THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Identification of new prognostic biomarkers for triple negative breast cancer

Ashwini Raghavendra

Master of Molecular Biology

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

School of Medicine

Abstract

Breast cancer is a complex, heterogeneous disease encompassing many different tumour entities. Management decisions are based on variables such as tumour size, histological grade, expression of hormone (oestrogen and progesterone) receptors (ER/PR), and expression of human epidermal growth factor receptor-2 (HER2). Clinical outcomes are highly variable amongst patients with 'triple-negative' breast cancer (TNBC; ER/PR/HER2-negative), but we currently cannot accurately predict prognosis in this group, thus inadequate disease control and over-treatment are challenges in the clinic. Furthermore, new therapeutic options are needed for TNBC, as there are currently no molecular-targeted therapies available.

Published data suggest that expression of the SRY (sex determining region Y)-box 10 (SOX10) transcription factor is associated with TNBC, particularly those exhibiting molecular similarity to the basal/myo-epithelium of the normal breast ('basal-like' breast cancer; BLBC). SOX10 belongs to the SOX family of transcription factors. It is an important regulator of early neural crest development and of melanocyte differentiation, in which expression is highest in melanocyte progenitors and is silenced in mature melanocytes. SOX10 is expressed in 87-97% of melanomas, and is used as a diagnostic marker clinically

The aims of the current study were to:

1. Investigate *SOX10* mRNA levels and mechanisms controlling its expression in breast cancer;
2. Investigate the prognostic significance of SOX10 in breast cancer;
3. Evaluate SOX10 protein expression in normal and tumour-adjacent normal breast;
4. Use gene expression profiling and *in vitro* and *in vivo* functional assays to investigate the possible functional roles of SOX10 in breast cancer; and
5. Investigate the prognostic significance of two cancer-testis antigens (MAGE-A and NY-ESO-1) in TNBC.

Analysis of published breast cancer genomic datasets showed that *SOX10* mRNA expression was inversely correlated with gene methylation (r^2 0.567; $p<00001$), suggesting epigenetic regulation is important in breast cancer. A subgroup of BLBCs (~65%) demonstrated high expression and significantly lower *SOX10* methylation than non-BLBCs (χ^2 $p<0.0001$). *SOX10* copy-number gains were also more frequent in BLBCs (χ^2 $p=0.0002$). *SOX10* mRNA expression was not associated with the clinical outcomes tested, including response to chemotherapy or overall survival, though immunohistochemical (IHC) analysis of breast cancer tissue microarrays showed that SOX10 protein expression was strongly associated with markers of poor outcome (*e.g.* histological grade,

$p<0.0001$) and predicted poor survival (hazard ratio (HR): 3.66; 95% CI: 1.78-7.51; $p<0.0001$). Critically, SOX10 protein expression also predicted poor survival in basal-like TNBCs (HR 3.20; 95% CI: 1.28-8.03; $p=0.0168$; n=52).

Little is known about the roles of SOX10 in either normal or neoplastic breast tissue. To investigate the roles of SOX10 in the human mammary epithelium, we analysed its expression *in situ* in reduction mammoplasty specimens (n=18) and tumour-associated normal breast tissue (n=19). SOX10 was consistently expressed in the myoepithelium and heterogeneously expressed in luminal cells. In luminal epithelia, there was a strong correlation between expression of SOX10 and the luminal progenitor marker c-kit, and an inverse association with ER and cytokeratin 8/18. *In silico* analysis of published gene expression data from mammary epithelial compartments showed *SOX10* was most highly expressed in luminal progenitor cells, and the least in mature luminal cells (χ^2 $p<0.0029$). In light of its role in differentiation in other tissues (*e.g.* melanocytes), these data suggest that SOX10 may regulate lineage differentiation in the mammary gland.

In order to investigate possible roles of SOX10 in breast cancer, we employed *in vitro* and *in vivo* breast cancer models. In-house screening experiments demonstrated infrequent expression of SOX10 in breast cancer cell lines. Only 2 of 39 basal-like lines (~5%) had detectable levels of *SOX10* mRNA. Short hairpin RNA (shRNA)-mediated knockdown of *SOX10* in basal-like MDA-MB-435 breast cancer cells did not alter their proliferation rate, migratory behaviour or sphere forming capacity *in vitro*, but did reduce colony formation on plastic ($p=0.028$), consistent with the idea that it regulates progenitor activity. *In vivo* tumourigenicity experiments showed that *SOX10* knockdown reduced tumour formation in the mouse mammary fat pad. Differential gene expression microarray analysis of these cells coupled with gene ontology analysis of *SOX10*-high and –low BLBCs showed a significant association between SOX10 and immune response genes.

