	>5
P53	Nucleus	>5	>140
BRCA1	Nucleus	>60	>148
LKB1	Cytoplasm	>30	>160
Ki67	Nucleus	>30	>25
BCL-2	Cytoplasm	>0	>223
CK 5/6	Cytoplasm	>0	>7
CK 7/8	Cytoplasm	>0	>7
MUC1	Cytoplasm	>30	>190

Table 16: Summary of 15 biomarkers for comparative analysis between CNB TMA and SE TMA

*The scoring system using for HER2 is outlined in Section 4.6.2 of this thesis.

Abbreviations: BCL – B-cell lymphoma; BRCA – breast cancer gene; CK – cytokeratin; CNB- core needle biopsy; ER – oestrogen receptor; EGFR – epidermal growth factor

receptor; HER – human epidermal growth factor receptor; LK – liver kinase; MUC – mucin; PR – progesterone receptor; SE – surgical excision; TMA – tissue microarray; VEGF – vascular endothelial growth factor

Percentage concordance was calculated between 'positive' or 'negative' results. Concordance between the result generated on CNB TMA was compared to the result generated on SE TMA in the same patient using Spearman's correlation. The correlation coefficient (r-value) was interpreted as outlined in Section 4.6.3. Confidence intervals (CIs) were generated using Stata software (version 17, 2021, StataCorp LLC).

6.2.2 Results

Table 17 shows the estimated concordance between positive and negative expression and Table 18 shows the concordance between H-scores.

Biomarker	N		SE	SE +ve	Con %
ER	167	**CNB -ve**	11	26	81%
		CNB +ve	5	125	
PR	172	**CNB -ve**	37	18	84%
		CNB +ve	9	108	
EGFR	162	**CNB -ve**	146	8	93%
		CNB +ve	4	4	
HER2	170	**CNB -ve**	162	2	99%
		CNB +ve	0	6	
HER3	153	**CNB -ve**	2	117	23%
		CNB +ve	1	33	
HER4	167	**CNB -ve**	7	101	38%
		CNB +ve	3	56	
VEGF	123	**CNB -ve**	1	14	80%
		CNB +ve	11	97	
p53	162	**CNB -ve**	20	3	35%
		CNB +ve	103	36	
BRCA1	142	**CNB -ve**	62	1	44%
		CNB +ve	79	0	
LKB1	125	**CNB -ve**	15	69	40%
		CNB +ve	6	35	
Ki67	184	**CNB -ve**	93	38	65%
		CNB +ve	27	26	
BCL2	173	**CNB -ve**	4	3	93%
		CNB +ve	9	157	
CK5/6	179	**CNB -ve**	88	47	61%
		CNB +ve	22	22	
CK7/8	175	**CNB -ve**	0	2	99%
		CNB +ve	0	173	
MUC1	162	**CNB -ve**	2	24	83%
		CNB +ve	3	133	

Table 17: Concordance between CNB TMA and SE TMA based on two categories 'positive' (+ve) or 'negative'(-ve)

Biomarker	**N**	**R value**	**CI**
ER	166	0.353	0.212 – 0.480
PR	168	0.712	0.628 – 0.779
EGFR	162	0.378	0.238 – 0.503
HER2	176	0.483	0.361 – 0.589
HER3	165	0.020	-0.133 – 0.173
HER4	156	0.051	-0.107 – 0.207
VEGF	139	0.037	-0.131 – 0.202
P53	153	0.347	0.199 – 0.479
BRCA1	132	-0.150	-0.313 – 0.022
LKB1	128	0.032	-0.143 – 0.204
Ki67	184	0.250	0.109 – 0.381
BCL2	169	0.561	0.448 – 0.657
CK5/6	166	0.217	0.067 – 0.357
CK7/8	174	0.391	0.258 – 0.510
MUC1	159	0.224	0.071 – 0.367

Table 18: Concordance between CNB TMA and SE TMA based on continuous H-score

6.2.3 Interpretation of results

The percentage concordance rates between positive/negative cut-offs as shown in Table 17 are above 60% for the markers ER, PR, EGFR, HER2, VEGF, p53, Ki67, BCL2, CK 7/8, MUC1, The markers HER3, HER4, CK 5/6, BRCA1 and LKB1 had a percentage concordance of <45%. Ultimately, different definitions for positive/negative were used in the CNB TMA specimens compared to the SE TMA specimens. This was based upon results generated by X-tile software (section 4.7) for each individual data set, suggesting that the H-scores generated for each TMA type did differ. However, it is reassuring to see that the biomarkers which are most frequently studied in breast cancer research (ER, PR, EGFR, HER2, VEGF, p53, Ki67) did show good concordance. This may suggest that for those markers less frequently studied, there is disagreement or lack of guidance with interpretation of staining, leading to greater variation.

However, the concordance between H-scores as a continuous variable between SE TMA and CNB TMA, shows only weak correlation between all markers with the exception of PR and BCL2. In general, cases are more likely to be positive for a biomarker on SE TMA than CNB TMA. This could suggest that the SE TMA result is an overestimation of the true value and that CNB TMA result is more accurate or at least not likely to overestimate the result. Estimating a reference value for the concordance between CNB TMA and SE TMA in terms of biomarker expression is challenging as evidence from previous work is limited.

Current literature suggests that concordance for ER, PR, HER2, BCL2, Ki67 and p53 (371-374) between CNB and SE samples should be strong, however discrepancies between different centres have been noted for PR, HER2 and Ki67 (374-376). There is a

lack of evidence in the literature for comparison between specimen types for the other biomarkers in the present study.

Variations between SE and CNB scores in this present study can be justified by considering multiple factors including: selection of cancer sample for construction of TMA, representation of tissue obtained from the biopsy block, differences between researchers in laboratory processes and scoring, differing antibody clones.

To construct a TMA either from SE or CNB sample, a section of breast cancer is selected from the original specimen. Clearly, in SE specimens this presents a much larger selection of potential tissue to use. One of the limitations of TMA construction is that the punched samples from donor tissue may not always be representative of the entire tumour (377), given the heterogenous nature of breast cancer and there may be selection bias by the researcher constructing the TMA in the way they select the tissue to use from a very small sample (CNB) compared to a larger sample (SE). Obtaining a TMA that accurately represents what would be seen on whole tissue sections is challenging and ultimately can only be accomplished by comparing the staining of a marker on both the TMA and the whole tissue section, which defeats the point of the TMA. One suggestion to overcome this is to perform a number of pilot tests on small numbers of core TMAs before a larger TMA is constructed (378).

There may be differences in representation of tissue obtained from the biopsy TMA block, which includes concern of heterogeneity between cores obtained from different biopsies of each patient, and morphology of the original tissue in the construction process (349).

To date, no factor in specimen handling has been identified that does not impact quality when measured by the most stringent assays (378); despite the best efforts of the researchers, variability between researchers will occur (379). Tissue loss during sectioning, transfer and staining will occur to varying degrees on any given day in a laboratory (379).

The ASCO / College of American Pathologists (CAP) guidelines suggest using H-score to measure tumour hormone receptors due to acceptable interobserver agreement (380) and this method has been widely adopted for other biological markers (381), however the subjectivity of the scoring process can cause poor inter- and intra- rater concordance (382). The future of assessment of IHC staining is likely to be using a digital platform (382, 383) with current research focusing on the application of artificial intelligence in this setting (384).

Different antibody clones were used in the present study, compared to previous work by the same research group, if the previous clone was no longer available on the market. Variability between antibody clones for use in IHC biomarker studies has been noted (385, 386) and standard protocols should be used wherever possible to minimise this.

To further understand the relationship between biological analysis of CNB TMAs compared to SE TMAs, the next step in this work was to explore the relationship between the two variables using principal components analysis (PCA).

6.3 Comparison of CNB TMA with SE TMA results using principal components analysis (PCA)

Work so far in this chapter suggests that the correlation between biomarkers measured on CNB TMAs compared to SE TMAs is not precise and certainly not uniform across all

biomarkers. Despite this, previous work by the same research group found that both sets of TMAs contained the same biological clusters of breast cancer and therefore, should hypothetically include comparable expressions of individual biomarkers.

Cluster analysis in the SE TMAs using a panel of 25 of the biomarkers, revealed six biological clusters which represented the following subtypes: Luminal A (high hormone receptors, luminal cytokeratins, BCL2, MUC1 and HER3; low expression of HER2, p53, EGFR, CD44, E-cadherin and basal cytokeratins), Luminal B (differed from luminal A cluster only in lower PR expression), low ER luminal (high expression of luminal cytokeratins, MUC1, HER3 and low expression of the remaining biomarkers), normal-like (low expression of all biomarkers), basal-like (high expression of basal cytokeratins and HER3, low expression of luminal cytokeratins, HER2 and hormone receptors), HER2 overexpressing (high luminal cytokeratins, MUC1 and HER3 and negative hormone receptor expression). These clusters included the four main subtypes expected in any breast cancer cohort and a distinct novel cluster (low ER luminal), which appears to be unique to the older cohort compared to a younger population (3). Cluster analysis in the CNB TMAs constructed from patients with ER-positive breast cancer using 16 biomarkers, demonstrated three clusters: Luminal A, Luminal B, low ER-luminal (116). The absence of basal-like and HER2 overexpressing clusters was explained by the fact that this cohort was of older patients with ER-positive cancers only; they were very unlikely to contain basal-like or HER2 overexpressing tumours.

Therefore, the research group hypothesised that the biological cluster results were comparable for SE and CNB TMAs.

The next step of this current work was to use a method to try and understand the variability in the two data sets and establish whether or not a formula can be generated to predict SE TMA result, based on CNB TMA result alone.

6.3.1 Methods

PCA is an unsupervised learning technique that converts a set of observations (in this case biomarkers) into a set of values of linearly uncorrelated variables (principal components). PCA measures how each variable is associated with one another using a covariance matrix. The covariance matrix estimates the differences within each variable (variances) and between variables (covariances). The directions of the spread of the data is represented by an Eigenvector and the relative importance of each of these directions is represented by an Eigenvalue (387, 388). If an eigenvalue is close to zero, this component is unlikely to account for much variance in results and can be discarded.
The software Past (version 4.09, 2022, Hammer & Harper, University of Oslo) was used for PCA.

6.3.2 Results

Table 19 gives the eigenvalue and percentage variance for each of the 15 components examined in order of contribution to variance in the data. Each component represents a combination of features; they are not directly correlated to the 15 biomarkers.

Component	Eigenvalue	% variance
1	23135.1	27.3
2	16802.9	19.8
3	7368.9	8.7
4	7055.3	8.3
5	51978	6.1
6	4579.7	5.4
7	4179.5	4.9
8	3949.3	4.7
9	3001.6	3.5
10	2899.5	3.4
11	2437.3	2.9
12	1919.3	2.3
13	1225.3	1.4
14	1055.0	1.2
15	0.4	<0.1

Table 19: Eigenvalues and percentage variance for each component of principal components analysis

The first component is enough to explain up to 27.3% of the variance in the data. Components one and two together, account for approximately 47% of variance within the data. Figures 34 and 35 shows scatterplots to demonstrate the variance between each component. CNB TMAs cases are represented by red and SE TMAs by black.

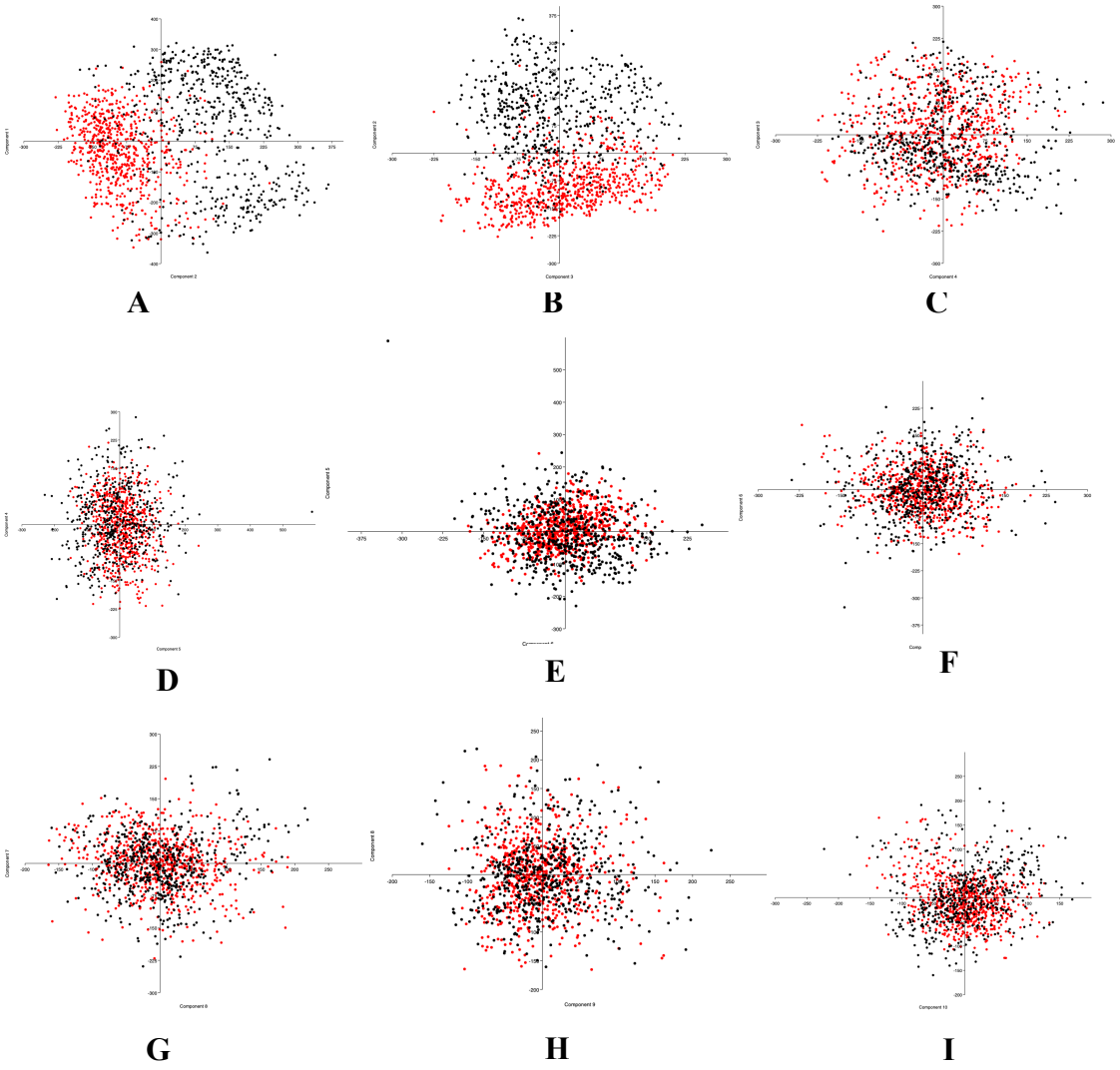


Figure 34: Scatterplots to demonstrate the variance between each component. CNB TMAs cases are represented by red and SE TMAs by black. A – component 1 (y-axis) versus component 2 (x-axis); B – component 2 (y) versus component 3 (x); C - component 3 (y) versus component 4 (x); D - component 4 (y) versus component 5 (x); E - component 5 (y) versus component 6 (x); F - component 6 (y) versus component 7 (x); G - component 7 (y) versus component 8 (x); H - component 8 (y) versus component 9 (x); I - component 9 (y) versus component 10 (x)

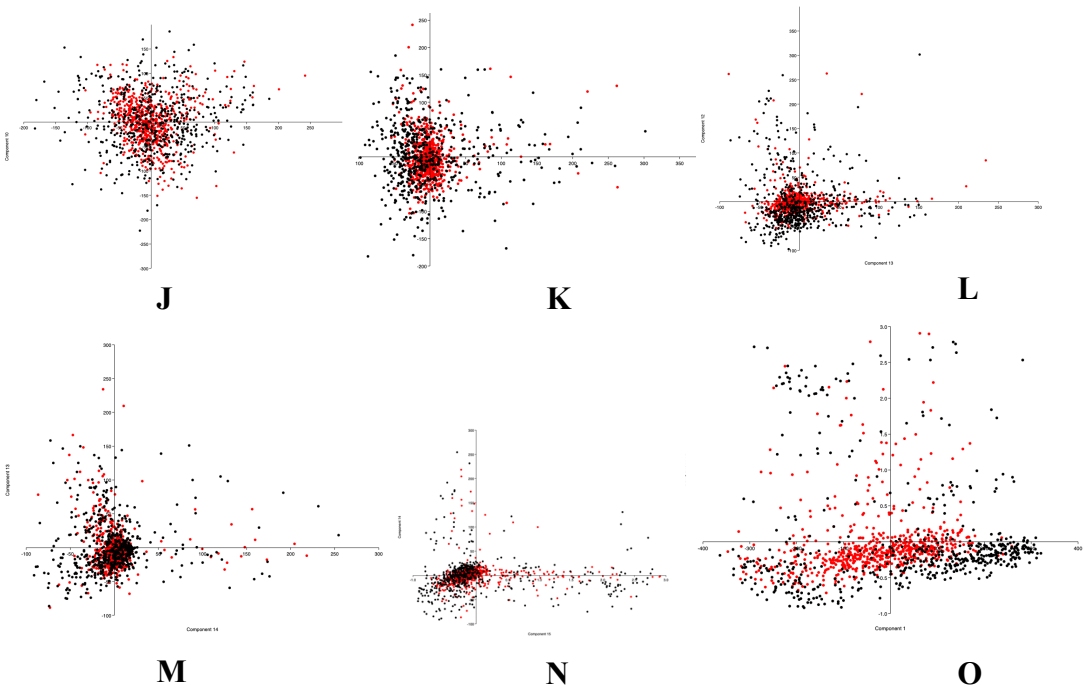


Figure 35: Scatterplots to demonstrate the variance between each component. CNB TMAs cases are represented by red and SE TMAs by black. J - component 10 (y-axis) versus component 11 (x-axis); K - component 11 (y) versus component 12 (x); L - component 12 (y) versus component 13 (x); M - component 13 (y) versus component 14 (x); N – component 14 (y) versus component 15 (x); O - component 15 (0x) versus component 1 (x)

The greatest variances between CNB TMA scores (red) and SE TMA scores (black) are seen in interactions with components 1 and 2 (Figure 33 A, B, C and O).

6.3.3 Interpretation of results

Despite the two cohorts of TMAs expressing the same biological clusters of breast cancer; there is no straightforward method of translating individual biomarker results derived from CNB TMA to predict SE TMA values; there are great variances between them. These findings may suggest that it is subtype of breast cancer that is important, rather than the individual biological profile of these subtypes; however, previous work in the research team has suggested that due to the heterogenous nature of breast cancer this is unlikely to be the case (3, 113, 146).

An alternative explanation may be that the present data set is not large enough to demonstrate the contribution of individual biomarkers. Furthermore, it is likely that some of the markers in the panel may not be significant contributors to the overall characterisation of the biology of primary breast cancer in older women and are at present, confounding the result.

6.4 Conclusions of this chapter and next steps

At the present time it is not possible within the present dataset to directly compare individual biomarker data between CNB and SE TMAs for all biomarkers in the panel. The biomarkers within the overall panel that have been most extensively studied in breast cancer research including ER, PR, Ki67, HER2, p53, BCL2 and VEGF, appear to show the greatest concordance between measurement on the two specimen sampling methods used for the TMAs. Poor concordance between the two tissue specimens, however, does not necessarily mean one sample is correct and one sample incorrect; simply that the two samples are not directly comparable.

Although there is much variation between the two TMAs, previous work of the research group has suggested that they each contain the expected number of biological clusters of breast cancer. Therefore, for the remainder of this thesis, the two TMA sets will continue to be analysed as two separate data sets.

As previously mentioned, cluster analysis has already been performed prior to this thesis in all the SE TMA cohort and the ER-positive cases in the CNB TMA cohort. The next step of this thesis is to complete cluster analysis of the whole CNB TMA cohort.

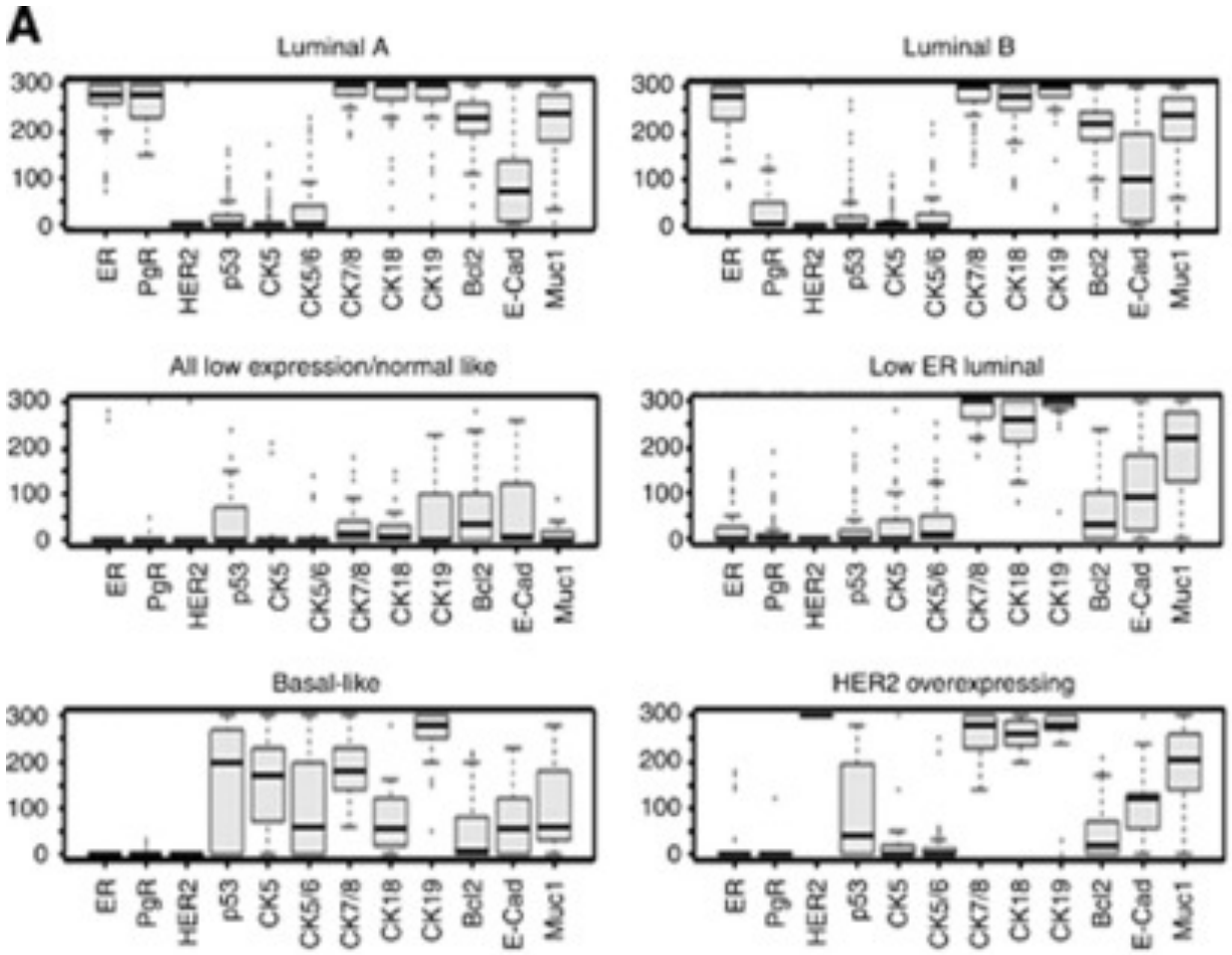
Chapter 7: Cluster analysis of all cases in core needle biopsy tissue microarrays (CNB TMAs)

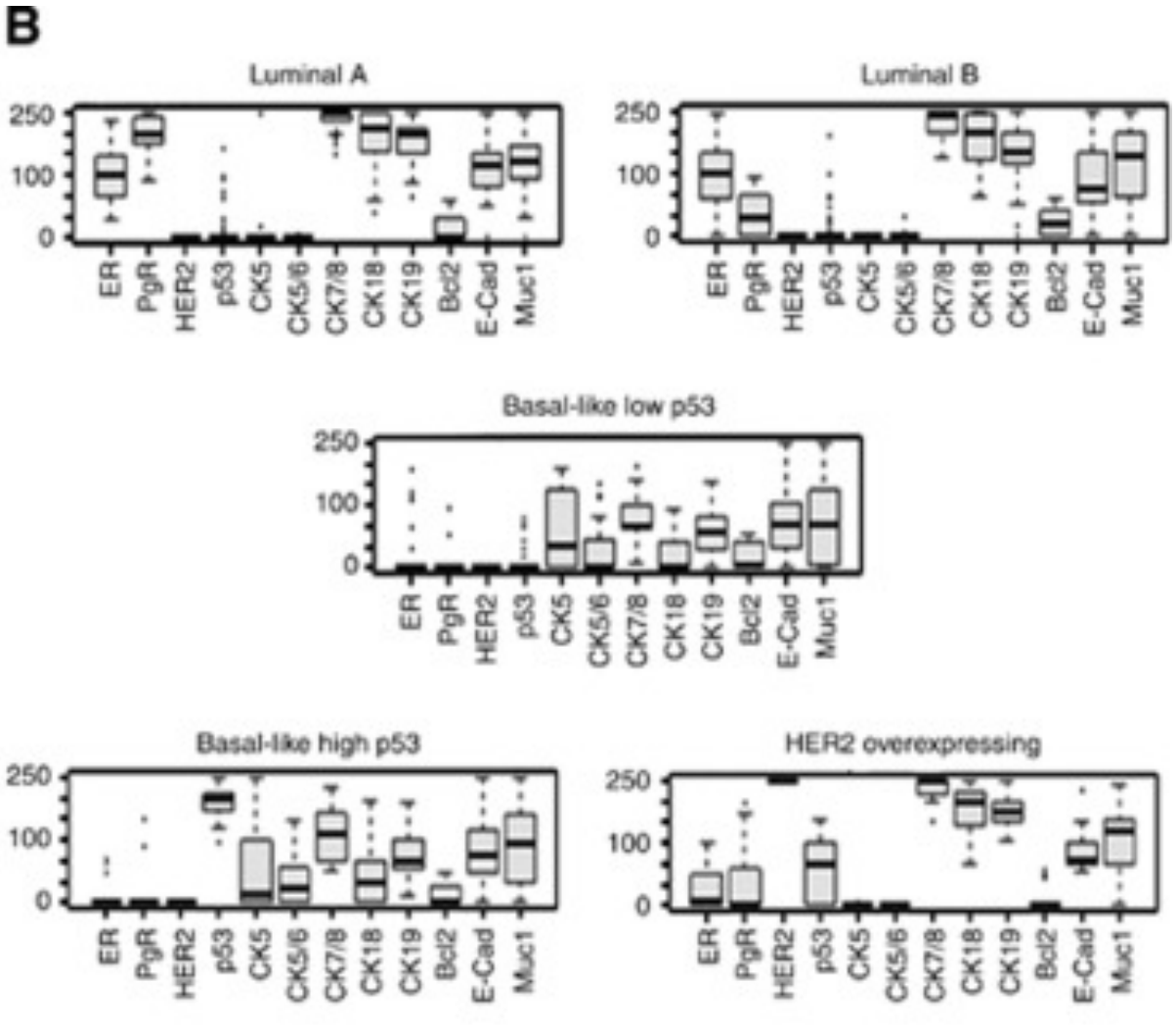
7.1 Introduction

This chapter will initially summarise the cluster analysis previously performed in the SE TMA cohort and ER-positive cases in the CNB TMA cohort and the anticipated findings in the CNB TMA cohort irrespective of ER status. Following this, the methodology of cluster analysis used in this present thesis will be described. The results of validity indices (how the number of clusters was decided) will then be given, followed by the results of cluster analysis itself, including a description of the clusters and key drivers. Comparison of BCSS between the different clusters will then be performed. The chapter will conclude with a discussion of how the findings fit into the complexity of the biology of breast cancer in older women.

7.2 Previous cluster analysis

Cluster analysis of a panel of 25 markers measured in the SE TMAs revealed six clusters (3). The same six clusters were reproduced by 12 biomarkers determined as key drivers in this panel (Figure 36A). Cluster analysis in a comparative group of younger women (Figure 36) demonstrated only five clusters; the low ER luminal cluster was unique to older women and had a distinct survival outcome (Figure 36C).


A
Luminal A
Luminal B
All low expression/normal like
Low ER luminal
Basal-like
HER2 overexpressing
300
200
100
0
ER
PgR
HER2
p53
CK5
CK5/6
CK7/8
CK18
CK19
Bcl2
E-Cad
Muc1


B
Luminal A
Luminal B
Basal-like low p53
Basal-like high p53
HER2 overexpressing
250
100
0
ER
PgR
HER2
p53
CK5
CK5/6
CK7/8
CK18
CK19
Bcl2
E-Cad
Muc1

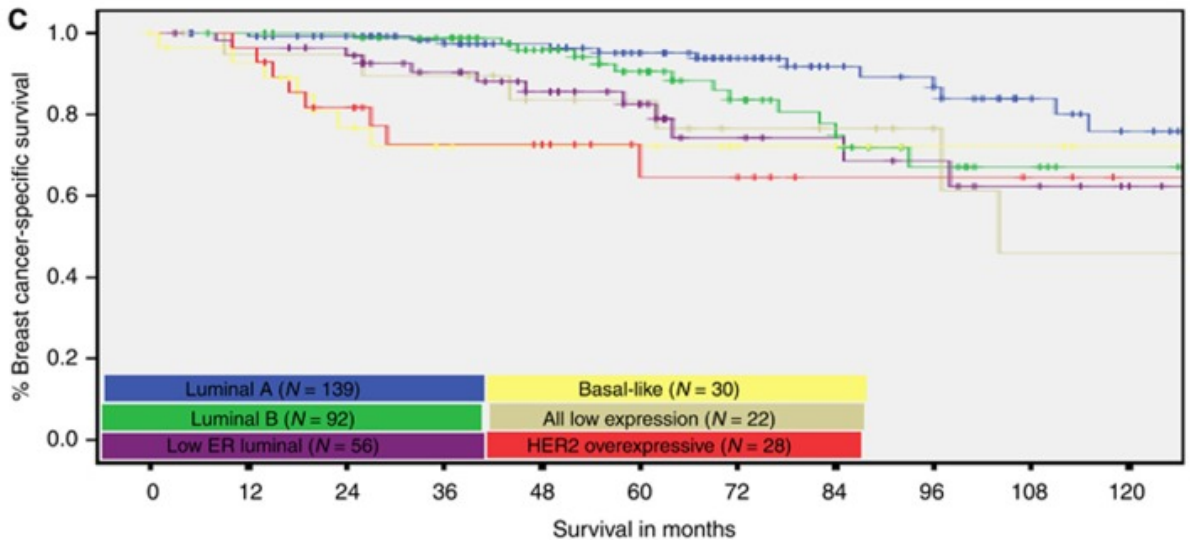


Figure 36: Biological classes of early operable primary breast cancer in SE TMAs – A) ≥70 years versus B) <70 years women. C) BCSS of older women with early operable primary breast cancer according to the biological clusters (3) – permission to use not required

Cluster analysis of 17 biomarkers in ER-positive CNB TMA cases revealed three clusters (Figure 37) (113). Cluster 1 was consistent with luminal A-type tumours; this cluster had high levels of expression of hormone receptors, luminal cytokeratins, BCL2 and MUC1 and low expression of HER2, Ki67, EGFR and basal cytokeratins. Cluster 2 was consistent with luminal-B type tumours; this cluster had the same characteristics as cluster 1 but much lower levels of PR expression. Cluster 3 was consistent with the low-ER luminal cluster; this cluster has a much lower expression of ER and PR and MUC1. The HER2 overexpressing and basal-like clusters were absent. This is a cohort of older women with ER-positive breast cancers only; therefore, patients with these cancers are more likely to have PR-positive, HER2-negative disease. In a cohort containing patients

irrespective of ER status, it would be expected that the HER2 overexpression and basal-like clusters would also appear.

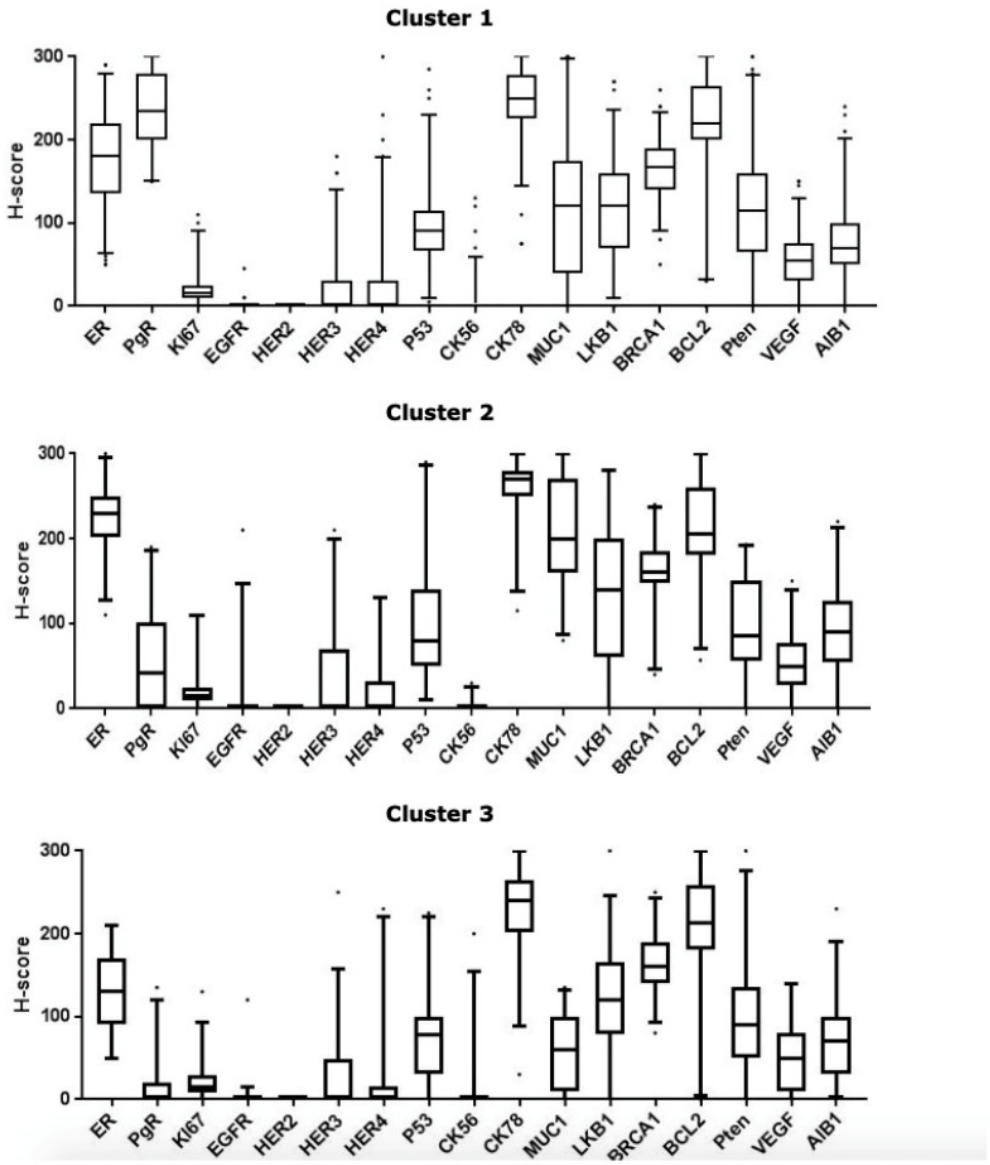


Figure 37: Biological classes of early operable primary breast cancer in CNB TMAs (113) – permission to use not required

Again, the low ER-luminal cluster had a distinct pattern of BCSS compared to the luminal A and B subtypes (Figure 38).

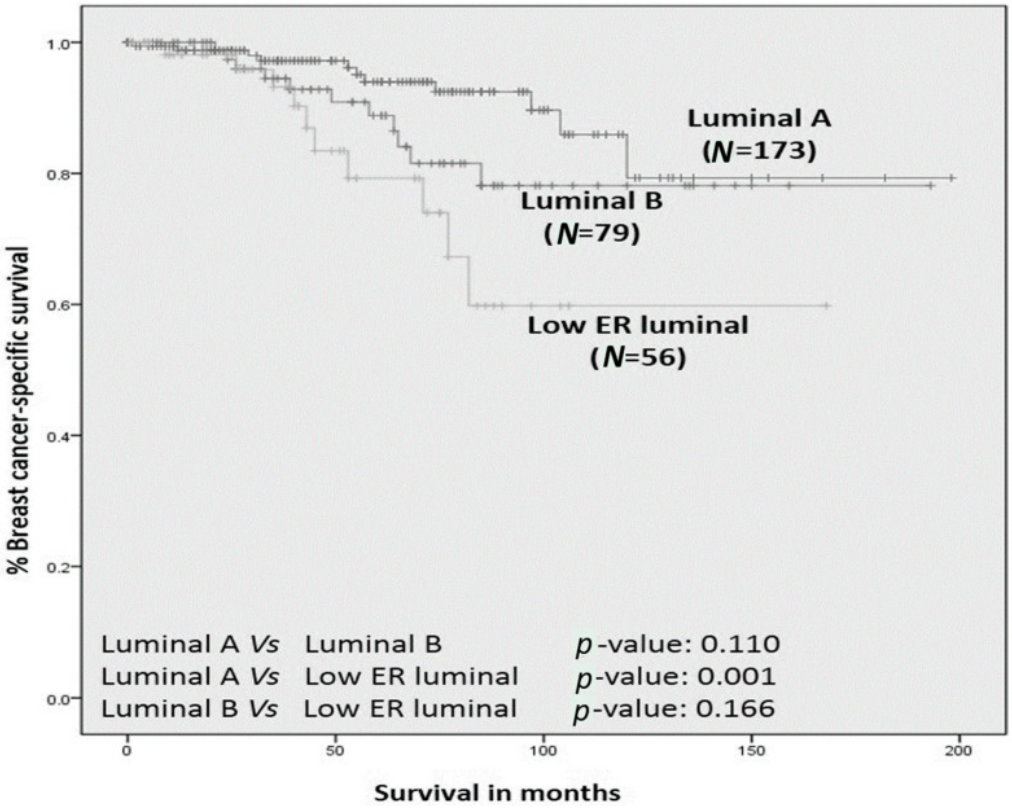


Figure 38: BCSS of biological cluster in ER-positive CNB TMAs in older women – permission to use not required

The next step of this work is to perform cluster analysis in the whole cohort of CNB TMAs (N = 693) (irrespective of ER status). It is expected that this analysis will reveal the same six clusters as seen in the SE TMAs as all subtypes of breast cancer should be represented; all CNBs were taken at time of diagnosis, prior to any treatment.

7.3 Methods

A panel of 17 biomarkers (Table 10) and associated clinical data was available for this analysis in all CNB TMAs: ER, PR, HER2, Ki67, EGFR, HER3, HER4, p53, CK 5/6, CK 7/8, MUC1, LKB1, BRCA1, BCL2, PTEN, VEGF, AIB1. These 17 biomarkers had been

stained prior to initiation of this thesis (see section 4.11). The step by step process of clustering used in this thesis is described below and follows the methodology described by Soria et al for use in IHC data from breast cancer patients (355). There are multiple recognised methods for clustering and choosing which method to use is not an easy task, as different clustering techniques identify "groupings" in slightly different ways. Consequently, it has been demonstrated that the use of multiple methods is often preferable in order to extract as much information as possible from the data. When using more than one algorithm, it is then common to define a consensus across the results in order to integrate diverse sources of similarly clustered data and to provide a stability of the results obtained from different techniques. Over a number of years a framework for clustering has been developed within the Nottingham research team, to elucidate a set of core groups in a dataset containing breast cancer patients, which has been validated in several datasets as described by Soria et al (355).

7.3.1 Handling of missing values

As discussed, (Table 12 and 13) there is variation in the number of TMA cores available for staining of an individual biomarker in a large TMA such as that used in this thesis. Cluster analysis can only be performed where entries have complete data sets. The same technique as was used in previous work by the same research team (3, 113) was used to fill missing data. Missing values were replaced by the median of the existing values for each biomarker. The median was chosen instead of the mean value, as the values are not normally distributed and this was in line with previous work from the Nottingham research group, to allow for direct comparisons (166, 167). Data was missing at random

depending on number of specimens stained from previous work; missing data for each biomarker ranged from 5% - 26% (167).

7.3.2 Validity Indices

The validity indices were used to obtain the appropriate number of clusters in the study cohort, separately for K-means and partition around medoids (PAM) algorithms. These techniques differ from each other as they have different measurement to define proximity of the data instances in order to establish the clusters. This means that they group the data looking at different characteristics present in the data. Six indices were used, namely the Calinski index, Hartigan index, Scott index, Marriot index, TraceW index, TraceW^(-1)B Index. K-means, PAM and the six indices were selected as they are among the most widely used methods in data mining.

7.3.3 Characterisation of clusters

To confirm and characterise the suggested appropriate number of clusters, bi-plots were constructed for all possible numbers of clusters starting from three in order to get the best possible agreement. The bi-plots defined the combination of variables by grouping them together to present a cluster and assigning a different colour to each cluster. Bi-plots demonstrate cluster distribution when a multi-dimensional dataset (each dimension is one biomarker) is projects in two dimensions, after it being clustered. Box-plots were constructed to demonstrate the pattern of each marker within each biological cluster.

7.3.4 Cluster validation

The characteristic groups were computed independently by K-means and PAM algorithms and the cluster patterns were validated by kappa-coefficient. The common clusters were constructed by combining the results of the two methods.

7.3.5 Identification of key drivers and reconstruction of clusters

The key drivers of classification were derived from the larger panel using supervised classification decision trees, with an accuracy of 95% using cross-validation.

7.3.6 Breast cancer-specific survival (BCSS)

Survival analysis was performed using Kaplan-Meier methods with application of log-rank as this is the most standard accepted method.

7.4 Results

There were a total of 681 complete datasets for clustering after filling in missing data.

7.4.1 Characterisation of clusters

Validity indices for K-means identified the possible number of clusters ranging from 3 – 15 (Figure 39).

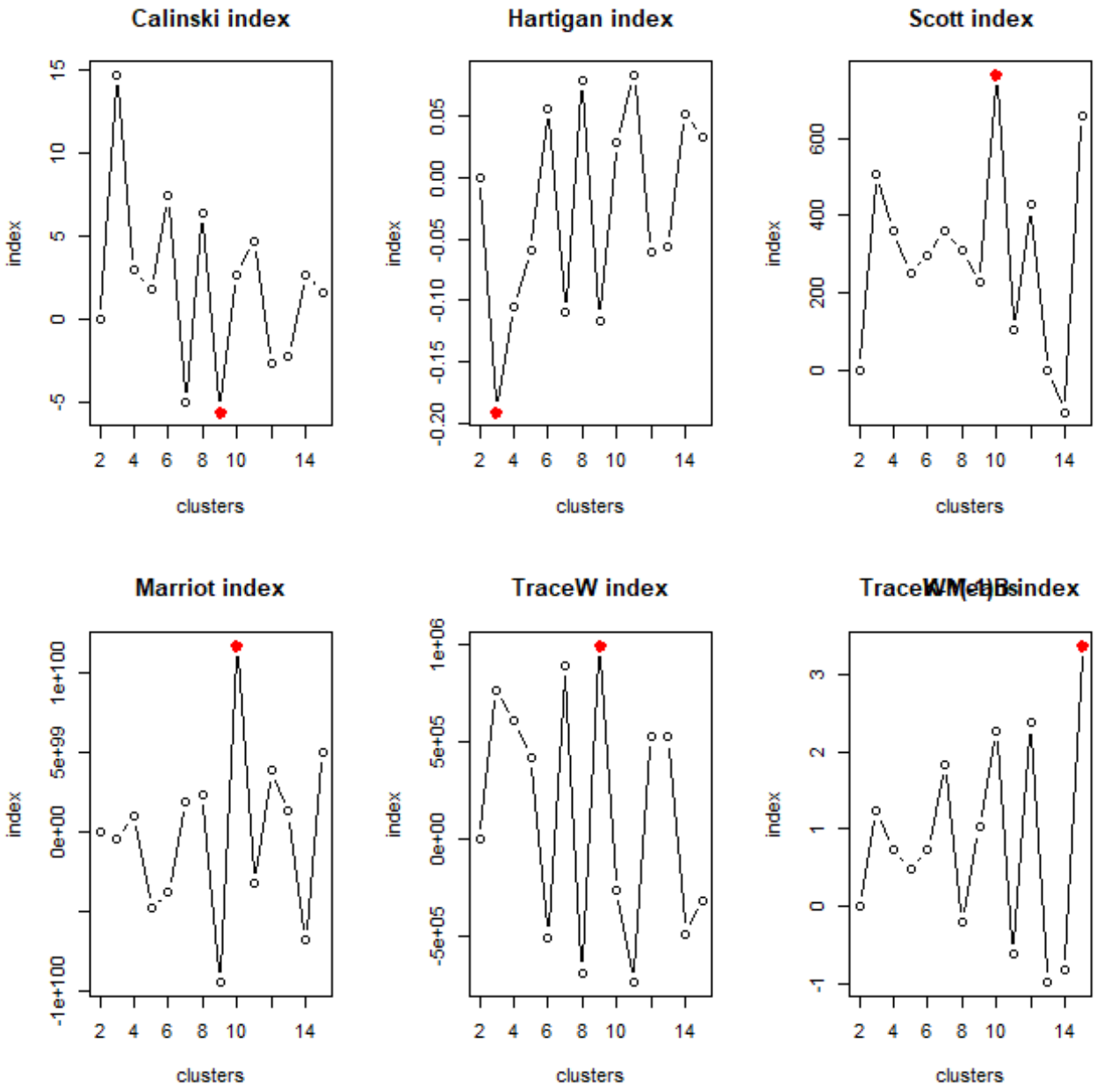


Figure 39: Validity indices for K-means showing the possible number of clusters using six of the most widely used indices in data mining. The red dot indicates the suggested optimal number of clusters for each method.

Validity indices for PAM identified the possible number of clusters ranging from 3 – 6 (Figure 40).

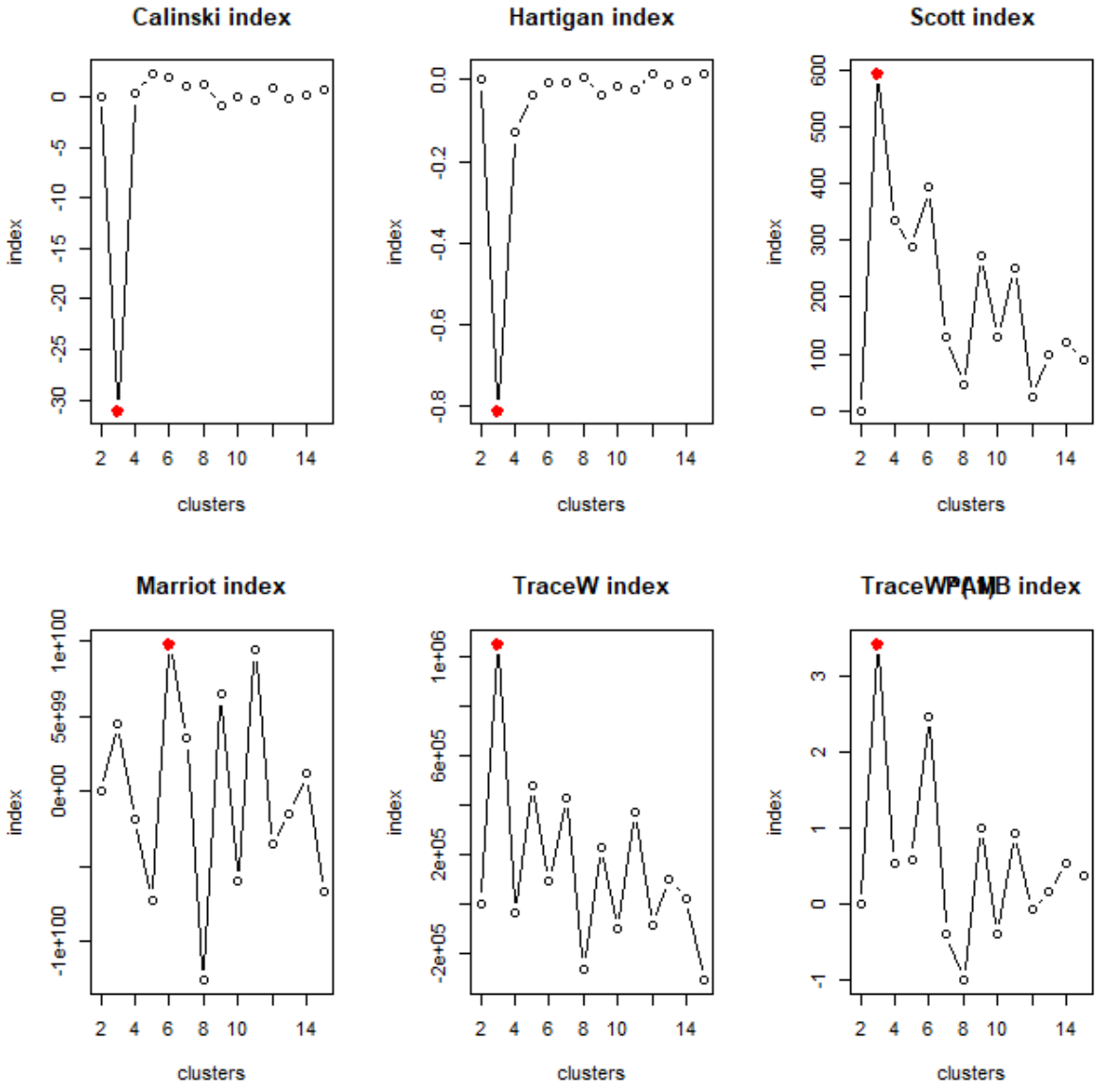


Figure 40: Validity indices for PAM showing the possible number of clusters using six of the most widely used indices in data mining. The red dot indicates the suggested optimal number of clusters for each method.

7.4.2 Characterisation of clusters and validation

The bi-plots defined the combination of variables by clumping them together to represent a cluster and assigning a different colour to each cluster, shown in Figure 41.



Figure 41: Bi-plots for validity indices showing a pattern of four clusters – three true clusters (red, yellow and purple dots) and one group which was unclustered (blue dots). The contribution of each biomarker is given by the blue lines, demonstrating that ER, PR and MUC1 are important in determining cluster allocation.

Using this method, three clusters appeared most prominent and showed substantial agreement, with kappa of 68.8% between different clustering methods.

The number of cases allocated to each cluster is given in Table 20.

Cluster	**Number of patients**
1 (red dots Figure 41)	217
2 (yellow dots)	125
3 (purple dots)	199
Unclassified (blue dots)	140

Table 20: Number of cases allocated to each cluster in CNB TMAs

7.4.3 Box-plots demonstrating the biomarker contribution to each cluster

Following on from looking at the bi-plot of the clusters (Figure 41), Figure 42 demonstrates the contribution of each biomarker to the three clusters through box-plots.

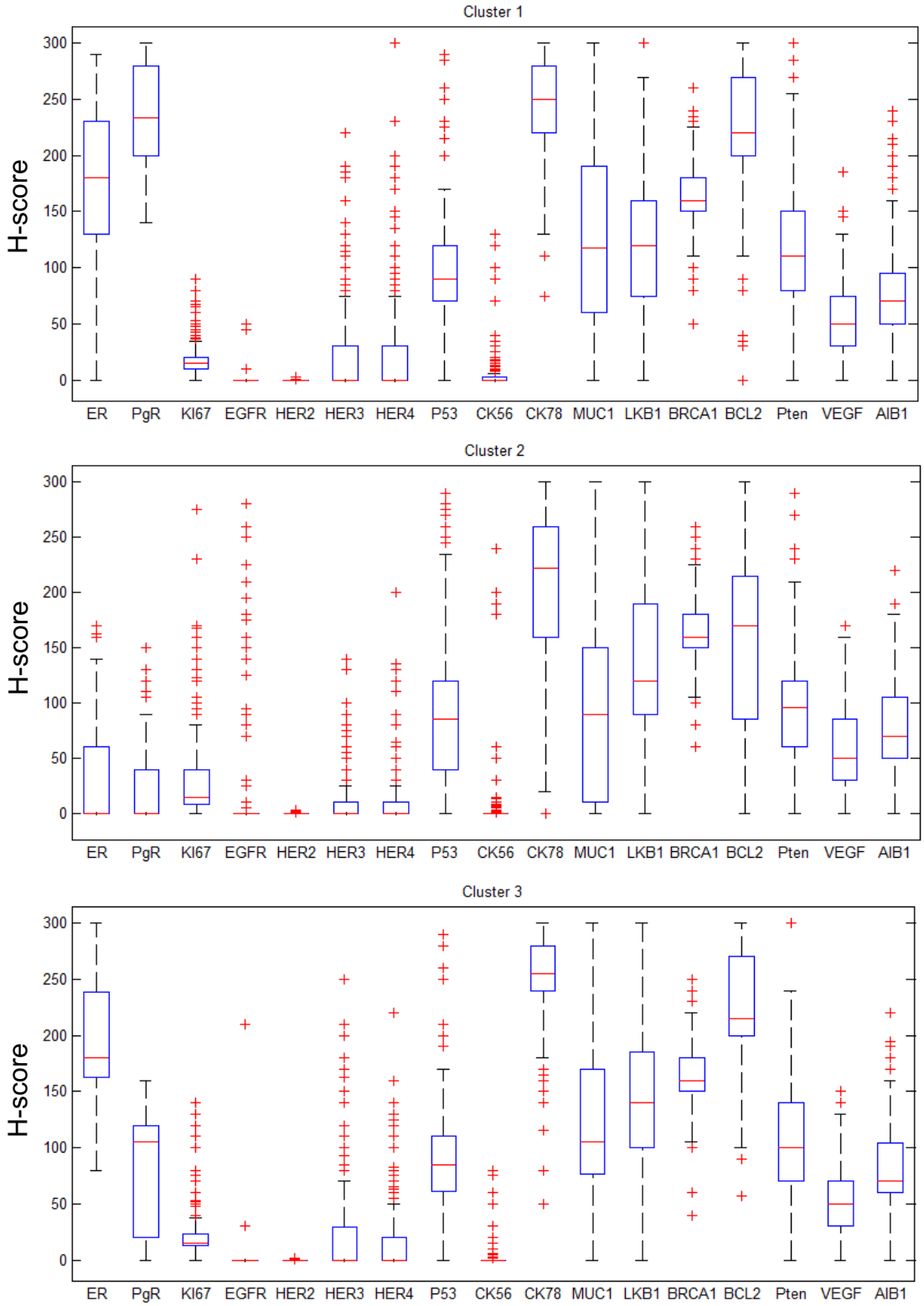


Figure 42: The contribution of each marker in the individual clusters.

The central line within the box represents the median value. The box represents the upper and lower quartiles of the spread of the data. The whiskers represent the overall spread of data.

Cluster 1 in Figure 42 shows high (≥150 H-score) expression of ER, PR, CK 7/8, BRCA1 and BCL2. There is low expression (<150 H-score) of Ki67, EGFR, HER2, HER3, HER4, p53, CK 5/6, MUC1, LKB1, PTEN, VEGF and AIB1.

Cluster 2 differs from cluster 1 in the expression pattern of PR and ER; this cluster has much lower levels of hormone receptors.

Cluster 3 differs from cluster 1 in expression of PR; this cluster has a much lower expression of PR but maintains high expression of ER.

The biomarker array of Cluster 1 is consistent with luminal A-type tumours and Cluster 3, consistent with luminal B-type tumours. Cluster 2 is similar to the low-ER luminal cluster; which has been previously demonstrated in SE TMAs and in the ER-positive cases within the CNB TMAs (113). The cluster distribution by each biomarker is given in Appendix 3, which confirms that the greatest difference between clusters is in expression of ER and PR.

7.4.4 Identification of key drivers

The key drivers were derived using decision trees (Appendix 4). The decision tree demonstrates the biomarkers which have the greatest role in determining which cluster a case will belong to; nodes on the top of the tree are the main drivers. As you move to the bottom of the tree, these nodes are less important. Therefore PR appears to be the most

important driver, followed by ER, then BCL2, then CK 5/6. The next most important markers (at the next level of the decision tree) are PTEN, BRCA1, p53, LKB1, HER2, followed by VEGF, Ki67, HER4, MUC1 at the lowest levels. The biomarkers EGFR, HER3, CK 7/8 and AIB1 do not feature in the decision tree.

7.4.5 Association between clusters and BCSS

Figure 43 compares the BCSS for all patients in luminal A cluster (Cluster 1), luminal B cluster (Cluster 3) and low-ER luminal cluster (Cluster 2) which was possible for 370 cases.

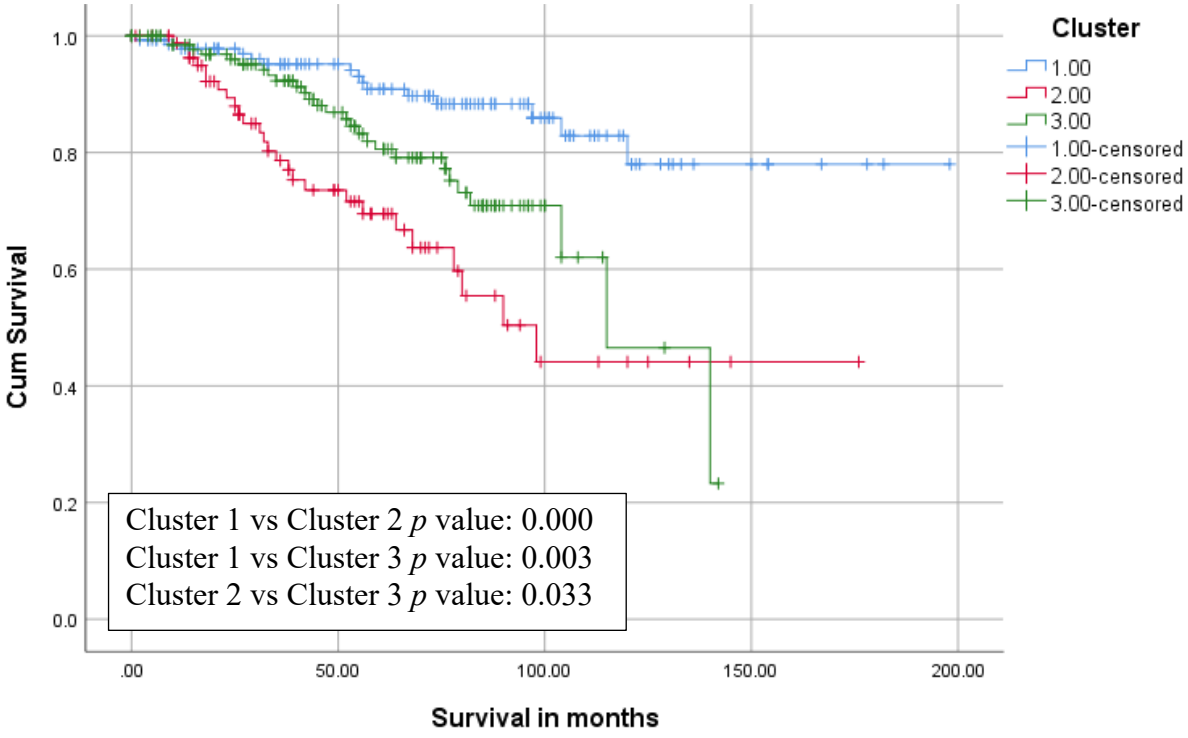


Figure 43: BCSS of biological clusters in all CNB TMAs

Survival analysis shows that those cases in Cluster 1 have the best BCSS and those in Cluster 2 the poorest BCSS. The cases in Cluster 3 have better BCSS compared to Cluster 2, but worse BCSS compared to Cluster 1.

7.5 Explanation of findings

Cluster analysis in the whole cohort of CNB TMAs has revealed three biological clusters which are consistent with those found in prior analysis of ER-positive cases only (113). It was anticipated that there would be three further clusters found: HER2-overexpressing, basal-like and normal-like.

Cases were reviewed individually to determine whether HER2-positive and TNBC cases had been allocated to one of the three clusters, or remained unallocated, in an attempt to explain these findings. Upon case review, it was evident that there were a significant number of cases which measured positive for HER2 in the SE TMAs, compared to CNB TMAs; 11% of all cases were positive for HER2 in the SE TMAs compared to only 2% in the CNB TMAs. This low number of HER2 positive cases in the CNB TMAs, means that the HER2-overexpressing and basal-like clusters would not have been prominent enough to develop additional clusters and at present are within the group that are unclustered during this analysis. This is a limitation in the work as it also means that survival analyses could be biased if HER2-positve and TNBC cases have been incorrectly allocated into one of the other clusters (rather than unclustered).

This adds to the findings of Chapter 6; the H-scores allocated to biomarkers measured on SE TMAs compared to CNB TMAs cannot simply be compared like-for-like.

Within the three clusters detected, luminal A (Cluster 1) and B clusters (Cluster 3) contained the most cases (Table 20), consistent with previous findings and had the same

pattern of biomarker expression as seen previously. The cluster termed low-ER luminal had some differences in this present thesis. Compared to the same cluster detected in ER-positive CNB TMAs alone, the low-ER luminal cluster in the present study, had a lower average ER expression (median H-score 144 compared to 174); which is to be expected; previous work excluded ER-negative cases which are included in the present study. Compared to the same cluster detected in SE TMAs, the low-ER luminal cluster in the present thesis had a similar expression of ER, however, lower expression of MUC1 and higher expression of BCL2. It is likely therefore, that the low-ER luminal cluster is a truly unique biological cluster but is still to be accurately defined or may contain further subclusters.

The low-ER luminal cluster had a distinct survival pattern compared to the other two clusters; cases in the low-ER luminal cluster had poorer survival. This may suggest that patients in the low-ER luminal cluster should be treated more aggressively; for example, even in patients who have ER-positive breast cancers, perhaps surgery should be more strongly recommended. Given that over 23% of all cases that were clustered, belonged to the low-ER luminal cluster, this is a significant proportion of patients and would have a substantial impact on clinical practice.

Around 20% of cases in this cohort were unable to be characterised into a cluster. As hypothesised in this thesis, this may be because breast cancer is such a heterogenous disease; there will always be cases that do not fit into pre-defined boxes. Despite a large panel of biomarkers participating in this present cluster analysis, the most significant markers contributing to the three cluster types, remain as ER and PR. However, presumably ER and PR do not play such a significant role in the cases which were

unclustered; other biomarker patterns may emerge here, but are not yet of a large enough size to generate a further cluster. This provides further support for development of a prognostic algorithm, based on an individual's unique breast cancer biology to help guide treatment of an individual patient.

According to decision tree analysis, PR is the most important marker in determining which cluster a case will belong to, followed by ER, BCL2 and CK 5/6. The markers BCL2 and CK 5/6 certainly warrant further investigation in this setting and may provide the key in determining additional biological clusters of breast cancer in older women, beyond routine subtypes. This will be of particular interest in Chapter 8 when a prognostic algorithm will be developed.

7.6 Conclusions of this chapter and next steps

This chapter has outlined cluster analysis performed using all CNB TMA cases. Three biological clusters have been identified which is consistent with previous work: luminal A-type; luminal B-type; low-ER luminal. The HER2-overepressing and basal-like clusters were absent due to the overall low expression of HER2 measured in these samples. The luminal A-type cluster had high levels of expression of hormone receptors, luminal cytokeratins, BCL2 and MUC1 and low expression of HER2, Ki67, EGFR and basal cytokeratins. The luminal B-type cluster had the same characteristics as luminal-A cluster but much lower levels of PR expression. The low-ER luminal cluster had a much lower expression of ER and PR and MUC1, compared to the other two clusters.

This chapter adds further support for the low-ER luminal cluster as a unique biological cluster with specific survival outcome. Furthermore, once again the same biological clusters were demonstrated in the CNB TMA group that were found in the SE TMA

group. Therefore, although as demonstrated in chapter 6, direct correlation between CNB and SE TMAs scores is not clear, the two cohorts do share the same biological clusters of breast cancer.

The next step in this thesis is to further examine the panels of biomarkers measured in both the CNB TMAs and the SE TMAs as separate cohorts, using bioinformatics analysis to attempt to generate a prognostic algorithm.

Chapter 8: Generation of a prognostic algorithm

8.1 Introduction

As highlighted throughout this thesis, breast cancer is a highly heterogeneous disease and the field of breast cancer research has embraced the potential of high-throughput analysis. The use of bioinformatics principles and analysis has been established at the genomic, transcriptomic (389) and proteomic (390) level for breast cancer.

The aim of the work outlined in this chapter, is to use bioinformatics analysis to generate a prognostic algorithm using the panel of biomarkers measured in this thesis (Table 10) in both the CNB TMA and SE TMA cohorts. According to the literature, this current project is unique in its application of bioinformatics analysis to the largest panel of biological markers measured by IHC in older women with primary breast cancer.

First, this chapter will discuss in greater detail the precise bioinformatics methodology employed and describe the disease outcomes to be explored. Then, the results of attempts at developing a prognostic algorithm for each disease outcome will be given using the data from SE and CNB TMAs separately. The prognostic algorithms generated which show potential correlation between the result generated compared to the actual result, will then be interpreted in further detail to understand how each marker contributes to the algorithm. Using this information, the chapter will conclude with a discussion of potential avenues to refine and develop the algorithm(s) further.

8.2 Methods

8.2.1 Software utilised

Data was stored in Microsoft Excel. The software package NeuroSolutions (version

7.1.1.1, 2015, NeuroDimension, USA) was used for data handling and bioinformatics analysis.

8.2.2 Handling of missing data

Missing biomarker measurements for particular samples occurred for a number of reasons:

i) The tissue core had lost structure over time and it was not possible to measure any biomarker reliably on that sample

ii) The tissue section cut from the TMA did not contain >15% tumour cells or was damaged in the process of sectioning and could not be scored

iii) The tissue section was damaged during the process of IHC (human error) and could not be scored.

In the instance of i) it was not possible to measure any biomarkers on these cores and therefore, these patients were discarded from further analysis. In the instance of ii) and iii) these samples were used for further analysis with all available results.

8.2.2.1 Method for filling missing data

Missing values were dealt with by using the K-nearest neighbour (KNN) method to complete the values, for each clinical feature observed. When missing data values are in a data set the simplest way of dealing with missing values is to discard the observations that contain them. However, this method is not practical in this context as the pattern of missing data is such that a large number of samples would be excluded. There are a few missing cores which have destabilised over time within the TMA block and therefore, it is

not possible to measure any biomarker on this particular core, however, for the majority of missing cases, it is during the laboratory process of testing for a particular biomarker, that the core sample has been damaged or degraded and can not be accurately measured and the pattern of this missing data is random.

It is commonplace in analysis of large data to fill missing data values using data from similar cases or missing data can be filled using data from neighbouring persons or dimension reduction techniques whereby the whole data set is decomposed and reconstructed to form an entire filled dataset (391, 392). Investigation of different methods to fill missing data has shown that use of different methods has a negligible impact on cluster analysis and classification downstream (391).

The KNN method is understood to be superior to other nearest neighbour methods (393) and is commonly accepted as a way to fill missing data. The method is based on finding the k closest neighbours. The missing value is replaced by the weight average of the k-neighbours, where the weights are proportional to the similarity between samples. The method uses Euclidean distances as its similarity measures. Following published guidance, the number of neighbours k was set to 20 (394).

8.2.3 Disease outcomes explored

At this stage each of the CNB TMAs and SE TMAs were analysed as two separate groups. The following data fields were extracted from the overall database: BCSS, OS, time to metastasis (TTM), time to progression of disease or local recurrence (TTP), time to regional recurrence (TRR).

8.2.4 Data mining

The interpolated data generated after applying KNN algorithms were randomly split into three subsets: 60% for training, 20% for validation to ensure model performance during the training process and 20% for blind testing of the original model based on widely accepted models (395).

8.2.4.1 Artificial Neural Network

Artificial Neural Network (ANN) database mining approach was used to build a classifier using the measured biomarkers and identify patterns associated with each disease outcome. This will comprise orthogonal features identified in a stepwise fashion from across the parameter set. Multiple randomly split sub datasets will be analysed in parallel 20 times to 20 marker additions to assess the stability of the identified marker set given the number of cases provided.

8.2.4.2 Architecture

The ANN modelling undertaken used a supervised learning approach applied to a three-layer multi-layer perceptron architecture. The initial weights matrix was randomised with a standard deviation of 0.1 to reduce the risk of overfitting the data. The ANN architecture utilised a range of hidden nodes (2 to 120) in a single hidden layer, as this level of depth of architecture has been shown to provide high quality results without significantly extending analysis time. During training, weights were updated by a Levenburg Marquet algorithm and /or backprop Gradient Descent. Learning rate and momentum were set at 0.1 and 0.5 respectively. The output node represents time to event associated with the outcome.

Prior to testing, a Monte Carlo cross-validation procedure was applied to avoid data overfitting and false discovery. Documentation of such an approach has proven to out-perform the commonly used leave-one-out cross validation (396).

8.2.4.3 Formulation of the prognostic algorithm

Model results were cross-validated with a systematic stepwise developed Cox proportional Hazards model building from a univariate to a multivariate model. The regression coefficient of the model was compared with the regression coefficient of the ANN model and performance, parameter inclusion and significance compared. From these analyses prognostic model formulae were extracted.

8.2.4.4 Validation within the existing Nottingham dataset

Comparison of outcome measures between the predicted value (as calculated by the algorithm) and the actual value (from the historic dataset) were made. Survival measures between predicted and actual values were evaluated by log rank. Comparison of other outcome measures were evaluated by parametric statistics (T-Test and ANOVA).

8.3 Results

Each disease outcome was explored in turn for both the SE TMAs and the CNB TMAs. An interpretation of the correlation between actual and predicted result for each disease outcome is given based on the correlation coefficient (r value) which is covered in section 4.6.3.

Briefly, r-values were interpreted as follows:

r 0.7 – 1 = strong correlation

r 0.5 – 0.7 = moderate correlation

r <0.4 = weak or no correlation.

8.3.1 Surgical excision (SE) tissue microarrays (TMAs)

8.3.1.1 SE TMAs – breast cancer specific survival

A neural network consisting of five hidden nodes was determined to be the most accurate predictor of time to BCSS.

Table 21 shows the best performance of the algorithm on the test data.

Performance	**Breast cancer specific survival (in months)**
Root Mean Squared Error (RMSE)	33.8
Normalised Root Mean Squared Error (NRMSE)	0.2
Mean Absolute Error (MAE)	22.7
Normalised Mean Absolute Error (NMAE)	0.1
Minimum Absolute Error (Min AE)	0.7
Maximum Absolute Error (Max AE)	102.4
Linear Correlation Coefficient (r value)	0.742

Table 21: Performance of the algorithm designed to predict BCSS in SE TMAs

Figure 44 depicts the association between the estimated prediction of BCSS as determined by the algorithm compared to actual data.

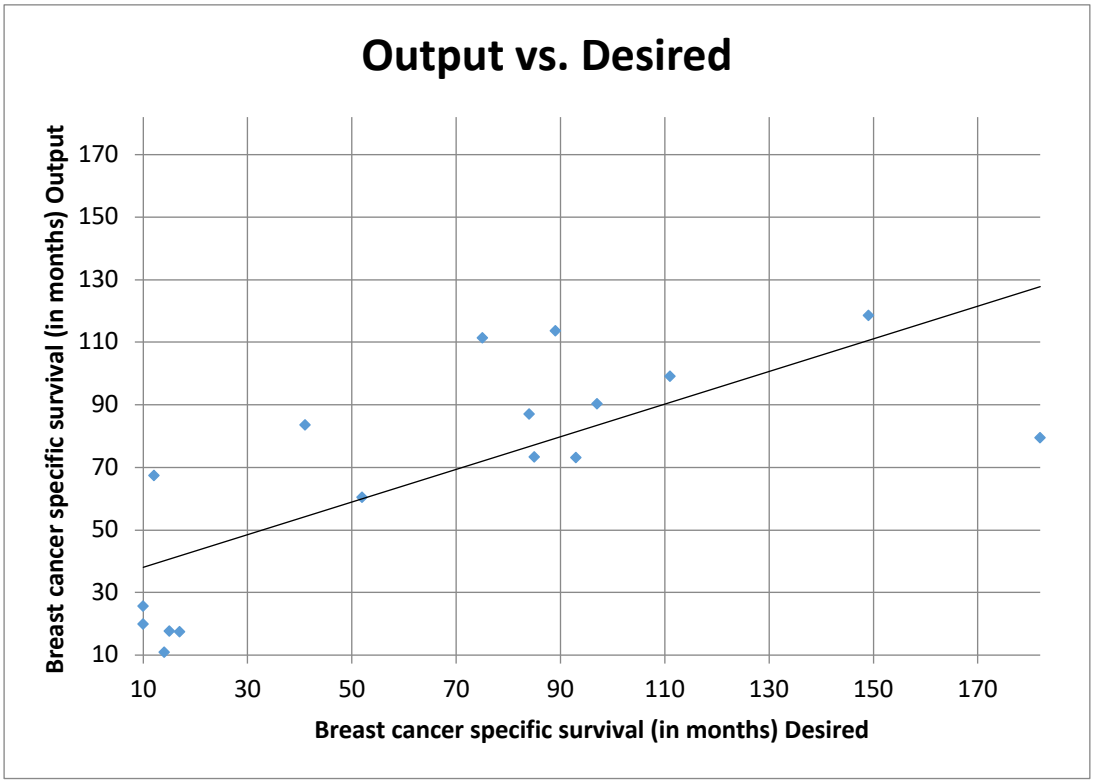


Figure 44: Actual BCSS compared to predicted BCSS in SE TMAs

There is a strong correlation between the prognostic model for BCSS measured in SE TMAs and actual result.

8.3.1.2 SE TMAs – overall survival

A neural network consisting of five hidden nodes was determined to be the most accurate predictor of OS.

Table 22 shows the best performance of the algorithm on the test data.

Performance	Overall survival in months
RMSE	43.21
NRMSE	0.2
MAE	33.8
NMAE	0.1
Min Abs Error	0.16
Max Abs Error	157.8
r	0.379

Table 22: Performance of the algorithm designed to predict OS in SE TMAs

Figure 45 depicts the association between the estimated prediction of OS as determined by the algorithm compared to actual data.

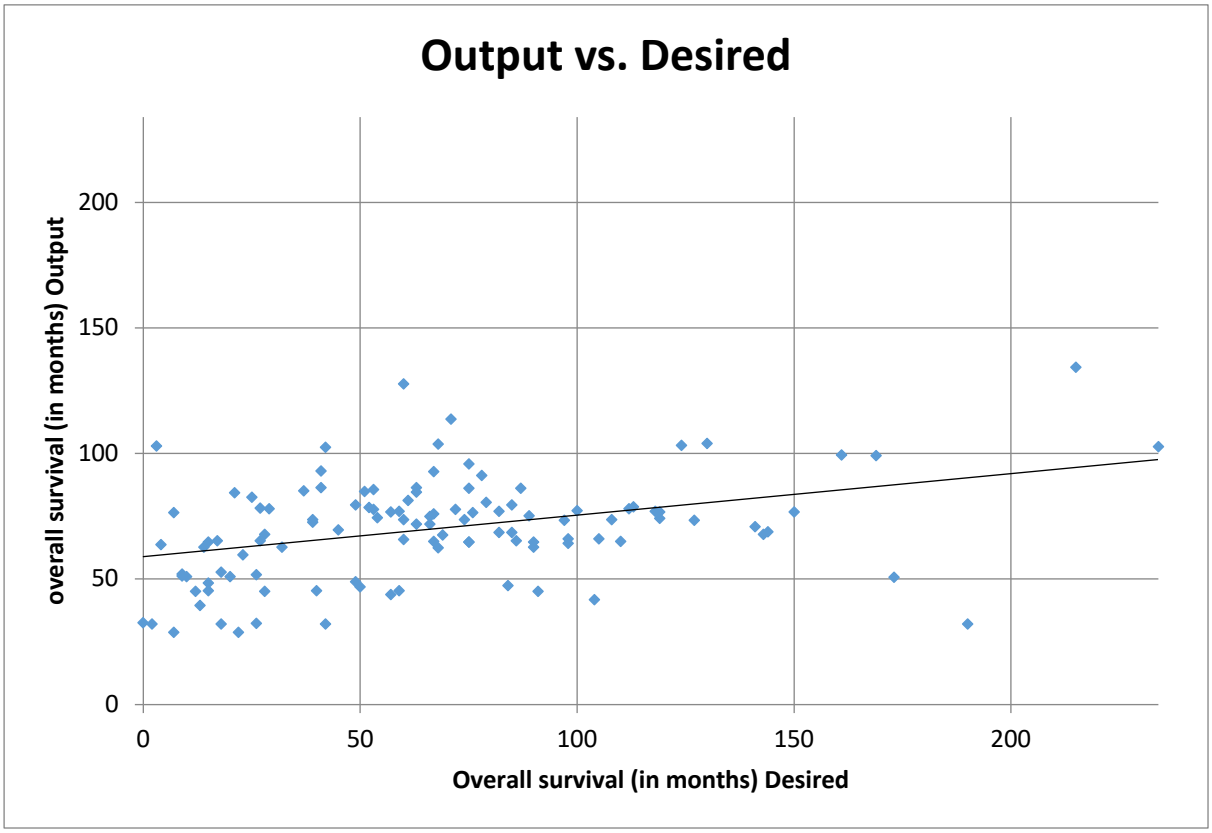


Figure 45: Actual OS compared to predicted OS in SE TMAs

There is a weak correlation between the prognostic model for OS measured in SE TMAs and actual result.

8.3.1.3 SE TMAs – time to metastasis

A neural network consisting of two hidden nodes was determined to be the most accurate predictor of TTM.

Table 23 shows the best performance of the algorithm on the test data.

Performance	Time to metastasis (in months)
RMSE	47.2
NRMSE	0.2
MAE	34.2
NMAE	0.2
Min Abs Error	0.6
Max Abs Error	179.6
r	-0.044

Table 23: Performance of the algorithm designed to predict time to metastasis in SE TMAs

Figure 46 depicts the association between the estimated prediction of TTM as determined by the algorithm compared to actual data.

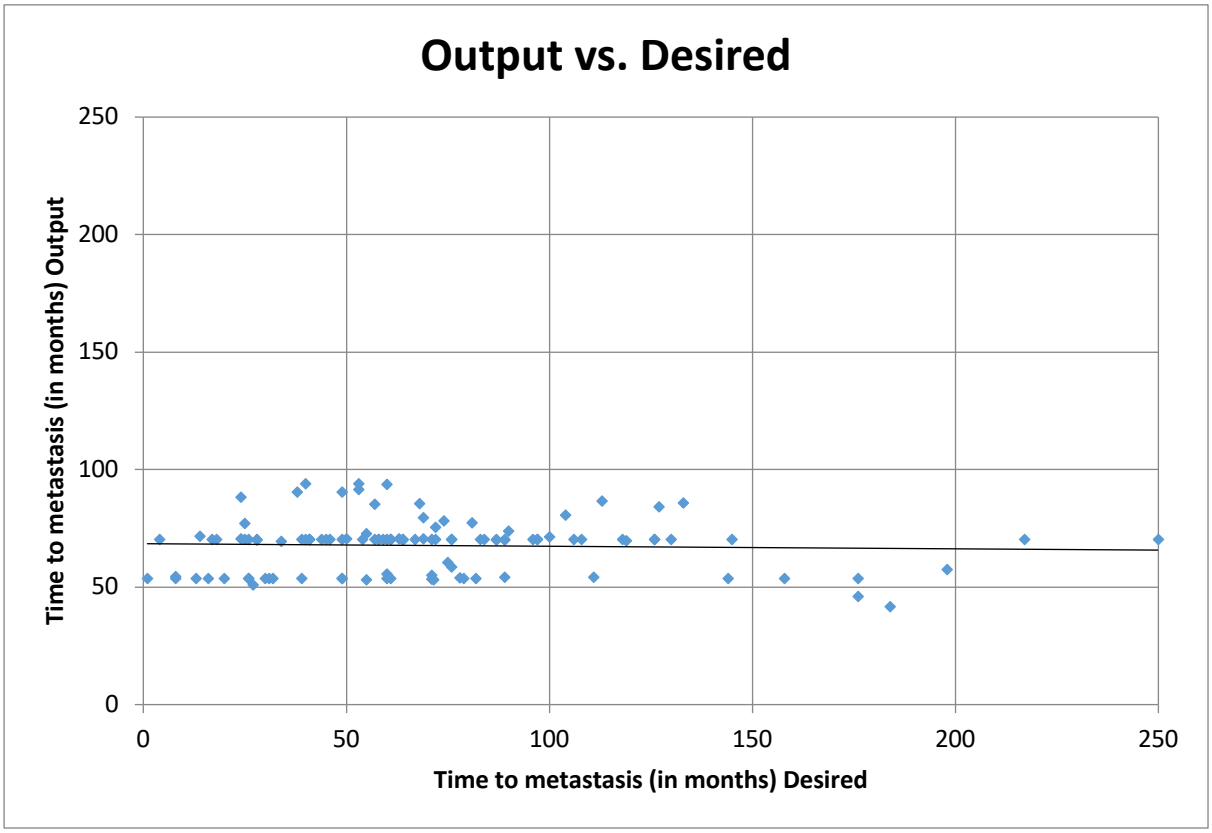


Figure 46: Actual TTM compared to predicted TTM in SE TMAs

There is a weak correlation between the prognostic model for TTM measured in SE TMAs and actual result.

8.3.1.4 SE TMAs – time to regional recurrence

A neural network consisting of three hidden nodes was determined to be the most accurate predictor of TRR.

Table 24 shows the best performance of the algorithm on the test data.

Performance	**Time to regional recurrence (in months)**
RMSE	46.1
NRMSE	0.1
MAE	36.3
NMAE	0.1
Min Abs Error	0.02
Max Abs Error	159.0
r	0.102

Table 24: Performance of the algorithm designed to predict TRR in SE TMAs

Figure 47 depicts the association between the estimated prediction of time to RR as determined by the algorithm compared to actual data.

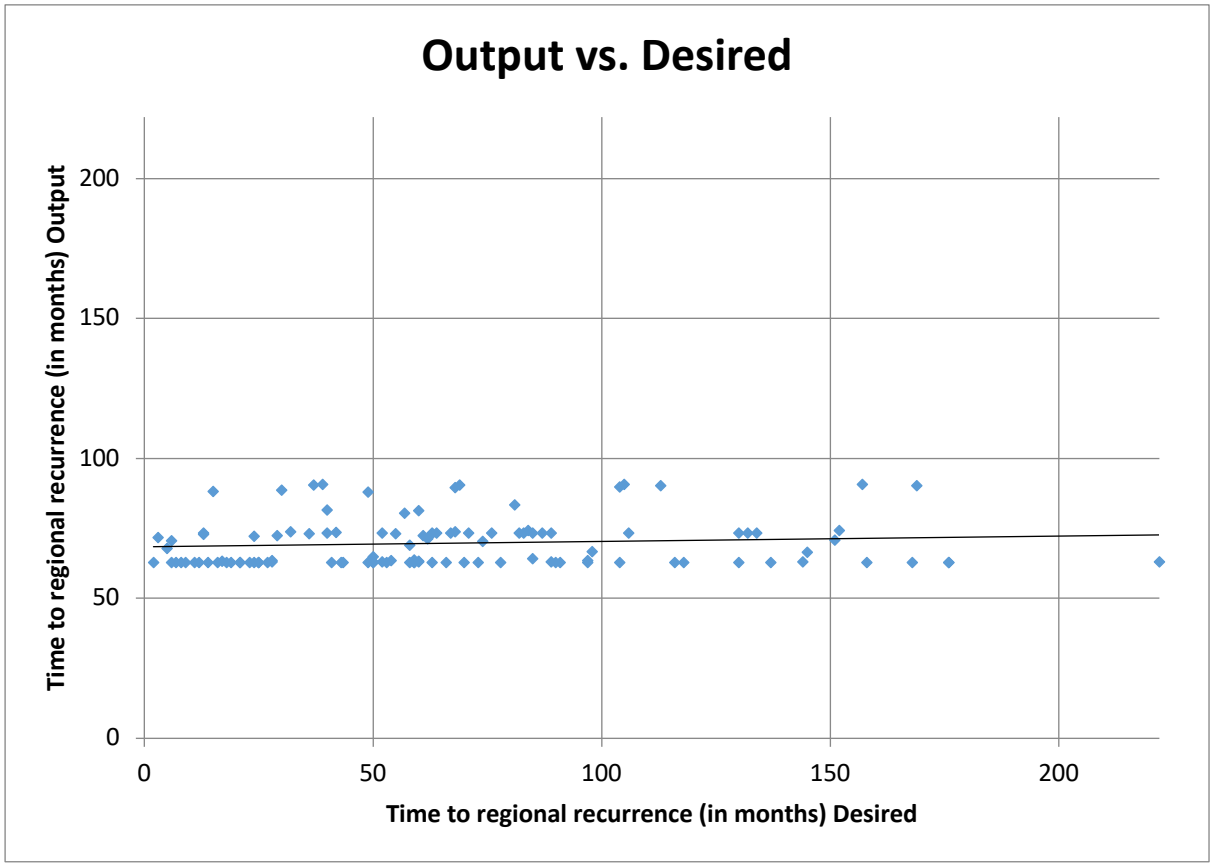


Figure 47: Actual time to RR compared to predicted TRR in SE TMAs

There is a weak correlation between the prognostic model for time to RR measured in SE TMAs and actual result.

8.3.1.5 SE TMAs – time to progression of disease

A neural network consisting of three hidden nodes was determined to be the most accurate predictor of time to TTP.

Table 25 shows the best performance of the algorithm on the test data.

Performance	**Time to progression of disease (in months)**
RMSE	45.1
NRMSE	0.1
MAE	31.2
NMAE	0.1
Min Abs Error	0.2
Max Abs Error	186.1
r	0.037

Table 25: Performance of the algorithm designed to predict time to disease progression in SE TMAs

Figure 48 depicts the association between the estimated TTP as determined by the algorithm compared to actual data.

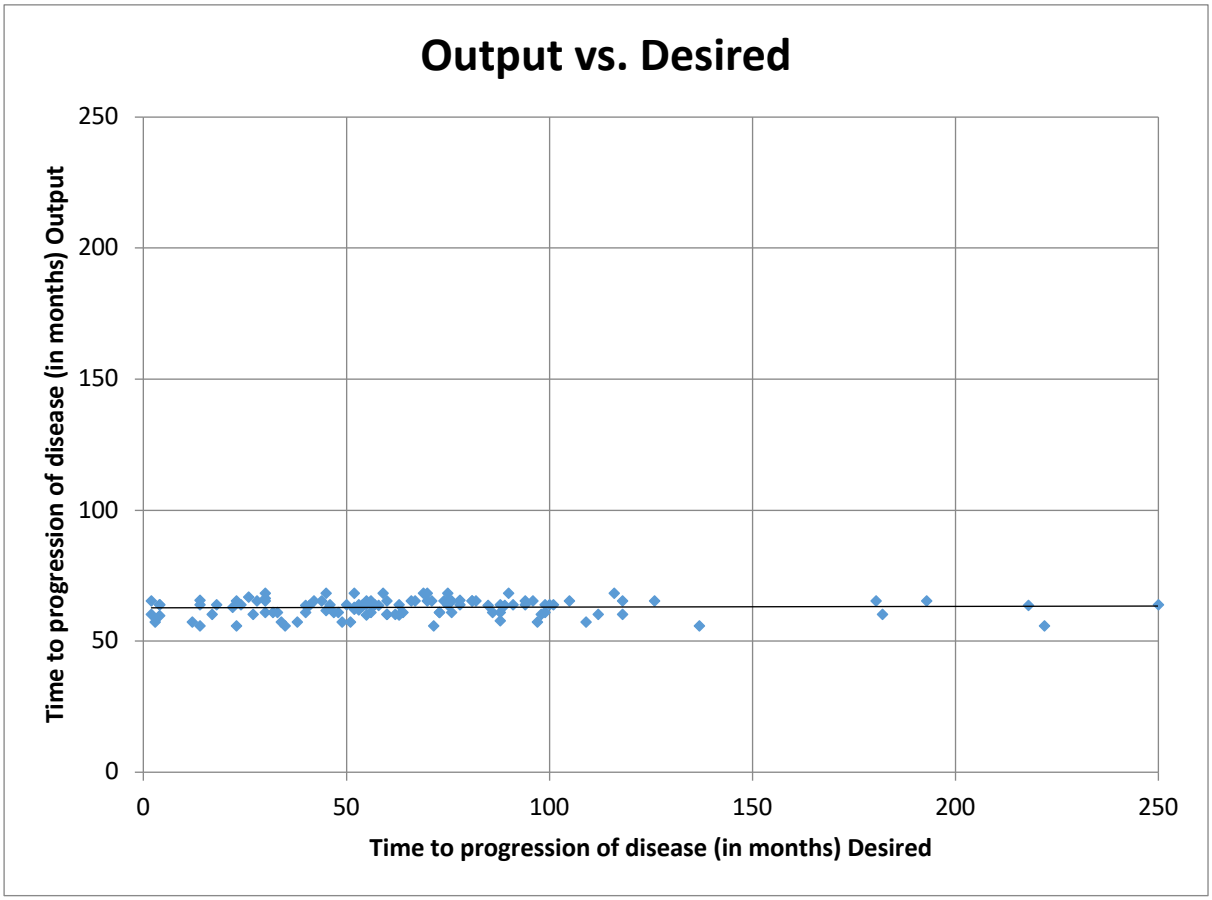


Figure 48: Actual TTP compared to predicted TTP in SE TMAs

There is a weak correlation between the prognostic model for TTP measured in SE TMAs and actual result.

8.3.2 Core needle biopsy (CNB) tissue microarrays (TMAs)

8.3.2.1 CNB TMAs – breast cancer specific survival

A neural network consisting of two hidden nodes was determined to be the most accurate predictor of BCSS.

Table 26 shows the best performance of the algorithm on the test data.

Performance	**Breast cancer specific survival (in months)**
Root Mean Squared Error (RMSE)	26.0
Normalised Root Mean Squared Error (NRMSE)	0.2
Mean Absolute Error (MAE)	17.5
Normalised Mean Absolute Error (NMAE)	0.2
Minimum Absolute Error (Min AE)	0.7
Maximum Absolute Error (Max AE)	64.2
Linear Correlation Coefficient (r value)	0.716

Table 26: Performance of the algorithm designed to predict BCSS in CNB TMAs

Figure 49 depicts the association between the estimated prediction of BCSS as determined by the algorithm compared to actual data.

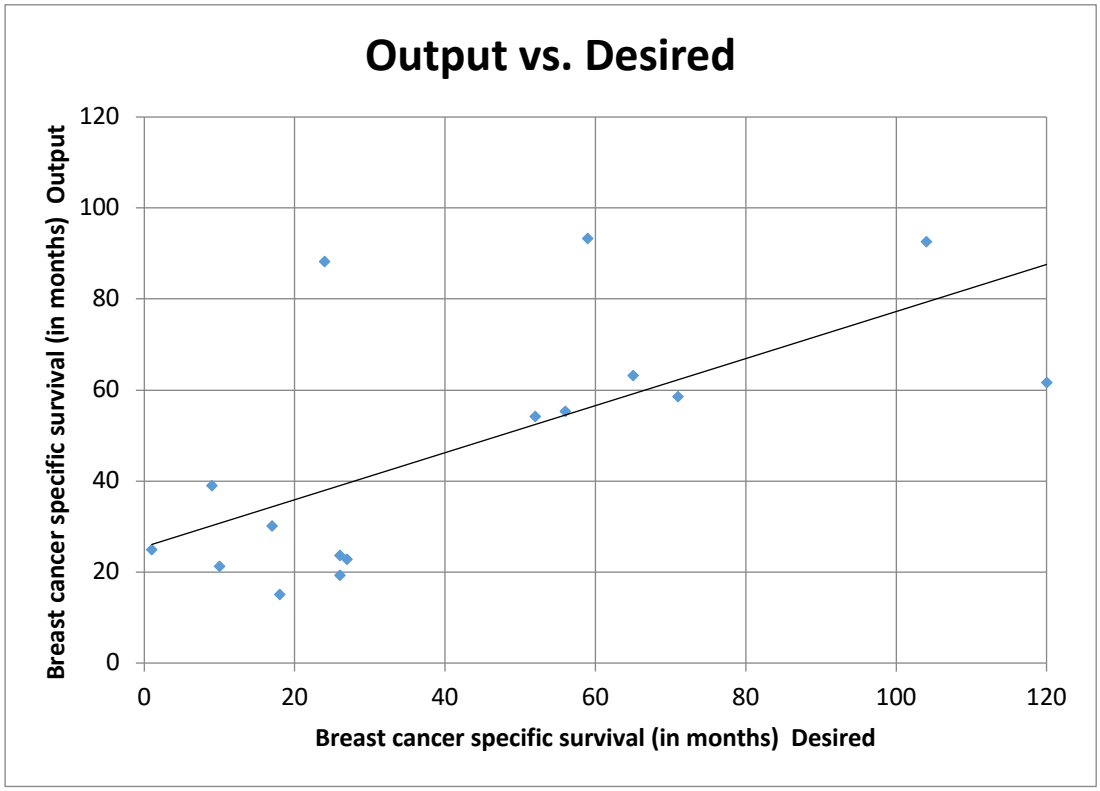


Figure 49: Actual BCSS compared to predicted BCSS in CNB TMAs

There is a moderate correlation between the prognostic model for BCSS measured in CNB TMAs and actual result.

8.3.2.2 CNB TMAs – overall survival

A neural network consisting of 22 hidden nodes was determined to be the most accurate predictor of OS.

Table 27 shows the best performance of the algorithm on the test data.

Performance	**Overall survival (in months)**
RMSE	40.1
NRMSE	0.2
MAE	31.0
NMAE	0.2
Min Abs Error	0.3
Max Abs Error	137.5
r	0.1

Table 27: Performance of the algorithm designed to predict OS in CNB TMAs

Figure 50 depicts the association between the estimated prediction of OS as determined by the algorithm compared to actual data.

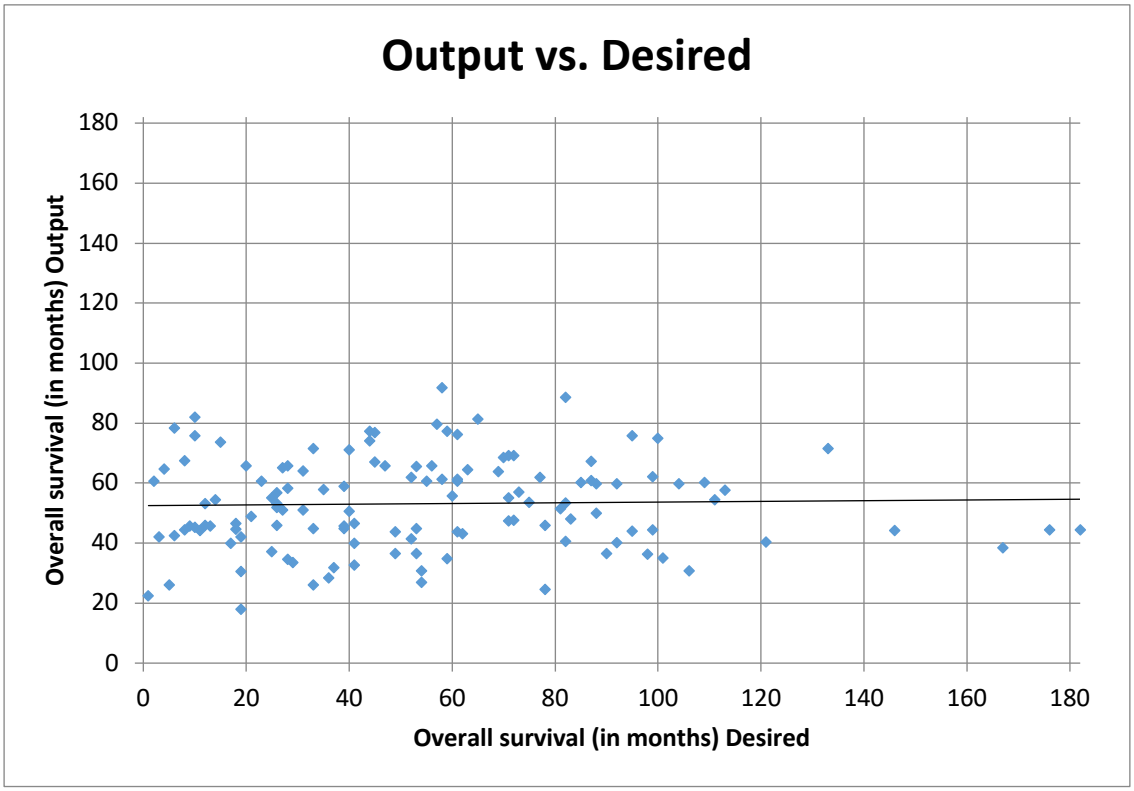


Figure 50: Actual OS compared to predicted OS in CNB TMAs

There is a poor correlation between the prognostic model for OS measured in CNB TMAs and actual result.

8.3.2.3 CNB TMAs – time to metastasis

A neural network consisting of 11 hidden nodes was determined to be the most accurate predictor of TTM.

Table 28 shows the best performance of the algorithm on the test data.

Performance	Time to metastasis (in months)
RMSE	36.5
NRMSE	0.2
MAE	29.7
NMAE	0.2
Min Abs Error	1.0
Max Abs Error	129.0
r	0.073

Table 28: Performance of the algorithm designed to predict TTM in CNB TMAs

Figure 51 depicts the association between the estimated prediction of TTM as determined by the algorithm compared to actual data.

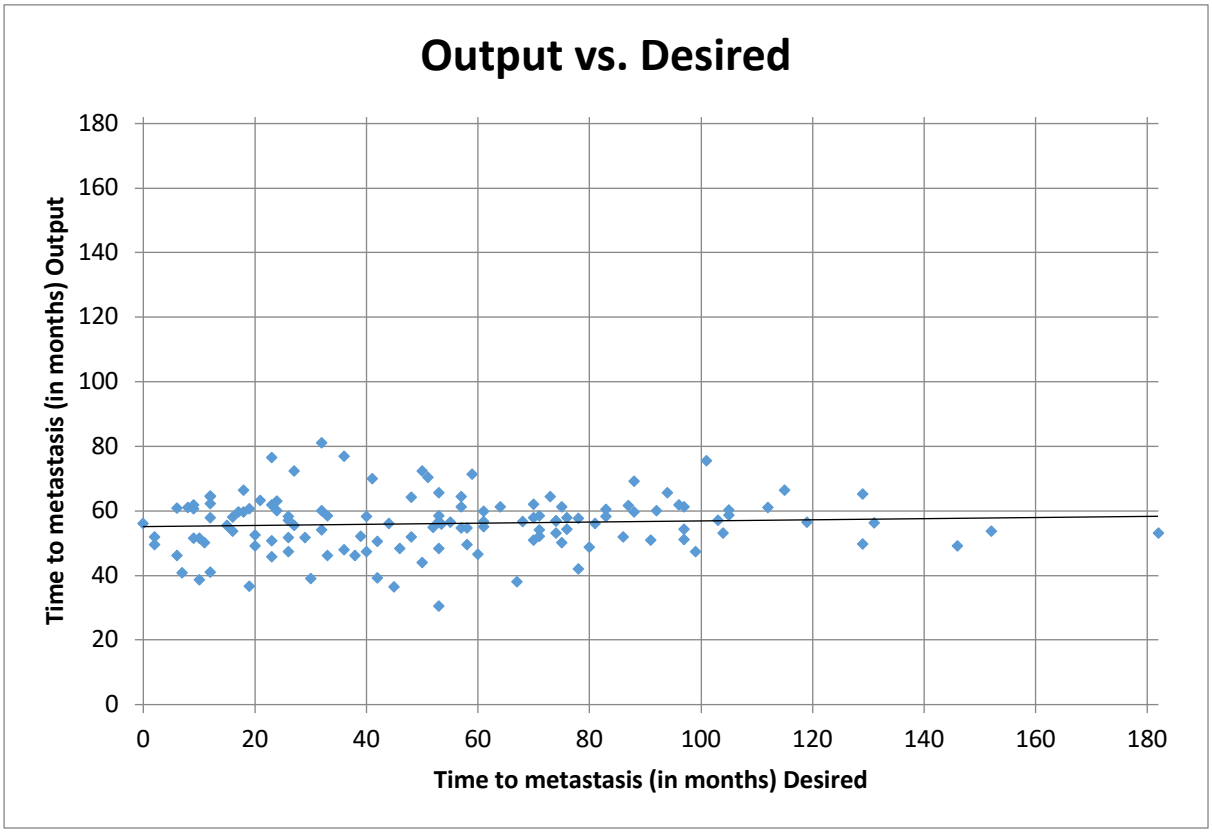


Figure 51: Actual TTM compared to predicted TTM in CNB TMAs

There is a poor correlation between the prognostic model for TTM measured in CNB TMAs and actual result.

8.3.2.4 CNB TMAs – time to regional recurrence

A neural network consisting of four hidden nodes was determined to be the most accurate predictor of TRR.

Table 29 shows the best performance of the algorithm on the test data.

Performance	Time to regional recurrence (in months)
RMSE	33.8
NRMSE	0.2
MAE	25.8
NMAE	0.1
Min Abs Error	0.3
Max Abs Error	114.3
r	0.281

Table 29: Performance of the algorithm designed to predict TRR in CNB TMAs

Figure 52 depicts the association between the estimated prediction of TRR as determined by the algorithm compared to actual data.