In light of its roles in coordinating differentiation in the neural crest and the melanocyte lineage, we hypothesise that SOX10 is involved in regulating mammary epithelial lineage differentiation and that its expression is inappropriately activated and/or maintained in sporadic basal-like TNBCs, and that the involvement of SOX10 contributes to poor outcome. Future studies are required to fully understand the functional role of SOX10 in breast cancer.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed papers

Junankar, S., L. A. Baker, D. L. Roden, R. Nair, B. Elsworth, D. Gallego-Ortega, P. Lacaze, A. Cazet, I. Nikolic, W. S. Teo, J. Yang, A. McFarland, K. Harvey, M. J. Naylor, S. R. Lakhani, P. T. Simpson, A. Raghavendra, J. Saunus, J. Madore, W. Kaplan, C. Ormandy, E. K. Millar, S. O'Toole, K. Yun and A. Swarbrick (2015). "ID4 controls mammary stem cells and marks breast cancers with a stem cell-like phenotype." Nat Commun **6**: 6548.

Al-Ejeh, F., P. T. Simpson, J. M. Sanus, K. Klein, M. Kalimutho, W. Shi, M. Miranda, J. Kutasovic, A. Raghavendra, J. Madore, L. Reid, L. Krause, G. Chenevix-Trench, S. R. Lakhani and K. K. Khanna (2014). "Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer." Oncogenesis **3**: e100.

Conference abstracts

Ashwini Raghavendra, Jodi M Saunus, Ana Cristina Vargas, Chanel E Smart, Sarah Song, Jason Madore, Lynne Reid, Mythily Mariasegaram, Colleen Niland, Darina Vuong, Sarah Sim, Margaret C Cummings, Peter T Simpson, Sunil R Lakhani: “SOX10 is an independent marker of poor prognosis in triple-negative and basal-like breast cancer”. Presented at The International Association of Breast Cancer Research (Sydney) September 2014.

A. Raghavendra, A C. Vargas, J. Madore, M. Mariasegaram, L. Reid, J. Kutasovic, C. Niland, M. Cumming, L. Da Silva1, E. Bhutada, S. Peyton, T M. Hauer, S. Song, P. Simpson, C E. Smart, J. Saunus, S. R Lakhani: “SOX10 as a prognostic biomarker in triple negative and basal-like breast cancer”. Presented at Australian Breast Cancer Conference (Melbourne), November 2013.

A. Raghavendra, A C. Vargas, J. Madore, M. Mariasegaram, L. Reid, J. Kutasovic, C. Niland, M. Cumming, L. Da Silva1, E. Bhutada, S. Peyton, T M. Hauer, S. Song, P. Simpson, C E. Smart, J. Saunus, S. R Lakhani: “SOX10 as a prognostic biomarker in triple negative and basal-like breast cancer”. Presented at Australian Society of Stem Cell Research (Brisbane), October 2013.

A. Raghavendra, A C. Vargas, J. Madore, M. Mariasegaram, L. Reid, J. Kutasovic, C. Niland, M. Cumming, L. Da Silva1, E. Bhutada, S. Peyton, T M. Hauer, S. Song, P. Simpson, C E. Smart, J. Saunus, S. R Lakhani: “Role of SOX10 in breast cancer”. Presented at Australian Society of Medical Research (Brisbane), May 2013.

Publications included in this thesis

"No publications included".

Contributions by others to the thesis

Several members of the research team have made important contributions to the work presented in this thesis as follows-

- Prof. Sunil Lakhani performed pathology review of breast carcinoma cases included in the study.
- Lynne Reid provided technical support for building the Queensland Follow-up (QFU) cohort.
- Jason Madore, Colleen Niland and Lynne Reid assisted in collecting pathological and clinical information of patients involved in this study.
- Ana Cristina Vargas, Leonard da Silva and Ekta Bhutada built some of the QFU TMAs (n=250).
- Figure 3.2: Jodi Saunus generated the bar graph.
- Figure 3.3-3.6: Samir Lal assisted in data collection and Jodi Saunus constructed the plots.
- Figure 4.11: Mariska Miranda and Fares Al-Ejeh performed the experiments and plotted the figure.