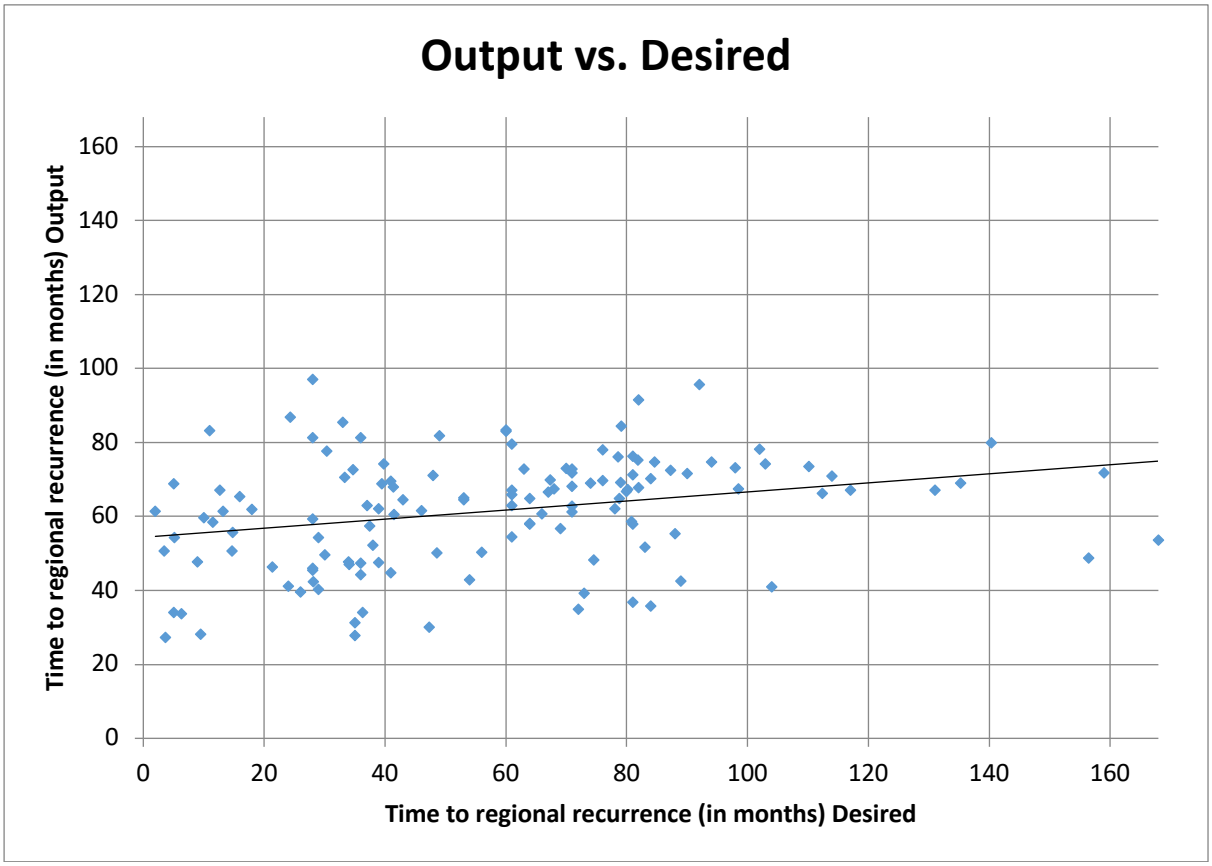


Figure 52: Actual TRR compared to predicted TRR in CNB TMAs

There is a poor correlation between the prognostic model for TRR measured in CNB TMAs and actual result.

8.3.2.5 CNB TMAs – time to progression of disease

A neural network consisting of three hidden nodes was determined to be the most accurate predictor of time to TTP.

Table 30 shows the best performance of the algorithm on the test data.

Performance	Time to disease progression (in months)
RMSE	35.3
NRMSE	0.2
MAE	27.8
NMAE	0.2
Min Abs Error	0.1
Max Abs Error	141.8
r	0.097

Table 30: Performance of the algorithm designed to predict TTP in CNB TMAs

Figure 53 depicts the association between the estimated prediction of TTP as determined by the algorithm compared to actual data.

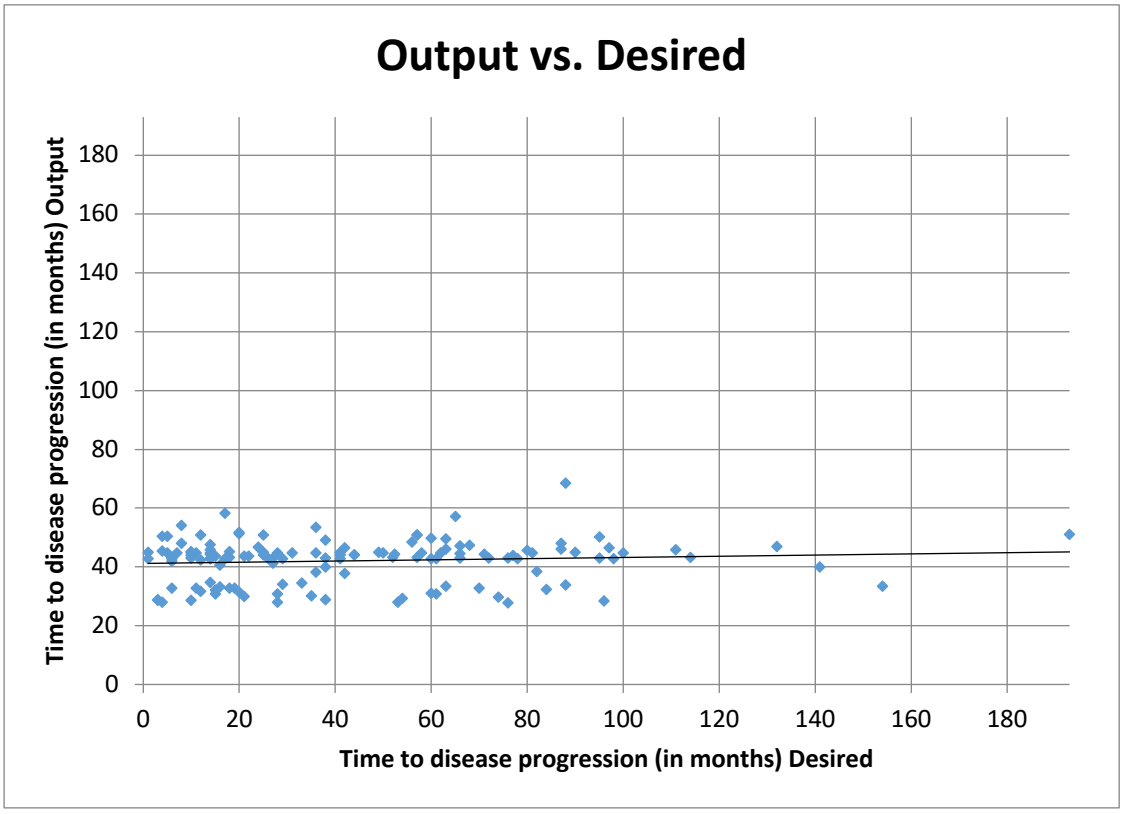


Figure 53: Actual TTP compared to predicted TTP in CNB TMA

There is a poor correlation between the prognostic model for time to RR measured in CNB TMAs and actual result.

8.4 Further results

The algorithms generated to predict BCSS in both the SE TMAs and CNB TMAs showed a strong correlation between the algorithm and real-life data. This is really exciting and important for older women with breast cancer as they often have competing comorbidities and therefore causes of death, which contribute to OS. Therefore, OS may not be the most appropriate marker of outcome to use in the older population; BCSS is a surrogate for tumour biology and may be more relevant when making treatment decisions. As the algorithms generated are based on tumour biology by a panel of biomarkers, that fact that the models for BCSS have a strong correlation compared to the other disease outcomes

measured, is reassuring and compounds this theory. I will now explore the two algorithms relating to BCSS in further detail.

8.4.1 Algorithm to predict BCSS in SE TMA BCSS

The formula for the algorithm is given in Appendix 5. Figure 54 shows how each biomarker within the model is contributing to the model and to what extent. The y-axis labelled 'sensitivity' demonstrates the relative impact on the model of removing an input; the greater the value, the more important the feature. For example, in figure 54, PR is 14x more important than p53 within the model. In other words, if PR was to be removed from the model, this would have a significant impact on the output, however, if p53 was to be removed from the model, this would have little impact.

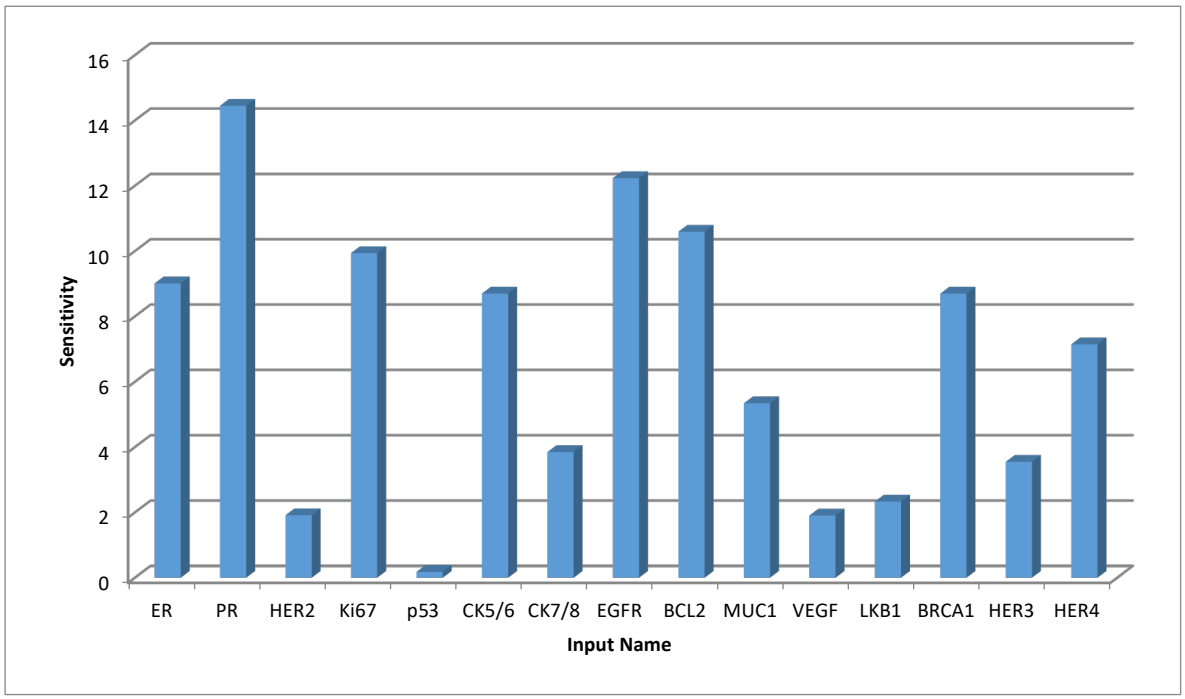


Figure 54: Contribution of each biomarker to model predicting BCSS in SE TMAs

Figures 55 – 69 show the relationship between each biomarker and BCSS in the model. The expression of the marker is given by H-score on the x-axis (scale 0-300) and BCSS is

given in months on the y-axis. These graphs are generated from 'test' data taken from the overall database and input into the algorithm; the overall number of data points is much smaller than the data used for 'training' of the algorithms. Therefore, BCSS data may not be presented at the extremes of H-scores, as the reliability of this data is at the present time, unpredictable.

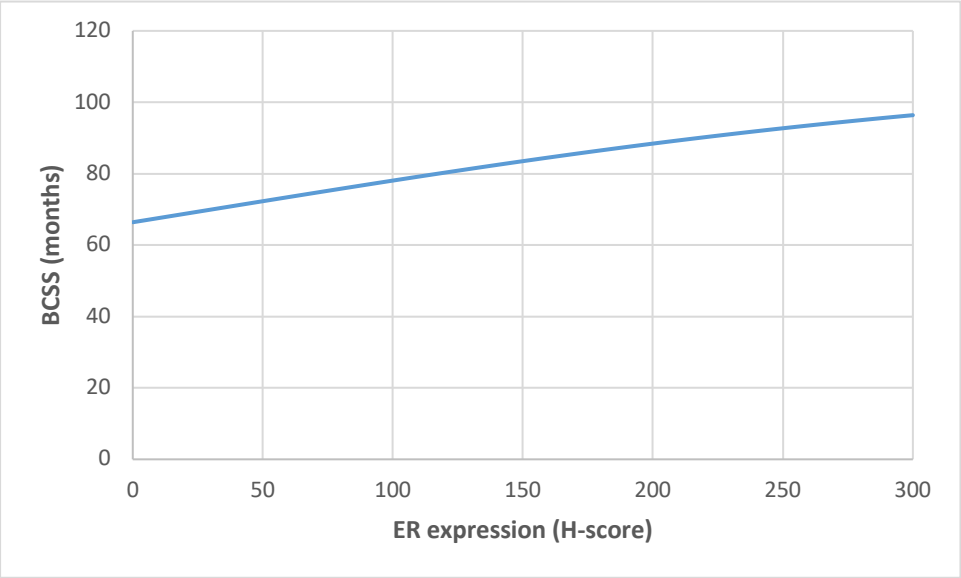


Figure 55: Relationship between ER expression and ability to predict BCSS within the model

Within this model, increasing ER positivity was associated with improved BCSS.

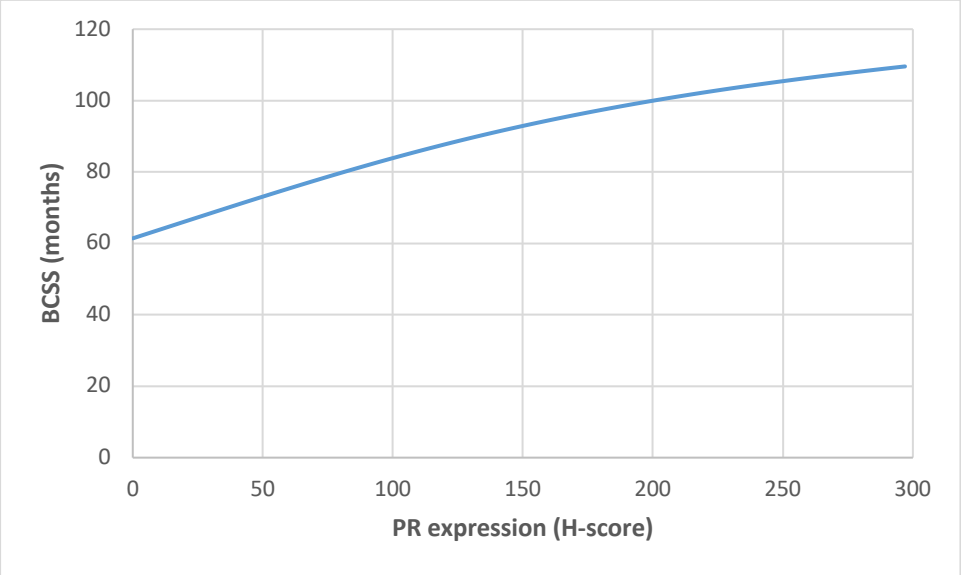


Figure 56: Relationship between PR expression and ability to predict BCSS within the model

Within this model, increasing PR positivity was associated with improved BCSS.



Figure 57: Relationship between HER2 expression and ability to predict BCSS within the model

Within this model, HER2 expression did not significantly contribute to BCSS.

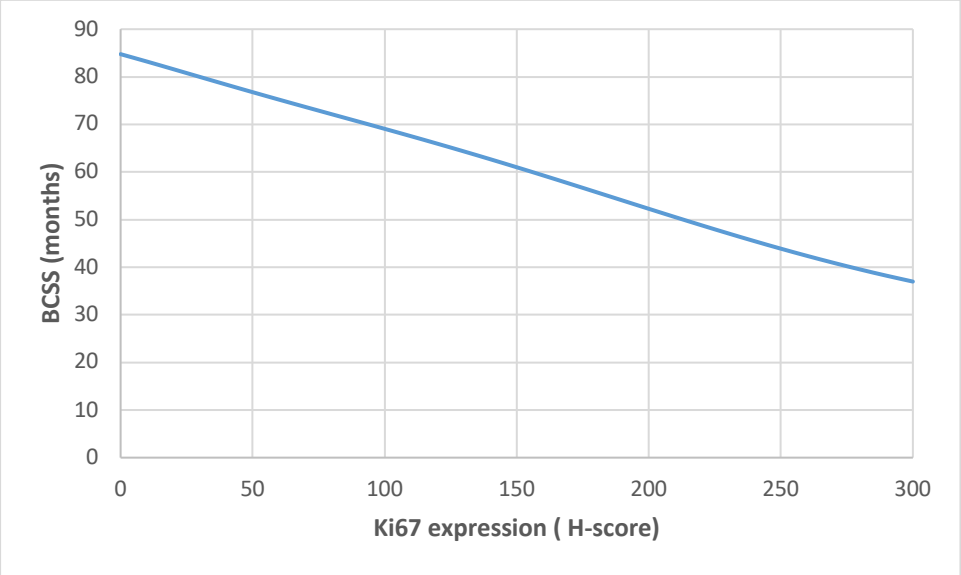


Figure 58: Relationship between Ki67 expression and ability to predict BCSS within the model

Within this model, increasing Ki67 positivity was associated with reduced BCSS.

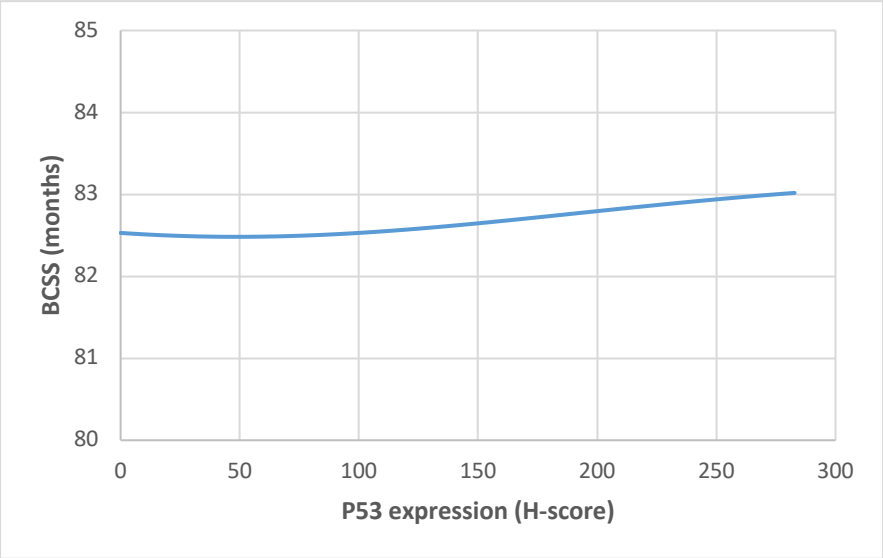


Figure 59: Relationship between P53 expression and ability to predict BCSS within the model

Within this model, expression of P53 did not have a significant association with BCSS.

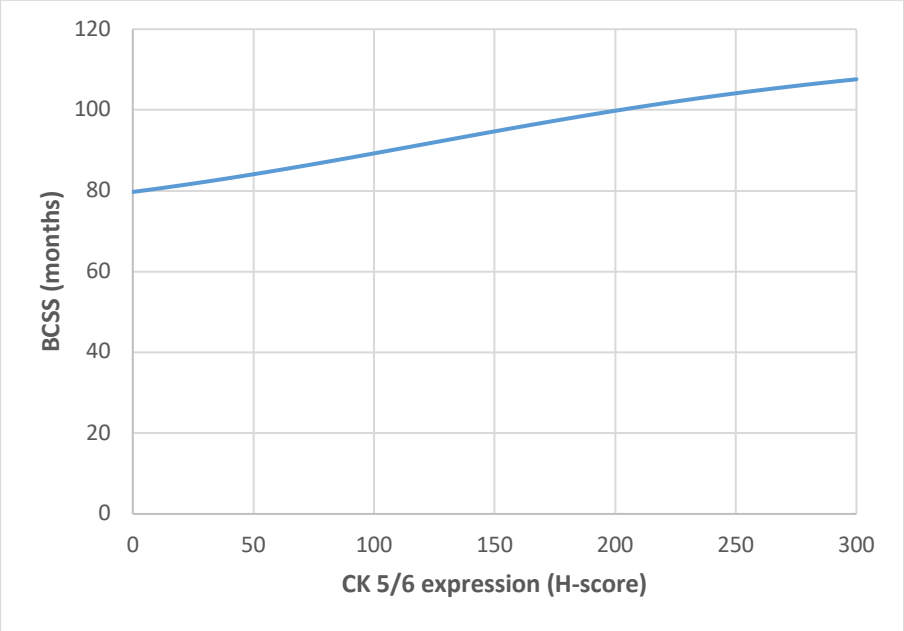


Figure 60: Relationship between CK 5/6 expression and ability to predict BCSS within the model

Within this model, increasing CK 5/6 positivity was associated with improved BCSS.

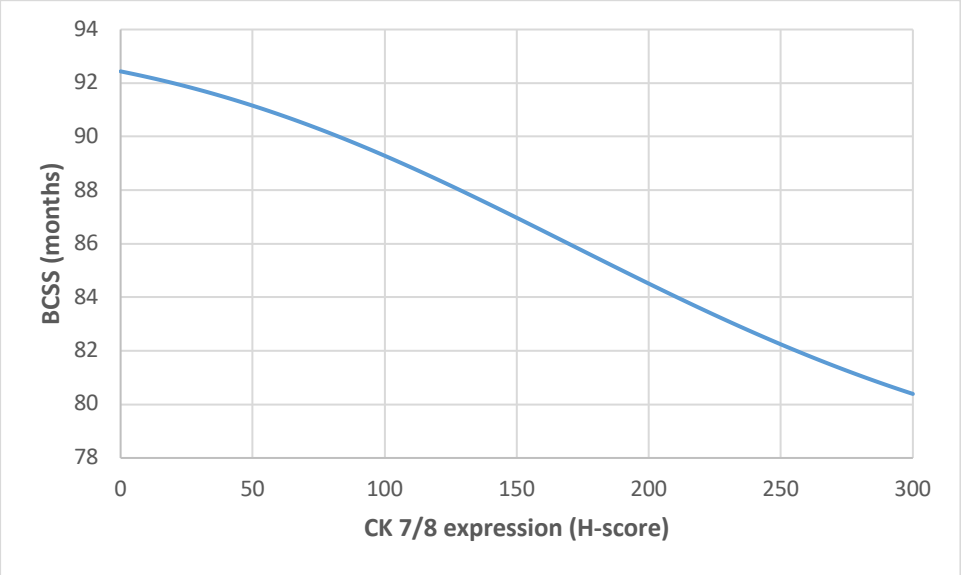


Figure 61: Relationship between CK 7/8 expression and ability to predict BCSS within the model

Within this model, increasing CK 7/8 positivity was associated with a decrease in BCSS.

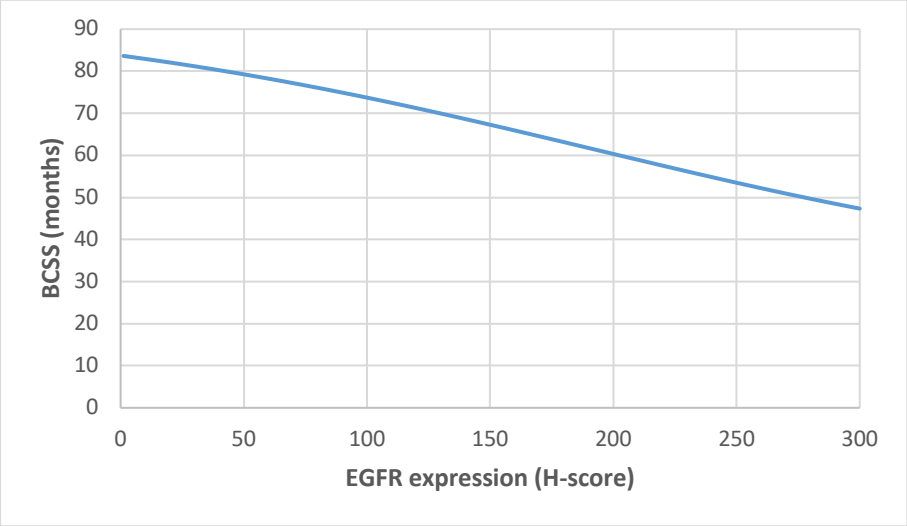


Figure 62: Relationship between EGFR expression and ability to predict BCSS within the model

Within this model, increasing EGFR positivity was associated with a decrease in BCSS.

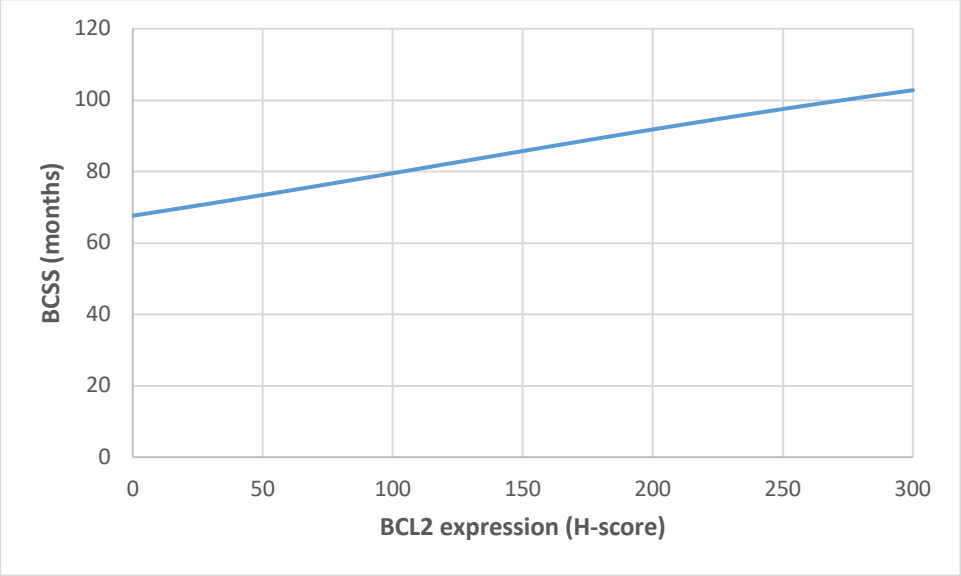


Figure 63: Relationship between BCL2 expression and ability to predict BCSS within the model

Within this model, increasing BCL2 positivity was associated with an increase in BCSS.

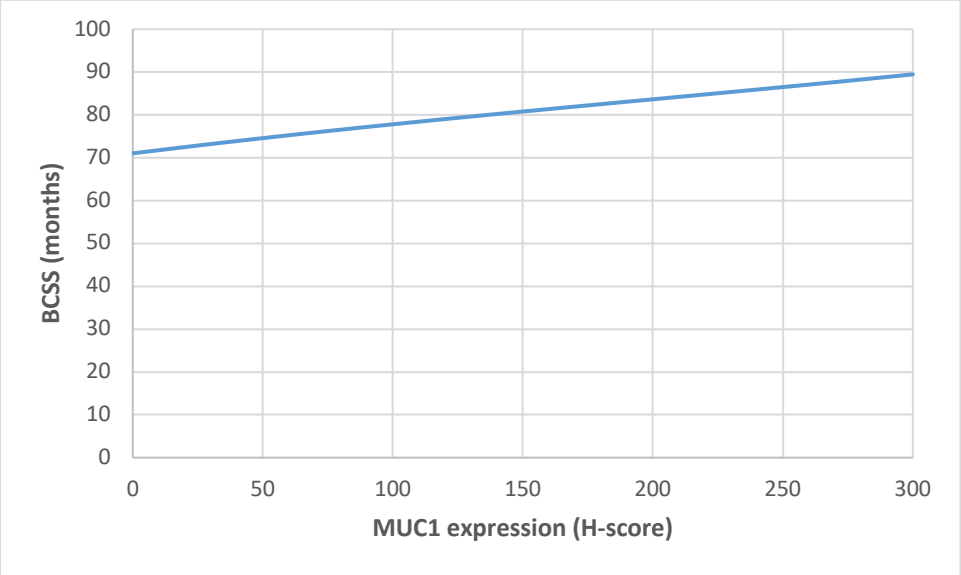


Figure 64: Relationship between MUC1 expression and ability to predict BCSS within the model

Within this model, increasing MUC1 positivity was associated with an increase in BCSS.

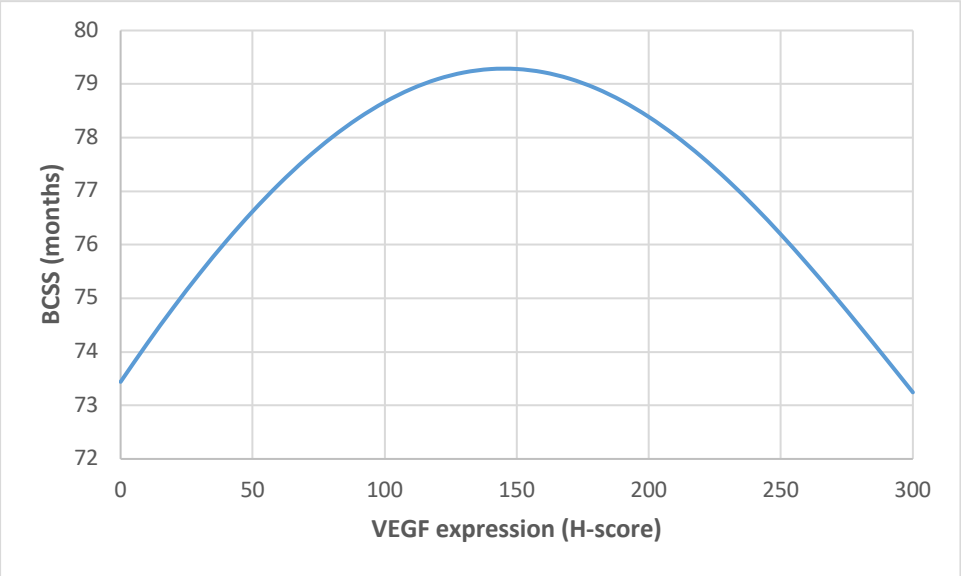


Figure 65: Relationship between VEGF expression and ability to predict BCSS within the model

Within this model, increasing VEGF positivity was associated with an increase in BCSS up to a specific point after which further VEGF positivity was associated with a decreased BCSS.

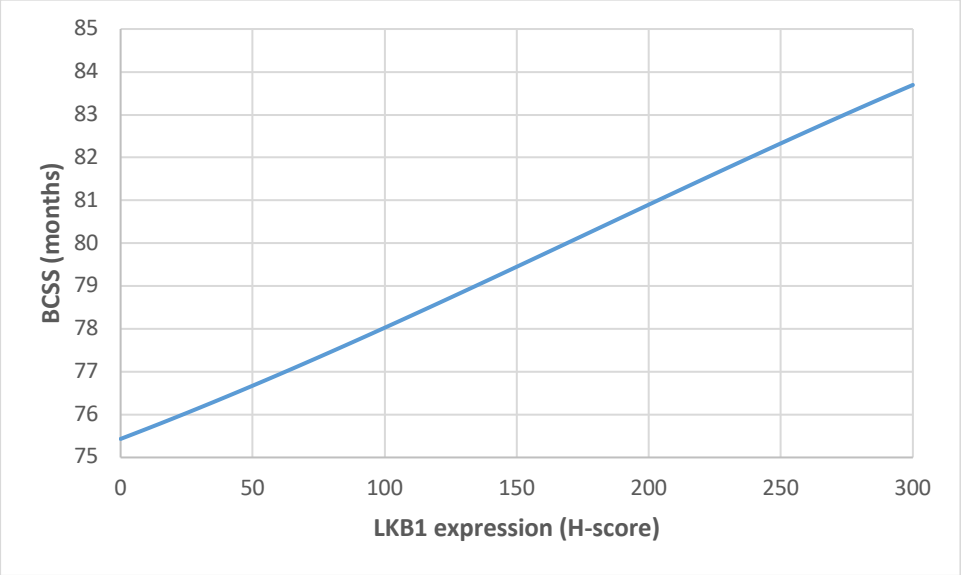


Figure 66: Relationship between LKB1 expression and ability to predict BCSS within the model

Within this model, increasing LKB1 positivity was associated with an increase in BCSS.

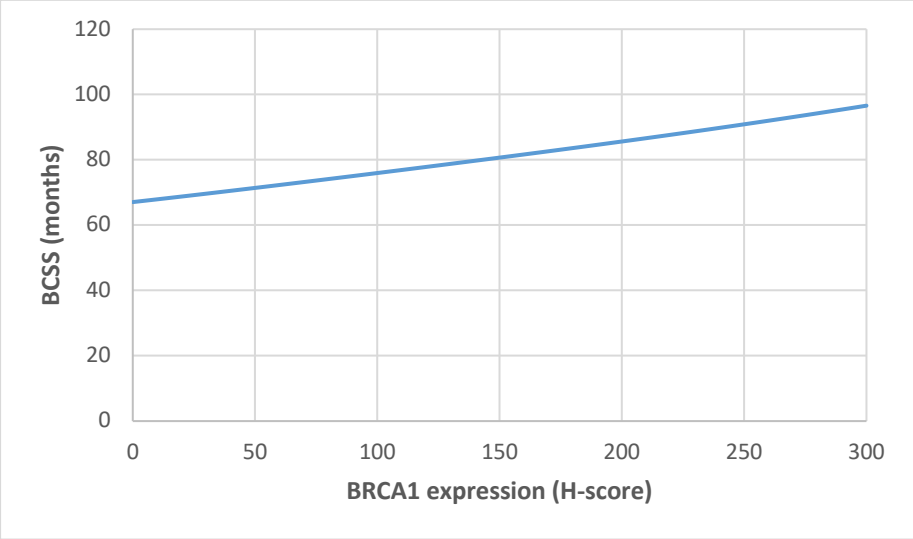


Figure 67: Relationship between BRCA1 expression and ability to predict BCSS within the model

Within this model, increasing BRCA1 positivity was associated with an increase in BCSS.

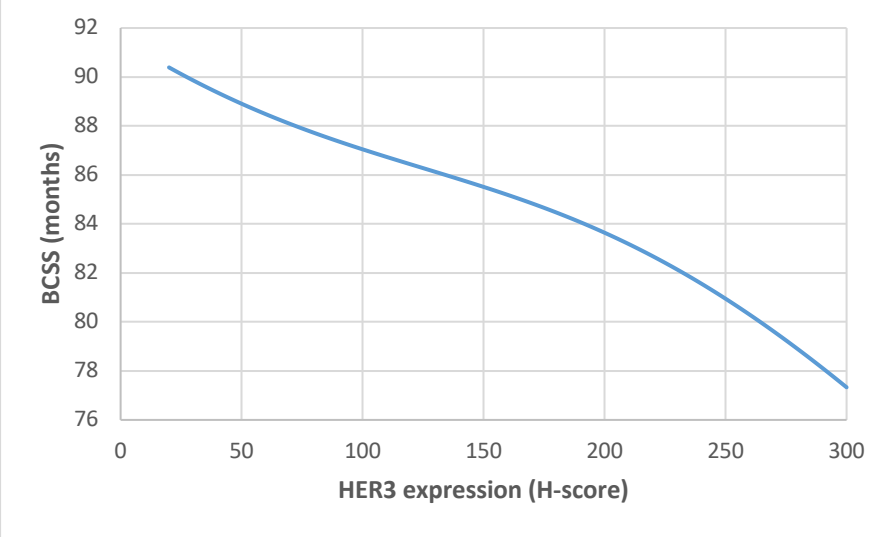


Figure 68: Relationship between HER3 expression and ability to predict BCSS within the model

Within this model, increasing HER3 positivity was associated with a decrease in BCSS.

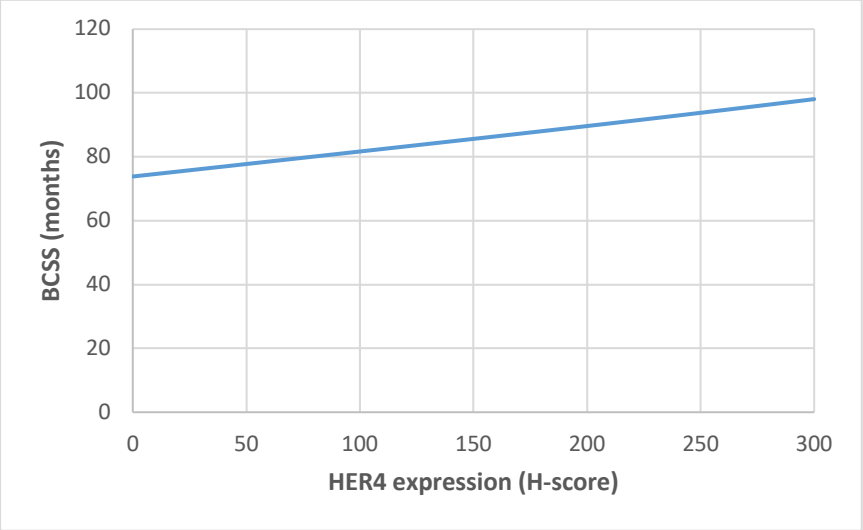


Figure 69: Relationship between HER4 expression and ability to predict BCSS within the model

Within this model, increasing HER4 positivity was associated with an increase in BCSS.

8.4.1 Algorithm to predict BCSS in CNB TMA

The formula for the algorithm is given in Appendix 6.

Figure 70 shows how each biomarker within the model is contributing to the model and to what extent.

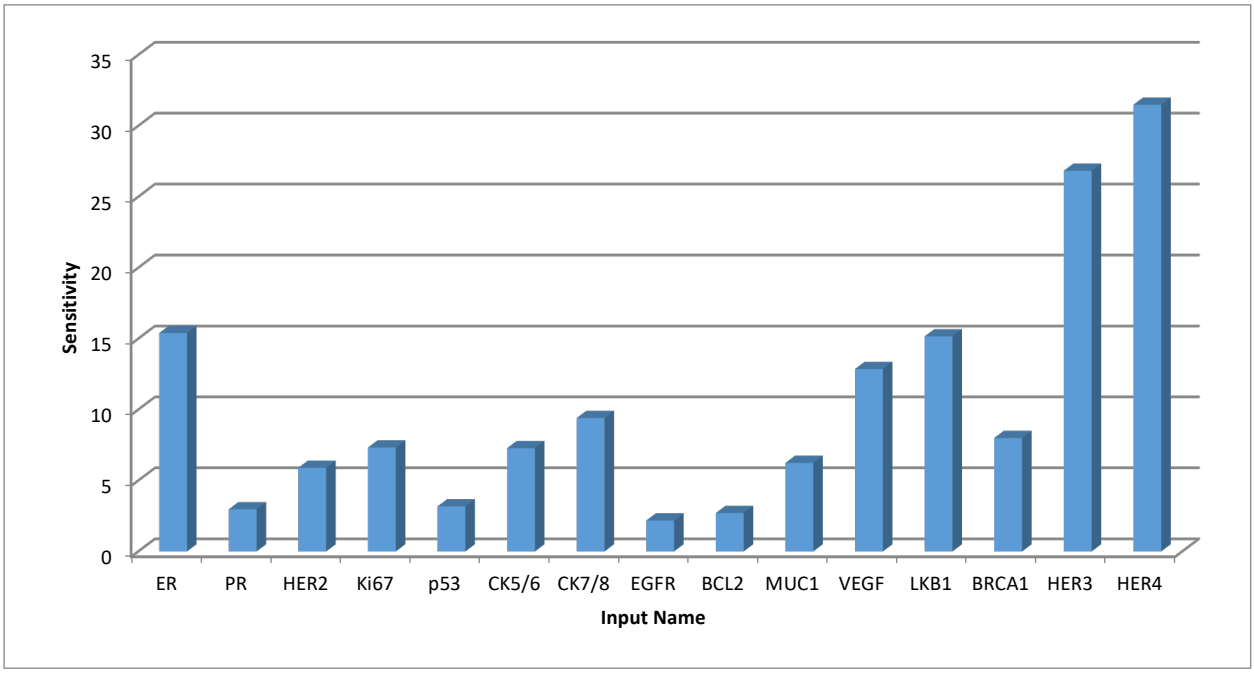


Figure 70: Contribution of each biomarker to model predicting BCSS in CNB TMAs

Figures 71 – 85 show the relationship between each biomarker and BCSS in the model.

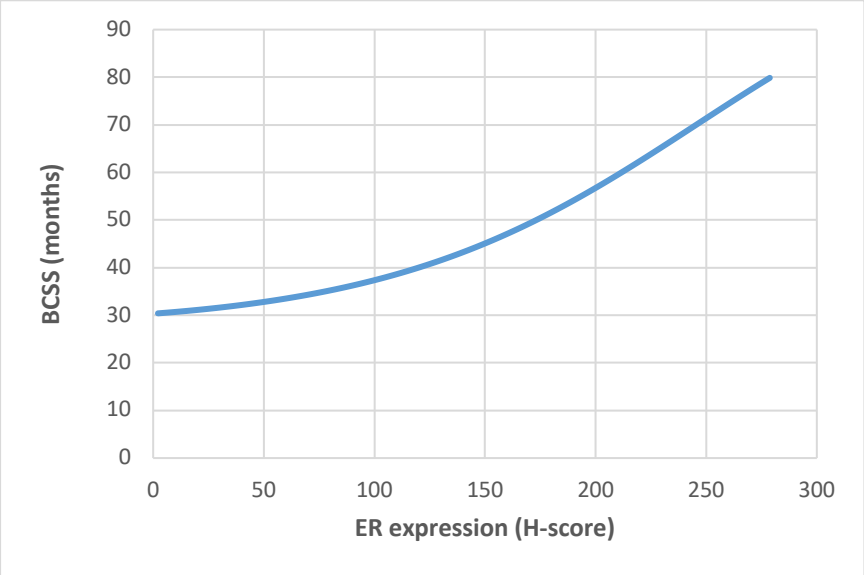


Figure 71: Relationship between ER expression and ability to predict BCSS within the model

Within this model, increasing ER positivity was associated with improved BCSS.

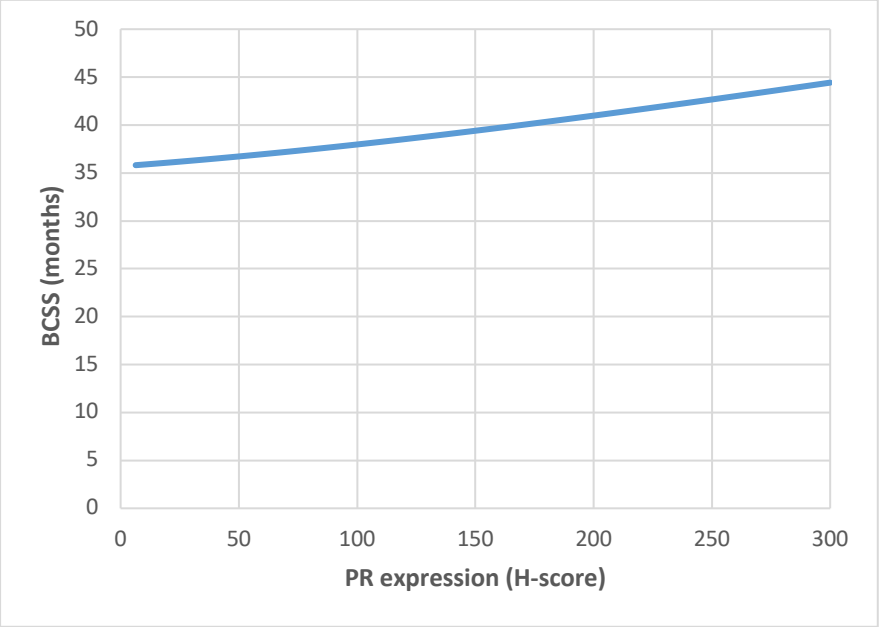


Figure 72: Relationship between PR expression and ability to predict BCSS within the model

Increasing ER positivity was associated with a small improvement in BCSS.



Figure 73: Relationship between HER2 expression and ability to predict BCSS within the model

Within this model, increasing HER2 positivity was associated with a decrease in BCSS.

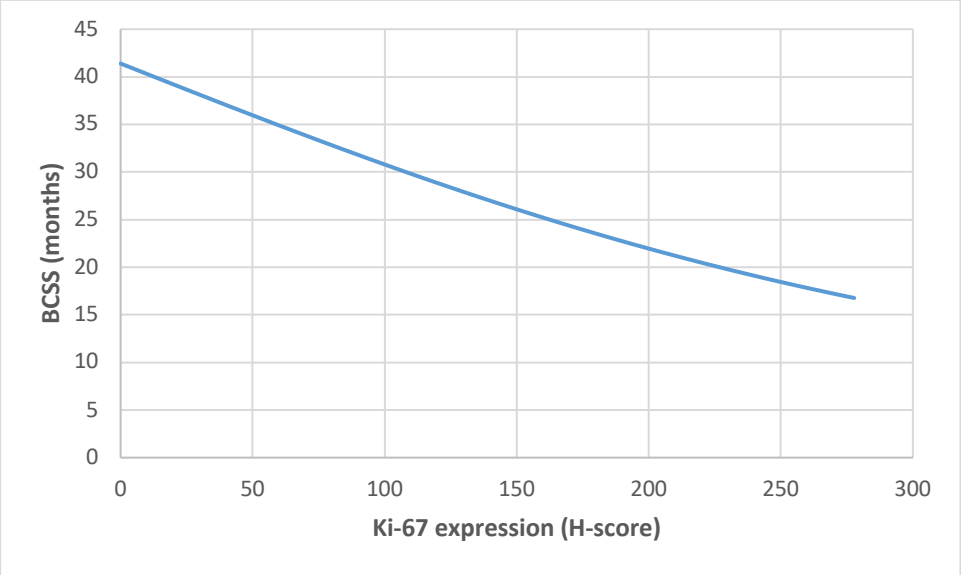


Figure 74: Relationship between Ki-67 expression and ability to predict BCSS within the model

Within this model, increasing Ki-67 positivity was associated with a decrease in BCSS.

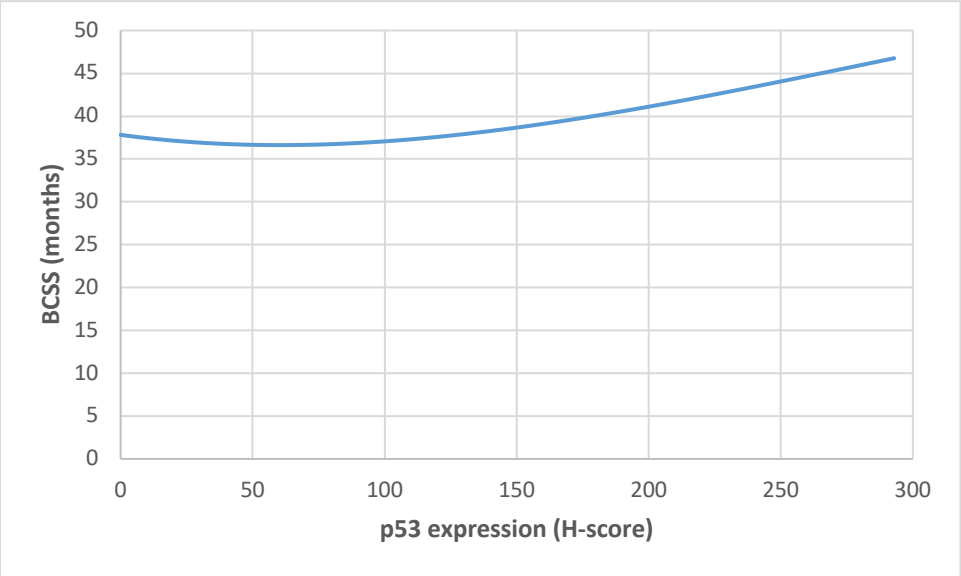


Figure 75: Relationship between p53 expression and ability to predict BCSS within the model

Increase in p53 expression is associated with a slight increase in BCSS.

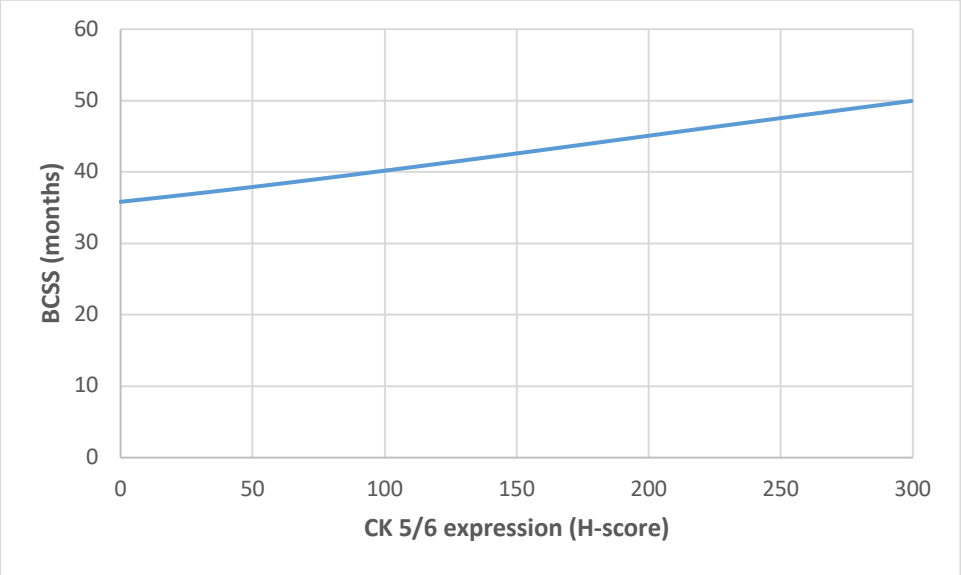


Figure 76: Relationship between CK 5/6 expression and ability to predict BCSS within the model

An increase in CK 5/6 expression was associated with increased BCSS.

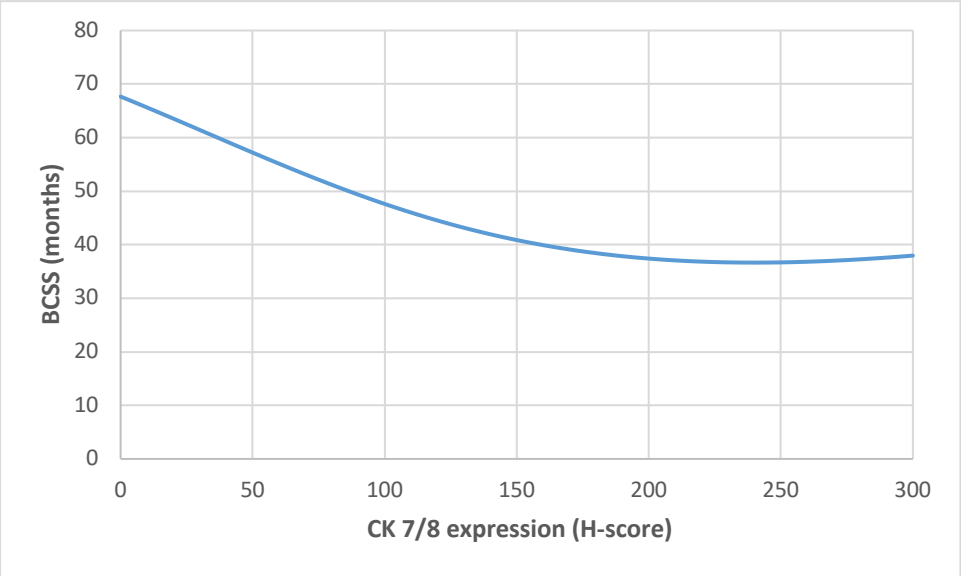


Figure 77: Relationship between CK 7/8 expression and ability to predict BCSS within the model

An increase in CK 7/8 expression resulted in a slight decrease in BCSS.

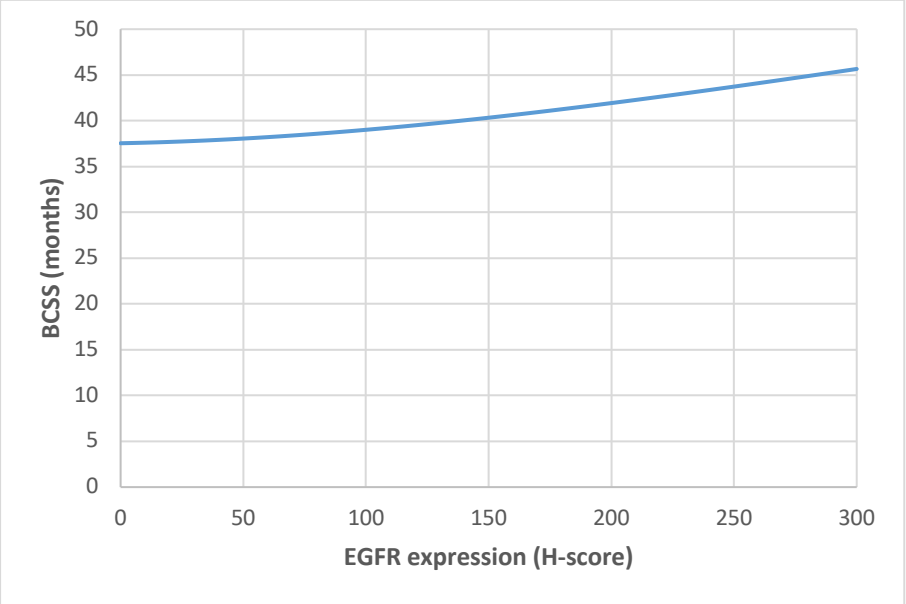


Figure 78: Relationship between EGFR expression and ability to predict BCSS within the model

An increase in EGFR expression was associated with improved BCSS within the model.

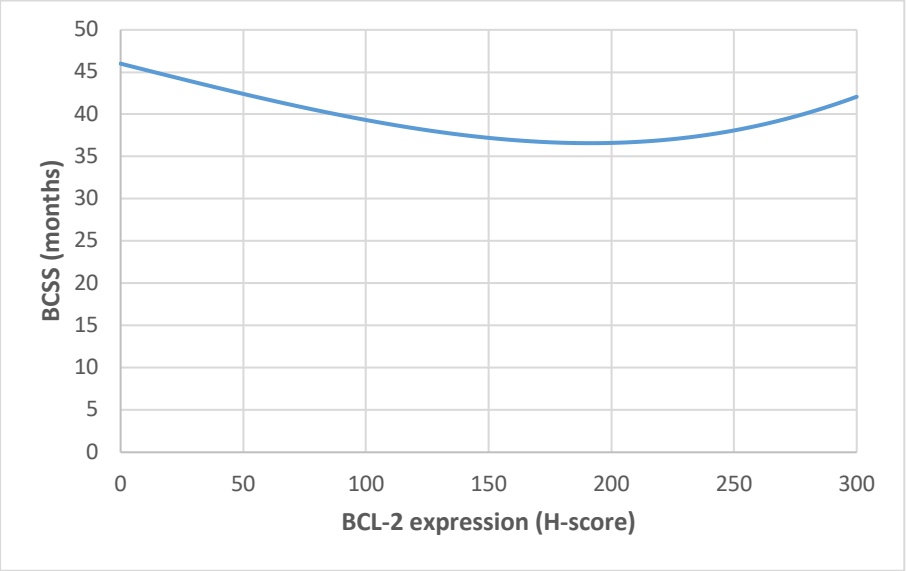


Figure 79: Relationship between BCL-2 expression and ability to predict BCSS within the model

BCSS decreased with increased expression of BCL-2 to a certain point and then had the inverse outcome.

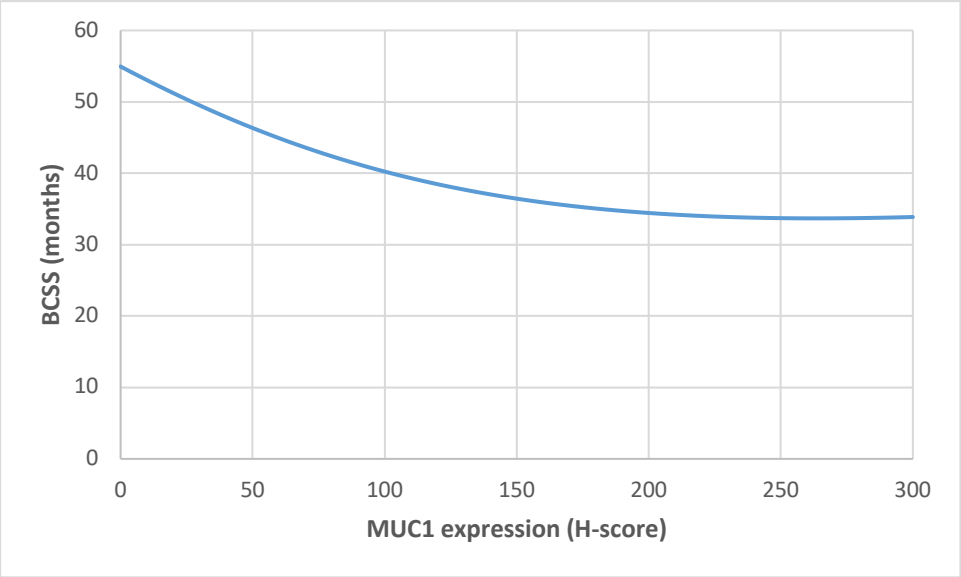


Figure 80: Relationship between MUC1 expression and ability to predict BCSS within the model

An increase in MUC1 expression was associated with a slight decrease in BCSS.

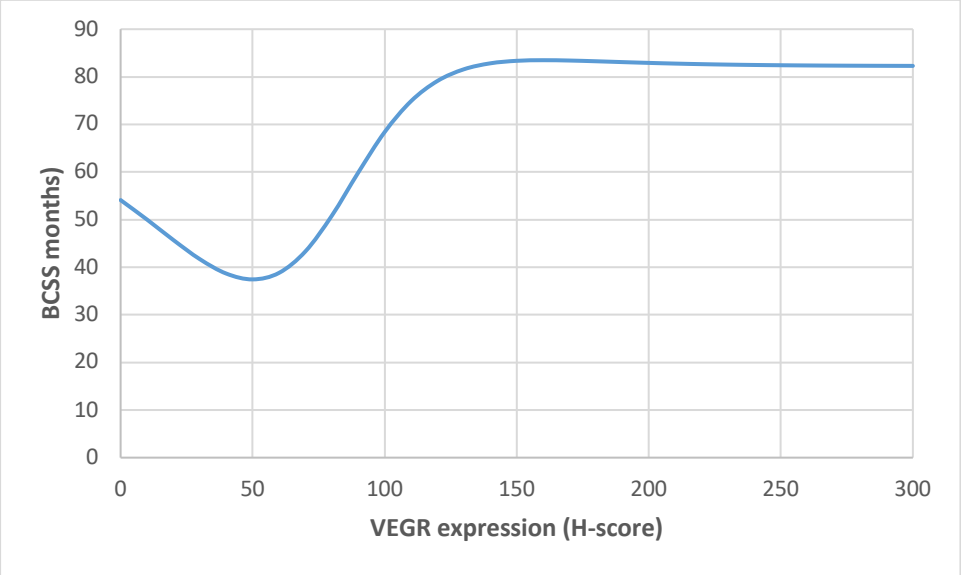


Figure 81: Relationship between VEGF expression and ability to predict BCSS within the model

The model indicated that an increase in VEGF resulted in a decrease in BCSS up to a certain H-score and then saw the inverse effect which plateaued at the higher range of H-scores.

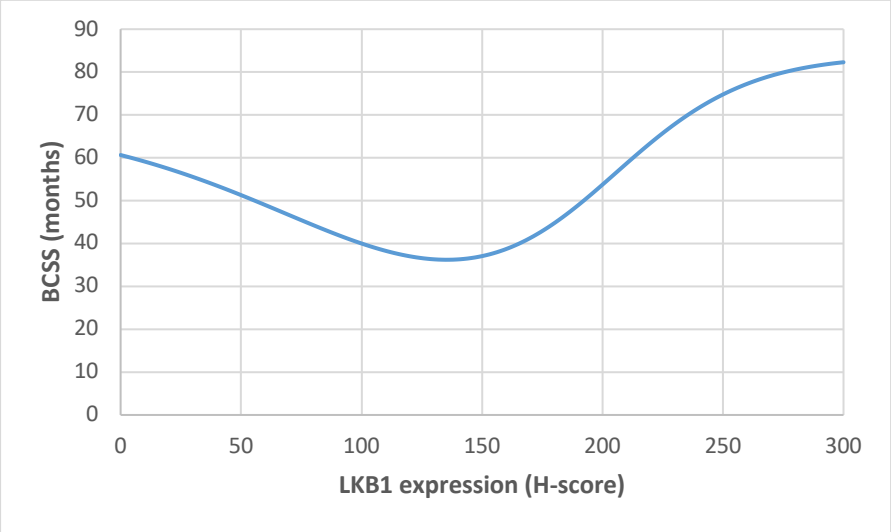


Figure 82: Relationship between LKB1 expression and ability to predict BCSS within the model

An increase in LKB1 expression was associated with improved BCSS within the model to a certain H-score after which the inverse was seen.

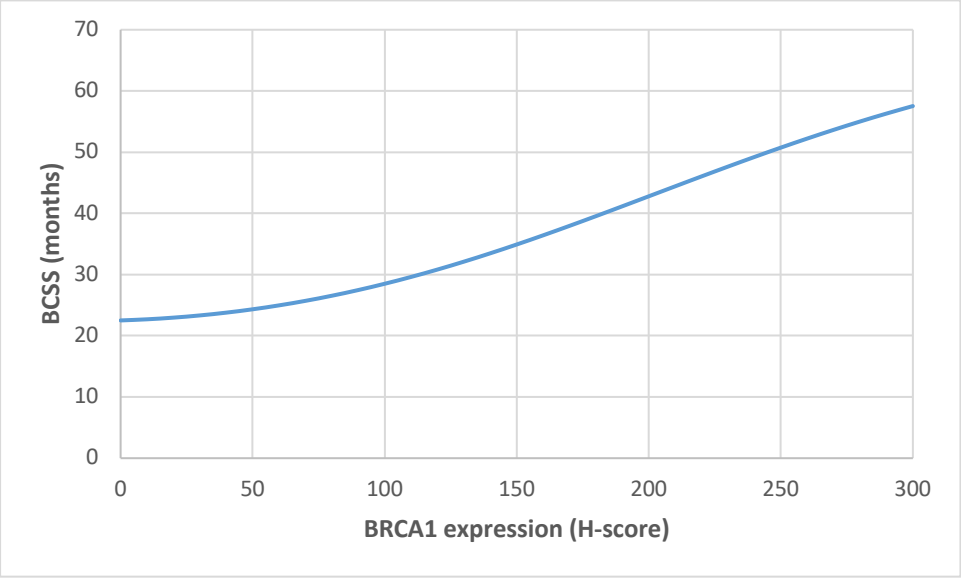


Figure 83: Relationship between BRCA1 expression and ability to predict BCSS within the model

An increase in BRCA1 expression was associated with increased BCSS.

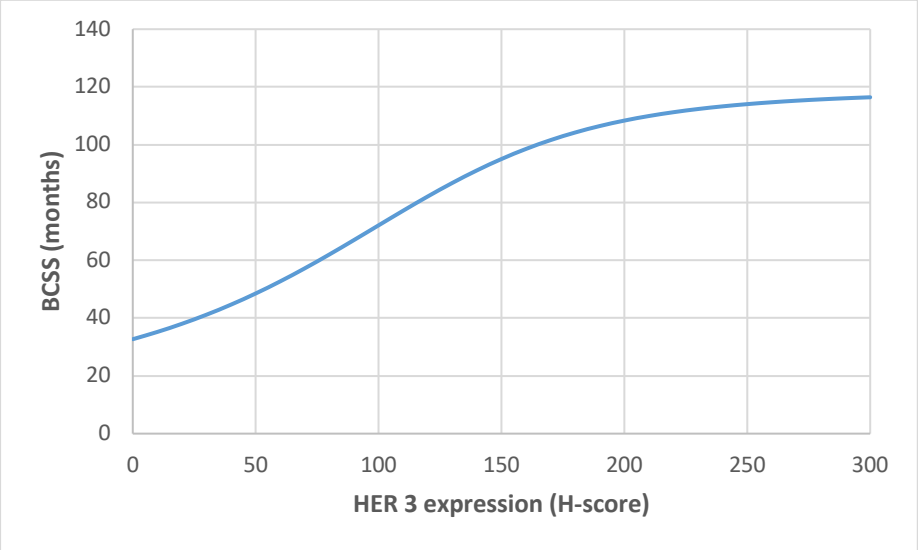


Figure 84: Relationship between HER3 expression and ability to predict BCSS within the model

An increase in HER3 expression in the model was associated with increased BCSS.

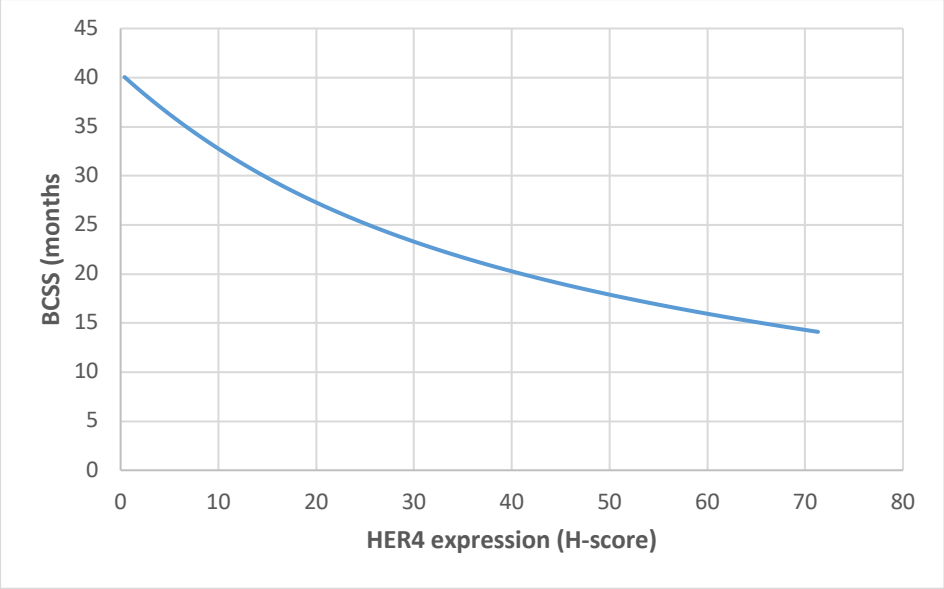


Figure 85: Relationship between HER4 expression and ability to predict BCSS within the model

An increase in HER4 expression was associated with decreased BCSS within the model.

8.5 Explanation of results

Table 31 summarises the findings and whether they are consistent with the current literature (as detailed in Chapter 3).

Biomarker	Association between +ve expression and BCSS on SE TMA	Findings consistent with the literature?	Association between +ve expression and BCSS on CNB TMA	Findings consistent with the literature?
ER	Increased BCSS	Yes	Increased BCSS	Yes
PR	Increased BCSS	Yes	Increased BCSS	Yes
HER2	Decreased BCSS	Yes	Decreased BCSS	Yes
Ki67	Decreased BCSS	Yes	Decreased BCSS	Yes
P53	Unclear	-	Increased BCSS	Yes
CK 5/6	Increased BCSS	Yes	Increased BCSS	Yes
CK 7/8	Decreased BCSS	Yes	Decreased BCSS	Yes
EGFR	Decreased BCSS	Yes	Increased BCSS	No
BCL2	Increased BCSS	No	Unclear	-
MUC1	Increased BCSS	No	Decreased BCSS	Yes
VEGF	Unclear	-	Unclear	-
LKB1	Increased BCSS	No	Unclear	-
BRCA1	Increased BCSS	No	Increased BCSS	No
HER3	Decreased BCSS	No	Increased BCSS	No
HER4	Increased BCSS	Yes	Decreased BCSS	No

Table 31: Summary of the association between each biomarker within the prognostic algorithms generated and BCSS

In the CNB TMAs, the contribution of ER, PR, Ki67, p53 CK 5/6, CK 7/8 and MUC1 to the model that predicts BCSS were consistent with the current literature (as per Chapter 3), but EGFR, BCL2, VEGF, LKB1, BRCA1, HER3 and HER4 were not.

In the SE TMAs, the contribution of ER, PR, Ki67, CK 5/6, CK 7/8, EGFR and HER4 were consistent with the current literature (Chapter 3), whereas p53, BCL2, MUC1, VEGF, LKB1, BRCA1 and HER3 were not.

There are a number of potential explanations for this including both methodological and biological factors.

Methodologically, this could mean that a particular biomarker is not a significant contributor to the overall model and is confounding the overall result; the accuracy of the model may improve by removing these markers. This could be a possible explanation for BCL2 and EGFR within the CNB TMA model (Figure 54) and MUC1, HER2, LKB1, VEGF and p53 in the SE TMA model (Figure 70); they have the lowest contributions to the models, however, does not uniformly explain the discrepancies; HER3, HER4, LKB1 and VEGF contribute significantly to the CNB TMA model and BCL2 and BRCA1 to the SE TMA model.

Biologically, this could be related to the discrepancies in staining between CNB TMAs and SE TMAs as discussed in Chapter 6. Furthermore, it could mean that genuine differences between biomarker expressions do exist in older women, compared to younger women. As described in Chapter 3 of this thesis (Table 4) there are conflicting results in the literature regarding differing expression between younger and older women and association with survival for the markers HER3, HER4, LKB1, VEGF, BCL2 and

BRCA1. These markers have not been extensively investigated in breast cancer in older women and these findings warrant further research.

8.6 Conclusions of this chapter and next steps

This chapter has explained the bioinformatics process of generating a prognostic algorithm for breast cancer in this thesis. Prognostic algorithms have been generated in both the CNB TMAs and SE TMAs in an attempt to predict BCSS, OS, TTP, TRR and TTM. The models for prediction of BCSS in both the CNB TMAs and SE TMAs showed the greatest concordance and therefore, potential for development. This also makes sense in the context of older women with primary breast cancer – these patients have competing causes of death other than their breast cancer and measures such as OS and TTP may not be helpful to an individual patient when making treatment decisions.

The models for prediction of BCSS have been broken down and contributions of individual biomarkers explored. The next step of this thesis is to refine these algorithms to make them more accurate and therefore, potentially utilisable in a clinical setting.

Chapter 9: Refining the prognostic algorithm

9.1 Introduction

In Chapter 8 of this thesis, prototype prognostic algorithms to predict BCSS from a panel of 15 biomarkers were generated. The algorithms generated in both the SE TMAs and the CNB TMAs showed strong concordance between predicted and actual BCSS ($r = 0.742$ and 0.716); $r > 0.7$ considered 'strong' correlation.

As BCSS is particularly relevant to older women with breast cancer (compared to other measures such as OS, time to LR or RR), the next step of this thesis is to further refine these two algorithms, to try and improve accuracy. Following that, discussion on how to potentially improve the algorithm further, and next steps this research theme should take, will be presented.

9.2 Refining prognostic algorithms to predict BCSS

9.2.1 Introduction

The current algorithms have been designed from a panel of 15 biomarkers. As seen in Figures 54 and 70 not all of the biomarkers in this panel are contributing to the overall algorithms and are therefore referred to as 'noisy' biomarkers; markers which are interfering with the overall sensitivity of the algorithms.

One method to attempt to improve the accuracy of the algorithms, as measured by concordance between predicted and actual result (r value) is to optimise the panel of biomarkers to specifically look at those markers which are independently related to BCSS within the data sets.

9.2.2 Methods

Survival analysis was performed using Kaplan-Meier methods as described in section 4.7 of this thesis.

9.2.2.1 Calculating BCSS in SE TMAs

Syed et al (3) found that expression of ER, PR, BCL2 and MUC1 (Figure 86) was associated with significantly better BCSS in SE TMAs, as was the absence of HER2, EGFR, Ki67 and p53 (Figure 87) in the SE TMA cohort.

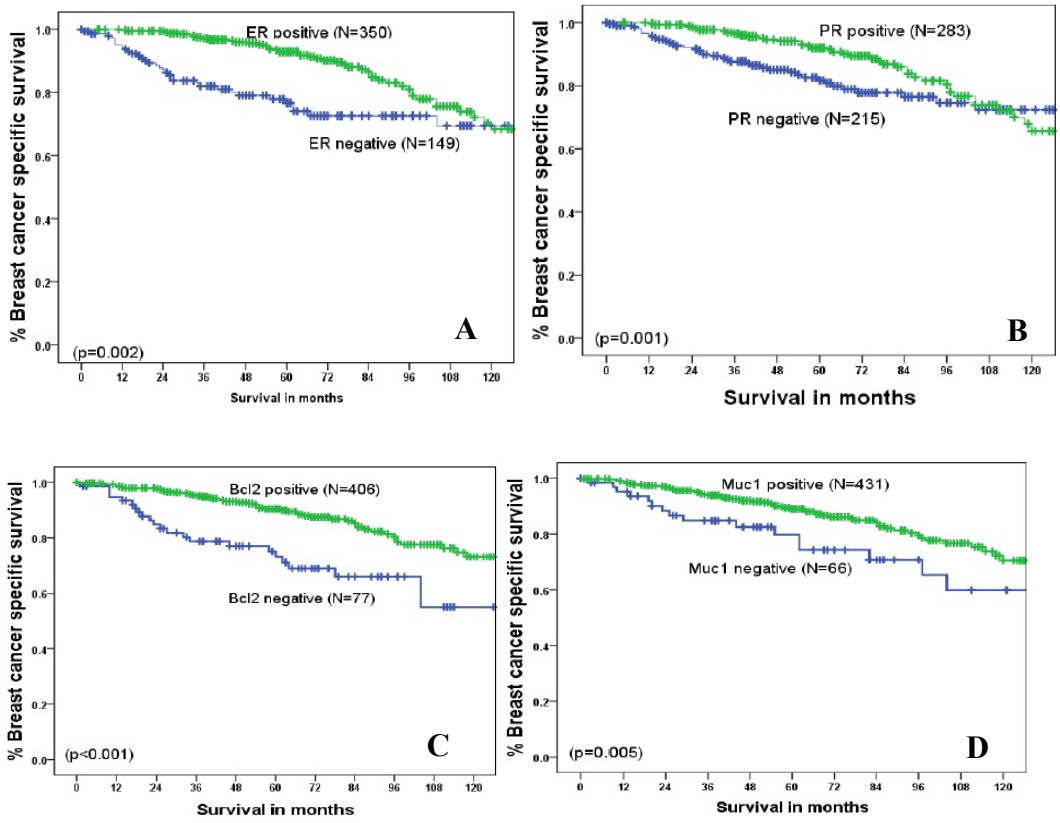


Figure 86: Prognostic significance of A) ER, B) PR, C) BCL2, D) MUC1 in early operable primary breast cancer treated by surgery – BCSS (3) - permission to use not required

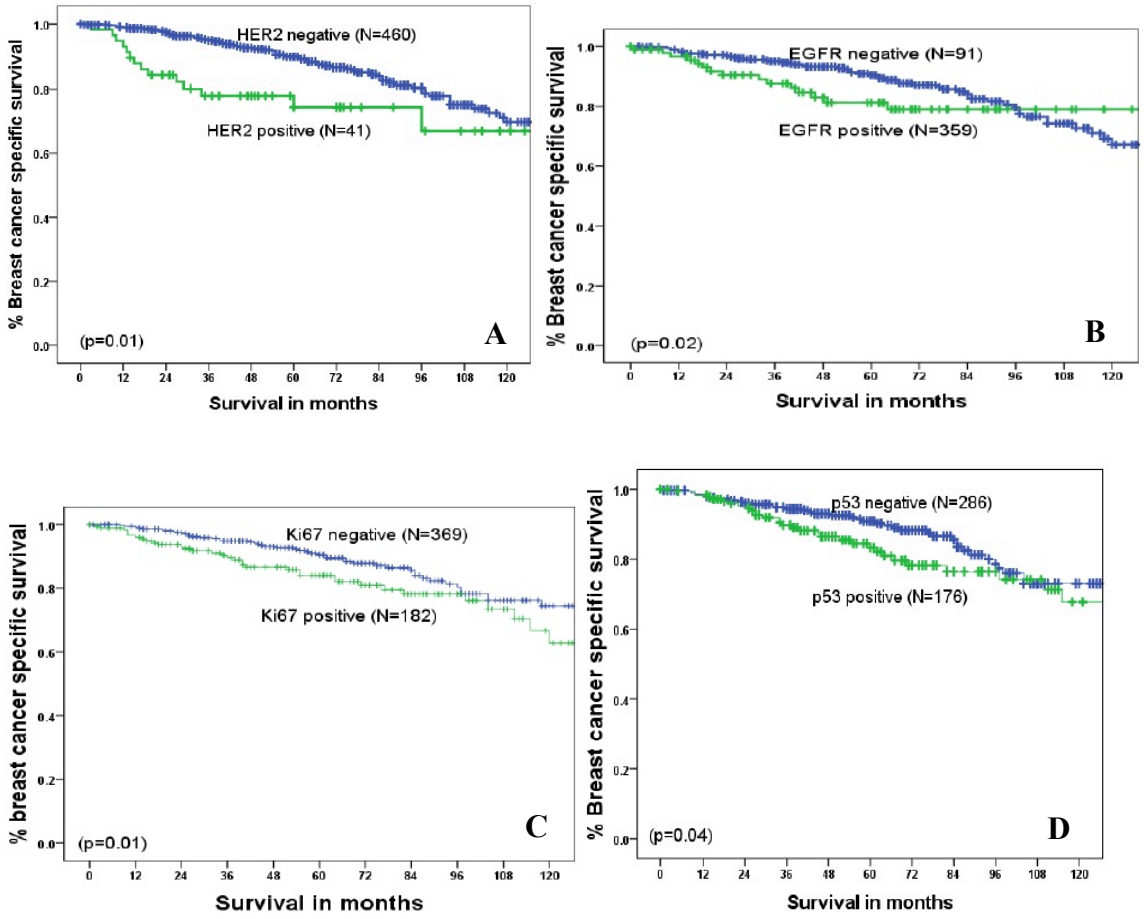


Figure 87: Prognostic significance of A) HER2, B) EGFR, C) Ki67, D) p53 in early operable primary breast cancer treated by surgery – BCSS (3) – permission to use not required

9.2.2.2 Calculating BCSS in CNB TMAs

As part of this PhD, BCSS in the whole CNB TMA cohort was calculated. Expression of ER, PR, BCL2 and VEGF (Figure 88) was associated with significantly better BCSS in CNB TMAs, as was the absence of Ki67, p53, EGFR, MUC1 (Figure 89).

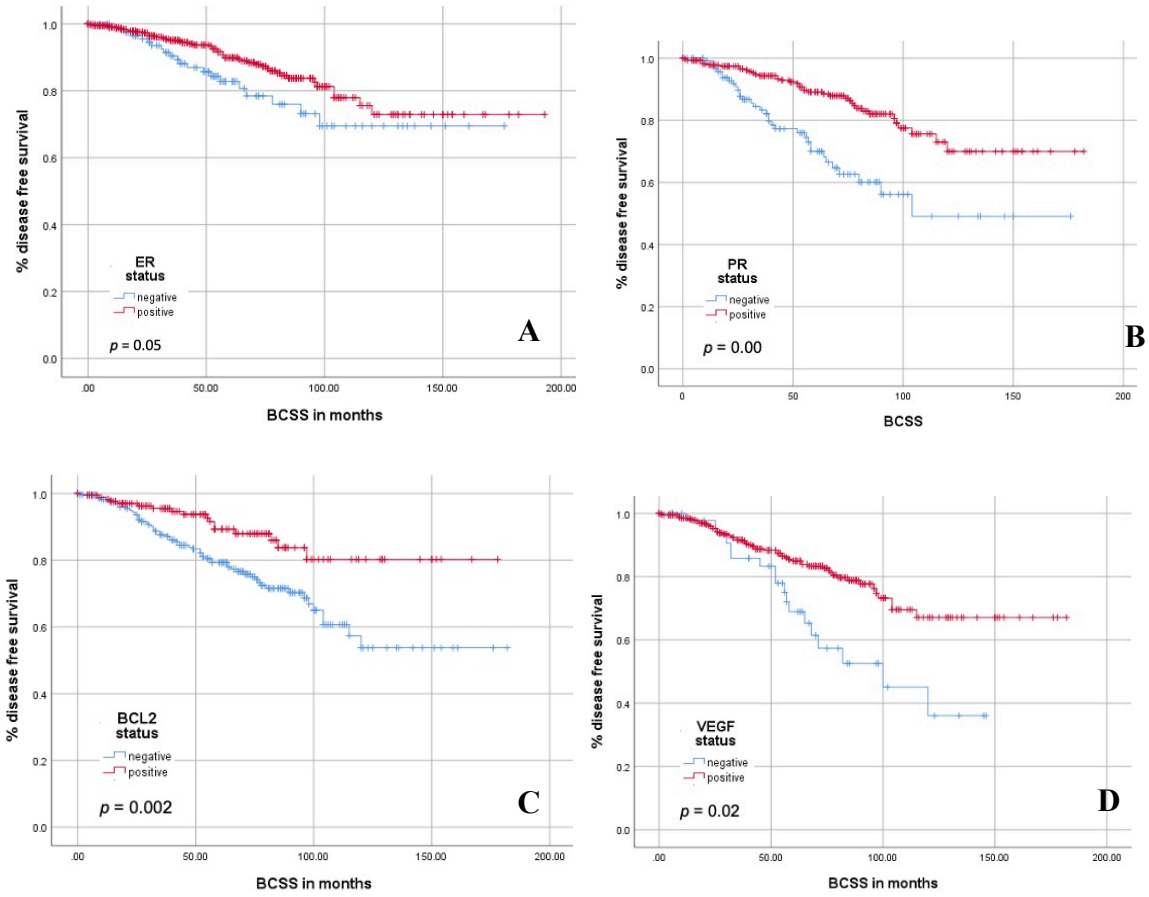


Figure 88: Prognostic significance of A) ER, B) PR, C) BCL2, D) VEGF in CNB TMAs in early operable primary breast cancer in older women– BCSS

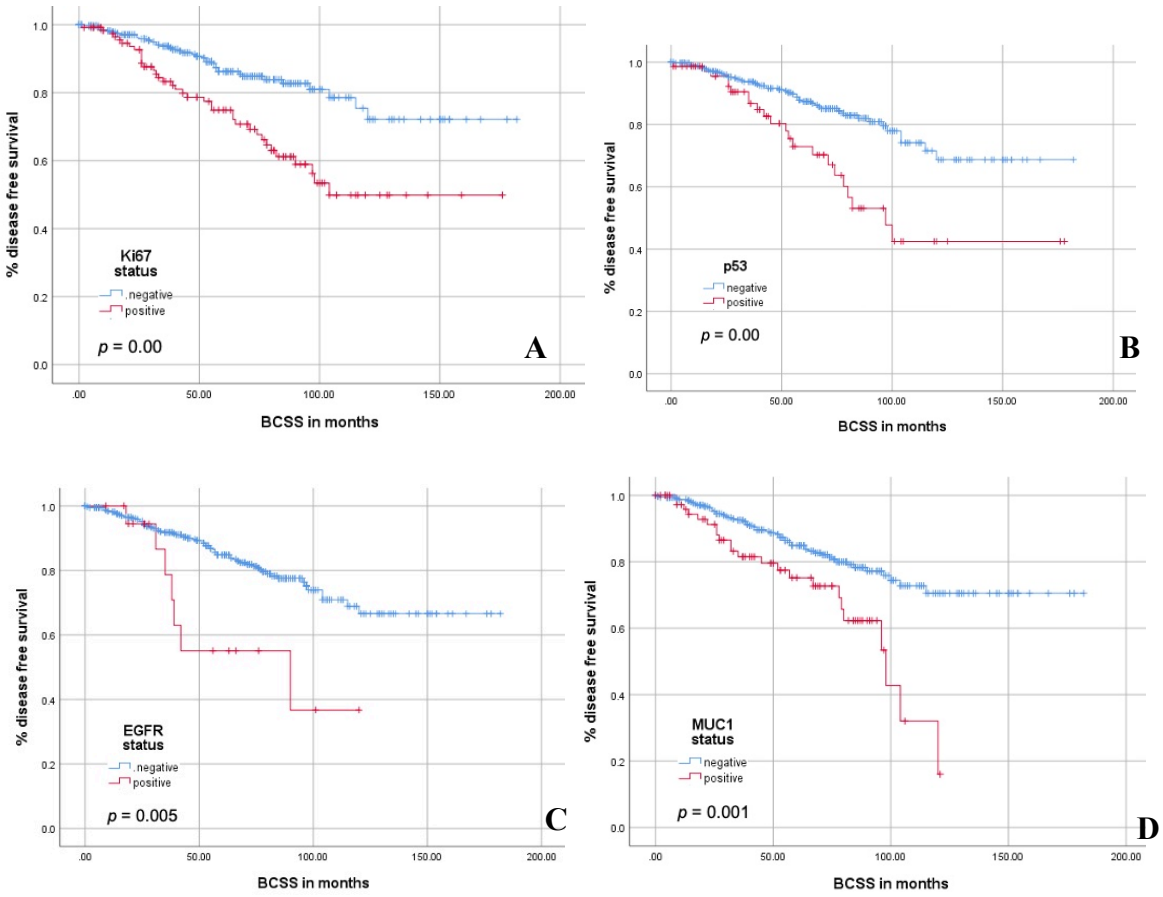


Figure 89: Prognostic significance of A) Ki67, B) p53, C) EGFR, D) MUC1 in CNB TMAs in early operable primary breast cancer in older women– BCSS

9.2.2.3 Optimised panel of biomarkers

Table 32 summarises the biomarkers which are independently prognostically significant for BCSS in each set of TMAs and therefore, will be taken forward to determine if a more accurate prognostic model can be formulated.

SE TMAs	CNB TMAs
ER	ER
PR	PR
EGFR	EGFR
P53	P53
Ki67	Ki67
BCL2	BCL2
MUC1	MUC1
HER2	-
-	VEGF

Table 32: The optimised panel of biomarkers for each set of TMAs taken from biomarkers which have shown independent significance for BCSS

9.2.3 Results

9.2.3.1 Optimised panel of biomarkers for SE TMAs - BCSS

A neural network consisting of three hidden nodes was determined to be the most accurate predictor of BCSS.

Table 29 shows the best performance of the algorithm on the test data.

Performance	Breast cancer specific survival (in months)
Root Mean Squared Error (RMSE)	29.1
Normalised Root Mean Squared Error (NRMSE)	0.2
Mean Absolute Error (MAE)	22.8
Normalised Mean Absolute Error (NMAE)	0.1
Minimum Absolute Error (Min AE)	1.4
Maximum Absolute Error (Max AE)	55.5
Linear Correlation Coefficient (r value)	0.622

Table 33: Performance of the algorithm designed to predict BCSS in SE TMAs using optimised panel of biomarkers

Figure 90 depicts the association between the estimated prediction of BCSS as determined by the algorithm compared to actual data.

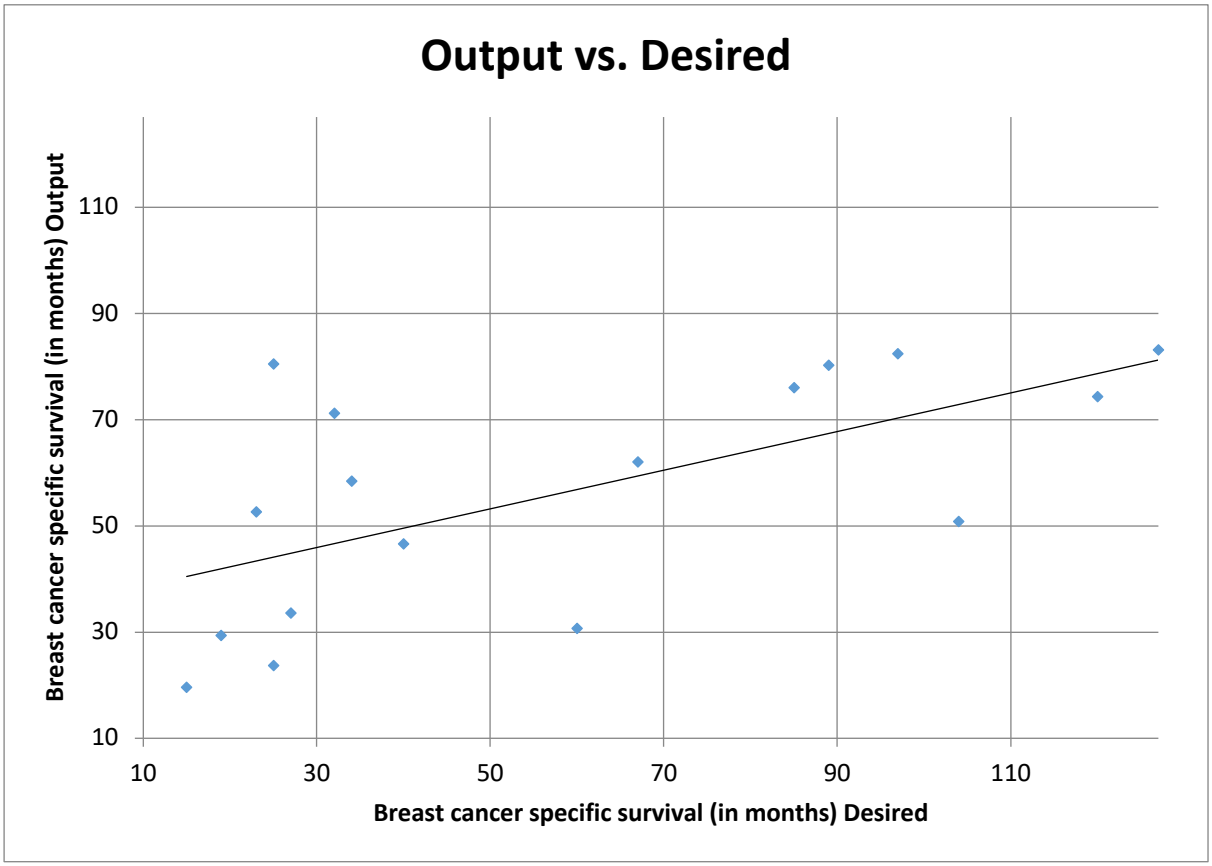


Figure 90: Actual BCSS compared to predicted BCSS in SE TMAs using optimised panel of biomarkers

There is a moderate correlation between the prognostic model for BCSS measured in CNB TMAs and actual result.

9.2.3.2 Optimised panel of biomarkers for CNB TMAs - BCSS

A neural network consisting of 120 hidden nodes was determined to be the most accurate predictor of BCSS.

Table 30 shows the best performance of the algorithm on the test data.

Performance	Time to progression of disease (in months)
RMSE	22.9
NRMSE	0.2
MAE	20.3
NMAE	0.2
Min Abs Error	1.9
Max Abs Error	38.1
r	0.364

Table 34: Performance of the algorithm designed to predict BCSS in CNB TMAs using optimised panel of biomarkers

Figure 91 depicts the association between the estimated prediction of BCSS as determined by the algorithm compared to actual data.

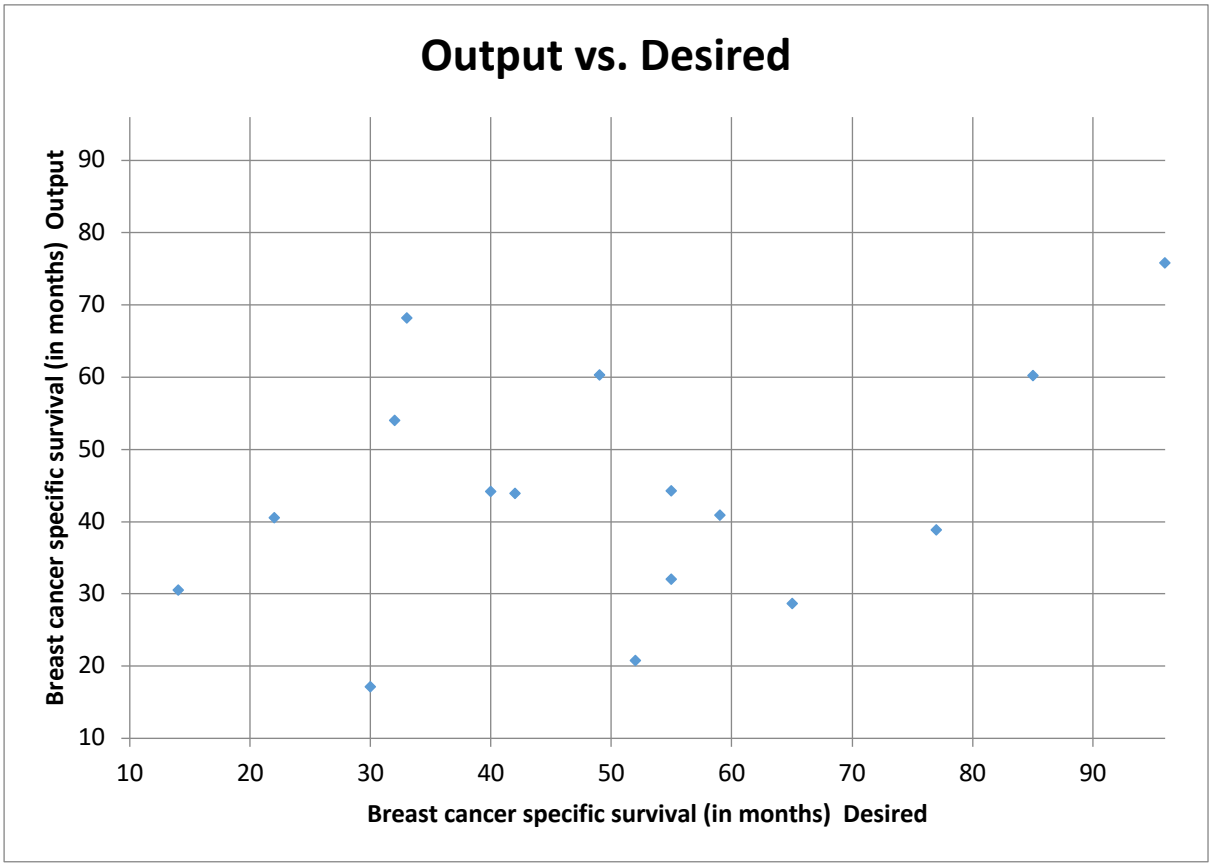


Figure 91: Actual BCSS compared to predicted BCSS in CNB TMAs using optimised panel of biomarkers

9.2.4 Explanation of results and potential next steps

Using the optimised panel of biomarkers based on those biomarkers which were independently significant for BCSS, did not improve the correlation of the estimated result generated by the model and actual BCSS in neither the CNB TMA or SE TMA cohort (Table 31).

Cohort of samples	Correlation coefficient (r-value) using panel of 15 biomarkers (Table 6)	Correlation coefficient (r-value) using optimised panel (Table 28)
SE TMAs	0.742	0.623
CNB TMAs	0.716	0.364

Table 35: Comparison of r-values for prognostic models generated for BCSS

Abbreviations: CNB – core needle biopsy; SE – surgical excision; TMA – tissue microarray

9.2.4.1 Biomarkers of independent prognostic significance for BCSS

Interestingly, the markers ER, PR, Ki67, p53, BCL2 and EGFR have the same prognostic significance in both the SE TMAs and the CNB TMAs (Table 28). MUC1 has good prognostic significance in SE TMAs but poor prognostic significance in the CNB TMAs. VEGF has prognostic significance in the CNB TMAs only and HER2 in the SE TMAs only. As described in section 7.5 of this thesis, the number of absolute cases of patients with HER2-positive breast cancer was extremely small when measured in the CNB TMAs and is likely to account for this finding.

The literature surrounding MUC1 as a prognostic marker is conflicting but overall, overexpression seems to indicate a poor prognosis (section 3.2.8.1). Therefore, the findings in this present thesis; that MUC1 has poor prognostic significance in CNB TMAs may be the more accurate result compared to the opposite findings in the SE TMA group from previous work by the research team (3). Rakha et al (169) importantly note that aberrant cytoplasmic and membranous localisation of MUC1 was associated with

poor outcomes, compared with apical localisation, which is the normal physiological site of expression. This should be a mandatory factor for researchers in this field to document when publishing laboratory studies measuring MUC1.

The finding of VEGF only having prognostic significance in the CNB TMA cohort is more difficult to explain. In the present study, VEGF was associated with increased BCSS, which is contrary to the present literature (section 3.2.2.5). On further evaluation of numbers of cases in each cohort who had positive expression of VEGF, no significant difference was seen; 87% of cases in the CNB TMAs compared to 83% in the SE TMAs were positive for VEGF expression. This finding could be a true difference between the CNB TMA and SE TMA cohorts, related to whether or not the patient received surgery or it could be due to any number of factors associated with the laboratory process of measuring VEGF, as discussed in section 6.2.3 of this thesis.

In the CNB TMA cohort there were 348 cases defined as positive for VEGF expression; of these 179 underwent primary surgical treatment and 169 non-surgical treatment. Patients who had surgery, had better BCSS that those who underwent non-surgical treatment (p=0.01) (Figure 92).

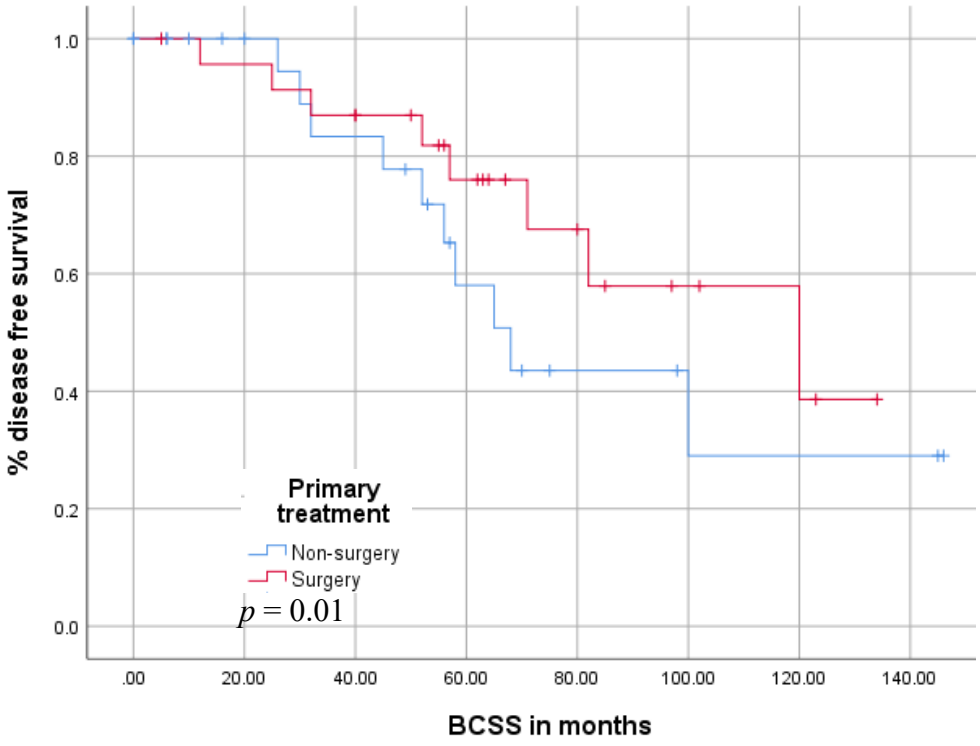


Figure 92: BCSS in cases of CNB TMAs with positive VEGF expression, comparison between primary surgery and non-surgical treatment

If this is a true difference, this could suggest that in older women with primary breast cancer who have positive expression of VEGF, have better BCSS with surgical, rather than non-surgical treatment. There is some evidence in the literature to support this with studies suggesting that VEGF may override the growth-suppressive effects of anti-hormonal therapy and alter response to therapy (397, 398). This finding is certainly worthy of further investigation in a larger cohort of patients.

9.2.4.2 Algorithms generated using optimised panel of biomarkers

Unfortunately, using the optimised panel of biomarkers, did not improve the correlation of the estimated result generated by the model and actual BCSS in neither the SE TMA or CNB TMA cohort (Table 31). This is not entirely surprising as the process of

bioinformatics analysis is complex and there is no 'one-size-fits-all' solution; training of the ANN to provide the optimum result often requires a 'trial-and-error' approach of many possible combinations (399), some of which will be described below.

In this present thesis the bioinformatics technique of ANN has been utilised as it has been widely used in genomic and proteomic microarray analysis to date, however, other machine learning methodologies are available and may produce differing results (356).

9.2.4.3 Potential next steps

There are a number of possible approaches to take in an attempt to improve the algorithms.

The most thorough method would be to try every different possible combination of the 15 biomarkers in the original algorithms, however this would total 15 to the power of 15, possible combinations and is clearly not a realistic option.

There are three potential realistic options which will now be discussed.

1) Focusing the algorithms further using biomarkers of independent prognostic significance for BCSS (Table 28). As discussed in section 9.2.4.1 there are some potentially unexpected findings within the two individual datasets (SE TMA and CNB TMA) with regards to MUC1 and VEGF and HER2 only has significance in SE TMAs due to low numbers of cases in the CNB TMA cohort. If these biomarkers are removed from the picture, we are left with six markers: ER, PR, p53, EGFR, Ki67 and BCL2 which have been found to have the same independent prognostic significance in both the SE TMAs and CNB TMAs separately. Attempting to generate algorithms using these six biomarkers may be more

successful if the findings due to MUC1 and VEGF have been created by laboratory error.

2) Looking at outlier cases to identify if they have particular features. This has been done initially been done in the SE TMA cohort by looking at the initial panel of 15 biomarkers and assessing individual where difference between predicted and actual BCSS was greater than six months. The biggest differences between cases were for the markers HER4, BRCA1, Ki67 and CK 5/6. If these biomarkers were to be removed from the original panel 11 biomarkers would remain: ER, PR, EGFR, HER2, HER3, VEGF, p53, LKB1, BCL2, CK 7/8, MUC1.
3) Assessing contributions to the initial model. Figures 54 and 70 show the contribution of each biomarker within the initial algorithms created for prediction of BCSS in the SE and CNB TMA cohorts respectively. The biomarkers which show the greatest contribution to the algorithm in the SE TMA cohort are: ER, PR, EGFR, HER4, BRCA1, Ki67, BCL2, CK 5/6. The biomarkers which show the greatest contribution to the algorithm in the CNB TMA cohort are: ER, HER3, HER4, VEGF, LKB1. These revised panels could be investigated.

The option of examining the algorithms in various subtypes of breast cancer to see if it performs better in some subtypes (for example in ER-positive breast cancer only) has been debated, however is not possible in the setting of BCSS as an endpoint. This is because numbers of cases for where BCSS information is available is smaller than the overall cohort due to patients being more likely to die from other causes, rather than their breast cancer. If the dataset was to be cut further, the models would be underpowered.

9.3 Conclusions of this chapter and next steps

Currently, the most accurate algorithms generated are using the initial panel of 15 biomarkers which generate strong correlation in both sets of TMAs in predicting BCSS. The r-value for the initial models developed in the SE TMAs and CNB TMAs were 0.742 and 0.716 respectively. Using a revised panel of biomarkers which were independently prognostically significant for BCSS in both the SE and CNB TMA cohorts, did not produce any improvement in the models. The r-values for the revised models were 0.623 and 0.364 respectively.

This chapter has outlined three potential options in an attempt to improve the accuracy of the algorithms which will be explored outside of this thesis. The final step of this thesis is to put these findings in context of the Nottingham research and the field of breast cancer research in older women in general and discuss next steps for research in this area.

GENERAL DISCUSSION

Chapter 10: General discussion

10.1 Introduction

This chapter will summarise the novel findings from each chapter of this thesis. The limitations of the work will then be discussed followed by a consideration of how this thesis adds to the knowledge base for older women with primary breast cancer and potential clinical application. The next steps for this particular research programme and possibilities for future projects of this kind will be discussed. Finally, the chapter will conclude with key take home messages from this thesis.

10.2 Key findings

This section will briefly summarise the key findings from each chapter of this thesis.