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

Firstly, I wish to offer my sincere gratitude to my principle advisor Prof Sunil Lakhani for his mentorship; I have been very fortunate to have had the opportunity to work under his supervision. I am also grateful to my co-supervisors Dr Jodi Saunus, Dr Chanel Smart and Dr Peter Simpson for giving me the opportunity to be a part of their research group. I have received continuous support, motivation, enthusiasm and great supervision from them throughout my candidature.

I would like to acknowledge Australia Awards Scholarship (AusAID) for providing financial support throughout my candidature. This scholarship also supported my living expenses, overseas health cover and family reunion fair once a year. The completion of my PhD would not have been possible without this financial security. I also thank Prof Sunil Lakhani for providing me additional financial support to attend The International Association of Breast Cancer Research Conference in Sydney (2014) and Australian Breast Cancer Conference in Melbourne (2013) during my candidature. I acknowledge Laurel T Mackinnon for professional editing of my thesis.

I have been very fortunate to have met some amazing people through my journey. I would like to extend my thank you to my wonderful friends Lucy Bennett, Jason Madore, Colleen Niland and her family for all their love and support. I really appreciate all the little things that you guys have done for me, and I thank you for your friendship.

Finally and most importantly, I am deeply and forever indebted to my family for their love, support and encouragement through the journey of my PhD and through life. I can't thank them enough for all the sacrifices they have made to help me get to where I am today. To my parents, this PhD would not have been possible without your love and support, and I thank you for believing in me. To my brother, I look forward to spending more quality time with you. To my husband, thank you for your support and for all the long-distance phone calls during my challenging times.

Keywords

SOX10, prognostic biomarker, triple negative, basal-like breast cancer, immunohistochemistry, normal breast, luminal progenitor, MAGE-A, NYESO-1

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 110316, Pathology, 60%

ANZSRC code: 060405, Gene Expression, 20%

ANZSRC code: 111201, Cancer Cell Biology, 20%

Fields of Research (FoR) Classification

FoR code: 1112, Oncology and Carcinogenesis, 80%

FoR code: 1103, Clinical Sciences, 20%

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

Ab	Antibody
Ag	Antigen
ANOVA	Analysis of Variance
ARC	Animal Recourse Centre
ASCO	American Society of Clinical Oncology
ATCC	American Type Culture Collection
BCSS	Breast Cancer Specific Survival
BL1	Basal-Like 1
BL2	Basal-Like 2
BLBC	Basal-Like Breast Cancer
BSA	Bovine Serum Albumin
CAP	College of American Pathologists
cDNA	Complementary Deoxyribonucleic Acid
CISH	Chromagen In-Situ Hybridization
CK	Cytokeratin
CN	Copy Number
CT	Cancer-Testis
DCIS	Ductal Carcinoma In-Situ
DCT	Dopachrome Tautomerase
DFS	Disease Free Survival
dHPLC	High Performance Liquid Chromatography
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
ECM	Extracellular Matrix
EDNRB	Endothelin Receptor-B
EGFR	Epidermal Growth Factor Receptor
ELF5	E74-Like Factor 5
EMC	Extra Cellular Matrix
EMT	Epithelial Mesenchymal Transition
ENS	Enteric Nervous System
ER	Oestrogen Receptor
FFPE	Formalin Fixed Paraffin Embedded
GDNF	Glial Cell Derived Neurotrophic Factor
H&E	Haematoxylin and Eosin
HER2	Human Epidermal Growth Factor 2
HMG	High Mobility Group
HR	Hormone Receptor
HR	Hazard Ratio
HRP	Horseradish Peroxidase
HRT	Hormone Replacement Therapy
ID4	Inhibitor of DNA Binding 4,
IC NST	Invasive Carcinoma No Special Type
IF	Immunofluorescence
IgG	Immunoglobulin G
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma

IM	Immune Modulatory
INPP4B	Inositol Polyphosphate-4-Phosphatase
ISH	In-Situ Hybridization
kConFab	Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer
LAR	Luminal Androgen Receptor
LB	Luria Broth
LN	Lymph Node
LP	Luminal Progenitor
LVI	Lympho Vascular Invasion
MAGE-A	Melanoma Associated Antigen A
MAPK	Mitogen Activated Protein Kinase
MaSC	Mammary Stem Cell
MBP	Myelin Basic Protein
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MHC	Major Histocompatibility Complex
MITF	Microphthalmia-Associated Transcript Factor
ML	Mature Luminal
MPSS	Massive Parallel Signature Sequencing
mRNA	Messenger Ribonucleic Acid
MSL	Mesenchymal Stem-Like
NC	Neural Crest
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
NTNC	Non-Target Negative Control
OS	Overall Survival
P0	Protein Zero
PARP	Poly-(ADP-Ribose)-Polymerase
pCR	Pathological Complete Response
PDGF-B	Platelet-Derived Growth factor B
PI3K	Phosphatidylinositol 3-Kinase
PLP	Proteolipid Protein
PNS	Peripheral Nervous System
PR	Progesterone Receptor
QCR	Queensland Cancer Registry
QFU	Queensland Follow-Up
qRT-PCR	quantitative Reverse Transcription Polymerase Chain Reaction
RBWH	Royal Brisbane Women's Hospital
RFS	Recurrence Free Survival
RNA	Ribonucleic Acid
RT	Room Temperature
SSP	Single Sample Predictor
TAD	Transactivation Domain
TBS	Tris Buffered Saline
TCGA	The Cancer Genome Atlas
TDLU	Terminal Ductal Lobular Units
TEB	Terminal End Buds

TMA	Tissue Microarray
TN	Triple Negative
TNBC	Triple Negative Breast Cancer
TNBL	Triple Negative Basal-Like
TR	Tamoxifen Resistant
WHO	World Health Organization

Chapter I

General Introduction

Chapter I: General Introduction

1.1 Epidemiology and natural history of breast cancer

Breast cancer is the most common cancer in Australian women and the second leading cause of cancer related death [1]. In 2010, the risk of developing breast cancer before the age of 85 years was 1 in 8 and the average age of diagnosis was 60 years [1]. The incidence of breast cancer in Australia is increasing (Figure 1.1A), and it is estimated that there will be 17,210 new cases diagnosed by 2020 [2]. However, the age-standardised mortality rate has decreased (Figure 1.1B) as a result of technological advances in the early diagnosis using mammographic screening and improvement in treatment modalities in the form of adjuvant therapy [2, 3]. Despite early detection and advances in systemic therapy, metastatic breast cancer is incurable and is responsible for about 90% of cancer-associated deaths [4].

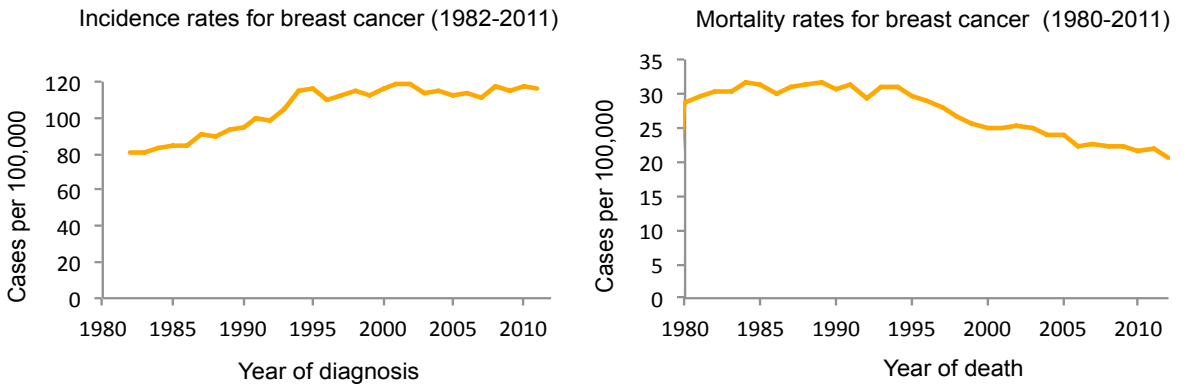


Figure 1.1: Age-specific **(A)** incidence and **(B)** mortality rates for breast cancer in Australian women [1].