Chapter 1

This chapter introduced the current challenges facing older women with primary breast cancer. Definitions of 'older' vary worldwide and chronological age does not necessarily equate to physiological age, however, it is a reasonable surrogate for this. Furthermore, recent work has shown that there is age related change in the biological characteristics of breast cancer, where 70 years is the milestone for a significant change; breast cancers above this age appear less aggressive. The incidence of breast cancer is increasing as our population is ageing and growing and therefore treatment of this group will become significantly more important. There is a disparity between the recommendations for treatment of primary breast cancer and the treatment that older women receive. Reasons why older women may have PET rather than surgery are complex and multifactorial. Traditional outcome measures such as OS may not be as useful to older women compared to their younger counterparts; maintaining quality of life may be the goal of treatment rather than prolonging survival. Current work by the Nottingham research group and

others, has shown that the biology of breast cancer in older women is important and extends to beyond those biomarkers routinely measured in clinical practice.

Chapter 2

This was a literature review of biomarkers relevant to primary breast cancer in older women, outside of those routinely measured in clinical practice of part of the standard breast cancer subtypes (ER, PR, HER2, Ki67, luminal A, luminal B, basal-like, HER2 overexpressing). Out of 23 studies identified, 15 involved measurement of biomarkers in tissue specimens. The most frequently reported biomarkers assessed by this method were: p53, BCL2, cyclin E, AR, EGFR, LKB1, MUC1. The likelihood of developing a p53 mutation increased with chronological age as did expression of BCL2. The markers AR and EGFR were associated with an increase in BCSS and MUC1, a decrease. LKB1 was found to potentially increase BCSS in those receiving adjuvant endocrine therapy. The markers identified were explored in further detail in chapter 3, alongside other biomarkers predicted to be important by the Nottingham research team.

Chapter 3

For each of the markers identified, the literature was searched to find any known existing association with age, the efficacy of the markers as a predictive target for use of endocrine therapy and any known association with survival. Markers which showed increased expression with age included: ER, PR, HER3, HER4, MDM2 and MDM4, E-cadherin and basal cytokeratins. Conversely, markers showing a decrease with age included: HER2, EGFR, Ki67 and luminal cytokeratins. Markers for which an association with age was unknown or evidence conflicting included: AR, VEGF, PTEN, BCL2, CD44, MUC1, AIB1, solute carriers associated with glutamine. In addition to ER, PR was

associated with a positive response to endocrine therapy, with weaker evidence to support BCL2, E-cadherin, CD44 and AIB. Expression of EGFR, HER3, HER4, p53, BRCA1, MUC1, PTEN and solute carriers have all been implicated in endocrine therapy resistance. The following markers were associated with increased survival: ER, PR, AR, HER4, BCL2, e-cadherin. The following markers were associated with decreased survival: HER2, HER3, VEGF, p53, MDM, PTEN, Ki67. The latter part of the chapter explored existing tissue based predictive and prognostic tools on the market. The evidence base for tools of this kind in older women is lacking and mainly focused on decision making of adjuvant treatment. Furthermore, the majority of these tools are based on biomarker measurement from SE specimens, therefore, may not be applicable to older women having PET or neoadjuvant treatment. There is a need for a specific prognostic tool to help older women with primary decision making at diagnosis of breast cancer.

Chapter 4

The Nottingham cohort of older women with primary breast cancer is a unique tissue resource and possibly the largest of its kind in the world from a single centre treated in a uniform manner. It contains breast cancer tissue for 1758 women aged ≥70 years diagnosed consecutively with primary breast cancer, with up to 37 years of follow-up data. Using this resource, TMAs can be constructed to allow high throughput IHC analysis of a large panel of biomarkers which can be associated with clinical data and analysed by standard statistical analysis, cluster analysis and bioinformatics analysis.

Chapter 5

For this present thesis successful IHC staining for the following markers was performed in all available SE TMAs: AR, PTEN, AIB1, SLC1A5, SLC7A5, SLC3A2; and in all

available CNB TMAS: AR, BRCA2, E-cadherin, CD44, MDM2, MDM4, CK5, CK14, CK17, CK18, CK 19. This brings the total number of biomarkers stained by the Nottingham research team in this cohort to 31 in the SE TMAs and 26 in the CNB TMAs. The majority of these markers have not been extensively measured outside of the present research group, however, results of staining were consistent with previous work. There was an apparent discrepancy between SE and CNB scores for the same markers.

Chapter 6

A total of 15 biomarkers had been assessed in both the SE and CNB TMAs from 202 cases and had clinical results available: ER, PR, EGFR, HER2, HER3, HER4, VEGF, p53, BRCA1, LKB1, Ki67, BCL2, CK 5/6, CK 7/8, MUC1. Although H-scores were not directly comparable, the same pattern of expression existed for ER, PR, HER2, VEGF, p53, Ki67, BCL2, CK 7/8 and MUC1, which are the markers most heavily researched in breast cancer. Comparable expression of HER3, HER4, BRCA1, LKB1 and CK 5/6, was not found. Overall, cases were more likely to be positive for a biomarker on SE rather than CNB TMA. PCA shows large variation between the CNB and SE TMA scores, ultimately meaning that at this stage, the results cannot be directly compared. Therefore, further analysis was performed simultaneously in the two groups for the rest of this thesis.

Chapter 7

Although comparison between CNB and SE scores in Chapter 6 suggests that the two datasets are difficult to compare, previous work by the Nottingham research group has shown that despite this, the same biological clusters of breast cancer were derived in the entire cohort of SE TMAs and the ER-positive cases of CNB TMAs. Cluster analysis using 17 biomarkers was performed in the entire cohort of CNB TMAs (no restriction by

ER status) and found the same biological clusters: luminal A type, luminal B type, low-ER luminal. As before, the low-ER luminal cluster has a unique survival pattern which has a worse survival rate compared to luminal A and B types.

Chapter 8

Using a panel of 15 biomarkers in both the CNB and SE TMA cohorts individually, an attempt was made to produce a mathematical model that could predict the following: time to local recurrence / disease progression, time to regional recurrence, time to distant recurrence, OS, BCSS. Both the models for BCSS showed strong concordance between the result predicted by the model and the actual result. The r-value for the models developed in the CNB TMAs and SE TMAs were 0.716 and 0.742 respectively. The r-values for the remainder of the models showed weak or no concordance. The findings in relation to the models for BCSS are particularly exciting for older women with breast cancer, who have competing causes of death other than their breast cancer which would impact their OS.

Chapter 9

In order to develop the mathematical models for BCSS further, the markers contributing independently to BCSS were explored. The markers ER, PR, Ki67, p53, BCL2 and EGFR have the same prognostic significance in both the SE TMAs and the CNB TMAs. MUC1 has good prognostic significance in SE TMAs but poor prognostic significance in the CNB TMAs. VEGF has prognostic significance in the CNB TMAs only and HER2 in the SE TMAs only. The mathematical models developed using these optimal panels of biomarkers, unfortunately had worse performance than the initial models. The markers

VEGF and MUC1 warrant further investigation in this setting, due to the conflicting results found.

10.3 Limitations

Comparability of studies within the literature (Chapter 2 and 3) and also with the results from this thesis is difficult due to the lack of research in older adults in general. Where biological studies have been performed, selection criteria and methodologies vary. Studies are mainly retrospective in nature, especially those with the largest numbers of cases. The majority of studies use tissue from SE specimens, not from CNB so there is potential bias towards those with ER-negative disease. These issues are always going to be a major hurdle in tissue based research of this kind; with evolving ethical processes we are unlikely to achieve a consecutive database of the size and nature of the Nottingham cohort again.

At the end of Chapter 4, I outline the major problem I have faced with the Nottingham Biobank and access to clinical data. Unfortunately, at the time of writing, the issue between Nottingham University Hospitals NHS Trust and the University of Nottingham has not been resolved, despite escalation. This has meant I have been unable to link IHC data generated by this thesis, to clinical information. Despite this, I have been able to conduct this thesis using pre-existing data and provided proof of concept at all stages.

One biomarker which has shown very exciting results in the SE TMAs by the Nottingham research group is cytoplasmic cyclin-E; preliminary work has suggested that expression of cytoplasmic cyclin-E shows improved BCSS up to 20 years. Unfortunately, I was not able to repeat staining of this marker in the CNB TMA group as part of this thesis as currently cytoplasmic cyclin-E IHC staining is only performed at the MD Anderson

Cancer Center in Texas, USA and requires shipping of TMA sections to this centre. This endeavour will require a separate grant application for funding in the future.

Some of the antibody clones which had been used previously by the research team were no longer manufactured in the UK. Where possible, the same clone was used by a different manufacturer. Where this was not possible, the most well-published alternative clones were used. This will cause potential differences in IHC staining of these biomarkers, however, it was felt this would be adequate for the purposes of this thesis in order to preserve the TMA specimens and not unnecessarily repeat staining of biomarkers. This is something to be addressed in the future (see section 10.5) by the research team.

During the IHC staining process (Chapter 5) a loss of cores was noted for some biomarkers. This could be related either to human error in the staining process or may be related to degradation of the TMA blocks or the tissue specimens within, over time. The first TMA blocks were constructed in 2009 and the oldest tissue specimen dates back to 1973. Work on these TMA samples have been conducted since 2009 by several different researchers, in different laboratories, using different equipment, therefore, variances will occur. This is something to be addressed in the future (see section 10.5) by the research team.

This thesis has shown that it is not yet possible to directly compare the biomarker results obtained from CNB and SE samples (Chapter 6). This is disappointing because all older women with breast cancer will have a CNB at diagnosis, but not all of them will go on to have surgery. Despite this, the same biological clusters of breast cancer do appear in each cohort (Chapter 7) and a prognostic model was developed in each cohort independently

for BCSS (Chapter 8). Therefore, it is likely that results generated in each cohort are comparable but measured on different scales.

As alluded to, within the cohort much smaller numbers of women died from breast cancer, rather than other causes. Therefore in creation of the prognostic algorithms (Chapter 8) a much smaller sample size was available for exploration of BCSS compared to the other factors. This can cause bias within the process of generation and testing of the models, which are currently derived from this dataset only. Again, this is an area to be addressed in the future.

At the present time, I have not been able to improve upon the correlation between BCSS generated by the prognostic models and what was seen in 'real-life'. A different approach is required to further refine these.

10.4 Addition to knowledge and change in practice

This thesis has comprehensively examined the biology of primary breast cancer in older women. To the best of the author's knowledge, this work represents the largest cohort of older women with breast cancer in the literature, in whom bioinformatics principles have been applied to measurement of a large panel of biological markers; this is a major strength of the work.

Out of a large panel of biological markers examined, the following markers seem to be key in predicting BCSS in this cohort: ER, PR, HER2, Ki67, p53, BCL2, EGFR, MUC1, VEGF. This thesis further supports that statement that breast cancer is a biologically heterogeneous disease and categorisation by substandard subtypes is no longer adequate to predict survival. Particularly in the older population with concurrent comorbidity, it is

important that we discuss this when making treatment plans; there is no 'one size fits all' approach and patients should be treated as individuals.

This thesis has started a journey to developing a prognostic algorithm specific to older women with primary breast cancer at diagnosis and could result in a change in clinical practice in the next 5-10 years, which will be discussed in the next section.

10.5 Informing future research

This will be broken down into the immediate next steps of this present piece of work, concurrent research in the biology arm of the research programme, concurrent work in the geriatric assessment arm of the research programme and potential application beyond the research programme.

10.5.1 Next steps of this present work

Although the prognostic models generated in this thesis have shown great promise in predicting BCSS in this cohort, there are still a number of steps to be performed before this knowledge could potentially change clinical practice, which are now described.

1) The reduced panels of biomarkers did not demonstrate any superiority at accuracy of BCSS prediction than the initial panel. The next step of the project is to explore further steps to refine the model and also to explore if perhaps the model performs better in a particular category of patients. In addition to improving the accuracy of the algorithms, a reduced panel of biomarkers has further benefits of resource preservation in terms of laboratory processes, time to complete the panel and costs.

2) Following refinement of the algorithm, it will need to be validated in an external cohort of patients with similar characteristics. The Bridging the Age Gap (BTAG) study included women >70 years with operable breast cancer (i.e. the same profile of population as the Nottingham cohort), recruited from 56 UK centres, with histological and follow-up data. Samples from consenting BTAG patients (N=3072) will be collected and transported to a central laboratory. Results of IHC staining will be input into the prognostic algorithm which will calculate predicted BCSS. This data will be correlated with 'real-life' outcome data from the BTAG study. A grant application in collaboration with the BTAG team is currently being drafted to support this.
3) If validation is successful, the algorithm could be commercialised for application in a clinical setting to help treatment decision making.

10.5.2 Concurrent work in the biology arm of the research programme

This thesis has highlighted a number of limitations (section 10.3) which should be addressed by future work. It is envisaged that this would happen alongside the steps in section 10.5.1.

1) To minimise methodological errors potentially encountered by different team members working on this project to date, IHC staining of the final refined algorithms should be repeated on freshly cut sections from the SE and CNB TMA cohorts. This should be performed by the same researcher, in the same laboratory, using the same antibody clones and laboratory equipment. During this thesis, IHC was conducted manually by the researcher. The Nottingham research group has since acquired an automated IHC machine which will ensure even greater

reproducibility and standard operating procedures for this work. By re-staining and rescoring the specimens, we may be able to determine how to compare results generated from CNB specimens with SE specimens and have a more accurate definition of cut-offs.

2) It is hoped that at some point in the near future the research team will be allowed access to the Biobank resource once again. This would give access to clinical data that would allow assessment of the larger panel of 26 biomarkers which have been stained in both the CNB and SE TMAs. Using the methodology outlined in this thesis, the steps to create a prognostic algorithm could be repeated, to see if any additional biomarkers are significant.
3) As mentioned, measurement of cytoplasmic cyclin-E was not possible as part of this present thesis. A separate grant application is currently being drafted to enable measurement of this marker in the CNB TMAs. If significant, the results could be integrated into the prognostic models as above.
4) Multiplex IHC is a powerful investigative tool that can demonstrate multiple staining characteristics on a single slide. This enables high throughput analysis, reduces sample preparation and increased work efficacy. Furthermore, additional information and insight in identification of cell types and function is possible. Multiplex IHC is an avenue worthy of further exploration in the setting of older women with primary breast cancer and has not yet been explored.
5) Even within one breast cancer sample from an individual patient, there is heterogeneity of biomarker expression. A research project investigating the

heterogeneity of biomarker expression within breast cancer samples in older women is planned.

10.5.3 Geriatric assessment arm of the research programme

As outlined in Section 1.6.1 of this thesis, the Nottingham research programme has a second arm to the overall investigation of primary breast cancer in older women and this is concerning geriatric assessment (GA).

Briefly, this is a prospective study recruiting newly diagnosed older women to complete a comprehensive GA validated for use in cancer as well as quality of life questionnaires. The assessments are conducted within six weeks of diagnosis and again at six months. We have published the results of our pilot study of this intervention in 47 patients in Nottingham (400). Comprehensive GA determined that increasing age (≥80 years), greater comorbidity (≥4), greater number of daily medications (≥4) and slower (≥19secs) timed up and go score, were significantly related to non-surgical treatment at six weeks after diagnosis. Baseline QOL scores were generally good and remained stable at six months' follow-up, however, there were differences in scores between patients who had mastectomy compared to BCS (401). The pilot study confirmed the feasibility of conducting CGA in a research setting which appeared to have value in assessing this patient population. To date we have recruited 180 patients to the study and recruitment is ongoing. Econometric evaluation has determined 1258 as the required sample size of patients completing all elements of the study, to generate conclusive results.

After the sample size has been reached, we will be able to conclusively determine which aspects of CGA/QOL questionnaires are most able to predict outcome in this cohort. In the future we hope that the results of this arm of the research programme can be added to

the prognostic models developed from the biological arm of the programme (this thesis) to develop a Bio-Geriatric Assessment (BGA). This is a novel concept which will consider an individual patient's own cancer biology as well as their unique circumstances. In the future, we envisage that all older women presenting with primary breast cancer would undergo BGA alongside discussion with the multi-disciplinary team in order to plan the most appropriate treatment for that patient.

10.5.4 Other potential applications of this work

The work presented in this thesis can potentially be utilised in other areas.

1) This thesis has discussed initial treatment decision making in primary breast cancer. The same principles apply to decision of adjuvant therapy and there is sufficient information within the Nottingham database to develop prognostic models in the setting of adjuvant therapy.
2) The concept of using biology to develop a prognostic model as demonstrated in this thesis could potentially be adapted to other cancer types and other diseases afflicting older adults. For example, in the field of prostate cancer, which is also heavily influenced by hormone expression, a similar application could exist.

10.6 The future of treatment for older women with primary breast cancer

This thesis has highlighted the need for individualised treatment plans for patients based on the unique biology of their individual cancer. This is becoming more and more recognised by patients and healthcare professionals alike and will only increase in the future. Patients want to and need to be more involved in understanding their own breast cancer, potential treatment options and outcomes. For older women, one major hurdle exists in encouraging them to participate in clinical trials and research projects. In the

future, treatment plans for older women with breast cancer will be individualised based on specific tumour biology and patient circumstances, rather than using the same broad guidelines irrespective of age.

10. 7 Conclusions of this thesis

This thesis is unique in that it describes the largest consecutive cohort of older women with primary breast cancer described in the literature, on whom an extensive panel of biological markers has been studied with long-term follow-up data available with application of bioinformatics. Primary breast cancer in older women can no longer be treated the same as in younger women.

A prognostic algorithm has been developed which can be applied in older women with primary breast cancer at diagnosis and can potentially be exploited for treatment decision making at that point (e.g. surgery versus non-operative therapy or neoadjuvant therapy or for adjuvant therapy in the future). The algorithm has not yet been fully evaluated in all these settings, but proof of concept has been confirmed in correlation with BCSS.

Given our ageing and expanding population, the potential impact of this work cannot be overstated; by 2040 over 1 million new breast cancer cases per year worldwide will be in women ≥70 years. The algorithms created as part of this thesis could be applied to all in this cohort.

References

1. International Agency for Research on Cancer (World Health Organisation). 2019. Available at: https://www.iarc.fr [Accessed: 25th July 2022].

2. CRUK. Cancer Research UK. Breast cancer incidence (invasive) statistics. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-One [Accessed: 29th May 2020]. 2016.

3. Syed BM, Green AR, Paish EC, Soria D, Garibaldi J, Morgan L, et al. Biology of primary breast cancer in older women treated by surgery: with correlation with long-term clinical outcome and comparison with their younger counterparts. British Journal of Cancer. 2013;108(5):1042-51.

4. NHS. Improving care for older people. Available at: https://www.england.nhs.uk/ourwork/clinical-policy/older-people/improving-care-for-older-people/ [Accessed: 29th May 2020]. 2019.

5. Office for National Statistics. How would you support our ageing population? 2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/howwouldyousupportourageingpopulation/2019-06-24 [Accessed: 25th July 2022].

6. United Nations. World Population Ageing 2019: Highlights. Department of Economic and Social Affairs, Population Division (2019). 2019;ST/ESA/SER.A/430.

7. Park J, Cho B, Kwon H, Lee C. Developing a biological age assessment equation using principal component analysis and clinical biomarkers of aging in Korean men. Archives of gerontology and geriatrics. 2009;49(1):7-12.

8. Syed BM, Green AR, Rakha EA, Morgan DAL, Ellis IO, Cheung K-L. Age-Related Biology of Early-Stage Operable Breast Cancer and Its Impact on Clinical Outcome. Cancers. 2021;13(6):1417.

9. American Institue for Cancer Research: World Cancer Reserach Fund. Breast cancer statistics. Available at: https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics [Accessed: 25th July 2022]. 2018.

10. American Institute for Cancer Research: World Cancer Research Fund. Breast cancer [online]. Available at: https://www.wcrf.org/dietandcancer/breast-cancer [Accessed: 25th July 2022]. 2018.

11. World Health Organisation. International Agency for Research on Cancer. Available at: https://www.iarc.fr [Accessed: 24th June 2019]. 2019.

12. Macmillan Cancer Support publication: Breast cancer in older women [online]. 2012. Available at: https://www.macmillan.org.uk/documents/aboutus/research/keystats/breastcancerinolderwomen.pdf [Accessed: 25th July 2022].

13. International Agency for Research on Cancer (World Health Organization). Cancer Tomorrow. Copyright (2020). Available at: https://gco.iarc.fr/tomorrow/en/dataviz/trends?mode=cancer&group_populations=0&multiple_populations=0&multiple_cancers=1&cancers=20&sexes=2&types=0&age_start=14 [Accessed: 25th July 2022].

14. (WHO) WHO. Population ageing. World Health Organization online. 2010.

15. United Nations. World Population Prospects [online]. 2019. Available at: https://www.un.org/en/sections/issues-depth/population/index.html [Accessed: 19th May 2020].

16. United Nations (UN), Department of Economics and Social Affairs, Population Division. World Population Prospects 2019. Data booklet (ST/ESA/SER.A/424). 2019.

17. United Nations. Economic and Social Affairs Publication: World Population Ageing. Highlights. 2017.

18. Office for National Statistics. How would you support our ageing population? Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/howwouldyousupportourageingpopulation/2019-06-24 [Accessed: 29th April 2020]. 2019.

19. Macmillan Cancer Support publication: Breast cancer in older women. Key statistics. 2012.

20. Biganzoli L, Battisti NML, Wildiers H, McCartney A, Colloca G, Kunkler IH, et al. Updated recommendations regarding the management of older patients with breast cancer: a joint paper from the European Society of Breast Cancer Specialists (EUSOMA) and the International Society of Geriatric Oncology (SIOG). The Lancet Oncology. 2021.

21. NICE. National Institute for Health and Care Excellence. Early and locally advanced breast cancer: diagnosis and management. Nice Guideline. 2018;101.

22. Parks RM JS, Cheung KL Chapter 7.14 Adjuvant endocrine therapy. Oxford Textbook of General Surgery. 2020 (accepted for publication).

23. Senkus E, on behalf of the EGC, Kyriakides S, on behalf of the EGC, Ohno S, on behalf of the EGC, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Annals of Oncology. 2015;26(suppl_5):v8-v30.

24. National Comprehensive Cancer Network. Breast Cancer: Early Stage. 2016.

25. Kaufmann M, Morrow M, von Minckwitz G, Harris JR. Locoregional treatment of primary breast cancer. Cancer. 2010;116(5):1184-91.

26. Hind D, Wyld L, Beverley C, Reed MW. Surgery versus primary endocrine therapy for operable primary breast cancer in elderly women (70 years plus). Cochrane Database of Systematic Reviews. 2006(1).

27. Morgan JL, Reed MW, Wyld L. Primary endocrine therapy as a treatment for older women with operable breast cancer - a comparison of randomised controlled trial and cohort study findings. Eur J Surg Oncol. 2014;40(6):676-84.

28. Jordan VC. Tamoxifen: catalyst for the change to targeted therapy. Eur J Cancer. 2008;44(1):30-8.

29. Weiss A, Noorbakhsh A, Tokin C, Chang D, Blair SL. Hormone receptor-negative breast cancer: undertreatment of patients over 80. Ann Surg Oncol. 2013;20(10):3274-8.

30. Harris JR, Hellman S. Primary radiation therapy for early breast cancer. Cancer. 1983;51(S12):2547-52.

31. Syed BM, Morgan DAL, Setty T, Green AR, Paish EC, Ellis IO, et al. Oestrogen receptor negative early operable primary breast cancer in older women—Biological characteristics and long-term clinical outcome. PLoS One. 2017;12(12):e0188528.

32. Syed BM, Morgan D, Setty T, Green AR, Paish EC, Ellis IO, et al. Oestrogen receptor negative early operable primary breast cancer in older women-Biological characteristics and long-term clinical outcome. PLoS One. 2017;12(12):e0188528-e.

33. NICE. National Institute for Health and Care Excellence (NICE) guidance NG 101. Early and locally advanced breast cancer: diagnosis and management. 2018.

34. National Institute for Health and Care Excellence. Adjuvant therapy for early and locally advanced breast cancer. NICE pathway. 2019.

35. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Research. 2010;12(1):R1.

36. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. New England Journal of Medicine. 2015;373(21):2005-14.

37. Murray Brunt A, Haviland JS, Wheatley DA, Sydenham MA, Alhasso A, Bloomfield DJ, et al. Hypofractionated breast radiotherapy for 1 week versus 3 weeks (FAST-Forward): 5-year efficacy and late normal tissue effects results from a multicentre, non-inferiority, randomised, phase 3 trial. The Lancet. 2020;395(10237):1613-26.

38. Inoue H, Hirano A, Ogura K, Kamimura M, Hattori A, Yukawa H, et al. Breast-conserving surgery without radiation in elderly women with early breast cancer. Surgical Oncology. 2019;31:22-5.

39. Clement Z, McLeay W, Hoffmann C, Shin P, Kiu A, Eaton M. Role of radiotherapy in women over the age of 65 after breast conserving surgery for breast cancer: A 5-year retrospective study. Breast disease. 2018;37(4):197-205.

40. Tang L, Matsushita H, Jingu K. Controversial issues in radiotherapy after breast-conserving surgery for early breast cancer in older patients: a systematic review. Journal of Radiation Research. 2018;59(6):789-93.

41. Wickberg Å, Liljegren G, Killander F, Lindman H, Bjöhle J, Carlberg M, et al. Omitting radiotherapy in women ≥ 65 years with low-risk early breast cancer after breast-conserving surgery and adjuvant endocrine therapy is safe. Eur J Surg Oncol. 2018;44(7):951-6.

42. Zhong Y, Xu Y, Zhou Y, Mao F, Lin Y, Guan J, et al. Omitting radiotherapy is safe in breast cancer patients ≥ 70 years old after breast-conserving surgery without axillary lymph node operation. Scientific Reports. 2020;10(1):19481.

43. The Royal College of Radiologists. Faculty of Clinical Oncology. Postoperative radiotherapy for breast cancer: UK consensus statements. Nov 2016. London: The Royal College of Radiologists.

44. Salerno K. NCCN Guidelines Update: Evolving Radiation Therapy Recommendations for Breast Cancer. Journal of the National Comprehensive Cancer Network. 2017;15(5.5):682-4.

45. Early Breast Cancer Trialists' Collaborative G. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. The Lancet. 2011;378(9793):771-84.

46. Walsh EM, Nunes R, Wilkinson MJ, Santa-Maria CA. Extended Endocrine Therapy for Early-Stage Breast Cancer: How Do We Decide? Current Oncology Reports. 2020;22(12):123.

47. Goss PE, Ingle JN, Pritchard KI, Robert NJ, Muss H, Gralow J, et al. A randomized trial (MA.17R) of extending adjuvant letrozole for 5 years after completing an initial 5 years of aromatase inhibitor therapy alone or preceded by tamoxifen in postmenopausal women with early-stage breast cancer. Journal of Clinical Oncology. 2016;34(18_suppl):LBA1-LBA.

48. Tjan-Heijnen VCG, van Hellemond IEG, Peer PGM, Swinkels ACP, Smorenburg CH, van der Sangen MJC, et al. Extended adjuvant aromatase inhibition after sequential endocrine therapy (DATA): a randomised, phase 3 trial. The Lancet Oncology. 2017;18(11):1502-11.

49. Mamounas EP, Bandos H, Lembersky BC, Jeong J-H, Geyer CE, Jr., Rastogi P, et al. Use of letrozole after aromatase inhibitor-based therapy in postmenopausal breast cancer (NRG Oncology/NSABP B-42): a randomised, double-blind, placebo-controlled, phase 3 trial. The Lancet Oncology. 2019;20(1):88-99.

50. ABS. Association of Breast Surgery. Update on optimal duration of adjuvant antihormonal therapy. 2015.

51. Burstein HJ, Lacchetti C, Anderson H, Buchholz TA, Davidson NE, Gelmon KA, et al. Adjuvant Endocrine Therapy for Women With Hormone Receptor–Positive Breast Cancer: ASCO Clinical Practice Guideline Focused Update. Journal of Clinical Oncology. 2018;37(5):423-38.

52. Early Breast Cancer Trialists' Collaborative G. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100000 women in 123 randomised trials. The Lancet. 2012;379(9814):432-44.

53. Fisher R and Smith I. Chapter 12: The role of adjuvant systemic therapy in patients with operable breast cancer. In: Dixon JM, editor. Breast Surgery (A Companion To Specalist Surgical Practice): Fifth edition, Elsevier Limited; 2014.

54. Cadoo KA, Morris PG, Cowell EP, Patil S, Hudis CA, McArthur HL. Adjuvant Chemotherapy and Trastuzumab Is Safe and Effective in Older Women With Small, Node-Negative, HER2-Positive Early-Stage Breast Cancer. Clinical breast cancer. 2016;16(6):487-93.

55. Vaz-Luis I, Keating NL, Lin NU, Lii H, Winer EP, Freedman RA. Duration and toxicity of adjuvant trastuzumab in older patients with early-stage breast cancer: a population-based study. J Clin Oncol. 2014;32(9):927-34.

56. Parks RM, Holmes HM, Cheung K-L. Current Challenges Faced by Cancer Clinical Trials in Addressing the Problem of Under-Representation of Older Adults: A Narrative Review. Oncology and Therapy. 2021;9(1):55-67.

57. Wyld L, Reed MWR, Morgan J, Collins K, Ward S, Holmes GR, et al. Bridging the age gap in breast cancer. Impacts of omission of breast cancer surgery in older women with oestrogen receptor positive early breast cancer. A risk stratified analysis of survival outcomes and quality of life. European Journal of Cancer. 2021;142:48-62.

58. National Audit of Breast Cancer in Older Patients. Part of the National Clinical Audit and Patients Outcomes Programme. 2020 Annual Report. Published by: the National Audit of Breast Cancer in Older Patients, Clinical Effectiveness Unit, The Royal College of Surgeons of England.

59. Lee CM TV, Tan BK, Madhukumar P, Yong WS, Wong CY, Ong KW. Perioperative Outcomes of Therapeutic Breast Surgery in the Elderly. Ann Acad Med Singapore. 2016;45(6):261-3.

60. Smith LI, Dayal S, Murray J, Lannigan A. Attitudes towards breast conservation in patients aged over 70 with breast cancer. Springerplus. 2016;5:478-.

61. Hamelinck VC, Bastiaannet E, Pieterse AH, Merkus JWS, Jannink I, den Hoed IDM, et al. A prospective comparison of younger and older patients' preferences for breast-conserving surgery versus mastectomy in early breast cancer. Journal of Geriatric Oncology. 2018;9(2):170-3.

62. Bowman CC, Lennox PA, Clugston PA, Courtemanche DJ. Breast reconstruction in older women: should age be an exclusion criterion? Plastic and reconstructive surgery. 2006;118(1):16-22.

63. Wyld L, Garg DK, Kumar ID, Brown H, Reed MWR. Stage and treatment variation with age in postmenopausal women with breast cancer: compliance with guidelines. British Journal of Cancer. 2004;90(8):1486-91.

64. Oh DD, Flitcroft K, Brennan ME, Spillane AJ. Patterns and outcomes of breast reconstruction in older women – A systematic review of the literature. European Journal of Surgical Oncology (EJSO). 2016;42(5):604-15.

65. Parks R. Impact of breast cancer surgery on functional status and independence. Journal of Geriatric Oncology. 2019;10(6):S1.

66. Tang V, Zhao S, Boscardin J, Sudore R, Covinsky K, Walter LC, et al. Functional Status and Survival After Breast Cancer Surgery in Nursing Home Residents. JAMA surgery. 2018;153(12):1090-6.

67. Singh R, Ansinelli H, Katz H, Jafri H, Gress T, Tirona MT. Factors Associated with Functional Decline in Elderly Female Breast Cancer Patients in Appalachia. Cureus. 2018;10(5):e2612-e.

68. Mandelblatt JS, Edge SB, Meropol NJ, Senie R, Tsangaris T, Grey L, et al. Sequelae of axillary lymph node dissection in older women with stage 1 and 2 breast carcinoma. Cancer. 2002;95(12):2445-54.

69. Girotto JA, Schreiber J, Nahabedian MY. Breast reconstruction in the elderly: preserving excellent quality of life. Annals of plastic surgery. 2003;50(6):572-8.

70. Harrison CA, Parks RM, Cheung KL. The impact of breast cancer surgery on functional status in older women; A systematic review of the literature. European Journal of Surgical Oncology. 2021;47(8):1891-9.

71. Jatoi I, Kunkler IH. Omission of sentinel node biopsy for breast cancer: Historical context and future perspectives on a modern controversy. Cancer. 2021;127(23):4376-83.

72. Society of Surgical Oncology. Choosing Wisely® guidelines (updated July 2021). Available at: https://www.choosingwisely.org/societies/society-of-surgical-oncology/ [Accessed: 5th February 2022].

73. Sun J, Mathias BJ, Sun W, Fulp WJ, Zhou JM, Laronga C, et al. Is it Wise to Omit Sentinel Node Biopsy in Elderly Patients with Breast Cancer? Ann Surg Oncol. 2021;28(1):320-9.

74. Carleton N, Zou J, Fang Y, Koscumb SE, Shah OS, Chen F, et al. Outcomes After Sentinel Lymph Node Biopsy and Radiotherapy in Older Women With Early-Stage, Estrogen Receptor–Positive Breast Cancer. JAMA Network Open. 2021;4(4):e216322-e.

75. Castelo M, Hansen BE, Paszat L, Baxter NN, Scheer AS. Omission of Axillary Staging and Survival in Elderly Women With Early Stage Breast Cancer: A Population-Based Cohort Study. Annals of Surgery Open. 2022;3(2):e159.

76. Brackstone M, Baldassarre FG, Perera FE, Cil T, Gregor MCM, Dayes IS, et al. Management of the Axilla in Early-Stage Breast Cancer: Ontario Health (Cancer Care Ontario) and ASCO Guideline. Journal of Clinical Oncology. 2021;39(27):3056-82.

77. Khan K, Seymour, M., Ring, A. & Johnson, P. in Problem solving in older cancer patients (eds. Ring A., Mansi J., Harari D., Kalsi T., Selby P. J.) (Clinical Publishing, 2015).

78. Rossi A, Colantuoni G, Maione P, Ferrara C, Airoma G, Barzelloni ML, et al. Chemotherapy of Breast Cancer in the Elderly. Current Medicinal Chemistry. 2005;12(3):297-310.

79. Battisti NML, De Glas N, Soto-Perez-de-Celis E, Liposits G, Bringuier M, Walko C, et al. Chemotherapy and gene expression profiling in older early luminal breast cancer patients: An International Society of Geriatric Oncology systematic review. Eur J Cancer. 2022;172:158-70.

80. Ring A, Battisti NML, Reed MWR, Herbert E, Morgan JL, Bradburn M, et al. Bridging The Age Gap: observational cohort study of effects of chemotherapy and trastuzumab on recurrence, survival and quality of life in older women with early breast cancer. British Journal of Cancer. 2021;125(2):209-19.

81. Brain E, Viansone AA, Bourbouloux E, Rigal O, Ferrero J-M, Kirscher S, et al. Final results from a phase III randomized clinical trial of adjuvant endocrine therapy ± chemotherapy in women ≥ 70 years old with ER+ HER2- breast cancer and a high

genomic grade index: The Unicancer ASTER 70s trial. Journal of Clinical Oncology. 2022;40(16_suppl):500-.

82. Muss HB, Polley M-YC, Berry DA, Liu H, Cirrincione CT, Theodoulou M, et al. Randomized Trial of Standard Adjuvant Chemotherapy Regimens Versus Capecitabine in Older Women With Early Breast Cancer: 10-Year Update of the CALGB 49907 Trial. J Clin Oncol. 2019;37(26):2338-48.

83. Muss HB, Woolf S, Berry D, Cirrincione C, Weiss RB, Budman D, et al. Adjuvant chemotherapy in older and younger women with lymph node-positive breast cancer. Jama. 2005;293(9):1073-81.

84. U.S. National Library of Medicine. National Surgery Adjuvant Study of Breast Cancer (N-SAS BC) 07 (RESPECT). Available at: https://clinicaltrials.gov/ct2/show/study/NCT01104935 [Accessed 5th April 2020]. ClinicalTrialsgov. 2019.

85. Sawaki M, Taira N, Uemura Y, Saito T, Baba S, Kobayashi K, et al. Randomized Controlled Trial of Trastuzumab With or Without Chemotherapy for HER2-Positive Early Breast Cancer in Older Patients. J Clin Oncol. 2020;38(32):3743-52.

86. Hagen KB, Aas T, Kvaløy JT, Søiland H, Lind R. Adherence to adjuvant endocrine therapy in postmenopausal breast cancer patients: A 5-year prospective study. The Breast. 2019;44:52-8.

87. Robinson B, Dijkstra B, Davey V, Tomlinson S, Frampton C. Adherence to Adjuvant Endocrine Therapy in Christchurch Women with Early Breast Cancer. Clinical Oncology. 2018;30(1):e9-e15.

88. Younus J, Kligman L. Management of aromatase inhibitor-induced arthralgia. Curr Oncol. 2010;17(1):87-90.

89. Franco P, Iorio GC, Bartoncini S, Airoldi M, De Sanctis C, Castellano I, et al. De-escalation of breast radiotherapy after conserving surgery in low-risk early breast cancer patients. Medical Oncology. 2018;35(5):62.

90. Ghignone F, van Leeuwen BL, Montroni I, Huisman MG, Somasundar P, Cheung KL, et al. The assessment and management of older cancer patients: A SIOG surgical task force survey on surgeons' attitudes. European Journal of Surgical Oncology. 2016;42(2):297-302.

91. Morgan JL, Burton M, Collins K, Lifford KJ, Robinson TG, Cheung KL, et al. The balance of clinician and patient input into treatment decision-making in older women with operable breast cancer. Psycho-oncology. 2015;24(12):1761-6.

92. National Breast Cancer Coalition. Outcome Measures or Endpoints for Clinical Trials. Available at: http://www.breastcancerdeadline2020.org/breast-cancer-information/understanding-research/outcome-measures-or-endpoints.html [Accessed: 30th January 2020]. 2013.

93. Divo MJ, Martinez CH, Mannino DM. Ageing and the epidemiology of multimorbidity. Eur Respir J. 2014;44(4):1055-68.

94. Derks MGM, van de Velde CJH, Giardiello D, Seynaeve C, Putter H, Nortier JWR, et al. Impact of Comorbidities and Age on Cause-Specific Mortality in Postmenopausal Patients with Breast Cancer. The Oncologist. 2019.

95. Park KU, Gregory M, Bazan J, Lustberg M, Rosenberg S, Blinder V, et al. Neoadjuvant endocrine therapy use in early stage breast cancer during the covid-19 pandemic. Breast Cancer Res Treat. 2021;188(1):249-58.

96. Tonneson JE, Hoskin TL, Day CN, Durgan DM, Dilaveri CA, Boughey JC. Impact of the COVID-19 Pandemic on Breast Cancer Stage at Diagnosis, Presentation, and Patient Management. Ann Surg Oncol. 2022;29(4):2231-9.

97. Puntmann VO. How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease. Postgraduate Medical Journal. 2009;85(1008):538.

98. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene. 2007;26(10):1507-16.

99. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res. 2015;5(10):2929-43.

100. Dai XC, Hongye; Bai, Zhonghu; Li, Jia. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. Journal of Cancer 2017;8(16):3131-41.

101. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How many etiological subtypes of breast cancer: two, three, four, or more? Journal of the National Cancer Institute. 2014;106(8):dju165.

102. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SAW, et al. Comprehensive genomic analysis identifies novel subtypes and targets of

triple-negative breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2015;21(7):1688-98.

103. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750-67.

104. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol. 2014;5(3):412-24.

105. Kerlikowske K, Gard CC, Tice JA, Ziv E, Cummings SR, Miglioretti DL, et al. Risk Factors That Increase Risk of Estrogen Receptor–Positive and –Negative Breast Cancer. JNCI: Journal of the National Cancer Institute. 2016;109(5).

106. Rhodes A, Jasani B, Balaton AJ, Barnes DM, Miller KD. Frequency of oestrogen and progesterone receptor positivity by immunohistochemical analysis in 7016 breast carcinomas: correlation with patient age, assay sensitivity, threshold value, and mammographic screening. J Clin Pathol. 2000;53(9):688-96.

107. Cheung KL, Wong AWS, Parker H, Li VWY, Winterbottom L, Morgan DAL, et al. Pathological features of primary breast cancer in the elderly based on needle core biopsies—A large series from a single centre. Critical Reviews in Oncology/Hematology. 2008;67(3):263-7.

108. Diab SG ER, Clark GM. Tumor characteristics and clinical outcome of elderly women with breast cancer. J Natl Cancer Inst. 2000;92(7):550-6.

109. Syed BM, Al-Khyatt W, Johnston SJ, Wong DWM, Winterbottom L, Kennedy H, et al. Long-term clinical outcome of oestrogen receptor-positive operable primary breast

cancer in older women: a large series from a single centre. British Journal of Cancer. 2011;104(9):1393-400.

110. Dimitrakopoulos F-ID, Kottorou A, Antonacopoulou AG, Makatsoris T, Kalofonos HP. Early-Stage Breast Cancer in the Elderly: Confronting an Old Clinical Problem. J Breast Cancer. 2015;18(3):207-17.

111. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JFR, et al. 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. International Journal of Cancer. 2005;116(3):340-50.

112. Syed BM, Johnston SJ, Wong DWM, Green AR, Winterbottom L, Kennedy H, et al. Long-term (37 years) clinical outcome of older women with early operable primary breast cancer managed in a dedicated clinic. Annals of Oncology. 2012;23(6):1465-71.

113. Parks RM, Albanghali M, Syed BM, Green AR, Ellis IO, Cheung KL. Biology of Oestrogen-Receptor Positive Primary Breast Cancer in Older Women with Utilisation of Core Needle Biopsy Samples and Correlation with Clinical Outcome. Cancers (Basel). 2020;12(8).

114. Syed BM, Green AR, Ellis IO, Cheung KL. Human epidermal growth receptor-2 overexpressing early operable primary breast cancers in older (≥70 years) women: biology and clinical outcome in comparison with younger (<70 years) patients. Annals of Oncology. 2014;25(4):837-42.

115. Syed BM, Green AR, Nolan CC, Morgan DAL, Ellis IO, Cheung K-L. Biological Characteristics and Clinical Outcome of Triple Negative Primary Breast Cancer in Older Women – Comparison with Their Younger Counterparts. PLoS One. 2014;9(7):e100573.

116. Parks RM, Albanghali MA, Syed BM, Green AR, Ellis IO, Cheung KL. Patterns of biomarker expression in patients treated with primary endocrine therapy - a unique insight using core needle biopsy tissue microarray. Breast Cancer Res Treat. 2021;185(3):647-55.

117. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. New England Journal of Medicine. 2004;351(27):2817-26.

118. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1-.

119. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). JNCI: Journal of the National Cancer Institute. 2005;97(16):1180-4.

120. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1).

121. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med. 2012;9(5):e1001216.

122. Brouwers B, Dalmasso B, Hatse S, Laenen A, Kenis C, Swerts E, et al. Biological ageing and frailty markers in breast cancer patients. Aging. 2015;7(5):319-33.

123. Brouwers B, Hatse S, Dal Lago L, Neven P, Vuylsteke P, Dalmasso B, et al. The impact of adjuvant chemotherapy in older breast cancer patients on clinical and biological aging parameters. Oncotarget. 2016;7(21):29977-88.

124. Di Benedetto A, Ercolani C, Pizzuti L, Angelucci D, Sergi D, Marinelli C, et al. Prognostic relevance of DNA damage and repair biomarkers in elderly patients with hormone-receptor-positive breast cancer treated with neoadjuvant hormone therapy: evidence from the real-world setting. Therapeutic Advances in Medical Oncology. 2019;11.

125. Dubot C, Bourbouloux E, Mir O, Kirscher S, Rigal O, Ferrero JM, et al. ASTER 70s UNICANCER phase III trial: Can a genomic prognosticator help tailoring adjuvant systemic treatment for luminal breast carcinoma in elderly women? Cancer Research. 2015;75(9 SUPPL. 1).

126. Engels CC, Kiderlen M, Bastiaannet E, van Eijk R, Mooyaart A, Smit VT, et al. The clinical value of HER-2 overexpression and PIK3CA mutations in the older breast cancer population: a FOCUS study analysis. Breast Cancer Res Treat. 2016;156(2):361-70.

127. Garimella V, Hussain T, Agarwal V, Radhakrishna S, Fox JN, Kneeshaw PJ, et al. Clinical response to primary letrozole therapy in elderly patients with early breast cancer: possible role for p53 as a biomarker. Int J Surg. 2014;12(8):821-6.

128. Kozlova L, Francijanc E, Vereskunova M, Vladimirova L, Shatova Y, Kuchkina L. Tissue fibrinolysis in menopausal patients with different molecular types of breast cancer. European Journal of Cancer. 2015;51(SUPPL. 3):S273-S4.

129. Lumachi F, Norberto L, Marino F, Basso SMM, Marzano B, Chiara GB. Hormone receptor rate, MIB-1 score and serum tumour markers CEA and CA 15-3 relationship in elderly women with pT1-2 breast cancer. Anticancer research. 2010;30(11):4701-4.

130. Mieog JD, de Kruijf EM, Bastiaannet E, Kuppen PJK, Sajet A, de Craen AJM, et al. Age determines the prognostic role of the cancer stem cell marker aldehyde dehydrogenase-1 in breast cancer. BMC Cancer. 2012:42.

131. Syed BM, Green AR, Morgan DAL, Ellis IO, Cheung K-L. Liver Kinase B1-A Potential Therapeutic Target in Hormone-Sensitive Breast Cancer in Older Women. Cancers. 2019;11(2):149.

132. Johnston SJ, Syed BM, Parks RM, Monteiro CJ, Caruso JA, Green AR, et al. Cytoplasmic cyclin E is an independent marker of aggressive tumor biology and breast cancer-specific mortality in women over 70 years of age. Cancers. 2020;12(3):712.

133. Cheung KL, Albanghali MA, Syed BM, Green AR, Rakha EA, Ellis IO. Biology and clinical outcomes of estrogen receptor positive (ER+) early primary breast cancer (BC) in older (>=70 years) women-A study based on core needle biopsies (CNB). Journal of Clinical Oncology. 2018;36(15 Supplement 1).

134. Parks R, Abdi J, Rakha E, Ellis I, Green A, Cheung K. The role of the androgen receptor in primary breast cancer in older women. Journal of Geriatric Oncology. 2019;10 (6):S48.

135. Aft R, Cherian M, Frith A, Suresh R, Glover-Collins K, Naughton M, et al. Endocrine treatment alone as primary treatment for elderly patients with estrogen receptor positive good prognosis operable breast cancer: A single arm phase II, single institution study. Cancer Research. 2019;79(4 Supplement 1).

136. Extermann M, Leeuwenburgh C, Samiian L, Sehovic M, Xu J, Cubitt C, et al. Impact of chemotherapy on medium-term physical function and activity of older breast cancer survivors, and associated biomarkers. Journal of geriatric oncology. 2017;8(1):69-75.

137. Muss HB, Krishnamurthy J, Alston SM, Lacy AC, Jolly TA, Williams G, et al. P16INK4aexpression after chemotherapy in older women with early-stage breast cancer. Journal of Clinical Oncology. 2011;29(15 SUPPL. 1).

138. Selenica P, Pareja F, Tadros A, Ferrando L, Brown DN, Zhang H, et al. Genomic landscape of breast cancer occurring in elderly individuals. Cancer Research. 2020;80(4 Supplement 1).

139. Lu G-W, Chou Y-E, Jin W-L, Su X-B. Usefulness of postoperative serum translocator protein as a predictive marker for delirium after breast cancer surgery in elderly women. J Int Med Res. 2020;48(6):300060520910044.

140. Ma D, Jiang YZ, Xie MD, Xiao Y, Zhao S, Shao ZM. Multi-omics profiling reveals distinct molecular features in young and elderly triple negative breast cancer. Cancer Research. 2019;79(4 Supplement 1).

141. Okumura Y, Osako T, Nakano M, Fujisue M, Arima N, Nishimura R. Investigation of the clinicopathological and prognostic risk factors for the recurrence of

breast cancer in elderly patients over 75 years of age. Breast (Edinburgh, Scotland). 2019;44(Supplement 1):S140-S1.

142. Engels CC, Kiderlen M, Bastiaannet E, Mooyaart AL, van Vlierberghe R, Smit VTHBM, et al. The clinical prognostic value of molecular intrinsic tumor subtypes in older breast cancer patients: A FOCUS study analysis. Molecular Oncology. 2016;10(4):594-600.

143. Richardson RB. p53 mutations associated with aging-related rise in cancer incidence rates. Cell Cycle. 2013;12(15):2468-78.

144. Ungerleider NA, Rao SG, Shahbandi A, Yee D, Niu T, Frey WD, et al. Breast cancer survival predicted by TP53 mutation status differs markedly depending on treatment. Breast Cancer Research. 2018;20(1):115.

145. Schon K, Tischkowitz M. Clinical implications of germline mutations in breast cancer: TP53. Breast cancer research and treatment. 2018;167(2):417-23.

146. Syed BM, Green AR, Paish EC, Soria D, Garibaldi J, Morgan L, et al. Biology of primary breast cancer in older women treated by surgery: with correlation with long-term clinical outcome and comparison with their younger counterparts. Br J Cancer. 2013;108(5):1042-51.

147. Daidone MG, Coradini D, Martelli G, Veneroni S. Primary breast cancer in elderly women: biological profile and relation with clinical outcome. Critical Reviews in Oncology/Hematology. 2003;45(3):313-25.

148. Alexander A, Karakas C, Chen X, Carey JPW, Yi M, Bondy M, et al. Cyclin E overexpression as a biomarker for combination treatment strategies in inflammatory breast cancer. Oncotarget. 2017;8(9):14897-911.

149. Hunt KK, Karakas C, Ha MJ, Biernacka A, Yi M, Sahin AA, et al. Cytoplasmic Cyclin E Predicts Recurrence in Patients with Breast Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017;23(12):2991-3002.

150. Kensler KH, Regan MM, Heng YJ, Baker GM, Pyle ME, Schnitt SJ, et al. Prognostic and predictive value of androgen receptor expression in postmenopausal women with estrogen receptor-positive breast cancer: results from the Breast International Group Trial 1-98. Breast Cancer Res. 2019;21(1):30-.

151. Speers C, Zhao SG, Chandler B, Liu M, Wilder-Romans K, Olsen E, et al. Androgen receptor as a mediator and biomarker of radioresistance in triple-negative breast cancer. NPJ breast cancer. 2017;3:29-.

152. Dieci MV, Tsvetkova V, Griguolo G, Miglietta F, Mantiero M, Tasca G, et al. Androgen Receptor Expression and Association With Distant Disease-Free Survival in Triple Negative Breast Cancer: Analysis of 263 Patients Treated With Standard Therapy for Stage I-III Disease. Front Oncol. 2019;9:452-.

153. Bhattarai S, Klimov S, Mittal K, Krishnamurti U, Li XB, Oprea-Ilies G, et al. Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study. Cancers. 2019;11(7):995.

154. Venema CM, Bense RD, Steenbruggen TG, Nienhuis HH, Qiu S-Q, van Kruchten M, et al. Consideration of breast cancer subtype in targeting the androgen receptor. Pharmacology & Therapeutics. 2019;200:135-47.

155. Rimawi MF, Shetty PB, Weiss HL, Schiff R, Osborne CK, Chamness GC, et al. Epidermal growth factor receptor expression in breast cancer association with biologic phenotype and clinical outcomes. Cancer. 2010;116(5):1234-42.

156. Bernsdorf M, Ingvar C, Jörgensen L, Tuxen MK, Jakobsen EH, Saetersdal A, et al. Effect of adding gefitinib to neoadjuvant chemotherapy in estrogen receptor negative early breast cancer in a randomized phase II trial. Breast Cancer Research and Treatment. 2011;126(2):463-70.