1.2 Breast cancer development

1.2.1 Breast cancer risk factors

Breast cancer aetiology is multifactorial, with interplay between genetic and environmental factors [5]. The strongest breast cancer risk factor is family history of the disease, with several high-risk susceptibility genes, and multiple low-risk/risk modifier genes implicated (as reviewed in [6, 7]). Women with one or two first-degree relatives (*i.e.* mother, sister, daughter) with breast cancer are at two to three-fold elevated risk. The risk is even greater when: i) more relatives are affected; ii) if the breast cancer occurred before menopause; and iii) if it was bilateral. The genetic basis for familial predisposition to breast cancer remained unknown until the discovery of breast cancer susceptibility gene known as *BRCA1* [8]. Hall and colleagues studied 146 breast cancer cases with 23 extended families and found strong evidence for linkage of breast cancer susceptibility to D17S74 anonymous marker located on chromosome 17q21. This linkage was found specifically in 40% of families with early onset of breast cancer [9, 10]. The following year (1993), Narod and colleagues

extended this research and determined that hereditary breast cancer in 3 of 5 large families was linked to the same marker D17S74 [11], which was later formally named as '*BRCA1'* [12]. In 1994, a second breast cancer susceptibility locus, *BRCA2*, was found by linkage analysis of 15 high-risk breast cancer families to be localised to 13q12-13 [13]. Familial breast cancers account for only about 5-10% of all breast cancers [14].

In addition to genetic predisposition, hormonal factors also contribute to risk. For example, early age of menarche (<12 years), late pregnancy, late menopause (>55 years), late age at first full-term pregnancy (>35 years) and nulliparity are all associated with increased risk [15]. These relationships are likely related to prolonged exposure to endogenous oestrogens [16-19]. Consistent with this idea, breast-feeding reduces the risk of breast cancer, with an estimated 4.3% risk reduction for every year of breast-feeding, in addition to a 7% decrease for each birth [17, 18, 20]. Finally, age has a major impact on breast cancer risk, with incidence doubling every 10 years until menopause (<50 years), after which the rate of increase plateaus between 70-80 years [17, 18, 21].

1.2.2 Histogenesis of the breast

The human breast is characterised by a branching network of ducts and lobules that end in clusters (terminal ductal lobular units; TDLUs), which comprise two major epithelial cell types: the inner secretory luminal cells that surround a central lumen and the outer contractile myoepithelial cells that are basally located adjacent to basement membrane (Figure 1.2) [22, 23]. Mammary epithelial cells are dynamic, undergoing cyclical morphogenetic changes during puberty, pregnancy and lactation. Evidence suggests the existence of an epithelial differentiation hierarchy within human breast, with undifferentiated mammary stem cells (MaSC) at the apex, which are capable of self-renewal and generation of daughter cells that differentiate into committed progenitor cells, luminal and myoepithelial lineages (Figure 1.2) [23]. Lim and colleagues functionally characterised the human mammary hierarchy in normal breast specimens based on cell surface markers, and isolated four subpopulations using fluorescence-activated cell sorting (FACS): luminal progenitor (LP), mature luminal (ML), myoepithelial and MaSC-enriched and stromal (S) populations. Analysis of the gene expression profiles of these subpopulations revealed similarities to the intrinsic subtypes of breast cancer (See section 1.4.1), raising the possibility that different breast cancer subtypes either arise from, and/or adopt phenotypic features of particular normal breast cell types [24, 25]. For example, basal-like breast tumours showed greatest transcriptomic similarity to luminal progenitor cells of the normal breast (Figure 1.2C) [23].