157. Nakai K, Hung M-C, Yamaguchi H. A perspective on anti-EGFR therapies targeting triple-negative breast cancer. Am J Cancer Res. 2016;6(8):1609-23.

158. Xia C, Ye F, Hu X, Li Z, Jiang B, Fu Y, et al. Liver kinase B1 enhances chemoresistance to gemcitabine in breast cancer MDA-MB-231 cells. Oncol Lett. 2014;8(5):2086-92.

159. Sengupta S, Nagalingam A, Muniraj N, Bonner MY, Mistriotis P, Afthinos A, et al. Activation of tumor suppressor LKB1 by honokiol abrogates cancer stem-like phenotype in breast cancer via inhibition of oncogenic Stat3. Oncogene. 2017;36(41):5709-21.

160. Jing X, Liang H, Hao C, Yang X, Cui X. Overexpression of MUC1 predicts poor prognosis in patients with breast cancer. Oncol Rep. 2019;41(2):801-10.

161. McGuckin MA, Walsh MD, Hohn BG, Ward BG, Wright RG. Prognostic significance of muc1 epithelial mucin expression in breast cancer. Human Pathology. 1995;26(4):432-9.

162. Samuels Y, Waldman T. Oncogenic mutations of PIK3CA in human cancers. Curr Top Microbiol Immunol. 2010;347:21-41.

163. Stirrups R. Alpelisib plus fulvestrant for PIK3CA-mutated breast cancer. The Lancet Oncology. 2019;20(7):e347.

164. Fan H, Li C, Xiang Q, Xu L, Zhang Z, Liu Q, et al. PIK3CA mutations and their response to neoadjuvant treatment in early breast cancer: A systematic review and meta-analysis. Thorac Cancer. 2018;9(5):571-9.

165. Jakubowski DM, Bailey H, Abran J, Blacklock A, Ciau N, Mies C, et al. Molecular characterization of breast cancer needle core biopsy specimens by the 21-gene Breast Recurrence Score test. J Surg Oncol. 2020;122(4):611-8.

166. Syed BM. A study of biological characteristics of early operable primary breast cancer in older women and correlation with clinical outcome. University of Nottingham, Thesis. January 2012.

167. Albanghali M. Biology and clinical outcomes of early primary breast cancer in older women - a study based on core needle biopsy. PhD Thesis, University of Nottingham. November 2016.

168. Rakha EA, El-Sheikh SE, Kandil MA, El-Sayed ME, Green AR, Ellis IO. Expression of BRCA1 protein in breast cancer and its prognostic significance. Human Pathology. 2008;39(6):857-65.

169. Rakha EA, Boyce RWG, Abd El-Rehim D, Kurien T, Green AR, Paish EC, et al. Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. Modern Pathology. 2005;18(10):1295-304.

170. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, Robertson JFR, et al. Expression of luminal and basal cytokeratins in human breast carcinoma. The Journal of Pathology. 2004;203(2):661-71.

171. El-Ansari R, Craze ML, Alfarsi L, Soria D, Diez-Rodriguez M, Nolan CC, et al. The combined expression of solute carriers is associated with a poor prognosis in highly proliferative ER+ breast cancer. Breast Cancer Research and Treatment. 2019;175(1):27-38.

172. CRUK. Cancer Research UK. Breast cancer survival. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Three [Accessed: 10th April 2020]. 2010-11.

173. Waters EA, McNeel TS, Stevens WM, Freedman AN. Use of tamoxifen and raloxifene for breast cancer chemoprevention in 2010. Breast cancer research and treatment. 2012;134(2):875-80.

174. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for Prevention of Breast Cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. JNCI: Journal of the National Cancer Institute. 1998;90(18):1371-88.

175. Eiermann W, Paepke S, Appfelstaedt J, Llombart-Cussac A, Eremin J, Vinholes J, et al. Preoperative treatment of postmenopausal breast cancer patients with letrozole: A randomized double-blind multicenter study. Annals of Oncology. 2001;12(11):1527-32.

176. Preece PE, Wood RA, Mackie CR, Cuschieri A. Tamoxifen as initial sole treatment of localised breast cancer in elderly women: a pilot study. Br Med J (Clin Res Ed). 1982;284(6319):869-70.

177. Caudle AS, Gonzalez-Angulo AM, Hunt KK, Pusztai L, Kuerer HM, Mittendorf EA, et al. Impact of progression during neoadjuvant chemotherapy on surgical management of breast cancer. Annals of surgical oncology. 2011;18(4):932-8.

178. Monypenny I. UK Symptomatic Breast Audit 1.4.2001-31.3.2002. British Association of Surgical Oncology. 2003.

179. Sopik V, Sun P, Narod SA. The prognostic effect of estrogen receptor status differs for younger versus older breast cancer patients. Breast Cancer Research and Treatment. 2017;165(2):391-402.

180. Putti TC, El-Rehim DMA, Rakha EA, Paish CE, Lee AHS, Pinder SE, et al. Estrogen receptor-negative breast carcinomas: a review of morphology and immunophenotypical analysis. Modern Pathology. 2005;18(1):26-35.

181. Vocka M, Zimovjanova M, Bielcikova Z, Tesarova P, Petruzelka L, Mateju M, et al. Estrogen Receptor Status Oppositely Modifies Breast Cancer Prognosis in BRCA1/BRCA2 Mutation Carriers Versus Non-Carriers. Cancers. 2019;11(6):738.

182. Kwapisz D. Cyclin-dependent kinase 4/6 inhibitors in breast cancer: palbociclib, ribociclib, and abemaciclib. Breast Cancer Research and Treatment. 2017;166(1):41-54.

183. Johnston SRD, Toi M, O'Shaughnessy J, Rastogi P, Campone M, Neven P, et al. Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer (monarchE): results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. The Lancet Oncology. 2023;24(1):77-90.

184. Rossi L, McCartney A, Risi E, De Santo I, Migliaccio I, Malorni L, et al. Cyclin-Dependent Kinase 4/6 Inhibitors in Neoadjuvant Endocrine Therapy of Hormone Receptor-Positive Breast Cancer. Clin Breast Cancer. 2019;19(6):392-8.

185. Helwick C. A Tale of Two CDK4/6 Inhibitors in Early Breast Cancer [online]. 2021. Available at: https://ascopost.com/issues/january-25-2021/a-tale-of-two-cdk46-inhibitors-in-early-breast-cancer/ [Accessed: 31st July 2022]. The ASCO Post. 2021.

186. Dai X, Xiang L, Li T, Bai Z. Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. Journal of Cancer. 2016;7(10):1281-94.

187. Thakkar JP, Mehta DG. A Review of an Unfavorable Subset of Breast Cancer: Estrogen Receptor Positive Progesterone Receptor Negative. The Oncologist. 2011;16(3):276-85.

188. Clark GM, Osborne CK, McGuire WL. Correlations between estrogen receptor, progesterone receptor, and patient characteristics in human breast cancer. Journal of Clinical Oncology. 1984;2(10):1102-9.

189. Nordenskjöld A, Fohlin H, Fornander T, Löfdahl B, Skoog L, Stål O. Progesterone receptor positivity is a predictor of long-term benefit from adjuvant tamoxifen treatment of estrogen receptor positive breast cancer. Breast cancer research and treatment. 2016;160(2):313-22.

190. Dowsett M, Cuzick J, Wale C, Howell T, Houghton J, Baum M. Retrospective Analysis of Time to Recurrence in the ATAC Trial According to Hormone Receptor Status: An Hypothesis-Generating Study. Journal of Clinical Oncology. 2005;23(30):7512-7.

191. Park S, Park B-W, Kim TH, Jeon CW, Kang H-S, Choi J-E, et al. Lack of Either Estrogen or Progesterone Receptor Expression Is Associated with Poor Survival Outcome among Luminal A Breast Cancer Subtype. Annals of Surgical Oncology. 2013;20(5):1505-13.

192. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Progesterone receptor loss identifies Luminal B breast cancer subgroups at higher risk of relapse. Annals of Oncology. 2013;24(3):661-8.

193. Van Asten K, Slembrouck L, Olbrecht S, Jongen L, Brouckaert O, Wildiers H, et al. Prognostic Value of the Progesterone Receptor by Subtype in Patients with Estrogen Receptor-Positive, HER-2 Negative Breast Cancer. The Oncologist. 2019;24(2):165-71.

194. Michmerhuizen AR, Spratt DE, Pierce LJ, Speers CW. ARe we there yet? Understanding androgen receptor signaling in breast cancer. npj Breast Cancer. 2020;6(1):47.

195. Tokunaga E, Hisamatsu Y, Taketani K, Yamashita N, Akiyoshi S, Okada S, et al. Differential impact of the expression of the androgen receptor by age in estrogen receptor-positive breast cancer. Cancer Med. 2013;2(6):763-73.

196. Witzel I, Loibl S, Wirtz R, Fasching PA, Denkert C, Weber K, et al. Androgen receptor expression and response to chemotherapy in breast cancer patients treated in the neoadjuvant TECHNO and PREPARE trial. British Journal of Cancer. 2019;121(12):1009-15.

197. Chintamani, Kulshreshtha P, Chakraborty A, Singh L, Mishra AK, Bhatnagar D, et al. Androgen receptor status predicts response to chemotherapy, not risk of breast cancer in Indian women. World J Surg Oncol. 2010;8:64.

198. Chen M, Yang Y, Xu K, Li L, Huang J, Qiu F. Androgen Receptor in Breast Cancer: From Bench to Bedside. Frontiers in Endocrinology. 2020;11(573).

199. Kono M, Fujii T, Lim B, Karuturi MS, Tripathy D, Ueno NT. Androgen Receptor Function and Androgen Receptor–Targeted Therapies in Breast Cancer: A Review. JAMA Oncology. 2017;3(9):1266-73.

200. Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: biology driving targeted therapeutics. Cell Mol Life Sci. 2008;65(10):1566-84.

201. Ayoub NM, Al-Shami KM, Yaghan RJ. Immunotherapy for HER2-positive breast cancer: recent advances and combination therapeutic approaches. Breast Cancer (Dove Med Press). 2019;11:53-69.

202. Asati V, Mahapatra DK, Bharti SK. PI3K/Akt/mTOR and Ras/Raf/MEK/ERK signaling pathways inhibitors as anticancer agents: Structural and pharmacological perspectives. European Journal of Medicinal Chemistry. 2016;109:314-41.

203. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. Annu Rev Med. 2011;62:233-47.

204. Tsutsui S, Kataoka A, Ohno S, Murakami S, Kinoshita J, Hachitanda Y. Prognostic and Predictive Value of Epidermal Growth Factor Receptor in Recurrent Breast Cancer. Clinical Cancer Research. 2002;8(11):3454.

205. G G. Prognostic value of vascular endothelial growth factor in breast cancer. Oncologist. 2000;5(Suppl 1):37-44.

206. Johnston S, Trudeau M, Kaufman B, Boussen H, Blackwell K, LoRusso P, et al. Phase II Study of Predictive Biomarker Profiles for Response Targeting Human

Epidermal Growth Factor Receptor 2 (HER-2) in Advanced Inflammatory Breast Cancer With Lapatinib Monotherapy. Journal of Clinical Oncology. 2008;26(7):1066-72.

207. Soto-Perez-De-Celis E, Loh KP, Baldini C, Battisti NML, Chavarri-Guerra Y, De Glas NA, et al. Targeted agents for HER2-positive breast cancer in older adults: current and future perspectives. Expert Opinion on Investigational Drugs. 2018;27(10):787-801.

208. Support MC. HER2 positive breast cancer. Macmillan Cancer Information and support. 2016.

209. Pinhel I, Hills M, Drury S, Salter J, Sumo G, A'Hern R, et al. ER and HER2 expression are positively correlated in HER2 non-overexpressing breast cancer. Breast Cancer Research. 2012;14(2):R46.

210. Laird-Fick HS, Gardiner JC, Tokala H, Patel P, Wei S, Dimitrov NV. HER2 status in elderly women with breast cancer. Journal of Geriatric Oncology. 2013;4(4):362-7.

211. Cameron D, Piccart-Gebhart MJ, Gelber RD, Procter M, Goldhirsch A, de Azambuja E, et al. 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. Lancet. 2017;389(10075):1195-205.

212. Cianfrocca M, Goldstein LJ. Prognostic and Predictive Factors in Early-Stage Breast Cancer. The Oncologist. 2004;9(6):606-16.

213. National Institute for Health and Care Excellence. Pertuzumab for adjuvant treatment of HER2-postive early stage breast cancer. Technology appraisal guidance [TA569]. 2019.

214. NICE. National Institute for Health and Care Excellent. Neratinib for extended adjuvant treatment of hormone receptor-positive, HER2-positive early stage breast cancer after adjuvant trastuzumab. Technology appraisal guidance TA612. Nov 2019. .

215. Diéras V, Miles D, Verma S, Pegram M, Welslau M, Baselga J, et al. Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial. Lancet Oncol. 2017;18(6):732-42.

216. NICE. National Institute for Health and Care Excellence. Trastuzumab deruxtecan for treating HER2-positive unresectable or metastatic breast cancer after 2 or more anti-HER2 therapies. Technology appraisal guidance TA704. May 2021. .

217. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987;235(4785):177.

218. NICE. National Institute for Health and Care Excellence. Atezolizumab with nab-paclitaxel for untreated PD-L1-positive, locally advanced or metastatic, triple-negative breast cancer. Technology appraisal guidance TA639. July 2020.

219. Travis A, Pinder SE, Robertson JF, Bell JA, Wencyk P, Gullick WJ, et al. C-erbB-3 in human breast carcinoma: expression and relation to prognosis and established prognostic indicators. British journal of cancer. 1996;74(2):229-33.

220. Li Q, Zhang R, Yan H, Zhao P, Wu L, Wang H, et al. Prognostic significance of HER3 in patients with malignant solid tumors. Oncotarget. 2017;8(40):67140-51.

221. Mishra R, Patel H, Alanazi S, Yuan L, Garrett JT. HER3 signaling and targeted therapy in cancer. Oncol Rev. 2018;12(1):355-.

222. Mota JM, Collier KA, Barros Costa RL, Taxter T, Kalyan A, Leite CA, et al. A comprehensive review of heregulins, HER3, and HER4 as potential therapeutic targets in cancer. Oncotarget. 2017;8(51):89284-306.

223. Luhtala S, Staff S, Kallioniemi A, Tanner M, Isola J. Clinicopathological and prognostic correlations of HER3 expression and its degradation regulators, NEDD4-1 and NRDP1, in primary breast cancer. BMC Cancer. 2018;18(1):1045-.

224. Merrimack. A trial of preoperative MM-121 with paclitaxel in HER2-negative breast cancer. ClinicalTrialsgov. 2016;2016: NCT01421472.

225. Naresh A, Long W, Vidal GA, Wimley WC, Marrero L, Sartor CI, et al. The ERBB4/HER4 Intracellular Domain 4ICD Is a BH3-Only Protein Promoting Apoptosis of Breast Cancer Cells. Cancer Research. 2006;66(12):6412.

226. Sartor CI, Zhou H, Kozlowska E, Guttridge K, Kawata E, Caskey L, et al. Her4 mediates ligand-dependent antiproliferative and differentiation responses in human breast cancer cells. Mol Cell Biol. 2001;21(13):4265-75.

227. Witton CJ, Reeves JR, Going JJ, Cooke TG, Bartlett JMS. Expression of the HER1–4 family of receptor tyrosine kinases in breast cancer. The Journal of Pathology. 2003;200(3):290-7.

228. Naresh A, Thor AD, Edgerton SM, Torkko KC, Kumar R, Jones FE. The HER4/4ICD estrogen receptor coactivator and BH3-only protein is an effector of tamoxifen-induced apoptosis. Cancer research. 2008;68(15):6387-95.

229. Koutras AK, Kalogeras KT, Dimopoulos MA, Wirtz RM, Dafni U, Briasoulis E, et al. Evaluation of the prognostic and predictive value of HER family mRNA expression in high-risk early breast cancer: a Hellenic Cooperative Oncology Group (HeCOG) study. British journal of cancer. 2008;99(11):1775-85.

230. Suo Z, Risberg B, Kalsson MG, Willman K, Tierens A, Skovlund E, et al. EGFR family expression in breast carcinomas. c-erbB-2 and c-erbB-4 receptors have different effects on survival. The Journal of Pathology. 2002;196(1):17-25.

231. Kew TY, Bell JA, Pinder SE, Denley H, Srinivasan R, Gullick WJ, et al. c-erbB-4 protein expression in human breast cancer. British journal of cancer. 2000;82(6):1163-70.

232. Folkman J. Role of angiogenesis in tumor growth and metastasis. Seminars in Oncology. 2002;29(6, Supplement 16):15-8.

233. Ferrara N. Vascular Endothelial Growth Factor. Arteriosclerosis, Thrombosis, and Vascular Biology. 2009;29(6):789-91.

234. Liu Y, Tamimi RM, Collins LC, Schnitt SJ, Gilmore HL, Connolly JL, et al. The association between vascular endothelial growth factor expression in invasive breast cancer and survival varies with intrinsic subtypes and use of adjuvant systemic therapy: results from the Nurses' Health Study. Breast Cancer Research and Treatment. 2011;129(1):175-84.

235. Lähteenvuo J, Rosenzweig A. Effects of aging on angiogenesis. Circ Res. 2012;110(9):1252-64.

236. Foekens JA, Peters HA, Grebenchtchikov N, Look MP, Meijer-van Gelder ME, Geurts-Moespot A, et al. High Tumor Levels of Vascular Endothelial Growth Factor

Predict Poor Response to Systemic Therapy in Advanced Breast Cancer. Cancer Research. 2001;61(14):5407.

237. Robert NJ, Diéras V, Glaspy J, Brufsky AM, Bondarenko I, Lipatov ON, et al. RIBBON-1: Randomized, Double-Blind, Placebo-Controlled, Phase III Trial of Chemotherapy With or Without Bevacizumab for First-Line Treatment of Human Epidermal Growth Factor Receptor 2–Negative, Locally Recurrent or Metastatic Breast Cancer. Journal of Clinical Oncology. 2011;29(10):1252-60.

238. G G. Prognostic Value of Vascular Endothelial Growth Factor in Breaset Cancder. The Oncologist. 2000;5(1):8.

239. Roviello G, Bachelot T, Hudis CA, Curigliano G, Reynolds AR, Petrioli R, et al. The role of bevacizumab in solid tumours: A literature based meta-analysis of randomised trials. European Journal of Cancer. 2017;75:245-58.

240. Gasco M, Shami S, Crook T. The p53 pathway in breast cancer. Breast Cancer Res. 2002;4(2):70-6.

241. Pharoah PD, Day NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. British journal of cancer. 1999;80(12):1968-73.

242. Lowe SW, Bodis S, McClatchey A, Remington L, Ruley HE, Fisher DE, et al. p53 status and the efficacy of cancer therapy in vivo. Science. 1994;266(5186):807.

243. Fernandez-Cuesta L, Anaganti S, Hainaut P, Olivier M. p53 status influences response to tamoxifen but not to fulvestrant in breast cancer cell lines. International Journal of Cancer. 2011;128(8):1813-21.

244. Yamashita H, Toyama T, Nishio M, Ando Y, Hamaguchi M, Zhang Z, et al. p53 protein accumulation predicts resistance to endocrine therapy and decreased post-relapse survival in metastatic breast cancer. Breast Cancer Res. 2006;8(4):R48-R.

245. Thor AD, Moore DH, II, Edgerton SM, Kawasaki ES, Reihsaus E, Lynch HT, et al. Accumulation of p53 Tumor Suppressor Gene Protein: An Independent Marker of Prognosis in Breast Cancers. JNCI: Journal of the National Cancer Institute. 1992;84(11):845-55.

246. Duffy MJ, Synnott NC, Crown J. Mutant p53 in breast cancer: potential as a therapeutic target and biomarker. Breast Cancer Research and Treatment. 2018;170(2):213-9.

247. Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003;72(5):1117-30.

248. Zhu Y, Liu Y, Zhang C, Chu J, Wu Y, Li Y, et al. Tamoxifen-resistant breast cancer cells are resistant to DNA-damaging chemotherapy because of upregulated BARD1 and BRCA1. Nat Commun. 2018;9(1):1595-.

249. Wesolowski R, Shealy AG, Tao J, Moore HC. Differential outcomes in patients treated with endocrine therapy for early or locally advanced breast cancer based on BRCA mutation status. Journal of Clinical Oncology. 2009;27(15_suppl):e22065-e.

250. van den Broek AJ, Schmidt MK, van 't Veer LJ, Tollenaar RAEM, van Leeuwen FE. Worse breast cancer prognosis of BRCA1/BRCA2 mutation carriers: what's the evidence? A systematic review with meta-analysis. PloS one. 2015;10(3):e0120189-e.

251. Caulfield SE, Davis CC, Byers KF. Olaparib: A Novel Therapy for Metastatic Breast Cancer in Patients With a BRCA1/2 Mutation. J Adv Pract Oncol. 2019;10(2):167-74.

252. Geyer CE, Jr., Garber JE, Gelber RD, Yothers G, Taboada M, Ross L, et al. Overall survival in the OlympiA phase III trial of adjuvant olaparib in patients with germline pathogenic variants in BRCA1/2 and high-risk, early breast cancer. Ann Oncol. 2022;33(12):1250-68.

253. Kulkarni DA, Vazquez A, Haffty BG, Bandera EV, Hu W, Sun YY, et al. A polymorphic variant in human MDM4 associates with accelerated age of onset of estrogen receptor negative breast cancer. Carcinogenesis. 2009;30(11):1910-5.

254. Swetzig WM, Wang J, Das GM. Estrogen receptor alpha (ERα/ESR1) mediates the p53-independent overexpression of MDM4/MDMX and MDM2 in human breast cancer. Oncotarget. 2016;7(13):16049-69.

255. Haupt S, Vijayakumaran R, Miranda PJ, Burgess A, Lim E, Haupt Y. The role of MDM2 and MDM4 in breast cancer development and prevention. J Mol Cell Biol. 2017;9(1):53-61.

256. Saal LH, Gruvberger-Saal SK, Persson C, Lövgren K, Jumppanen M, Staaf J, et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. Nature genetics. 2008;40(1):102-7.

257. Dean SJR, Perks CM, Holly JMP, Bhoo-Pathy N, Looi L-M, Mohammed NAT, et al. Loss of PTEN Expression Is Associated With IGFBP2 Expression, Younger Age, and Late Stage in Triple-Negative Breast Cancer. American Journal of Clinical Pathology. 2014;141(3):323-33.

258. Shoman N, Klassen S, McFadden A, Bickis MG, Torlakovic E, Chibbar R. Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. Modern Pathology. 2005;18(2):250-9.

259. Lebok P, Kopperschmidt V, Kluth M, Hube-Magg C, Özden C, B T, et al. Partial PTEN deletion is linked to poor prognosis in breast cancer. BMC Cancer. 2015;15:963-.

260. Tanic N, Milovanovic Z, Tanic N, Dzodic R, Juranic Z, Susnjar S, et al. The impact of PTEN tumor suppressor gene on acquiring resistance to tamoxifen treatment in breast cancer patients. Cancer Biol Ther. 2012;13(12):1165-74.

261. Li S, Shen Y, Wang M, Yang J, Lv M, Li P, et al. Loss of PTEN expression in breast cancer: association with clinicopathological characteristics and prognosis. Oncotarget. 2017;8(19):32043-54.

262. Presti D, Quaquarini E. The PI3K/AKT/mTOR and CDK4/6 Pathways in Endocrine Resistant HR+/HER2– Metastatic Breast Cancer: Biological Mechanisms and New Treatments. Cancers. 2019;11(9):1242.

263. NICE. Everolimus with exemestane for treating advanced breast cancer after endocrine therapy. NICE Technology appraisal guidance. 2016.

264. Yardley DA, Noguchi S, Pritchard KI, Burris HA, Baselga J, Gnant M, et al. Everolimus Plus Exemestane in Postmenopausal Patients with HR+ Breast Cancer: BOLERO-2 Final Progression-Free Survival Analysis. Advances in Therapy. 2013;30(10):870-84.

265. Owusu-Brackett N, Evans KW, Akcakanat A, Yuca E, Tapia C, Rizvi YQ, et al. TAK228 enhances antitumor activity of eribulin in triple negative breast cancer. Oncotarget. 2019;10(49):5011-9.

266. Chen IC, Chang Y-C, Lu Y-S, Chung K-P, Huang C-S, Lu T-P, et al. Clinical Relevance of Liver Kinase B1(LKB1) Protein and Gene Expression in Breast Cancer. Scientific reports. 2016;6:21374-.

267. Guo Q, Liu Z, Jiang L, Liu M, Ma J, Yang C, et al. Metformin inhibits growth of human non-small cell lung cancer cells via liver kinase B-1-independent activation of adenosine monophosphate-activated protein kinase. Mol Med Rep. 2016;13(3):2590-6.

268. Brown JR, DiGiovanna MP, Killelea B, Lannin DR, Rimm DL. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. Laboratory Investigation. 2014;94(1):98-106.

269. Abubakar M, Orr N, Daley F, Coulson P, Ali HR, Blows F, et al. Prognostic value of automated KI67 scoring in breast cancer: a centralised evaluation of 8088 patients from 10 study groups. Breast Cancer Res. 2016;18(1):104-.

270. Alco G, Bozdogan A, Selamoglu D, Pilanci KN, Tuzlali S, Ordu C, et al. Clinical and histopathological factors associated with Ki-67 expression in breast cancer patients. Oncol Lett. 2015;9(3):1046-54.

271. Alba E, Lluch A, Ribelles N, Anton-Torres A, Sanchez-Rovira P, Albanell J, et al. High Proliferation Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Early Breast Cancer. The Oncologist. 2016;21(2):150-5.

272. Smith IE, Dowsett M, Ebbs SR, Dixon JM, Skene A, Blohmer JU, et al. Neoadjuvant Treatment of Postmenopausal Breast Cancer With Anastrozole, Tamoxifen, or Both in Combination: The Immediate Preoperative Anastrozole, Tamoxifen, or Combined With Tamoxifen (IMPACT) Multicenter Double-Blind Randomized Trial. Journal of Clinical Oncology. 2005;23(22):5108-16.

273. Petrelli F, Viale G, Cabiddu M, Barni S. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. Breast Cancer Research and Treatment. 2015;153(3):477-91.

274. de Azambuja E, Cardoso F, de Castro G, Jr., Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. British journal of cancer. 2007;96(10):1504-13.

275. Menon SS, Guruvayoorappan C, Sakthivel KM, Rasmi RR. Ki-67 protein as a tumour proliferation marker. Clinica Chimica Acta. 2019;491:39-45.

276. Eom YH, Kim HS, Lee A, Song BJ, Chae BJ. BCL2 as a Subtype-Specific Prognostic Marker for Breast Cancer. J Breast Cancer. 2016;19(3):252-60.

277. Syed BM, Green AR, Nolan CC, Morgan DAL, Ellis IO, Cheung K-L. Biological characteristics and clinical outcome of triple negative primary breast cancer in older women - comparison with their younger counterparts. PloS one. 2014;9(7):e100573-e.

278. Elledge RM, Green S, Howes L, Clark GM, Berardo M, Allred DC, et al. bcl-2, p53, and response to tamoxifen in estrogen receptor-positive metastatic breast cancer: a Southwest Oncology Group study. Journal of Clinical Oncology. 1997;15(5):1916-22.

279. Zhang G-J, Kimijima I, Onda M, Kanno M, Sato H, Watanabe T, et al. Tamoxifen-induced Apoptosis in Breast Cancer Cells Relates to Down-Regulation of bcl-2, but not bax and bcl-X_L, without Alteration of p53 Protein Levels. Clinical Cancer Research. 1999;5(10):2971.

280. Callagy GM, Webber MJ, Pharoah PDP, Caldas C. Meta-analysis confirms BCL2 is an independent prognostic marker in breast cancer. BMC Cancer. 2008;8:153-.

281. Cory S, Huang DCS, Adams JM. The Bcl-2 family: roles in cell survival and oncogenesis. Oncogene. 2003;22(53):8590-607.

282. Dawson SJ, Makretsov N, Blows FM, Driver KE, Provenzano E, Le Quesne J, et al. BCL2 in breast cancer: a favourable prognostic marker across molecular subtypes and independent of adjuvant therapy received. British journal of cancer. 2010;103(5):668-75.

283. Campbell KJ, Tait SWG. Targeting BCL-2 regulated apoptosis in cancer. Open Biol. 2018;8(5):180002.

284. Chen J, Jin S, Abraham V, Huang X, Liu B, Mitten MJ, et al. The Bcl-2/Bcl-X_L/Bcl-w Inhibitor, Navitoclax, Enhances the Activity of Chemotherapeutic Agents In Vitro and <em>In Vivo</em&gt. Molecular Cancer Therapeutics. 2011;10(12):2340.

285. Singhai R, Patil VW, Jaiswal SR, Patil SD, Tayade MB, Patil AV. E-Cadherin as a diagnostic biomarker in breast cancer. N Am J Med Sci. 2011;3(5):227-33.

286. Yang L, Wang X-W, Zhu L-P, Wang H-L, Wang B, Zhao Q, et al. Significance and prognosis of epithelial-cadherin expression in invasive breast carcinoma. Oncol Lett. 2018;16(2):1659-65.

287. Arthur LM, Turnbull AK, Webber VL, Larionov AA, Renshaw L, Kay C, et al. Molecular Changes in Lobular Breast Cancers in Response to Endocrine Therapy. Cancer Research. 2014;74(19):5371-6.

288. Wang Q, Gun M, Hong XY. Induced Tamoxifen Resistance is Mediated by Increased Methylation of E-Cadherin in Estrogen Receptor-Expressing Breast Cancer Cells. Sci Rep. 2019;9(1):14140.

289. Younis LK, El Sakka H, Haque I. The Prognostic Value of E-cadherin Expression in Breast Cancer. Int J Health Sci (Qassim). 2007;1(1):43-51.

290. Yu W, Yang L, Li T, Zhang Y. Cadherin Signaling in Cancer: Its Functions and Role as a Therapeutic Target. Front Oncol. 2019;9:989-.

291. Devemy E, Blaschuk OW. Identification of a novel N-cadherin antagonist. Peptides. 2008;29(11):1853-61.

292. Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. Journal of hematology & oncology. 2018;11(1):64-.

293. Kawai T, Iwata K, Shinotsuka Y, Kubo S, Masuoka H, Yabuta T, et al. CD44v8-10 and CD44s Are Age-dependently Expressed in Primary Cultured Papillary Thyroid Carcinoma Cells and Are Associated with Cell Proliferation. Kobe J Med Sci. 2019;65(1):E1-E9.

294. Deng Y, Chen Z-J, Lan F, He Q-T, Chen S-Y, Du Y-F, et al. Association of CD44 polymorphisms and susceptibility to HBV-related hepatocellular carcinoma in the Chinese population. J Clin Lab Anal. 2019;33(8):e22977-e.

295. Hong SP, Chan TE, Lombardo Y, Corleone G, Rotmensz N, Bravaccini S, et al. Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. Nat Commun. 2019;10(1):3840.

296. Kong Y, Lyu N, Wu J, Tang H, Xie X, Yang L, et al. Breast cancer stem cell markers CD44 and ALDH1A1 in serum: distribution and prognostic value in patients with primary breast cancer. Journal of Cancer. 2018;9(20):3728-35.

297. Wang H, Wang L, Song Y, Wang S, Huang X, Xuan Q, et al. CD44(+)/CD24(-) phenotype predicts a poor prognosis in triple-negative breast cancer. Oncol Lett. 2017;14(5):5890-8.

298. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. 2013;2(1):78-91.

299. Honeth G, Bendahl P-O, Ringnér M, Saal LH, Gruvberger-Saal SK, Lövgren K, et al. The CD44+/CD24- phenotype is enriched in basal-like breast tumors. Breast Cancer Res. 2008;10(3):R53-R.

300. Shao M-M, Chan SK, Yu AMC, Lam CCF, Tsang JYS, Lui PCW, et al. Keratin expression in breast cancers. Virchows Archiv. 2012;461(3):313-22.

301. Yang J, Gao S, Xu J, Zhu J. Prognostic value and clinicopathological significance of serum- and tissue-based cytokeratin 18 express level in breast cancer: a meta-analysis. Biosci Rep. 2018;38(2):BSR20171145.

302. Bozionellou V, Mavroudis D, Perraki M, Papadopoulos S, Apostolaki S, Stathopoulos E, et al. Trastuzumab Administration Can Effectively Target Chemotherapy-Resistant Cytokeratin-19 Messenger RNA–Positive Tumor Cells in the Peripheral Blood and Bone Marrow of Patients With Breast Cancer. Clinical Cancer Research. 2004;10(24):8185.

303. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. The American journal of pathology. 2002;161(6):1991-6.

304. Haupt B, Ro JY, Schwartz MR. Basal-like Breast Carcinoma: A Phenotypically Distinct Entity. Archives of Pathology & Laboratory Medicine. 2010;134(1):130-3.

305. Stergiou N, Nagel J, Pektor S, Heimes A-S, Jäkel J, Brenner W, et al. Evaluation of a novel monoclonal antibody against tumor-associated MUC1 for diagnosis and prognosis of breast cancer. Int J Med Sci. 2019;16(9):1188-98.

306. Haddon L, Hugh J. MUC1-mediated motility in breast cancer: a review highlighting the role of the MUC1/ICAM-1/Src signaling triad. Clinical & Experimental Metastasis. 2015;32(4):393-403.

307. Maeda T, Hiraki M, Jin C, Rajabi H, Tagde A, Alam M, et al. MUC1-C Induces PD-L1 and Immune Evasion in Triple-Negative Breast Cancer. Cancer Research. 2018;78(1):205.

308. Taylor-Papadimitriou J, Burchell JM, Graham R, Beatson R. Latest developments in MUC1 immunotherapy. Biochem Soc Trans. 2018;46(3):659-68.

309. Chang AK, Wu H. The role of AIB1 in breast cancer. Oncol Lett. 2012;4(4):588-94.

310. Lee K, Lee A, Song BJ, Kang CS. Expression of AIB1 protein as a prognostic factor in breast cancer. World journal of surgical oncology. 2011;9:139-.

311. Weiner M, Skoog L, Fornander T, Nordenskjöld B, Sgroi DC, Stål O. Oestrogen receptor co-activator AIB1 is a marker of tamoxifen benefit in postmenopausal breast cancer. Ann Oncol. 2013;24(8):1994-9.

312. Osborne CK, Bardou V, Hopp TA, Chamness GC, Hilsenbeck SG, Fuqua SAW, et al. Role of the Estrogen Receptor Coactivator AIB1 (SRC-3) and HER-2/neu in

Tamoxifen Resistance in Breast Cancer. JNCI: Journal of the National Cancer Institute. 2003;95(5):353-61.

313. Alkner S, Jensen MB, Rasmussen BB, Bendahl PO, Fernö M, Rydén L, et al. Prognostic and predictive importance of the estrogen receptor coactivator AIB1 in a randomized trial comparing adjuvant letrozole and tamoxifen therapy in postmenopausal breast cancer: the Danish cohort of BIG 1-98. Breast cancer research and treatment. 2017;166(2):481-90.

314. Lonard DM, O'Malley BW. Molecular Pathways: Targeting Steroid Receptor Coactivators in Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016;22(22):5403-7.

315. Song X, Zhang C, Zhao M, Chen H, Liu X, Chen J, et al. Steroid Receptor Coactivator-3 (SRC-3/AIB1) as a Novel Therapeutic Target in Triple Negative Breast Cancer and Its Inhibition with a Phospho-Bufalin Prodrug. PloS one. 2015;10(10):e0140011-e.

316. El Ansari R, McIntyre A, Craze ML, Ellis IO, Rakha EA, Green AR. Altered glutamine metabolism in breast cancer; subtype dependencies and alternative adaptations. Histopathology. 2018;72(2):183-90.

317. Alfarsi LH, El-Ansari R, Craze ML, Masisi BK, Mohammed OJ, Ellis IO, et al. Co-Expression Effect of SLC7A5/SLC3A2 to Predict Response to Endocrine Therapy in Oestrogen-Receptor-Positive Breast Cancer. Int J Mol Sci. 2020;21(4).

318. Alfarsi LH, El Ansari R, Craze ML, Mohammed OJ, Masisi BK, Ellis IO, et al. SLC1A5 co-expression with TALDO1 associates with endocrine therapy failure in estrogen receptor-positive breast cancer. Breast Cancer Res Treat. 2021;189(2):317-31.

319. El Ansari R, Craze ML, Miligy I, Diez-Rodriguez M, Nolan CC, Ellis IO, et al. The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours. Breast Cancer Res. 2018;20(1):21.

320. Ziliotto S, Gee JMW, Ellis IO, Green AR, Finlay P, Gobbato A, et al. Activated zinc transporter ZIP7 as an indicator of anti-hormone resistance in breast cancer. Metallomics. 2019;11(9):1579-92.

321. Wilkinson RDA, Burden RE, McDowell SH, McArt DG, McQuaid S, Bingham V, et al. A Novel Role for Cathepsin S as a Potential Biomarker in Triple Negative Breast Cancer. J Oncol. 2019;2019:3980273-.

322. Alfarsi LH, Ansari RE, Craze ML, Toss MS, Masisi B, Ellis IO, et al. CDC20 expression in oestrogen receptor positive breast cancer predicts poor prognosis and lack of response to endocrine therapy. Breast Cancer Research and Treatment. 2019;178(3):535-44.

323. Joseph C, Al-Izzi S, Alsaleem M, Kurozumi S, Toss MS, Arshad M, et al. Retinoid X receptor gamma (RXRG) is an independent prognostic biomarker in ER-positive invasive breast cancer. British Journal of Cancer. 2019;121(9):776-85.

324. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. BMC Cancer. 2019;19(1):230.

325. Ballman KV. Biomarker: Predictive or Prognostic? J Clin Oncol. 2015;33(33):3968-71.

326. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. British journal of cancer. 1982;45(3):361-6.

327. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer Program to Assist in Making Decisions About Adjuvant Therapy for Women With Early Breast Cancer. Journal of Clinical Oncology. 2001;19(4):980-91.

328. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ, et al. Long-term impact of the 70-gene signature on breast cancer outcome. Breast Cancer Res Treat. 2014;143(3):587-92.

329. Aalders K, Genbrugge E, Poncet C, Kuijer A, Pistilli B, Piccart M, et al. Young age and the risk of disease recurrence as assessed by the 70-gene signature - an analysis from the EORTC 10041/BIG 03-04 MINDACT trial. Cancer Research. 2018;78(4 Supplement 1).

330. Piccart M, van 't Veer LJ, Poncet C, Lopes Cardozo JMN, Delaloge S, Pierga J-Y, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. The Lancet Oncology. 2021;22(4):476-88.

331. Ross DT, Kim C-y, Tang G, Bohn OL, Beck RA, Ring BZ, et al. Chemosensitivity and Stratification by a Five Monoclonal Antibody Immunohistochemistry Test in the NSABP B14 and B20 Trials. Clinical Cancer Research. 2008;14(20):6602-9.

332. Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MCU, et al. Novel Prognostic Immunohistochemical Biomarker Panel for Estrogen Receptor–Positive Breast Cancer. Journal of Clinical Oncology. 2006;24(19):3039-47.

333. Bartlett JM, Bloom KJ, Piper T, Lawton TJ, van de Velde CJ, Ross DT, et al. Mammostrat as an immunohistochemical multigene assay for prediction of early relapse

risk in the tamoxifen versus exemestane adjuvant multicenter trial pathology study. J Clin Oncol. 2012;30(36):4477-84.

334. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. Ann Oncol. 2014;25(2):339-45.

335. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A New Molecular Predictor of Distant Recurrence in ER-Positive, HER2-Negative Breast Cancer Adds Independent Information to Conventional Clinical Risk Factors. Clinical Cancer Research. 2011;17(18):6012.

336. Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K, et al. Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. Breast Cancer Res. 2014;16(2):R38.

337. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. 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. Journal of Clinical Oncology. 2011;29(32):4273-8.

338. Bartlett JM, Christiansen J, Gustavson M, Rimm DL, Piper T, van de Velde CJ, et al. Validation of the IHC4 Breast Cancer Prognostic Algorithm Using Multiple

Approaches on the Multinational TEAM Clinical Trial. Arch Pathol Lab Med. 2016;140(1):66-74.

339. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. New England Journal of Medicine. 2018;379(2):111-21.

340. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, et al. 21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer. New England Journal of Medicine. 2021;385(25):2336-47.

341. Straver ME, Glas AM, Hannemann J, Wesseling J, van de Vijver MJ, Rutgers EJ, et al. The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. Breast Cancer Res Treat. 2010;119(3):551-8.

342. Bear HD, Wan W, Robidoux A, Rubin P, Limentani S, White RL, Jr., et al. Using the 21-gene assay from core needle biopsies to choose neoadjuvant therapy for breast cancer: A multicenter trial. J Surg Oncol. 2017;115(8):917-23.

343. Lee J, Lee EH, Park HY, Kim WW, Lee RK, Chae YS, et al. Efficacy of an RNA-based multigene assay with core needle biopsy samples for risk evaluation in hormone-positive early breast cancer. BMC Cancer. 2019;19(1):388.

344. Ensani F, Omranipour R, Jahanzad I, Jafari A, Nafarzadeh S, Aminishakib P. The Core Needle and Surgical Biopsy Concordance to Detect Estrogen, Progesterone, and Her-2 Receptors in Breast Cancer: A Comparative Study. Iran J Pathol. 2017;12(3):202-8.

345. Karakas C, Biernacka A, Bui T, Sahin AA, Yi M, Akli S, et al. Cytoplasmic Cyclin E and Phospho-Cyclin-Dependent Kinase 2 Are Biomarkers of Aggressive Breast Cancer. Am J Pathol. 2016;186(7):1900-12.

346. Thomson TA, Zhou C, Ceballos K, Knight B. Tissue microarray for routine clinical breast biomarker analysis. The British Columbia Cancer Agency 2008 experience. Am J Clin Pathol. 2010;133(6):909-14.

347. Camp RL CL, Rimm DL. Validation of tissue microarray technology in breast carcinoma. Lab Invest. 2000;80(12):1943-9.

348. Syed BM. A study of biological characteristics of early operable primary breast cancer in older women and correlation with clinical outcome. University of Nottingha, Thesis. January 2012.

349. Albanghali M, Green A, Rakha E, Aleskandarany M, Nolan C, Ellis I, et al. Construction of tissue microarrays from core needle biopsies – a systematic literature review. Histopathology. 2015;68(3):323-32.

350. Li K, Du H, Lian X, Yang S, Chai D, Wang C, et al. Characterization of β2-microglobulin expression in different types of breast cancer. BMC Cancer. 2014;14:750-.

351. McCarty KS Jr ML, Cox EB, Konrath J, McCary KS Sr. Estrogen receptor analysis. Correlation of biochemical and immunohistochemical methods using monoclonal anitreceptor antibodies. Archives of Pathology & Laboratory Medicine. 1985;109(8):716 - 21.

352. Vijayashree R, Aruthra P, Rao KR. A comparison of manual and automated methods of quantitation of oestrogen/progesterone receptor expression in breast carcinoma. J Clin Diagn Res. 2015;9(3):Ec01-5.

353. Kirkwood B, JAC S. Essential Medical Statistics. Hoboken: John Wiley & Sons Incoporated; 2003.

354. Camp RL, Dolled-Filhart M, Rimm DL. X-Tile. A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. 2004;10(21):7252-9.

355. Soria D, Garibaldi JM, Ambrogi F, Green AR, Powe D, Rakha E, et al. A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients. Computers in Biology and Medicine. 2010;40(3):318-30.

356. Lancashire LJ, Lemetre C, Ball GR. An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies. Briefings in Bioinformatics. 2009;10(3):315-29.

357. RM Parks MA, BM Syed, AR Green, IO Ellis and KL Cheung. Biology of oestrogen-receptor positive breast cancer in older women using biopsy. Submitted to Endocrine Related Cancers. 2020.

358. Karlsson C, Karlsson MG. Effects of long-term storage on the detection of proteins, DNA, and mRNA in tissue microarray slides. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2011;59(12):1113-21.

359. DiVito KA, Charette LA, Rimm DL, Camp RL. Long-term preservation of antigenicity on tissue microarrays. Laboratory Investigation. 2004;84(8):1071-8.

360. Albanghali MA, Rakha E, Green AR, Syed BM, Nolan C, Diez-Rodriguez MA, et al. Potential of a novel technique for constructing tissue microarrays from core needle biopsy as seen in older women with primary breast cancer. Journal of Clinical Oncology. 2016;34(Supplement 15).

361. Kim S-W, Roh J, Park C-S. Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. J Pathol Transl Med. 2016;50(6):411-8.

362. Voskuil J. Commercial antibodies and their validation. F1000Res. 2014;3:232-.

363. Gill KL, Machavaram KK, Rose RH, Chetty M. Potential Sources of Inter-Subject Variability in Monoclonal Antibody Pharmacokinetics. Clin Pharmacokinet. 2016;55(7):789-805.

364. Abdi J. Biological assessment - part of a novel tool to optimise treatment of primary breast cancer in older women: Focusing on AR as a potential prognostic marker. University of Nottingham dissertation for MSc in Oncology. 2018.

365. Bronte G, Bravaccini S, Ravaioli S, Puccetti M, Scarpi E, Andreis D, et al. Androgen Receptor Expression in Breast Cancer: What Differences Between Primary Tumor and Metastases? Transl Oncol. 2018;11(4):950-6.

366. Agrawal A, Ziolkowski P, Grzebieniak Z, Jelen M, Bobinski P, Agrawal S. Expression of Androgen Receptor in Estrogen Receptor-positive Breast Cancer. Applied immunohistochemistry & molecular morphology : AIMM. 2016;24(8):550-5.

367. Riaz N, Idress R, Habib S, Lalani E-N. Lack of Androgen Receptor Expression Selects for Basal-Like Phenotype and Is a Predictor of Poor Clinical Outcome in Non-Metastatic Triple Negative Breast Cancer. Front Oncol. 2020;10.

368. Collins LC, Cole KS, Marotti JD, Hu R, Schnitt SJ, Tamimi RM. Androgen receptor expression in breast cancer in relation to molecular phenotype: results from the Nurses' Health Study. Mod Pathol. 2011;24(7):924-31.

369. Park S, Koo J, Park HS, Kim JH, Choi SY, Lee JH, et al. Expression of androgen receptors in primary breast cancer. Annals of Oncology. 2010;21(3):488-92.

370. The Human Protein Atlas. BRCA [online]. Available at: https://www.proteinatlas.org/ENSG00000139618-BRCA2 [Accessed on: 31st July 2022].

371. Ricci MD, Calvano Filho CM, Oliveira Filho HR, Filassi JR, Pinotti JA, Baracat EC. Analysis of the concordance rates between core needle biopsy and surgical excision in patients with breast cancer. Revista da Associacao Medica Brasileira (1992). 2012;58(5):532-6.

372. Rakha EA, Ellis IO. An overview of assessment of prognostic and predictive factors in breast cancer needle core biopsy specimens. J Clin Pathol. 2007;60(12):1300-6.

373. Shanmugalingam A, Hitos K, Hegde S, Al-Mashat A, Pathmanathan N, Edirimmane S, et al. Concordance between core needle biopsy and surgical excision for breast cancer tumor grade and biomarkers. Breast Cancer Res Treat. 2022;193(1):151-9.

374. Kalvala J, Parks RM, Green AR, Cheung KL. Concordance between core needle biopsy and surgical excision specimens for Ki-67 in breast cancer - a systematic review of the literature. Histopathology. 2022;80(3):468-84.

375. Arnedos M, Nerurkar A, Osin P, A'Hern R, Smith IE, Dowsett M. Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). Annals of Oncology. 2009;20(12):1948-52.

376. Lebeau A, Turzynski A, Braun S, Behrhof W, Fleige B, Schmitt WD, et al. Reliability of human epidermal growth factor receptor 2 immunohistochemistry in breast core needle biopsies. J Clin Oncol. 2010;28(20):3264-70.

377. Zhang D, Salto-Tellez M, Putti TC, Do E, Koay ES-C. Reliability of Tissue Microarrays in Detecting Protein Expression and Gene Amplification in Breast Cancer. Modern Pathology. 2003;16(1):79-85.

378. Hewitt SM. Tissue microarrays as a tool in the discovery and validation of predictive biomarkers. Methods Mol Biol. 2012;823:201-14.

379. Voduc D, Kenney C, Nielsen TO. Tissue microarrays in clinical oncology. Semin Radiat Oncol. 2008;18(2):89-97.

380. Cohen DA, Dabbs DJ, Cooper KL, Amin M, Jones TE, Jones MW, et al. Interobserver Agreement Among Pathologists for Semiquantitative Hormone Receptor Scoring in Breast Carcinoma. American Journal of Clinical Pathology. 2012;138(6):796-802.

381. Kirkegaard T, Edwards J, Tovey S, McGlynn LM, Krishna SN, Mukherjee R, et al. Observer variation in immunohistochemical analysis of protein expression, time for a change? Histopathology. 2006;48(7):787-94.

382. Ram S, Vizcarra P, Whalen P, Deng S, Painter CL, Jackson-Fisher A, et al. Pixelwise H-score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. PLoS One. 2021;16(9):e0245638.

383. Bankhead P, Fernández JA, McArt DG, Boyle DP, Li G, Loughrey MB, et al. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. Laboratory Investigation. 2018;98(1):15-26.

384. Ibrahim A, Gamble P, Jaroensri R, Abdelsamea MM, Mermel CH, Chen P-HC, et al. Artificial intelligence in digital breast pathology: Techniques and applications. Breast (Edinburgh, Scotland). 2020;49:267-73.

385. Anagnostou VK, Welsh AW, Giltnane JM, Siddiqui S, Liceaga C, Gustavson M, et al. Analytic variability in immunohistochemistry biomarker studies. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(4):982-91.

386. Røge R, Nielsen S, Riber-Hansen R, Vyberg M. Impact of Primary Antibody Clone, Format, and Stainer Platform on Ki67 Proliferation Indices in Breast Carcinomas. Applied Immunohistochemistry & Molecular Morphology. 2019;27(10):732-9.

387. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202.

388. M Rekha. Eigen Decomposition and PCA [online]. 2019. Available at: https://blog.clairvoyantsoft.com/eigen-decomposition-and-pca-c50f4ca15501 [Accessed: 13th May 2022].

389. Sims AH. Bioinformatics and breast cancer: what can high-throughput genomic approaches actually tell us? J Clin Pathol. 2009;62(10):879-85.

390. Amjad E, Asnaashari S, Sokouti B, Dastmalchi S. Systems biology comprehensive analysis on breast cancer for identification of key gene modules and genes associated with TNM-based clinical stages. Scientific Reports. 2020;10(1):10816.

391. de Souto MCP, Jaskowiak PA, Costa IG. Impact of missing data imputation methods on gene expression clustering and classification. BMC Bioinformatics. 2015;16(1):64.

392. Salgado CM, Azevedo C, Proença H, Vieira SM. Missing Data. In: Data MITC, editor. Secondary Analysis of Electronic Health Records. Cham: Springer International Publishing; 2016. p. 143-62.

393. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. BMC Medical Informatics and Decision Making. 2016;16(3):74.

394. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. Bioinformatics. 2001;17(6):520-5.

395. Picard RR, Cook RD. Cross-Validation of Regression Models. Journal of the American Statistical Association. 1984;79(387):575-83.

396. Lancashire LJ, Powe DG, Reis-Filho JS, Rakha E, Lemetre C, Weigelt B, et al. A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. Breast Cancer Res Treat. 2010;120(1):83-93.

397. Liang Y, Brekken RA, Hyder SM. Vascular endothelial growth factor induces proliferation of breast cancer cells and inhibits the anti-proliferative activity of anti-hormones. Endocr Relat Cancer. 2006;13(3):905-19.

398. Manders P, Beex LV, Tjan-Heijnen VC, Span PN, Sweep CG. Vascular endothelial growth factor is associated with the efficacy of endocrine therapy in patients with advanced breast carcinoma. Cancer. 2003;98(10):2125-32.

399. Kenji Suzuki (Editor). Artificial Neural Networks: Architectures and Applications. IntechOpen. 2013.

400. Parks RM, Hall L, Tang S-W, Howard P, Lakshmanan R, Winterbottom L, et al. The potential value of comprehensive geriatric assessment in evaluating older women

with primary operable breast cancer undergoing surgery or non-operative treatment - A pilot study. Journal of Geriatric Oncology. 2015;6(1):46-51.

401. Parks* RM, Howard P, Tang SW, Hall L, Lakshmanan R, Winterbottom L, et al. Assessment of changes of quality of life (QOL) following initial treatment of older women with primary operable breast cancer. Journal of Geriatric Oncology. 2012;3:S42.

Appendices

Appendix 1: Ethical approval confirmation for study 340

Appendix 2: Protocol..... 346

Appendix 3: Cluster distribution by biomarker for cluster analysis of all CNB TMA samples, identifying three clusters (Cluster 0 are the cases which are unclustered) 380

Appendix 4: Decision tree model to determine key drivers of the classification used for cluster analysis in the CNB TMA cohort 390

Appendix 5: Prognostic algorithm designed to predict BCSS in SE TMA cohort 391

Appendix 6: Prognostic algorithm designed to predict BCSS in CNB TMA cohort..... 395

Appendix 1: Ethical approval confirmation for study

Yorkshire & The Humber - Leeds East Research Ethics Committee
NHSBT Newcastle
Blood Donor Centre
Holland Drive
Newcastle upon Tyne
NE2 4NQ

Telephone: 0207 1048 088

23 August 2019

Professor Emad A Rakha
Professor of Breast Pathology/Consultant Pathologist
University of Nottingham
Department of Pathology
Nottingham City Hospital NHS Trust
Nottingham
NG5 1PB

Dear Professor Rakha

Study title:	**Refining the prognosis of breast cancer and breast cancer precursor lesions for personalised therapy**
REC reference:	**19/YH/0293**
IRAS project ID:	**266925**

The Proportionate Review Sub-committee of the Yorkshire & The Humber - Leeds East Research Ethics Committee reviewed the above application by correspondence.

Ethical opinion

On behalf of the Committee, the sub-committee gave a **favourable** ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) should be sought from all NHS

organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.
Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

It is a condition of the REC favourable opinion that all clinical trials are registered on a publicly accessible database. For this purpose, clinical trials are defined as the first four project categories in IRAS project filter question 2. For clinical trials of investigational medicinal products (CTIMPs), other than adult phase I trials, registration is a legal requirement.

Registration should take place as early as possible and within six weeks of recruiting the first research participant at the latest. Failure to register is a breach of these approval conditions, unless a deferral has been agreed by or on behalf of the Research Ethics Committee (see here for more information on requesting a deferral: https://www.hra.nhs.uk/planning-and-improving-research/research-planning/research-registratio n-research-project-identifiers/

As set out in the UK Policy Framework, research sponsors are responsible for making information about research publicly available before it starts e.g. by registering the research project on a publicly accessible register. Further guidance on registration is available at: https://www.hra.nhs.uk/planning-and-improving-research/research-planning/transparency-responsibilities/

You should notify the REC of the registration details. We routinely audit applications for compliance with these conditions.

Publication of Your Research Summary

We will publish your research summary for the above study on the research summaries section of our website, together with your contact details, no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, make a request to defer, or require further information, please visit: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

After ethical review: Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study, including early termination of the study
- Final report

The latest guidance on these topics can be found at https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Approved documents

The documents reviewed and approved were:

Document	Version	Date
Covering letter on headed paper (Covering Letter IRAS266925)		26 July 2019
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) (Evidence of Insurance - University of Nottingham &/or Subsidiary Companies)		31 July 2018
IRAS Application Form (IRAS_Form_06082019)		06 August 2019
Letter from sponsor (19052 Sponsor Letter)		30 July 2019
Other (FINAL APPROVAL OF CHARITY RESEARCH APPLICATION REF: CRF-RAKHA-NOV 2017)	NA	15 May 2017
Other (Pathological Society Consultant's Pump-Priming Small Grant Application)	NA	28 June 2017
Other (Research Grant Application: The role of SEC14L In Breast Cancer)	NA	02 May 2017
Other (2018 Investigator Initiated Research Scheme - Award Confirmation)	NA	27 November 2017
Other (Life Cycle 6 funds Disbursement)	NA	27 March 2017
Participant consent form (NHSB Consent form v7.0 Final)	7.0	01 May 2018
Participant information sheet (PIS) (NHSB Information Leaflet NUH02168N v9.0)	9.0	18 May 2018
Participant information sheet (PIS) (Donation of surplus tissues for research - Patient Information Leaflet)	1.0	02 January 2013
Research protocol or project proposal (Refining prognosis of breast cancer for personalised therapy)	1.0	26 July 2019
Summary CV for Chief Investigator (CI) (Summary CV for CI Professor Emad Rakha)		26 July 2019
Summary CV for student (Summary CV for Students_IRAS 266925)	2.0	

Summary CV for supervisor (student research) (Summary CV for Supervisor D Andrew Green)		26 July 2019

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

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

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

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

19/YH/0293 Please quote this number on all correspondence

Yours sincerel y pp

Sarah Prothero

Dr Deborah Jane Fox Vice Chair

Email: nrescommittee.yorkandhumber-leedseast@nhs.net

Enclosures: List of names and professions of members who took part in the review

"After ethical review – guidance for researchers"

Copy to: Ms Angela Shone

Lead Nation - England: HRA.Approval@nhs.net

Yorkshire & The Humber - Leeds East Research Ethics Committee

Attendance at PRS Sub-Committee of the REC meeting held by correspondence

Committee Members:

Name	Profession	Present
Dr Deborah Jane Fox (Chair)	Assistant Director of Nursing and Patient Safety	Yes
Mr Aidan Hindley	Tissue Bank Coordinator	Yes
Mrs Ann Kay	Retired Special Needs Coordinator	Yes

Also in attendance:

Name	Position (or reason for attending)
Miss Sarah Prothero	Approvals Officer

Appendix 2: Protocol



FULL TITLE OF THE HUMAN TISSUE STUDY

Refining prognosis of breast cancer for personalised therapy

Draft 2.0 / Final Version 1.0
04/07/2019

Short title: **Refining prognosis of breast cancer for personalised therapy**

IRAS Project ID: *266925*

Study Sponsor: University of Nottingham

Sponsor reference: *19052*

Funding Source: The research projects are sponsored by the PhD students' funds and various grants

STUDY PERSONNEL AND CONTACT DETAILS

Sponsor:	University of Nottingham
Contact name	Ms Angela Shone Research and Innovation University of Nottingham East Atrium Jubilee Conference Centre

Triumph Road
Nottingham
NG8 1DH

Chief investigator:

Prof Emad A Rakha
Professor of Breast Pathology and
Consultant Pathologist
Division of Cancer and Stem Cells
School of Medicine, University of Nottingham
Nottingham City Hospital,
Nottingham, NG5 1PB
Phone: 0115 9691169 ext 56875
Email: emad.rakha@nottingham.ac.uk

Co-investigators	**Contact**
Ian O Ellis	UoN, Ian.Ellis@nottnigham.ac.uk
Srinivasan Madhusudan	UoN, srinivasan.madhusudan@nottingham.ac.uk
Andrew Green	UoN, Andrew.green@nottingham.ac.uk
Mohammed Aleskandarany	UoN, mohammed.aleskandarany@nottingham.ac.uk
Michael Toss	UoN, msxms8@exmail.nottingham.ac.uk
Islam Miligy	UoN, msxima@exmail.nottingham.ac.uk
Wakkas Fadhil	UoN, mszwf@exmail.nottingham.ac.uk
Chitra Joseph	UoN, mszcj@exmail.nottingham.ac.uk
Jaineeta Richardson	UoN, jaineeta.richardson@gmail.com
Holly Nicholls	UoN, mszhrn@exmail.nottingham.ac.uk

Students	**Qualification studying for**
Mansour Alsaleem	PhD, msxma39@exmail.nottingham.ac.uk
Lutfi Alfarsi	PhD, msxla8@nottingham.ac.uk
Maryam Althobiti	PhD, msxma32@nottingham.ac.uk
Rokaya El Ansari	PhD, msxre2@nottingham.ac.uk
Sara Abuelmaaty	PhD, mzxsra@exmail.nottingham.ac.uk
Yousif Kariri	PhD, msxyk2@exmail.nottingham.ac.uk
Sami Alsaeed	PhD, msxsa33@exmail.nottingham.ac.uk
Khloud Elsharawy	PhD, msxkae@exmail.nottingham.ac.uk

Abrar Aljohani	PhD, msxaa98@exmail.nottingham.ac.uk
Brenda Masisi	PhD, msxbkma@exmail.nottingham.ac.uk
Abdulbaqi Hamad	PhD, msxabh@exmail.nottingham.ac.uk
Bandar Alshreef	PhD, msxba13@exmail.nottingham.ac.uk

Statistician:

Professor Graham Ball
Phone:+44 (0)115 848 8308
Email: graham.ball@ntu.ac.uk

Professor Nigel Mongan
Phone:+44 (0) 115 951 6625
Email: nigel.mongan@nottingham.ac.uk

Study Coordinating Centre:

Department of Translational Oncology
Breast Pathology Research Group.
Division of Cancer and Stem Cells,
School of Medicine
The University of Nottingham
City Hospital Campus, Nottingham, NG5 1PB

STUDY SYNOPSIS

Title	**Refining the prognosis of breast cancer for personalised therapy**
Chief Investigator	Professor Emad Rakha
Study centre(s)	Department of Academic Oncology, Breast Cancer Research Group, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham, City Hospital Campus, Nottingham NG5 1PB
Objectives	In this study we aim to determine the key mechanisms and driver genes/proteins/features controlling breast cancer behaviour and prognosis.
Number of participants / tissue samples	8,000 breast cases including breast cancer and its precursors lesions

Diagnosis and main criteria for inclusion	In this application we aim to study breast cancer and its precursor lesions in patients with breast disease. These cases were prepared as part of the ethically approved Breast Pathology Database (REC reference 19/EM/0162) which comprises a cohort of 8000 samples from patients who presented to Nottingham University Hospitals NHS trust with breast lesions and received surgical treatment. The data are available to us as pseudoanonymised. Inclusion criteria into this study: • Patients presenting with breast cancer and its precursor lesions at the Nottingham University Hospitals NHS Trust between 1st January 1988 - 31st December 2017. • Age range: Lower limit 18 years – upper 100 years • Availability of tissue material surplus to diagnostic pathology needs. • Availability of linked metadata.
Duration of study	Minimum of 5 years, with intention to secure further funding to extend this study.
Description of interventions	This retrospective study will include the use of formalin-fixed and paraffin-embedded (FFPE) tissue samples and linked metadata. No prospective cases will be included in this study. Metadata was collected through other ethically approved project as above and is available to the research group in a pseudoanonymised form. No collection of patients' identifiable information or patients contact is proposed in this application.
Statistical methods	For statistical analysis of research data, appropriate statistical software packages will be used including SPSS (SPPS Inc., Chicago, IL, USA), R package (R Foundation for Statistical Computing, Vienna, Austria), X-tile bioinformatics (Yale University, USA) and GraphPad Prism (GraphPad Prism for Windows/Mac, California, USA). Relevant statistical tests provided by these packages will be used to interrogate the statistical associations with clinicopathological factors and outcome variables. We have experience with statistical analysis of more than 400 biomarkers over the last 20 years which resulted in publications of more than 650 articles in peer reviewed journals. We perform all statistical tests needed, but when more advanced computational analyses are indicated we then work with our computer scientist collaborators who have track record in such analyses.

Table of contents

STUDY PERSONNEL AND CONTACT DETAILS 1
STUDY SYNOPSIS 3
ABBREVIATIONS 6
STUDY BACKGROUND INFORMATION AND RATIONALE 7
STUDY OBJECTIVES AND PURPOSE 8
STUDY DESIGN 9
STUDY CONFIGURATION 9
STUDY MANAGEMENT 10
DURATION OF THE STUDY AND PARTICIPANT INVOLVEMENT 10
SELECTION AND WITHDRAWAL OF STUDY PARTICIPANTS 10
Recruitment 10
Eligibility criteria 11
Inclusion criteria 11
Exclusion criteria 11
Informed consent 11
STUDY TREATMENT AND REGIMEN 12
Sample size and justification 12
TRANSPORT AND STORAGE OF THE TISSUES 12
LABORATORY ANALYSES 13
STATISTICS 16
Statistical analysis 16
Primary Outcome Measure 17
ETHICAL AND REGULATORY ASPECTS 17
Ethical Issues 17
Informed Consent and Participant Information 17
Withdrawal 17
RECORDS 18
Sample Labelling 18
Source documents 18
Data Protection 18
QUALITY ASSURANCE & AUDIT 19
INSURANCE AND INDEMNITY 19
STUDY CONDUCT 19
STUDY DATA 20
RECORD RETENTION AND ARCHIVING 20

DISCONTINUATION OF THE STUDY BY THE SPONSOR 20
STATEMENT OF CONFIDENTIALITY 20
PUBLICATION POLICY 22
STUDY FINANCES 22
Funding source 22
Participant stipends and payments 22
REFERENCES 22

ABBREVIATIONS

AI	Artificial Intelligence
DNA	Deoxyribonucleic Acid
DRE	Database Record Entries
FFPE	Formalin-fixed and paraffin-embedded
H&E	Haematoxylin and Eosin
ICGC	International Caner Genome Consortium
IHC	Immunohistochemistry
ISH	In Situ Hybridisation
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
NGS	Next Generation Sequencing
NHSB	Nottingham Health Science Biobank
NUH	Nottingham University Hospitals
PCR	Polymerase Chain Reaction
REC	Research Ethics Committee
RNA	Ribonucleic Acid
SPSS	Statistical Package for Social Sciences
TCGA	The Cancer Genome Atlas
TMA	Tissue Microarray
UoN	University of Nottingham
WB	Western Blot

STUDY BACKGROUND INFORMATION AND RATIONALE

Breast cancer affects over 56,000 women in the UK and 2.4 million women worldwide each year. Despite advancements in breast cancer screening programs, early detection and increased availability of therapeutic options, the mortality remains high with 20-30% of patients dying within the first 10 years after diagnosis and a similar proportion of patients relapse and develop resistance to current therapies. Moreover, there is still lack of specific therapeutic regimens for some breast cancer subgroups, such as triple negative and basal subtypes.

Despite extensive studies of breast cancer biology and behaviour, significant gaps in our understanding of the disease still exist. This mainly stems from the inter-tumour and intra-tumour heterogeneity of molecular profiles, resulting in variable clinical presentation, morphology, behaviour and response to therapy. Moreover, at the individual tumour levels, the behaviour is not only related to the expression of single or multiple genes by the tumour cells, but also to the tumour microenvironment and to the capacity of the cancer cells to infiltrate within and invade the local vascular and lymphatic channels. Understanding these various mechanisms will help not only in risk stratification of patients but also in identifying novel therapeutic targets that will lead to improved patients' outcome.

The breast research group in Nottingham has a long standing national and international track record in breast cancer diagnosis, prognostic and predictive stratification since 1970s with more than 650 publications in peer reviewed journals and several books and book chapters based on its research output. The group currently comprises 12 PhD, and MSc students, several undergraduate students, and 9 staff members led by a Consultant Pathologist and Professor of Breast Pathology who are active in breast cancer research. Nottingham Breast Research Group has one of the largest cohorts of samples of breast cancer and its precursor lesions with long term follow-up and treatment data worldwide and carried out an extensive number of previous ethically approved studies utilising subsets of these cohorts. A recent ethics approval to construct the Breast Pathology Database was granted (REC reference 19/EM/0162) to collect material and linked metadata from the patients presented with breast lesions to the Nottingham Hospitals from 1988 to 2017 to be used in various ethically approved research studies as pseudoanonymised. In that database, tissue material surplus to diagnostic pathology needs were collected from patients diagnosed and treated at Nottingham from 1988 to 2006 by the Nottingham breast surgical team led by Professor Roger Blamey and patient consents were taken by the surgical team. After 2010 cases were collected by the Nottingham University Hospitals NHS Trust (NUH) Trust-wide Biobank staff using REC approved consent forms. Some of the samples collected in between were collected by Prof Ian Ellis Biobank Team. Ethical approval to use subsets of these cases in research was previously given by the Nottingham Research Ethics Committee (REC) 2 for studies involving aspects of molecular and phenotypic alterations in these archival breast tissue samples (Ref: REC202313). After 2014, breast research projects were approved by the Nottingham Health Science Biobank Biobank (NHSB) Access Committee under the generic NHSB biobank ethical approval by the North West - Greater Manchester Central REC (Ref: 15/NW/0685 and 10/H1008/72).
However, in April 2019 the Biobank activities including the ethical approval of the existing projects were suspended. Therefore, an ethics approval to construct the

Breast Pathology Database was submitted and approved (REC reference

19/EM/0162) to allow collect material and linked metadata. In this Breast Pathology Database application, a tissue acquisition team was set up to collect samples and metadata and to have access to patient identifiable information. The team will release the cases to the research team as pseudoanonymised. We herein apply for approval of our ongoing breast research study to use the tissue and data from that database by the Nottingham Breast Pathology Research Group which comprises research staff and involves undergraduate and postgraduate students and other collaborators. This project aims at refining the diagnosis and prognosis of breast cancer and its precursor lesions. Below we explain the methodology used and the various aims of the project. This project is funded by PhD students' funds and various research grants including those from the Nottingham Breast Cancer Research Centre (NBCRC), the Pathological Society of Great Britain and Ireland, Nottingham Life Cycle 6, Breast Cancer Research Trust, Nottingham University Hospitals Charity, and the National Breast Cancer Foundation (NBCF) of Australia

In this proposal, we aim to continue and expand our comprehensive characterisation of breast diseases utilising tissue samples from this large series of cases using the state-of-the-art research techniques to investigate genomic, transcriptomic, and phenotypic changes associated with breast lesions. This will help patients by identifying novel targets that can be used to stratify them for management decisions and refine the use of systemic therapies. Better understanding of the complex molecular processes underlying breast cancer progression can lead to better treatment and improve outcomes. In this project we will use the power of the recent technologies (listed below) that investigate several thousands of genes to identify those associated with specific features of breast cancer development and progression. We will then study these genes individually in our breast diseases' series to test and validate their ability to identify risk of breast cancer progression and predict response to specific therapies. We aim to use different techniques to test the most likely candidate genes and features implicated in different molecular and neoplastic processes contributing to breast cancer initiation and progression that can be reliably used to diagnose and predict outcome and response to therapy to help personalised approach of patient management. We are also developing the use of image analysis, digital pathology and artificial intelligence in pathology as cost-efficient methods for diagnosis and classification of breast diseases. This proposal forms part of a long-term programme aimed at studying breast cancer and its precursor lesions and to improve the diagnostic and therapeutic strategies.

STUDY OBJECTIVES AND PURPOSE

PURPOSE: Our aim is to determine the key mechanisms/roles and drivers controlling breast cancer behaviour and prognosis in an attempt to refine classification of breast diseases and to address gaps in current management strategies. In our endeavour to identify novel diagnostic, prognostic and predictive biomarker in breast cancer, we will interrogate publicly available molecular datasets to utilise the power of gene expression microarrays which provide data on tens of thousands of genes. We will use various statistical and bioinformatics approaches to decipher these databases and to identify novel targets for further investigation using our large and well characterised breast cohorts using different techniques. Whole slide images will also be used to identify the key morphological diagnostic and prognostic factors in breast lesions using image analysis and artificial intelligence

approaches with the help of internal and external computer scientists' collaborators.

PRIMARY OBJECTIVE:

The principal research question of this study is whether identification of distinctive morphological and biological characteristics of breast cancer and its precursors lesions can improve our understanding of the biology. Our primary objective is to better classify breast cancer and its precursors lesions and to provide robust methods to predict its behaviour, outcomes and response to therapy, using this to improve personalised management of patients with breast disease.

SECONDARY OBJECTIVES:

We will use the information gained from genomic, transcriptomic, and phenotypic changes associated with breast lesions to develop *in vitro* (cell culture) models for studying breast cancer, which will enable more detailed analysis of mechanisms than is possible using patient-derived materials alone.

We also intend to use digitised images of stained sections from patient samples to develop and refine AI tools for aiding clinical diagnostics and improving the efficiency of research investigations.

STUDY DESIGN

STUDY CONFIGURATION

This is a single-centre study with most of the work being conducted at the University of Nottingham.

Where appropriate, we will collaborate with other institutions and commercial partners both within the UK and overseas to give us access to specialised techniques and facilities, or to validate our results. The Chief Investigator will ensure that collaborative ventures will be in line with this application and fully covered by ethical approvals in place. Any transfer of data or tissue will be carried out in compliance with all existing regulations, using duly authorised Material and Data transfer agreements.

Samples

The study will utilise up to 8,000 patient cases of formalin fixed paraffin embedded (FFPE) breast pathology tissue samples which are currently in long-term storage in the Histopathology Department, Nottingham University Hospitals NHS Trust as part of the patient records. Studies will use various subsets of this cohort based on tumour subtypes, treatment given and the specific aims of the study. We will use the retrospectively collected and archived breast tissues, benign breast lesions and lymph node resection specimens, which have been surgically removed as part of their normal care. In this study, materials surplus to diagnostic materials will be used.

This research includes cases included in the ethically approved Breast Pathology Database (REC reference 19/EM/0162) and it does not include any patient involvement or access to identifiable patients' information and the tissue usage will not affect or compromise and further pathological assessment of the specimen.

STUDY MANAGEMENT

The Chief Investigator of this project will be responsible for the tissue samples used, and data generated from the studies. The Chief Investigator will be ultimately responsible for the design and conduct of all studies carried out by staff and students within the group and in association with external collaborators. Responsibility for individual studies within the scope of this application may be delegated to key members of staff reporting to the Chief Investigator. Material or Data transfer agreement will be completed before transferring or using any of the samples or data with other collaborators.

DURATION OF THE STUDY AND PARTICIPANT INVOLVEMENT

The study duration will be 5 years starting from the REC approval date; we may seek to extend or renew this if circumstances permit.
This is retrospective study using archival material; there are no patient participants.

Research samples and data (pseudoanonymised) may be retained beyond the end of this study for use in other ethically-approved projects in accordance with HTA requirements.

SELECTION AND WITHDRAWAL OF STUDY PARTICIPANTS

Recruitment

The study will include retrospectively collected breast tissues from patients who attended the Nottingham University Hospitals NHS Trust for breast diseases between 1988 and 2017. We will not be recruiting any new prospective patients.

This research does not include any patient involvement and the tissue usage will not affect or compromise pathological assessment of the specimen.

Participant Withdrawal

This application is based on using the ethically approved Breast Pathology Database (REC reference 19/EM/0162) which includes mechanisms of participant withdrawal. Data will be used as pseudoanonymised in the research projects. According to the ethics approval (REC reference 19/EM/0162), participants may withdraw their tissue from the study either at their own request or at the discretion of the Chief Investigator.

It was stated in the Breast Pathology Database application that the participants have been made aware that involvement in research will not affect their future healthcare. As part of the original consenting process by the NHSB Biobank, patients were offered information leaflets and consent forms that stated the right to withdraw consent and the explained method to withdraw the consent. It was also explained that they can withdraw at any time. Participants have been made aware (via the

information sheet and consent form) that, should they withdraw, any data collected to date cannot be erased from any research where it has already been used, but they will not be included in any further research.

As the tissue samples in this study are used pseudo-anonymised, we will rely on the Tissue Acquisition Team of the ethically approved Breast Pathology Database (REC reference 19/EM/0162, IRAS ID: 256309) to inform us whenever patients wish to withdraw their consent. Samples will not be used in any future studies. Also once pseudo-anonymised data has been entered onto the University of Nottingham secure network, data cannot be erased as backups cannot be tampered with. This is covered on clause 2 of the consent form and the withdrawal section of the participant information sheet template.

Eligibility criteria

The population base and the selection criteria for inclusion of samples was detailed in the Breast Pathology Database (REC reference 19/EM/0162) and it is summarised below:

All consented patients, male and female, presenting at the Nottingham University Hospitals Trust between 1st January 1988 and 31st December 2017 with breast lesions who had surgery for diagnostic purposes.

In this cohort, patients were primarily treated surgically, some being given adjuvant chemotherapy or hormonal therapy. These treatments were prescribed according to standard protocols.

Inclusion criteria

- Age range: 18 years - 100 years.
- Diagnostic or therapeutic excision of breast lesion.
- Availability of clinicopathological and outcome data.

Exclusion criteria

- Patients less than 18 years old at the time of surgery
- Patients with no clinicopathological data available.
- Patients with insufficient residual tissue for research.
- Patients who stated that they do not want their material or data to be used in research or who withdrew their consent.
- Patients presented after the suspension of the Nottingham University Hospitals NHS Trust Biobank consenting process.

Informed consent

This application is based on using the ethically approved Breast Pathology Database (REC reference 19/EM/0162) cases which includes information about consent. According to the approval (REC reference 19/EM/0162), details are as follows: As part of the routine surgical operation consent procedure, patients were informed that tissue samples will be taken for research use.

Patients' consent for research purposes prior to 2007 was obtained by Professor Roger Blamey. Between 2007 and 2010 consent was taken by Professor Ian Ellis. After 2010, consent was taken by the Nottingham University Hospitals NHS Trust Biobank staff according to the previous and existing ethical approvals. Consent forms after 2010 are kept in the Biobank, however, consents prior to 2007 were kept by Professor Roger Blamey and Professor Ian Ellis. Patients samples collected prior

2007 were included in several previously ethically approved research projects and consent forms were provided at the time. In addition, under the Human Tissue Act 2004 legislation, archived patient tissue collected prior to 1st of September 2006 is allowed for use in research studies without obtaining informed consent. After the 1st of September 2006, patients were consented to use of surplus material and data in research and consents are held in the Biobank. We will ensure that the whole series is used in a pseudo-anonymised form and that researchers will not have access to patients' identifiable information.
For those samples for which we do not have consent we are relying on the CAG support applied for through the Breast Pathology Database (REC reference 19/EM/0162, and CAG number: 19/CAG/0084) and that these patients' data will only be accessed by the research team in pseudo-anonymised form through the Breast Pathology Database.

STUDY TREATMENT AND REGIMEN

This study will only use archival FFPE tissue samples surplus to diagnostic requirements. No prospective collection of patients is proposed in this application.

Sample size and justification

The potential cohort of patients fulfilling the inclusion criteria constitutes up to 8,000 breast cancer and its precursors' lesions. As we aim to identify robust prognostic stratification for breast cancer, all relevant cases (those fulfilling the inclusion criteria) from consecutive patients attending the Nottingham University Hospitals NHS Trust between 1988 and 2017 will be used to allow subgroup analysis (for instance the expression of the protein in treated versus untreated cases as well as those treated by different modalities) and generate statistically reliable results. During the course of the study, power calculation and estimation of sample size (to ensure that an adequate and appropriate number of samples are included) will be performed for each individual study addressing specific research questions according to its prevalence in the population as shown in the literature.

TRANSPORT AND STORAGE OF THE TISSUES

All original Formalin fixed paraffin embedded (FFPE) tissue samples are stored at ambient temperature within the Histopathology Department, Nottingham University Hospitals NHS Trust. All tissue microarrays (TMAs) constructed using the original tissue samples are stored in Human Tissue Authority licensed premises at 4°C.

The Tissue Acquisition team will comprise dedicated University of Nottingham technical staff, who are not researchers. They have NHS Honorary Contracts, and they will be working under supervision of Professor Emad Rakha, Lead Breast Histopathologist, to ensure diagnostic patient records are not compromised.

The Tissue Acquisition Team will request and collect the FFPE samples from the Histopathology Department and prepare tissue sections from these as requested by

the researcher. All sections for research will be pseudo-anonymised by allocation of a unique identifying code. These codes will be generated by the data manager under a system that is subject to separate ethical review (Breast Pathology Database: REC reference 19/EM/0162; CAG number: 19/CAG/0084). The master database

including patient identifiable information such as surname and local designated NHS laboratory number will be held in password encrypted files in NHS IT system.

Pseudo-anonymised tissue sections will be provided to the researcher for analysis. The analysis of samples will take place at the University of Nottingham within the School of Medicine. None of the researchers will have access to any patient identifiable data. None of the staff involved in the research process will have access to any patient identifiable data. Any transfer of samples will contain a complete inventory, along with the name of the person responsible for release of research samples and a signatory of those receiving them. Tissue microarray blocks and other research tissues samples as well as all tissue sections on glass slides will be stored within University of Nottingham HTA licenced premises (DI Dr William Dunn- Licence Number 12265). High resolution digital images of the slides will be securely stored, with appropriate backups, on the University of Nottingham's IT system. Any samples being analysed outside of the University of Nottingham will be sent to collaborators, once the appropriate MTA/DTA has been completed, using University approved couriers with tracking.

LABORATORY ANALYSES

This study will include approaches to decipher the molecular profiles and to identify the nature of the breast lesion as well as to predict disease outcome. To develop and validate diagnostic/prognostic models, analysis will be utilised to extract the morphological criteria of breast cancer and its related lesions.

From each suitable FFPE tissue block, a fresh thin (4um) tissue section will be cut (using a microtome) and stained with haematoxylin and eosin (H&E). Conventional human eye and light microscope assessment of H&E stained sections will be used to assess the tumour burden in each sample and evaluation of samples for their suitability for research. This will be carried out in consultation with Professor Emad Rakha, Lead Breast Histopathologist. Each stained section will be digitally scanned at high resolution for storage. In addition, we will utilise these H&E sections for image analysis using AI and machine learning technology to investigate morphological patterns relating it to patient clinicopathological parameters and outcome.

Clinicopathological and outcome database

Pseudoanonymised clinicopathological and long term-follow up outcome data will be maintained Breast Pathology Database, which is the subject of separate review (REC reference 19/EM/0162). This will allow the tissue samples to be linked to data held in the database for downstream analysis. The key variables recorded include patient age, sex, clinical and histopathological data including menopausal status, laterality and multifocality, tumour size, nuclear grade, tumour type, vascular invasion and extent, presence of other precursor lesions, lymph node stage, treatment modality, site of recurrence, time to recurrence, breast cancer specific survival and overall survival.

Molecular Characterisation of the cohort

Basic molecular characterisation of the tissue samples using immunohistochemistry (IHC) will also be conducted where needed and not provided by the clinical records.

Staining for key biomarkers in breast cancer such as Oestrogen Receptor (ER),

Progesterone Receptor, (PR) Human Epidermal Growth Factor Receptor 2 (HER2) and the proliferation index using Ki67 will be performed. Equivocal HER2 cases will be confirmed using in situ hybridisation (ISH) to assess HER2 gene copy number and amplification status.

We also aim to apply several techniques to the tissue sections to explore additional biological characteristics as required for research objectives and validation of results using a number of different experimental analysis including: IHC, Polymerase Chain Reaction (PCR), NGS, ISH, and mass spectrometry. For high-throughput analysis we will utilise tissue microarray (TMAs). Image analysis with deep machine learning and AI will be used on high resolution images from these sections where appropriate.

The choice of techniques will be based on the function of the gene identified and the level of validation required although most the activity will involve TMAs and IHC.

Techniques to be used Tissue

Microarray

TMA is a technique where a limited amount of tissue will be taken to prepare FFPE tissue blocks harbouring several representative cores from separate cases. The aim of using TMAs is to preserve the original tissue samples and enable high-throughput analysis using standardised experimental methods. As part of our previously REC approved studies, we have constructed TMAs consisting of approximately 5,000 breast tissue samples.

In this study, we aim to continue using these and expand our high-precision TMA construction activities. There will also be a requirement to reconstruct existing TMAs through the course of the project as they become exhausted. We will construct new TMAs as and when required and will utilise digital H&E images from above as a guide to construct TMAs.

High quality TMAs will be constructed from appropriate breast tissues using an automated machine where small cylindrical cores (0.6-2mm in diameter) of breast tissue will be taken from donor FFPE samples. A 0.6mm TMA core from a tumour with a volume of $1cm^2$ typically represents just 0.3% of the original sample. These cores are placed into a recipient wax blocks allowing up to 150 cores of tissue per new block. Thin (4um) sections of this TMA can be used to screen biological characteristics using techniques such as IHC and in situ hybridisation. On average, around 200 sections can be used from each TMA block.

TMA technology enables a high throughput approach to analysing thousands of tissue specimens for biological characterisation in a single experiment. Advantages of TMAs include: 1) only small cores of donor tissue are used, the donor block is left relatively intact, allowing even rare or precious specimens to be sampled and tested many times; 2) the speed of biomarker analysis is increased and considerable time and money savings can be made through reduced usage of laboratory consumables; and 3) importantly, the TMA format almost eliminates technical sources of variation between samples. In addition, as TMA uses thousands of data points it effectively eliminates the variability of a single data point by the sheer size of numbers studied.

Using TMAs as a primary method of assessing protein expression is always limited by the probability of TMA core loss (which in some cases can be as high as 25% of

total TMA core number). The expression of candidate proteins will be subject to combinatorial expression that needs sufficient number of cases to be informative for the proteins under investigation.

Identification of candidate genes/proteins for the study

We will interrogate the large publicly available genomics/proteomics databases of breast cancer through bioinformatics approaches, which we have successfully used in previous studies, to identify potential key controlling genes and proteins that potentially affect breast disease. All identified proteins related to tumour heterogeneity, metabolic pathways, vascular invasion, intrinsic subtypes, key biological pathways and tumour microenvironment that show association with disease progression and outcome will be considered and potentially investigated in this study. These will be fully scrutinised in detail before taking them forward. We will also add more relevant markers, with relevance to cancer progression based on emerging literature, with the final aim to identify key markers that predict patient outcome and therapeutic response.

The expression of each target protein and gene will be assessed using standard and optimised IHC and ISH protocols in tissue sections. Prior to IHC, specificity of all antibodies will be confirmed through Western blotting and/or blocking experiments using specific peptides. Cell lysates from different breast cancer cell lines (e.g. MCF-7, MD-MBA-231, 293, SKBR3) will be used for this purpose. Negative (omission of the primary antibody) and positive controls will be included according to manufacturer datasheets of each antibody. Immunoreactivity and RNA/DNA expression will be quantitatively assessed by researchers under a light microscope according to standard operating protocols.

Image analysis

Image analysis software packages will also be used to achieve higher levels of reproducibility and to alleviate subjectivity of human inspection. These packages include ImageJ and Philips Xplore image analysis packages. Image analysis with deep machine learning and AI approaches will be further applied to identify tumour heterogeneity in terms of morphological variations, grade discordance, spatial distribution of microenvironmental structures including infiltrating lymphocytes, cancer associated fibroblast and other extracellular matrix components. A combination of different programming languages for all the analysis will be used. Image normalisation and H&E deconvolution to pre-process all the images will be performed. Classified regions of interest identified using the low-resolution scan can have its coordinates registered and cropped from the maximum magnification image for both additional texture analysis and nucleus segmentation and investigation. Texture features we use are well reviewed in image analysis literature and include entropy, grey level run length, fractal dimensions, and multiple filters. In addition, assessment of various studied biomarkers' distribution and level of expressions of IHC/ISH stained tissue sections will be also analysed using imaging and machine learning techniques. This includes better cost and time effective approaches by automatically classifying high risk regions, quantifying morphological features for thousands of cells and ducts and using unbiased features which may be invisible even to the highly trained human eye. Additionally, we hope to amend convolutional neural networks for both classification and feature extraction purposes. Classification of region and risk is done using a bagged tree classifier and will involve

data from both low and high-resolution images, spatial data, cell data, and morphological features. Invasive risk will be compared using a cox multivariate model and visualised with Kaplan-Meier survival curves.

Other techniques

Other techniques including DNA/RNA extraction from FFPE sections for further analysis of DNA or RNA using techniques such as PCR, mass spectrometry and NGS will be applied whenever needed according to studies' objective and generated preliminary results, either for validation of the results or for further data discovery.

STATISTICS

Statistical analysis

For statistical analysis of research data, appropriate statistical software packages will be used including SPSS (SPPS Inc., Chicago, IL, USA), R package (R Foundation for Statistical Computing, Vienna, Austria), and GraphPad Prism (GraphPad Prism for Windows/Mac, California, USA). Relevant statistical tests provided by these packages will be used to interrogate the statistical associations of protein expression against relevant clinicopathological and outcome variables. For instance, Spearman correlation test, Intraclass Correlation Coefficient and kappa statistic will be used to test the inter-observer reproducibility and the correlation between the scores of protein expression by different scorers. Statistical tests including, yet not limited to, Student t-test, one-way Analysis of Variance, Chi-square test will be used to test association between protein expression and clinicopathological parameters. Kaplan-Meier plots will be used to test for association of protein expression and patient outcome with differences in survival estimated using Log-rank tests. Multivariable Cox proportional hazards models will be fitted to test confounders and statistical independence and significance of the studied protein as predictors of patient outcome. Participants with missing outcome information will be censored at the time if their data is missing.

To determine cut-off points, X-tile bioinformatics (Yale University, USA) package will be used. For the purpose of generating classifiers that could be used for patient risk stratification and management based on protein expression, decision tree algorithms, artificial neural networks and learning techniques will be used.

For computational pathology and digital image analysis studies, appropriate software packages will be used including image analysis software, deep learning including artificial and conventional neural networks and machine learning techniques. Analytical results will be summarised in tables and graphs when appropriate for publication, conference dissemination and grant application(s).

Analysis of data for this project will be carried out by research staff and PhD students working in this project having authorised approval from the data manager. Data will also be generated by research collaborators in the UK and elsewhere who use our TMA and slides in collaborative and ethically approved research projects. Analysis will be performed by the study researchers. However; some AI and big data analysis will be performed by our computer scientist collaborators using its computers where statistical and bioinformatics software packages are installed.

Primary Outcome Measure

As part of this project, a large number of genes/proteins will be interrogated, and statistical approaches will be used to identify only novel prognostic factors of key role in disease progression and patient risk stratification. In this application, we also aim to expand our study and apply novel computation and multidimensional data analysis techniques to identify novel prognostic, and predictive algorithms and tools. The primary outcome for each studied gene/protein or a set of genes/proteins will be the association with breast cancer outcome.

ETHICAL AND REGULATORY ASPECTS

Ethical Issues

This application is based on using the ethically approved Breast Pathology Database (REC reference 19/EM/0162) cases. The proposed research project involves the use of retrospective human FFPE breast tissue that have been collected from patients attending the Nottingham University Hospitals NHS Trust between 1988 and 2017 as part of their normal care and routine surgical operations for their pathological breast conditions. The material used in any proposed research is in excess of what is required for routine patients care and will not affect or compromise routine diagnostic pathological assessment.

As this is a retrospective study using archived tissue surplus to diagnostic use, patients will not be approached or directly affected beyond the use of archived material in research. No results of research will be directly released to patients or affect their management pathways. Patients will not be re-contacted to obtain consent or to inform them of the results due to use of retrospective cohorts.

The study will be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, 1996; the principles of Good Clinical Practice and Good Laboratory Practice, the UK Department of Health Policy Framework for Health and Social Care, 2017 and in accordance with the Human Tissue Act, 2004.

The University of Nottingham Data Security and Protection Toolkit was last published 30/04/19 under the reference University of Nottingham (EE133856-RGD) with the assessment Standards Met.

Informed Consent and Participant Information

As part of the routine surgical operation consent procedure, patients were informed that tissue samples will be taken for research use.

Withdrawal

Details of withdrawal are included in the ethically approved Breast Pathology Database (REC reference 19/EM/0162). There is a process for patients to record their dissent and the Nottingham University Hospitals NHS Trust have ensured that a system is in place to allow patients being able to withdraw their consent if previously given or having indicated that they don't wish their tissue to be used in research. The participants have been made aware that this will not affect their future healthcare. Participants have also been made aware (via the information sheet and consent form) that should they withdraw, the data collected to date cannot be erased

from any research where it has already been used, but they will not be included in any further research.

RECORDS

Database Record Entries (DRE)

All information held on the research database will be pseudo-anonymised, with no patient identifiers relayed or being held by the researchers. Data will be treated as confidential and held securely in accordance with existing regulations and provisions of the General Data Protection Act 2018.
Pseudonymised data from study samples and linked, patient-derived, metadata will be kept on the University of Nottingham IT system within a limited access, password-protected, secure system, backed up regularly. Any data generated in the study will be delivered to the research data manager. It will be quality checked by the data manager and merged with existing research data. Audits related to the use of each sample and data, processing of tissues and transportation of tissues or samples will be maintained for inspection in compliance with HTA regulations. The group will ensure that the highest standard of recording is followed, and we will work with the Nottingham University Hospitals NHS Trust and University of Nottingham IT teams to ensure patient safety and safe and secure access to data in the pseudoanonymised form. We work closely with the Histopathology Department at the Nottingham University Hospitals NHS Trust to ensure that all local regulations are followed. Access to DREs shall be restricted to those personnel approved by the Chief Investigator and recorded as such in the database records.

Study de-identification codes will be kept in a password protected folder on a shared access server on the NUH IT system and access restricted to appropriate NHS staff and a Data manager and deputy employed by the University of Nottingham and holding an Honorary Contract with the NUH. No member of the research team will have access to this

Sample Labelling

Confidentiality will be ensured by using a unique identification sequential code corresponding to each case in the patient cohort by the Tissue Acquisition Team.. Tissue samples released to researchers for research experiments will only carry this anonymisation code with no patient identifiable information.

Source documents

A list of all patients with the linked codes and other relevant information are kept within the Nottingham University Hospitals NHS Trust by the Tissue Acquisition TEAM without access to researchers to maintain patients confidentially and anonymisation.

Data Protection

All staff and investigators will store, manage and use the research data in accordance with the General Data Protection Act, 2018. Any research data

generated will be held securely, and access to the information will be limited to the database staff and investigators and any relevant regulatory authorities. Computer held data will be held securely and password protected. All data will be stored on a secure dedicated server. Access will be restricted by user identifiers and passwords.

Access to the data file locations on the University of Nottingham servers is granted only by written request from a senior manager and limited to research staff and PhD students where required. Access to secure data file locations is managed by the University of Nottingham IT services and accessed by unique usernames and passwords. The secure data files will also be password controlled with a secondary password protection to ensure that the access to the database itself is restricted once logged into the University system.

The clinicopathological database will only contain pseudo-anonymised patient data. A local back up is also carried out on a daily basis, encrypted using AES256 bit encryption and backed up to an external hard drive located and locked in a secure room. Electronic data will be backed up every 24 hours to both local and remote media in encrypted format.

QUALITY ASSURANCE & AUDIT INSURANCE AND INDEMNITY

Insurance and indemnity for study participants and NHS staff is covered within the NHS Indemnity Arrangements for clinical negligence claims in the NHS, issued under cover of HSG (96)48. There are no special compensation arrangements, but study participants may have recourse through the NHS complaints procedures.

The University of Nottingham as research Sponsor indemnifies its staff, research participants and research protocols with both public liability insurance and clinical trials insurance. These policies include provision for indemnity in the event of a successful litigious claim for proven non-negligent harm.

STUDY CONDUCT

This is an ongoing study with a large research team comprising several investigators, research fellows, postgraduate and undergraduate research students and collaborators. This application aims to continue our research studies, which started in early 1970s and resulted in more than 650 publications in peer-reviewed journals. We have IT infrastructures and we are supported by the University of Nottingham IT department which enabled us to store and maintain all previous study results for long term. All generated data from the current study will be stored for an extended period (up to 20 years) on password protected locations within the University of Nottingham IT system, which is only accessible by authorised personnel. We will use grants' fund to support digital storage of all images and data generated during this study as we have done with previous studies.

Study conduct may be subject to systems audit of the study files for inclusion of

essential documents; permissions to conduct the study; CVs of study staff and training received; local document control procedures; consent procedures and

recruitment logs; adherence to procedures defined in the protocol (e.g. inclusion / exclusion criteria, timeliness of visits); and equipment calibration logs.

The Chief Investigator, or where required, a nominated designee, shall carry out a site system audit at least yearly and an audit report shall be made to the Study Sponsor.

STUDY DATA

Analysis of the research data generated throughout this project will be carried out within University of Nottingham premises using University computers where statistical and bioinformatics software packages are installed.

Any research data generated by the study will be delivered to the data manager to be quality checked and merged with the pseudo-anonymised data held at the dedicated secure University servers. Authorised researchers, on password protected University computer systems, will then be able to access the quality checked data. Analysis of data for this project will be carried out by research staff and PhD students working in this project and authorised by the Chief Investigator. Some parts of the research may be carried out by the NHS staff. Data will also be generated by research collaborators in the UK and elsewhere who use our TMA and slides in collaborative research projects. Analysis will be performed by the study researchers. However; some AI and 'big data' analysis may be performed by our computer scientist collaborators.
Monitoring of study data shall include; source data verification; data storage and data transfer procedures; local quality control checks and procedures, back-up and disaster recovery of any local databases and validation of data manipulation. Study data and evidence of monitoring and systems audits will be made available for inspection by the regulatory authority as required.

RECORD RETENTION AND ARCHIVING

The Chief Investigator will maintain all records and documents of the study. The study documents held by the Chief Investigator on behalf of the Sponsor shall be finally archived at secure archive facilities at the University of Nottingham. This archive shall include all study databases and associated metadata.

Encryption codes will be maintained on the NUH Trust IT system.

DISCONTINUATION OF THE STUDY BY THE SPONSOR

The Sponsor reserves the right to discontinue this study at any time for failure to meet expected enrolment goals, for safety or any other administrative reasons.

STATEMENT OF CONFIDENTIALITY

Cases comprising the cohort of patients suitable for this project were pseudo-anonymised by allocation of a unique identifying coding number. This code allows

linkage of a pseudo-anonymised case with known tumour variables and key clinical outcome measures to results of studied protein under investigation at the time of data analysis.

All information held on the pseudo-anonymised data will contain no patient identifiers. This data has been acquired and pseudo-anonymised by the Tissue

Acquisition Team within the NHS secure computer network as part of the Breast Pathology Database. The data acquisition is carried out under NUH NHS Trust protocols, governance and audits, using the NHS computer systems. Tissue Acquisition Team will de-identify the cases with pseudo-anonymised codes as a continuation of the previous process. Data will be released in a pseudo-anonymised form to the data manager for storage on University of Nottingham IT systems.

The Data Management Team will de-identify the cases and apply pseudo-anonymised codes as a continuation of the previous process. Data will be released in a pseudo-anonymised form to the data manager for storage on University of Nottingham IT systems.

Computer held pseudo-anonymised data will be kept securely and password protected. All data will be stored on a secure dedicated University web server at the University of Nottingham. A front end will be created to link to secure data files and encrypted to AES128bit system using a one-way encryption method. Access will be restricted by user identifiers and passwords.

Access to data file locations on the University servers is granted only by written request from the data manager and limited to research staff and PhD students working in the project where required. Following data manager approval, granting access to secure data file locations, is managed by University of Nottingham IT services and accessed by a University username and password. The secure data files will also be password controlled with a secondary password protection to ensure access to the database itself is restricted once logged into the University system.

All data stored on University Servers is backed up by University of Nottingham IT. Services on a daily basis. A local back up is also carried out on a daily basis, encrypted using AES256bit encryption and backed up to an external hard drive located and locked in a secure room. No duplicates of data will be generated by researchers or PhD students in any form or any transferrable data storage will be produced nor any downloads be permitted to personal computers owned by researchers or by PhD students.

Individual participant medical information obtained as a result of this study is considered confidential and disclosure to third parties is prohibited with the exceptions noted above.

Data generated as a result of this study will be available for inspection on request by the Sponsor, the REC and representatives of the Human Tissue Authority.

We have been carrying out this exercise for over 20 years and no breach of data governance and management incidents has been reported.

As this is a retrospective cohort study using archived material surplus to diagnostic use, patients will not be approached or directly affected beyond the use of material in research. No results of research will be directly released to patients or affect their management pathways.

PUBLICATION POLICY

We aim to publish results from this study in high impact scientific journals and presentations in national and international scientific conferences. We will also send out a newsletter with a lay research summary of progress/study results to various breast cancer charity groups, including a feature in the Nottingham Breast Cancer Research Centre newsletter.

Participants will not be identified in any data transfers or publications of research results where the data has been used.

STUDY FINANCES

Funding source

This study is funded by several sources including several PhD studentship' funds and various grants including Nottingham Breast Cancer Research Centre (NBCRC), Pathological Society of Great Britain and Ireland, Nottingham Life Cycle 6, Breast Cancer Research Trust, NUH Hospitals Charity, and The National Breast Cancer Foundation (NBCF) of Australia.

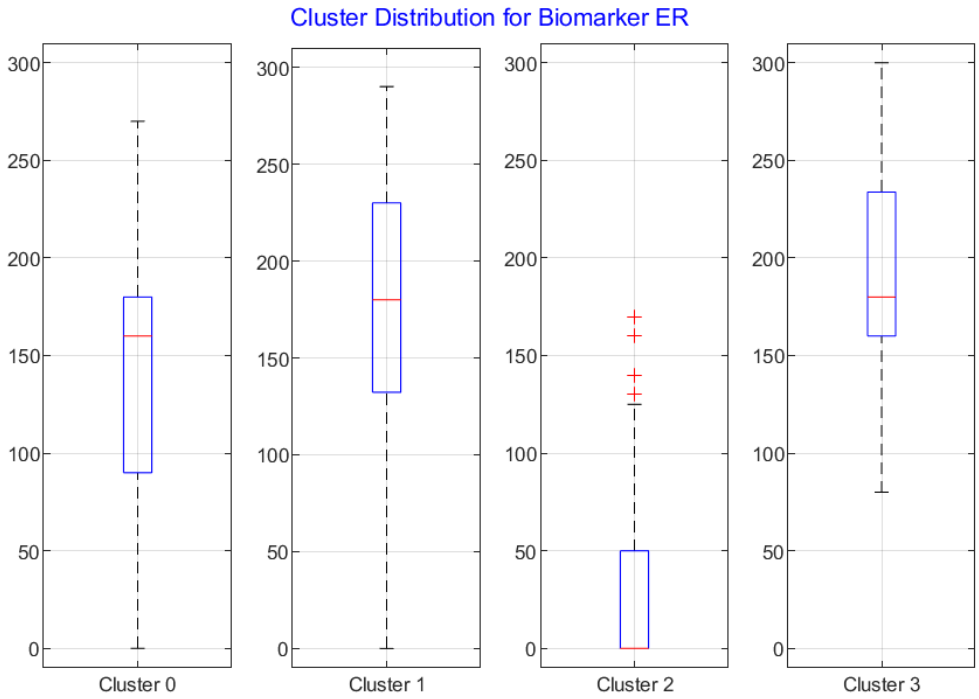
Participant stipends and payments

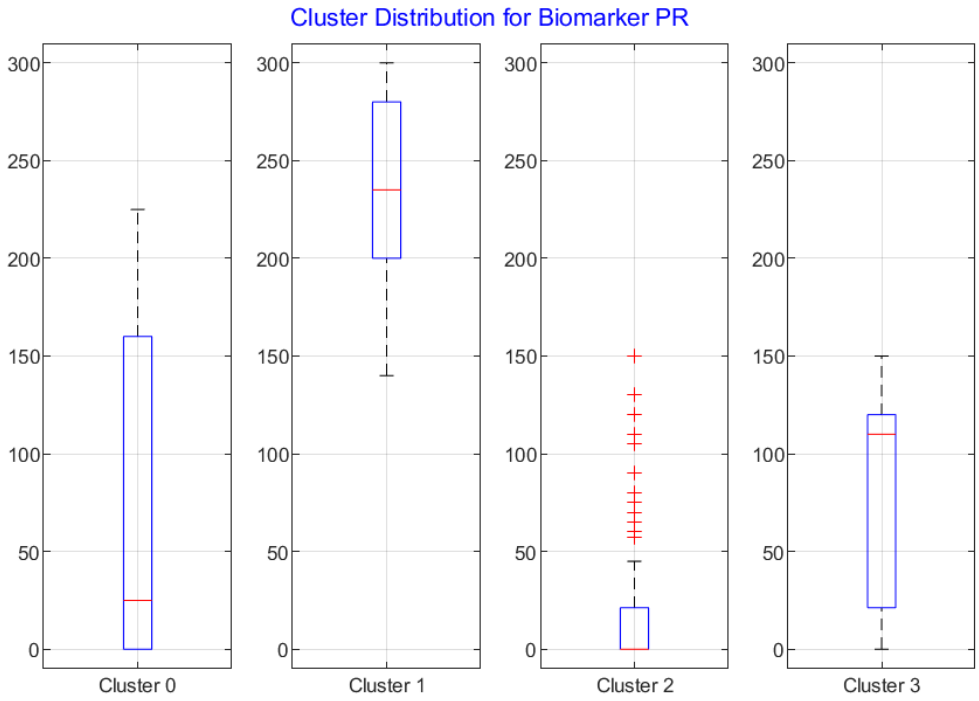
Participants will not be paid to participate in the study.

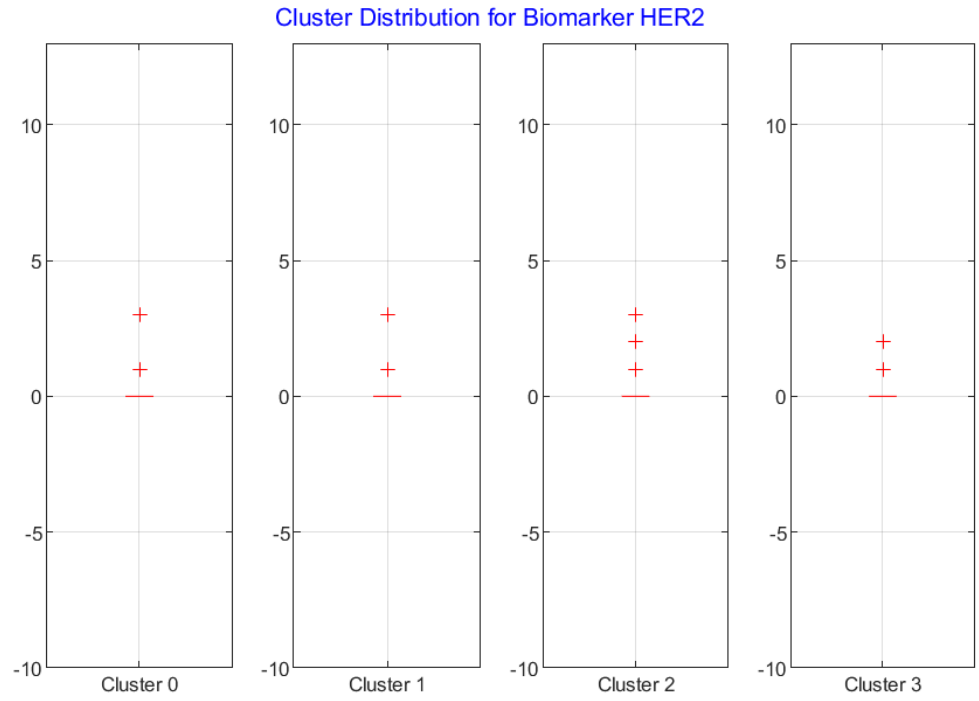
REFERENCES

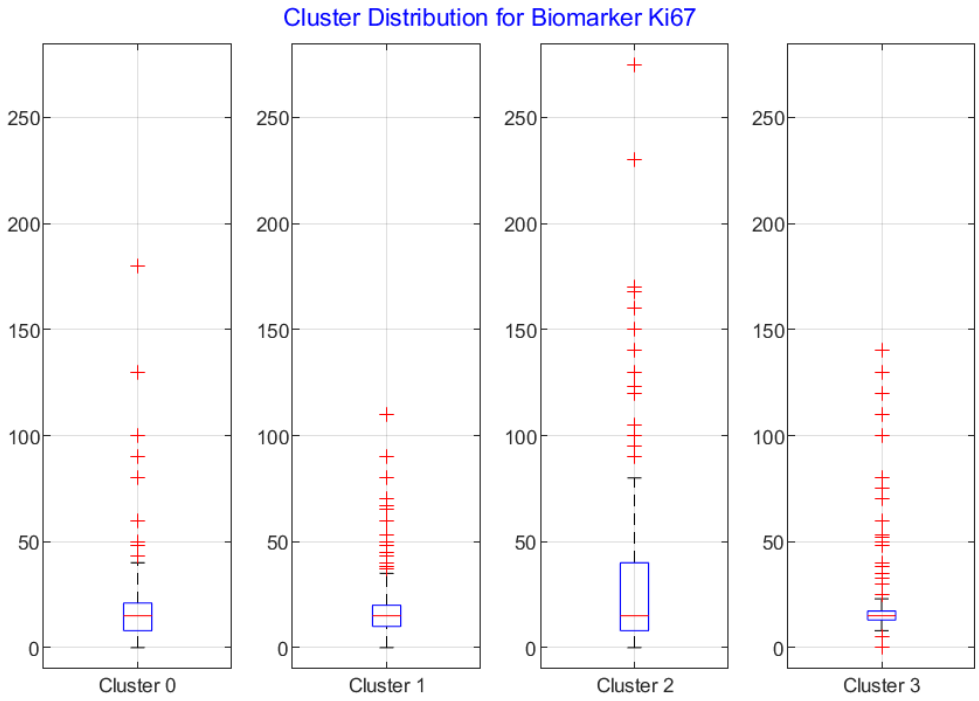
1. CRUK. Breast cancer statistics, Cancer Research UK 2014 [cited 2019 24/06/2019]. Available from: http://www.cancerresearchuk.org/health-professional/cancer- statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive. 2019.
2. Galea, M.H., et al., The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat, 1992. 22(3): p. 207-19.Blamey, R.W., et al., ONCOPOOL
   - a European database for 16,944 cases of breast cancer. Eur J Cancer, 2010. 46(1):
   p. 56-71.
3. Rakha, E.A., et al., Prognostic stratification of oestrogen receptor-positive HER2- negative lymph node-negative class of breast cancer. Histopathology, 2017. 70(4):
   p. 622-631.
4. Rakha, E.A., et al., Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer. Br J Cancer, 2014. 110(7): p. 1688-97.
5. Rakha, E.A., et al., Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res, 2010. 12(4): p. 207.
6.

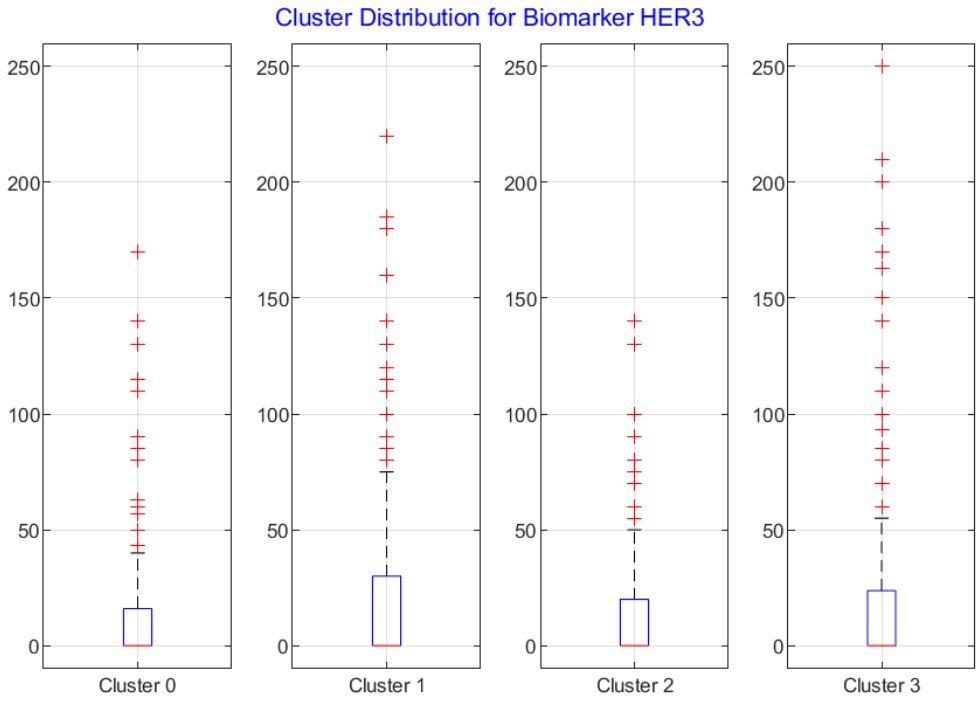
Appendix 3: Cluster distribution by biomarker for cluster analysis of all CNB TMA samples, identifying three clusters (Cluster 0 are the cases which are unclustered)

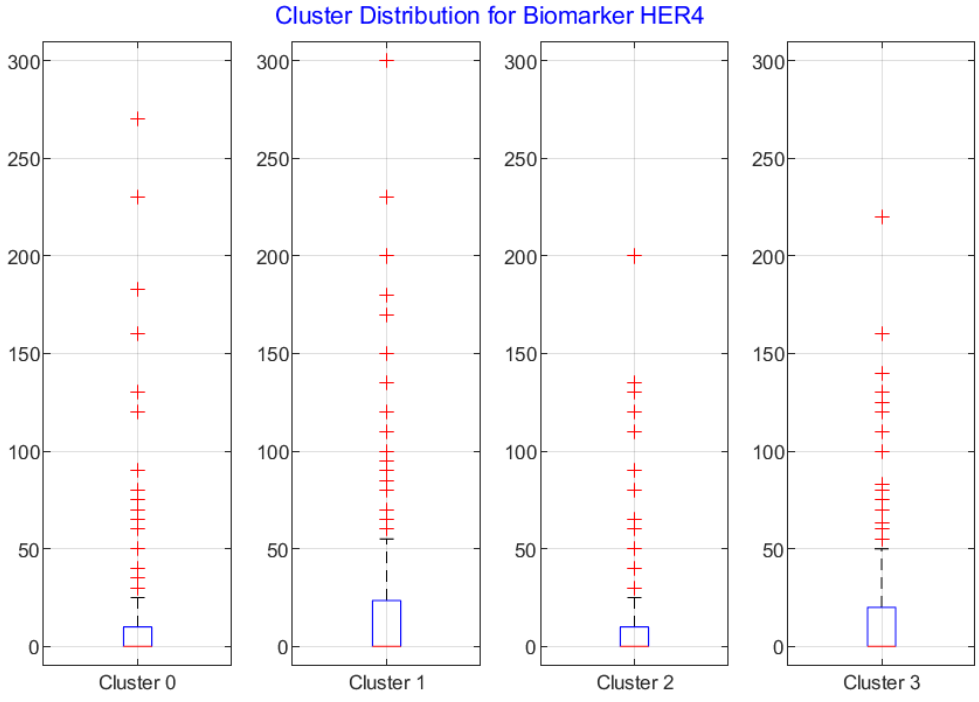


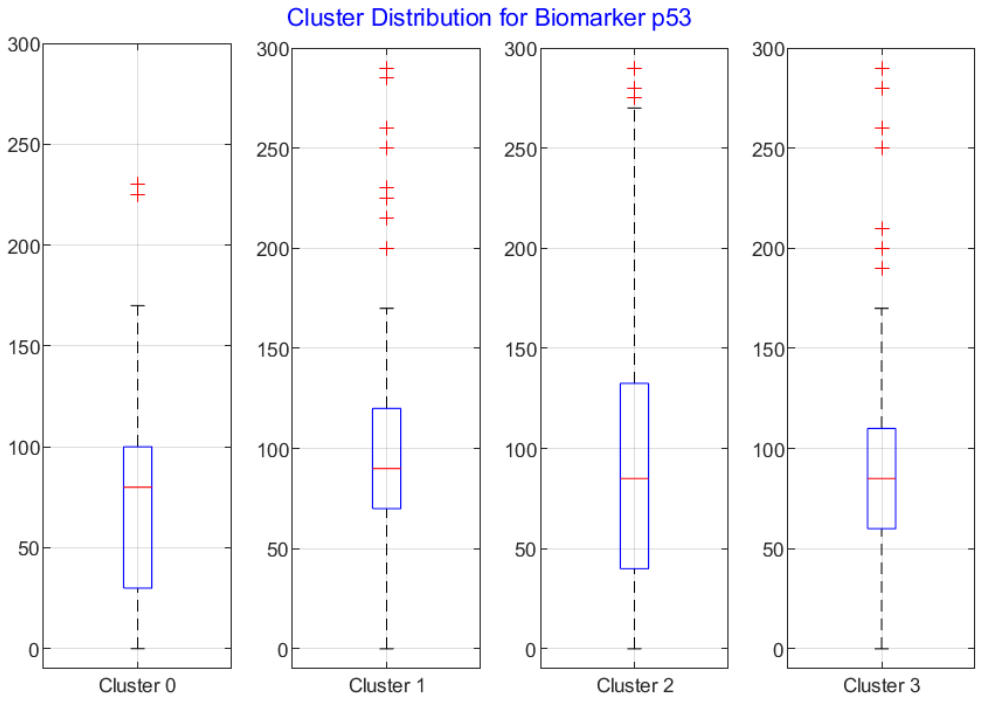

Cluster Distribution for Biomarker PR
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

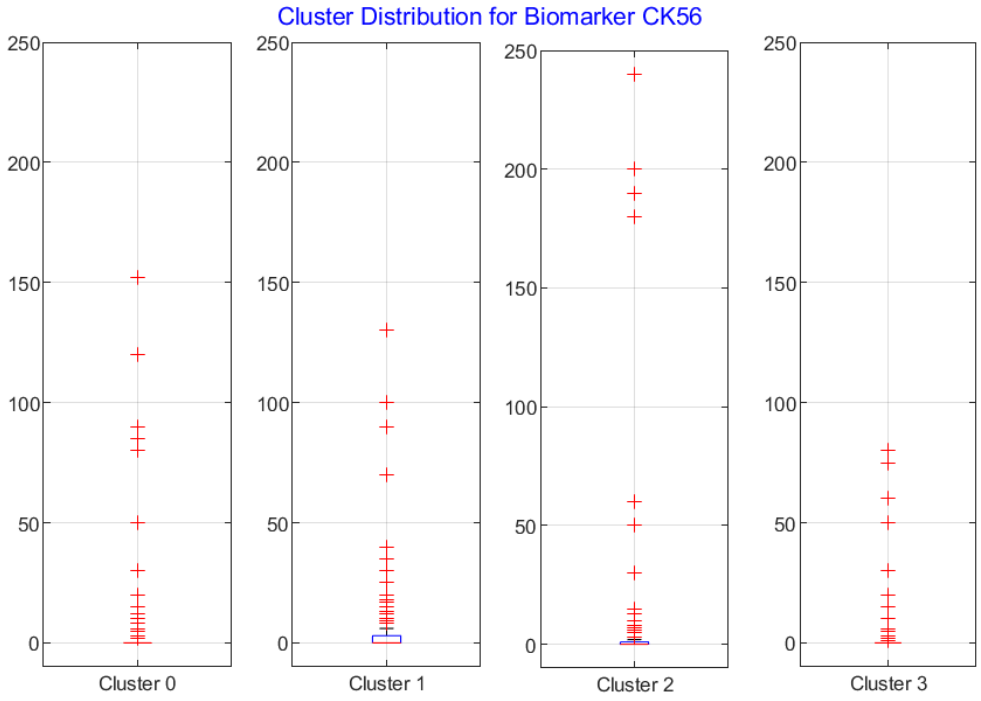

Cluster Distribution for Biomarker HER2
10
5
0
-5
-10
Cluster 0
Cluster 1
Cluster 2
Cluster 3

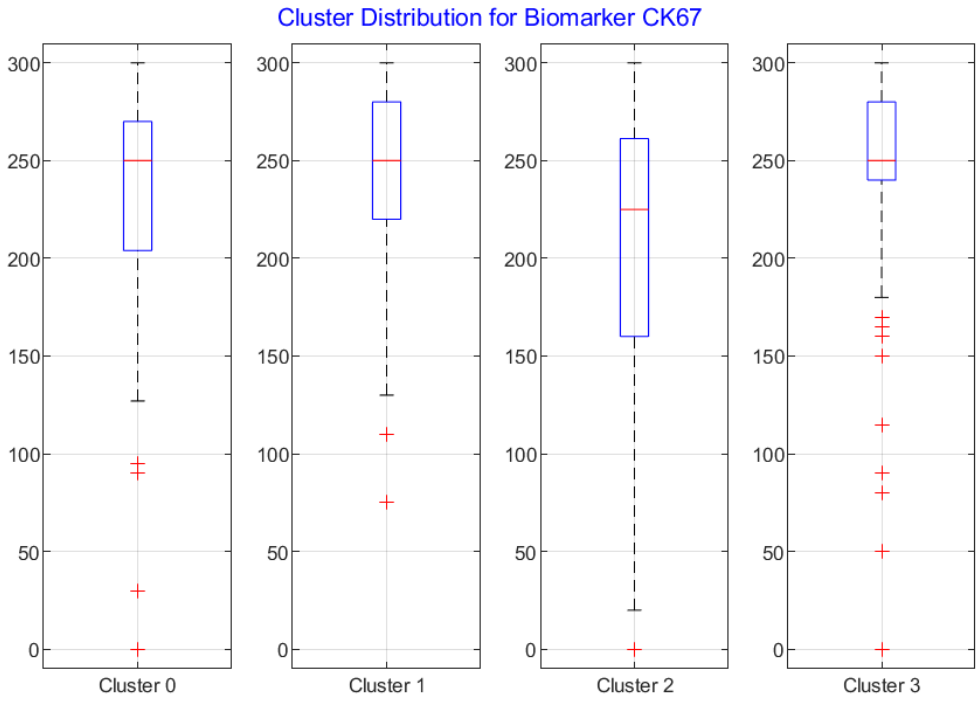

Cluster Distribution for Biomarker Ki67
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

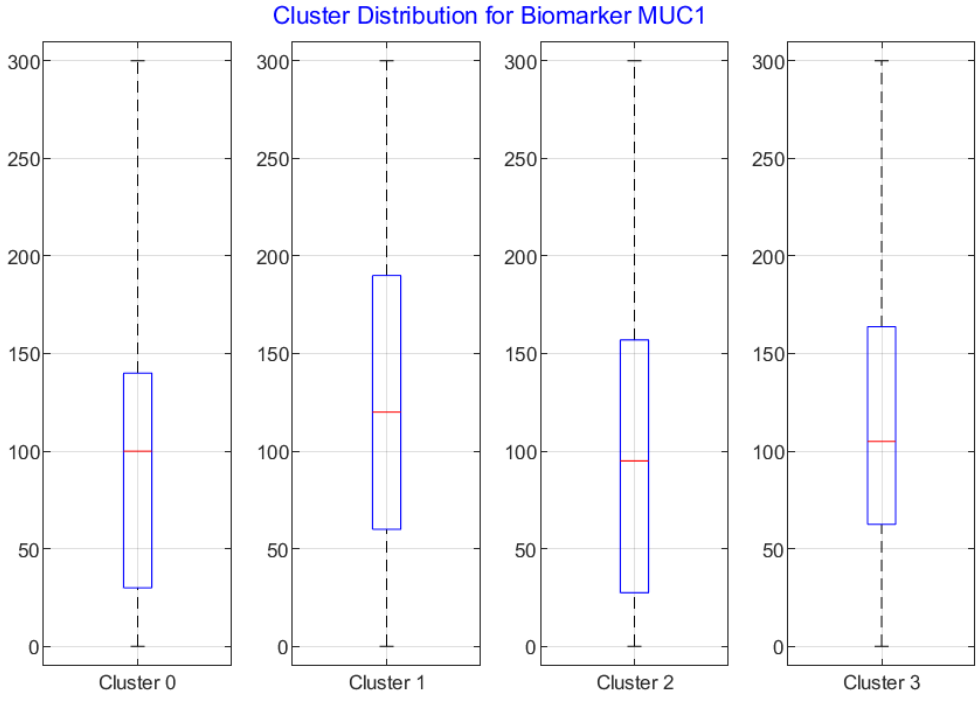

Cluster Distribution for Biomarker HER3
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

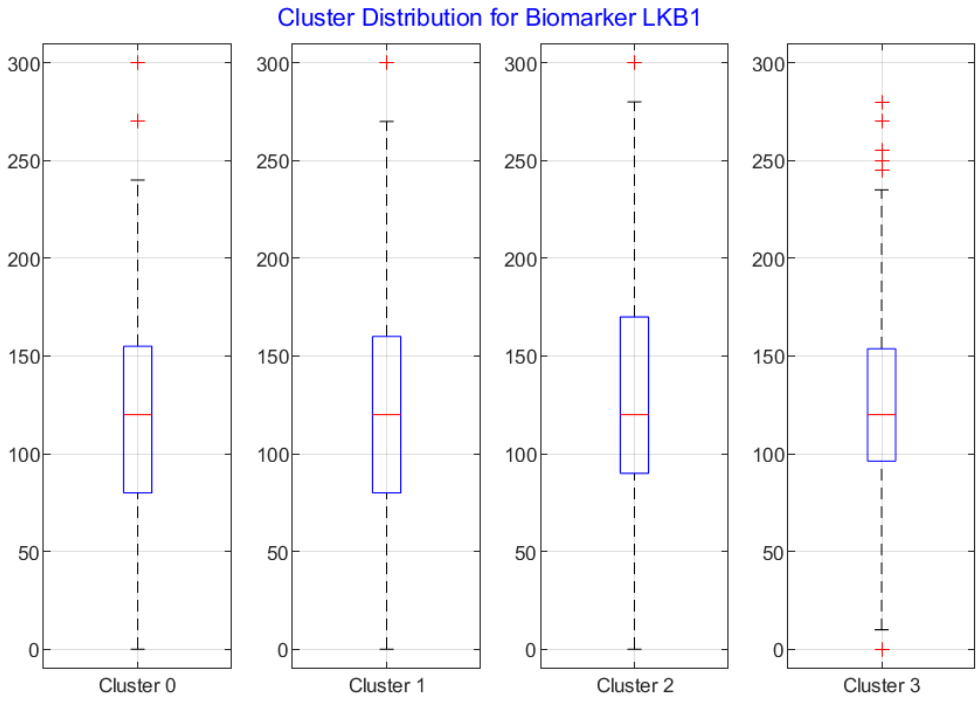

Cluster Distribution for Biomarker HER4
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

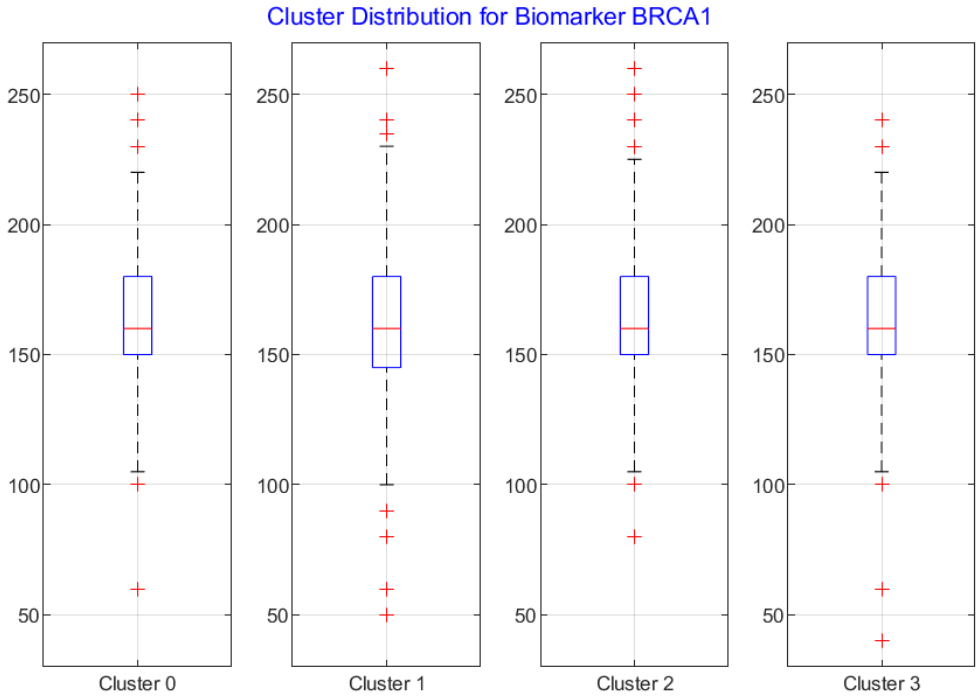

Cluster Distribution for Biomarker p53
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

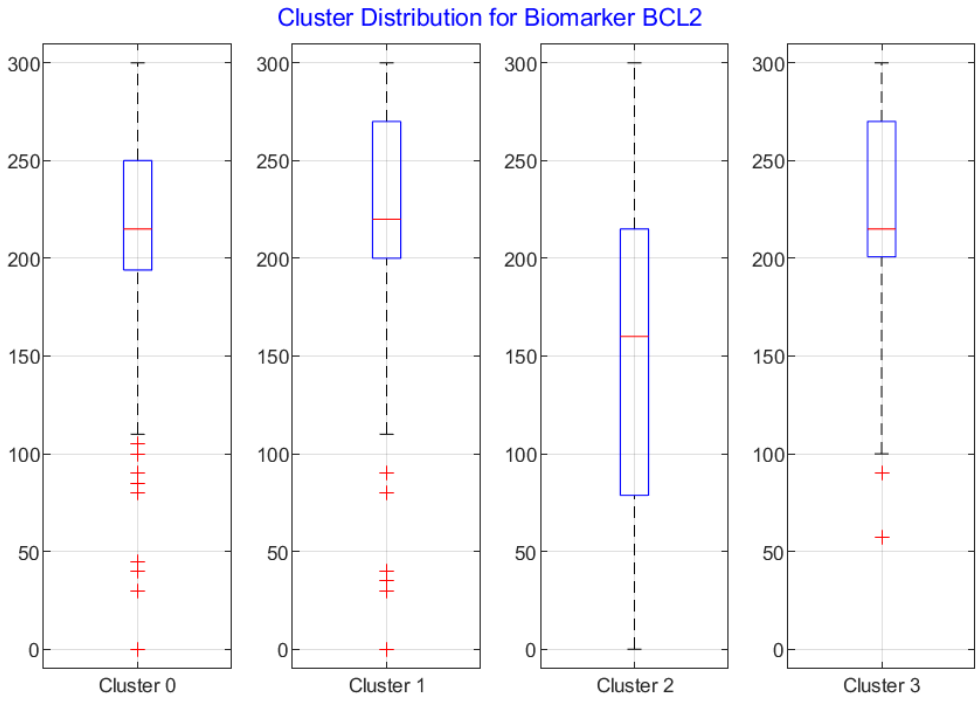

Cluster Distribution for Biomarker CK56
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

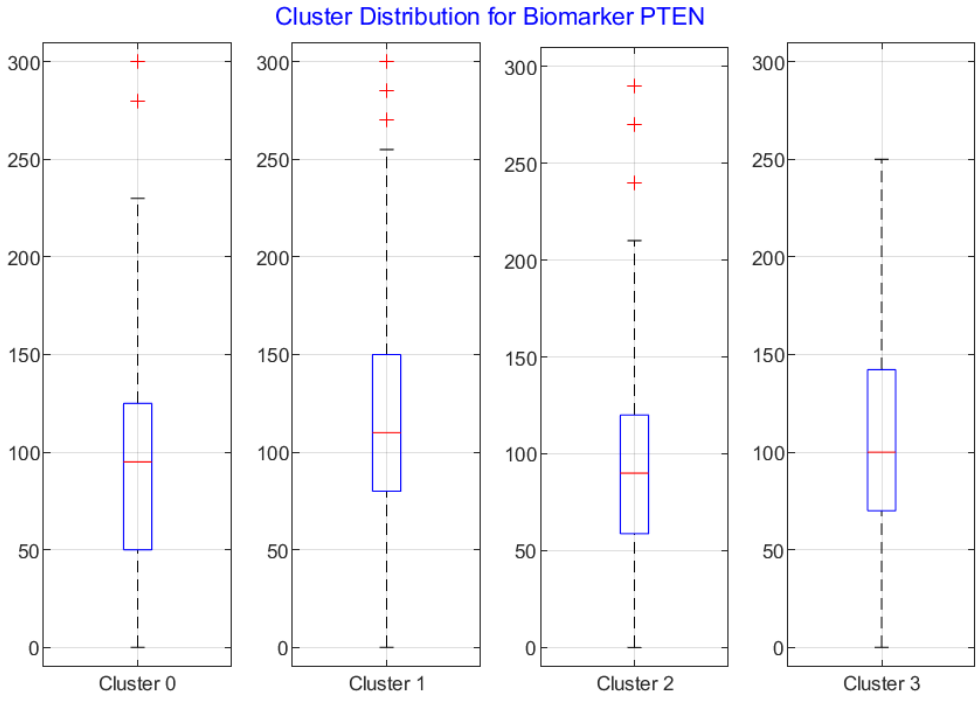

Cluster Distribution for Biomarker CK67
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

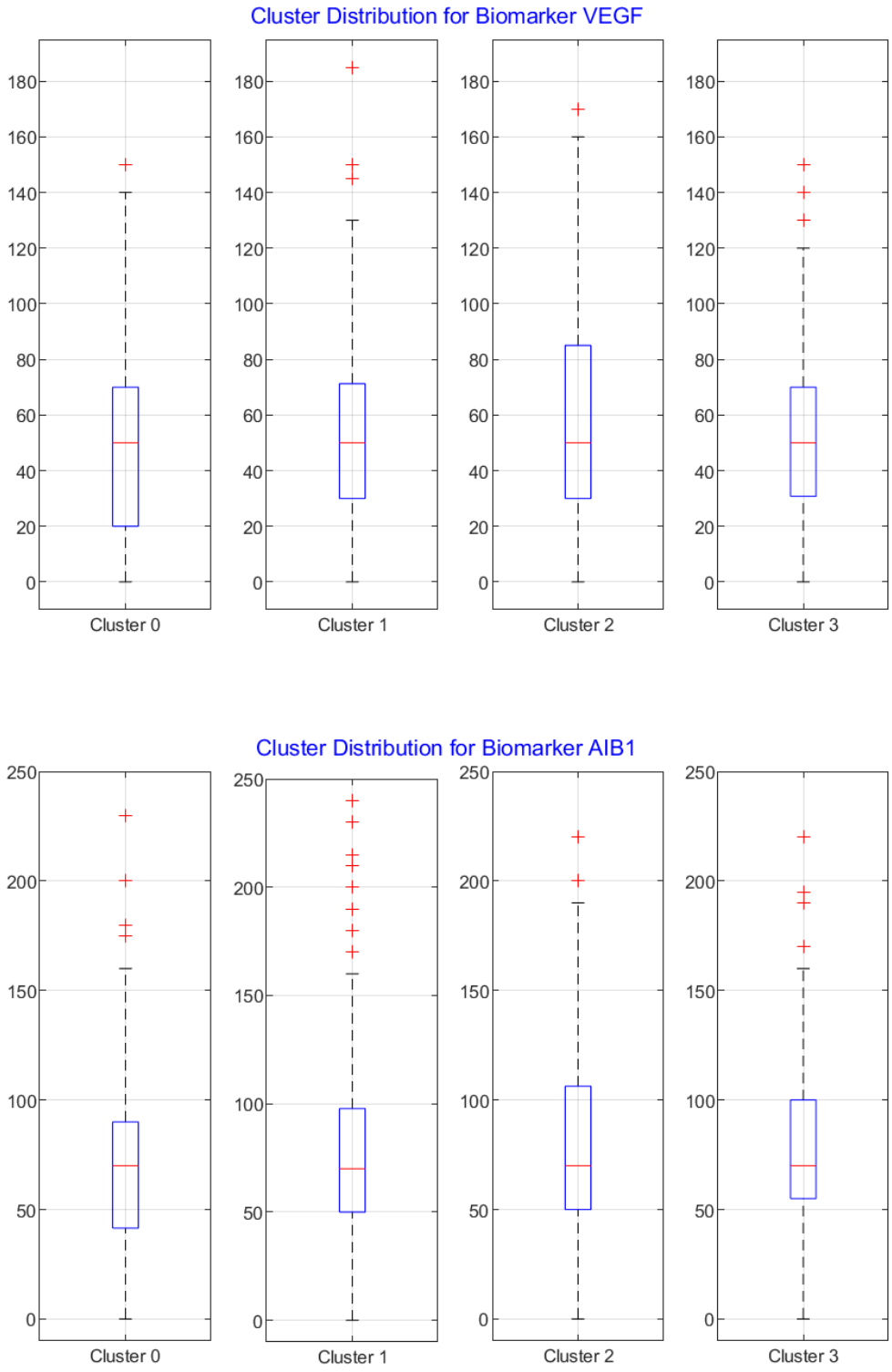

Cluster Distribution for Biomarker MUC1
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3


Cluster Distribution for Biomarker LKB1
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3


Cluster Distribution for Biomarker BRCA1
250
200
150
100
50
Cluster 0
Cluster 1
Cluster 2
Cluster 3


Cluster Distribution for Biomarker BCL2
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3

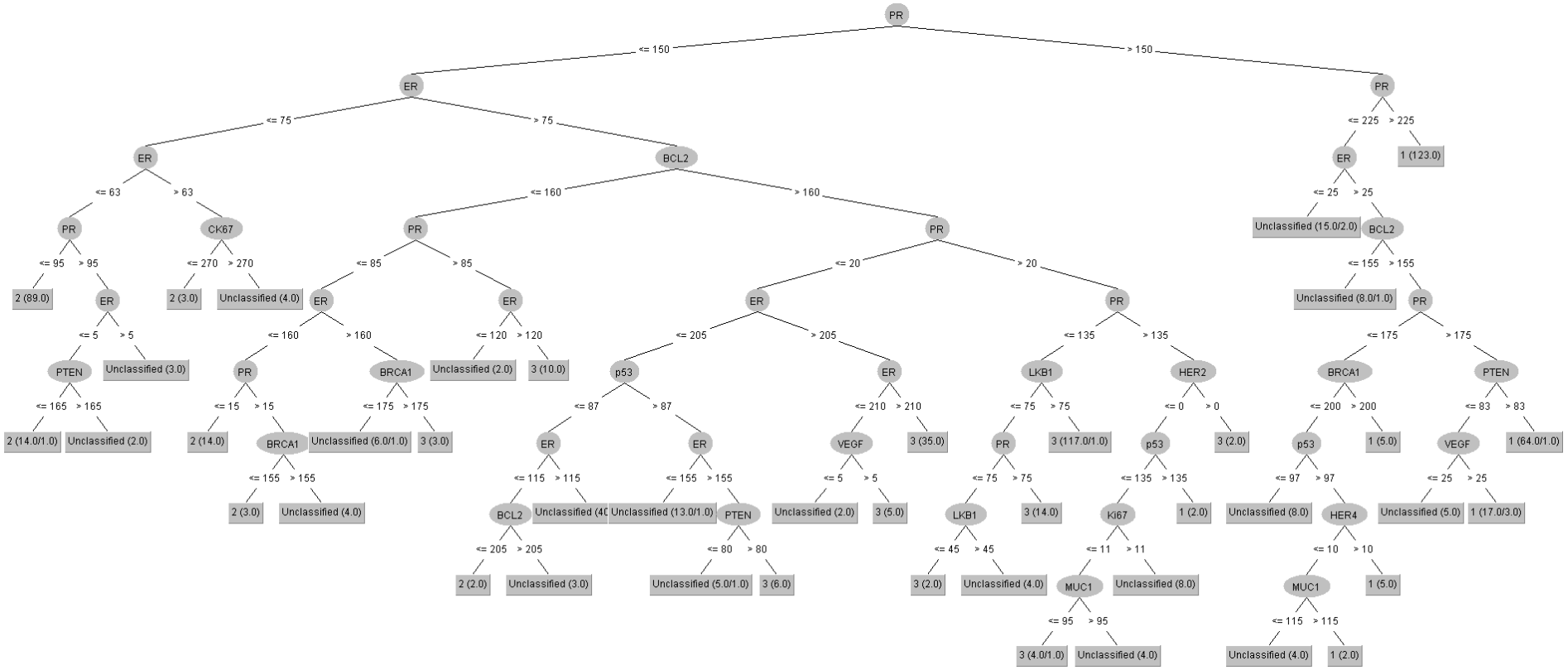

Cluster Distribution for Biomarker PTEN
300
250
200
150
100
50
0
Cluster 0
Cluster 1
Cluster 2
Cluster 3



Abbreviations: BCL – B-cell lymphoma; BRCA – breast cancer gene; CK – cytokeratin; ER – oestrogen receptor; EGFR – epidermal growth factor receptor; HER

– human epidermal growth factor receptor; LK – liver kinase; MUC – mucin; PR – progesterone receptor; VEGF – vascular endothelial growth factor

Appendix 4: Decision tree model to determine key drivers of the classification used for cluster analysis in the CNB TMA cohort



Appendix 5: Prognostic algorithm designed to predict BCSS in SE TMA cohort

```
Sub NSMacro1()

Dim i As Integer
Dim RowString As String
For i = 2 To 2
RowString = CStr(i)

Dim hidden1AxonPE1 As Double
hidden1AxonPE1 = tanh( ( (Range ("A" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
2.2955349203369080e-001 )+ ((Range ("B" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*1.2095329380757003e-001 )+ ((Range ("C" + RowString)*2.0000000000000001e-001 + -9.0000000000000002e-001
)*4.8652792312593507e-001 )+ ((Range ("D" + RowString)*9.4736842105263164e-003 + -9.0000000000000002e-001 )*-
2.7461552279217655e-001 )+ ((Range ("E" + RowString)*6.4285714285714285e-003 + -9.0000000000000002e-001 )*-
2.4285850078620597e-001 )+ ((Range ("F" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*1.5375933709619621e-001 )+ ((Range ("G" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
2.4679548746271646e-001 )+ ((Range ("H" + RowString)*5.2941176470588233e-003 + -6.8823529411764695e-001
)*2.9856202723447789e-001 )+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
3.3369903920084876e-001 )+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*3.6471729416296422e-001 )+ ((Range ("K" + RowString)*7.5000000000000006e-003 + -9.0000000000000002e-001
)*2.1354908154095131e-001 )+ ((Range ("L" + RowString)*6.9230769230769233e-003 + -9.0000000000000002e-001
)*2.5192305668798709e-002 )+ ((Range ("M" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*6.8130948692317228e-001 )+ ((Range ("N" + RowString)*6.4285714285714285e-003 + -1.0285714285714285e+000 )*-
3.1774922837174752e-001 )+ ((Range ("O" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*5.6997255969409777e-001 )+-3.6529181523358545e-001 )

Dim hidden1AxonPE2 As Double
hidden1AxonPE2 = tanh( ( (Range ("A" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*3.4245077788493727e-001 )+ ((Range ("B" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
```

)*7.2962616372601519e-001)+ ((Range ("C" + RowString)*2.0000000000000001e-001 + -9.0000000000000002e-001
)*1.1431213180717936e-001)+ ((Range ("D" + RowString)*9.4736842105263164e-003 + -9.0000000000000002e-001)*-
5.4492116498821574e-001)+ ((Range ("E" + RowString)*6.4285714285714285e-003 + -9.0000000000000002e-001)*-
1.0121657477641488e-002)+ ((Range ("F" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*3.5842886234466187e-003)+ ((Range ("G" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*7.7768193094674587e-002)+ ((Range ("H" + RowString)*5.2941176470588233e-003 + -6.8823529411764695e-001)*-
7.2803493347893355e-001)+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*3.9132484297783188e-001)+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.6883482371769281e-001)+ ((Range ("K" + RowString)*7.5000000000000006e-003 + -9.0000000000000002e-001)*-
5.5660043543024296e-001)+ ((Range ("L" + RowString)*6.9230769230769233e-003 + -9.0000000000000002e-001)*-
4.6741818876730901e-002)+ ((Range ("M" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
1.6018705433865787e-001)+ ((Range ("N" + RowString)*6.4285714285714285e-003 + -1.0285714285714285e+000)*-
2.6725716585727211e-001)+ ((Range ("O" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
2.2486398719266515e-001)+3.7667246346540728e-001)

Dim hidden1AxonPE3 As Double
hidden1AxonPE3 = tanh(((Range ("A" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
6.9627670745848635e-001)+ ((Range ("B" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
1.1443246622559833e+000)+ ((Range ("C" + RowString)*2.0000000000000001e-001 + -9.0000000000000002e-001
)*4.7152848791393731e-001)+ ((Range ("D" + RowString)*9.4736842105263164e-003 + -9.0000000000000002e-001)*-
3.4831126834481002e-001)+ ((Range ("E" + RowString)*6.4285714285714285e-003 + -9.0000000000000002e-001)*-
5.9567288398038232e-001)+ ((Range ("F" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*6.1395092551780861e-001)+ ((Range ("G" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*2.1626701757243894e-001)+ ((Range ("H" + RowString)*5.2941176470588233e-003 + -6.8823529411764695e-001
)*3.4910045561504655e-001)+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
7.1001640055025295e-001)+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001)*-
1.3270682689072949e-001)+ ((Range ("K" + RowString)*7.5000000000000006e-003 + -9.0000000000000002e-001)*-
3.3534853284517402e-001)+ ((Range ("L" + RowString)*6.9230769230769233e-003 + -9.0000000000000002e-001)*-
3.6003184559524704e-001)+ ((Range ("M" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.2690955500737959e-001)+ ((Range ("N" + RowString)*6.4285714285714285e-003 + -1.0285714285714285e+000
)*4.8154581030138011e-001)+ ((Range ("O" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*2.0426186695852511e-001)+-6.4160086499678082e-001)

```
Dim hidden1AxonPE4 As Double
hidden1AxonPE4 = tanh( ( (Range ("A" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*2.5459441017379736e-001 )+ ((Range ("B" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*1.8593386515535326e-002 )+ ((Range ("C" + RowString)*2.0000000000000001e-001 + -9.0000000000000002e-001 )*-
9.7663721492421973e-002 )+ ((Range ("D" + RowString)*9.4736842105263164e-003 + -9.0000000000000002e-001
)*7.1579405762423653e-001 )+ ((Range ("E" + RowString)*6.4285714285714285e-003 + -9.0000000000000002e-001
)*4.2909515326115566e-001 )+ ((Range ("F" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.9416250055178801e-002 )+ ((Range ("G" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.6845221087240851e-001 )+ ((Range ("H" + RowString)*5.2941176470588233e-003 + -6.8823529411764695e-001 )*-
2.3301272721894206e-001 )+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.6303951214153316e-002 )+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.4554477192801700e-002 )+ ((Range ("K" + RowString)*7.5000000000000006e-003 + -9.0000000000000002e-001
)*2.8185531616716031e-001 )+ ((Range ("L" + RowString)*6.9230769230769233e-003 + -9.0000000000000002e-001 )*-
1.5668711321705997e-001 )+ ((Range ("M" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*5.5510847607355440e-001 )+ ((Range ("N" + RowString)*6.4285714285714285e-003 + -1.0285714285714285e+000
)*8.1287526043860292e-001 )+ ((Range ("O" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*5.0919817006997603e-001 )+4.1966204005824659e-001 )

Dim hidden1AxonPE5 As Double
hidden1AxonPE5 = tanh( ( (Range ("A" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
5.6110178420221590e-002 )+ ((Range ("B" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
2.5491317342299663e-001 )+ ((Range ("C" + RowString)*2.0000000000000001e-001 + -9.0000000000000002e-001
)*7.6421471451339118e-001 )+ ((Range ("D" + RowString)*9.4736842105263164e-003 + -9.0000000000000002e-001
)*1.1171526977878998e+000 )+ ((Range ("E" + RowString)*6.4285714285714285e-003 + -9.0000000000000002e-001
)*5.6743817430643828e-001 )+ ((Range ("F" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
8.3580871704871118e-001 )+ ((Range ("G" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*7.2655741293584264e-001 )+ ((Range ("H" + RowString)*5.2941176470588233e-003 + -6.8823529411764695e-001 )*-
2.4272724864052242e-001 )+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
4.7331228805878445e-001 )+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.5715034794151221e-001 )+ ((Range ("K" + RowString)*7.5000000000000006e-003 + -9.0000000000000002e-001 )*-
2.2449583819080887e-001 )+ ((Range ("L" + RowString)*6.9230769230769233e-003 + -9.0000000000000002e-001 )*-
```

```
2.9100129454287899e-001 )+ ((Range ("M" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*2.6232472703607845e-003 )+ ((Range ("N" + RowString)*6.4285714285714285e-003 + -1.0285714285714285e+000
)*6.3235067226418851e-001 )+ ((Range ("O" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*4.2176166778001400e-002 )+1.2243160068019565e-001 )

Dim outputAxonPE1 As Double

outputAxonPE1 = tanh( ( hidden1AxonPE1*3.8095594691335588e-001 )+ ( hidden1AxonPE2*4.5062775227374896e-001 )+ (
hidden1AxonPE3*-9.3878836800313059e-002 )+ ( hidden1AxonPE4*3.1566534984624628e-001 )+ ( hidden1AxonPE5*-
2.4453532727505278e-001 )+-1.5676174737097570e-001 )

Range ("P" + RowString) = ( (outputAxonPE1)--9.1022727272727277e-001) / 1.0227272727272727e-002
Next i

End Sub

Function tanh(x As Double) As Double
tanh = (Exp(x) - Exp(-x)) / (Exp(x) + Exp(-x))
End Function

Function sigmoid(x As Double) As Double
sigmoid = 1.0 / (1.0 + Exp(-x))
End Function
Within the model A = expression of ER, B = PR, C = HER2, D = Ki67, E = p53, F = CK 5/6, G = CK 7/8, H = EGFR, I = BCL2, J =

MUC1, K = VEGF, L = LKB1, M = BRCA1, N = HER3, O = HER4
```

Appendix 6: Prognostic algorithm designed to predict BCSS in CNB TMA cohort

```
Sub NSMacro1()

Dim i As Integer
Dim RowString As String
For i = 2 To 2
RowString = CStr(i)

Dim hidden1AxonPE1 As Double
hidden1AxonPE1 = tanh( ( (Range ("A" + RowString)*4.8648648648646e-003 + -4.3783783783770e-001 )*-
1.1282146601213685e+000 )+ ((Range ("B" + RowString)*2.3529411764705885e-003 + 1.9411764705882351e-001
)*7.0258431402215360e-001 )+ ((Range ("C" + RowString)*1.8000000000000002e-002 + -9.0000000000000024e-001
)*2.5508147863041830e-001 )+ ((Range ("D" + RowString)*6.5454545454545453e-003 + -9.0000000000000002e-001 )*-
1.0922648359460663e-002 )+ ((Range ("E" + RowString)*6.2068965517241377e-003 + -8.9999999999999980e-001
)*6.5603095608130579e-001 )+ ((Range ("F" + RowString)*3.3395176252319111e-003 + -9.0000000000000002e-001
)*5.0955160425148072e-001 )+ ((Range ("G" + RowString)*5.4545454545454550e-003 + -9.0000000000000002e-001
)*9.6741922370932154e-001 )+ ((Range ("H" + RowString)*7.1999999999999998e-003 + -9.0000000000000002e-001
)*3.3523925139194388e-001 )+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
7.7641802537131110e-001 )+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*7.4309871423058294e-001 )+ ((Range ("K" + RowString)*1.5000000000000001e-002 + -9.0000000000000002e-001 )*-
1.7205796625734293e+000 )+ ((Range ("L" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
2.2794540913666927e+000 )+ ((Range ("M" + RowString)*1.0000000000000000e-002 + -1.6000000000000001e+000
)*2.7854992013325286e-001 )+ ((Range ("N" + RowString)*6.9230769230769233e-003 + -3.4615384615384615e-001 )*-
1.1060273109281480e+000 )+ ((Range ("O" + RowString)*1.3846153846153847e-002 + -6.9230769230769207e-002
)*1.6612244502337796e+000 )+7.9194129044803674e-001 )

Dim hidden1AxonPE2 As Double
hidden1AxonPE2 = tanh( ( (Range ("A" + RowString)*4.8648648648646e-003 + -4.3783783783770e-001 )*-
6.5198081146695702e-002 )+ ((Range ("B" + RowString)*2.3529411764705885e-003 + 1.9411764705882351e-001
```

```
)*5.4076751232925324e-001 )+ ((Range ("C" + RowString)*1.8000000000000002e-002 + -9.0000000000000024e-001 )*-
1.8821337291443163e-001 )+ ((Range ("D" + RowString)*6.5454545454545453e-003 + -9.0000000000000002e-001 )*-
3.4541219579437732e-001 )+ ((Range ("E" + RowString)*6.2068965517241377e-003 + -8.9999999999999980e-001
)*3.5330206261094593e-001 )+ ((Range ("F" + RowString)*3.3395176252319111e-003 + -9.0000000000000002e-001
)*5.0192178383909436e-001 )+ ((Range ("G" + RowString)*5.4545454545454550e-003 + -9.0000000000000002e-001
)*4.2214813726452227e-001 )+ ((Range ("H" + RowString)*7.1999999999999998e-003 + -9.0000000000000002e-001
)*1.9983621381700756e-001 )+ ((Range ("I" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
4.2507864186876221e-001 )+ ((Range ("J" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001
)*1.5166671642761920e-001 )+ ((Range ("K" + RowString)*1.5000000000000001e-002 + -9.0000000000000002e-001 )*-
1.0425620558439237e+000 )+ ((Range ("L" + RowString)*6.0000000000000001e-003 + -9.0000000000000002e-001 )*-
1.3843937118451779e+000 )+ ((Range ("M" + RowString)*1.0000000000000000e-002 + -1.6000000000000001e+000
)*4.6794678550958291e-001 )+ ((Range ("N" + RowString)*6.9230769230769233e-003 + -3.4615384615384615e-001
)*2.7112894109564861e-001 )+ ((Range ("O" + RowString)*1.3846153846153847e-002 + -6.9230769230769207e-002 )*-
3.6264914265409764e-001 )+-2.2739681539277484e-001 )

Dim outputAxonPE1 As Double

outputAxonPE1 = tanh( ( hidden1AxonPE1*-9.8488855655220897e-001 )+ ( hidden1AxonPE2*8.8290267051699034e-001
)+2.9618799655362205e-001 )

Range ("P" + RowString) = ( (outputAxonPE1)--9.3185840707964596e-001) / 1.5929203539823009e-002
Next i
End Sub

Function tanh(x As Double) As Double
tanh = (Exp(x) - Exp(-x)) / (Exp(x) + Exp(-x))
End Function

Function sigmoid(x As Double) As Double
sigmoid = 1.0 / (1.0 + Exp(-x))
End Function
```

Within the model A = expression of ER, B = PR, C = HER2, D = Ki67, E = p53, F = CK 5/6, G = CK 7/8, H = EGFR, I = BCL2, J = MUC1, K = VEGF, L = LKB1, M = BRCA1, N = HER3, O = HER4