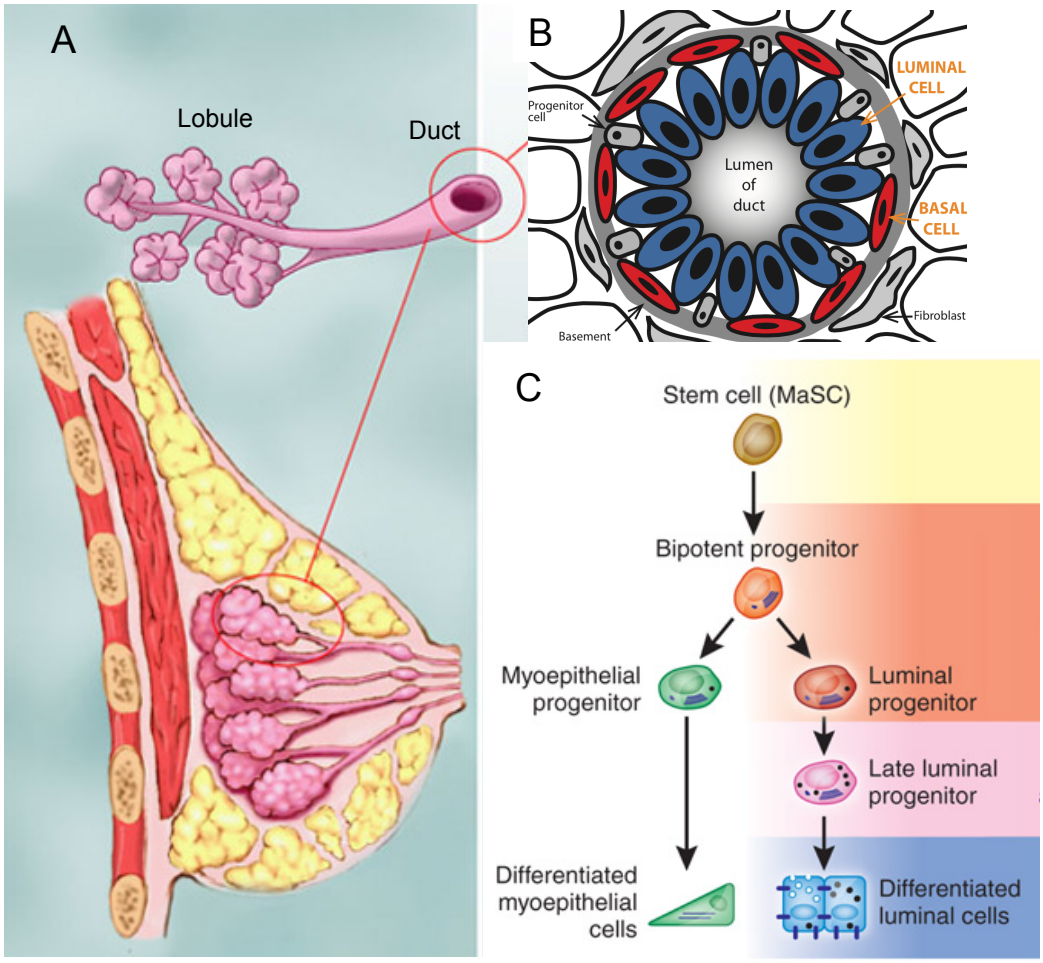


Figure1.2: Normal breast architecture and differentiation hierarchy **(A)** Structure of normal breast showing lobules and ducts, **(B)** Different cell types of normal breast: outer contractile myoepithelial (in red) and inner columnar luminal cells (in blue), luminal progenitor cells are shown to be located in between luminal cells adjacent to myoepithelial (in grey). **(C)** Model of differentiation hierarchy in normal breast. The common progenitor cells are also called as bipotent progenitor cells which are capable of differentiating into committed progenitor cells of luminal and myoepithelial lineages [25]. Permission to use figure C was obtained from the publisher through Copyright Clearance Center's RightsLink Service.

1.2.3 Breast cancer evolution

Breast cancer arises from the TDLUs of the breast through a series of morphological changes that are recognised as pre-invasive lesions. These lesions are characterised by clonal proliferation of epithelial cells that remain contained within the basement membrane and hence confined to the ductal structure of the breast. The most commonly identified precursor lesions are columnar cell lesions (CCL), atypical ductal hyperplasia (ADH), ductal carcinoma *in situ* (DCIS), and lobular carcinoma *in situ* (LCIS); with 'ductal' and 'lobular' breast cancers representing the two most common subtypes (see section 1.3.2).

The original model of development suggested that breast cancers evolve in a stepwise manner, beginning as hyperplastic benign lesions (hyperplasia of usual type) that progress to atypical

hyperplasia, followed by *in situ* carcinoma and finally invasive cancer, as cells penetrate the basement membrane and invade the local stroma. This process was thought to be separate for ductal and lobular types of cancer. Our understanding of this model has evolved over time with more sophisticated molecular analysis of pre-invasive lesions (*e.g.* sequencing, loss-of-heterozygosity (LOH) and comparative genomic hybridization (GGH) analyses). We now understand this to be a more complex multistep model. A key molecular finding that supported this multistep progression of breast cancer was loss of LOH identified at 16q and 17p in both DCIS and invasive, which was also seen at a similar frequency in ADH [26]. Lesions like hyperplasia of usual type are no longer considered as clonal precursor lesions, whereas columnar cell lesions are the likeliest and earliest morphological stage between normal epithelium and atypical hyperplasia [27]. Importantly, these pre-invasive lesions can be detected during early screening programs and are considered to be indicators of risk for invasive breast cancer [28].

1.3 Clinical management

Currently, breast examination, mammography and ultrasound are the standard screening methods for breast cancer [29]. Management and treatment of breast cancer requires a multidisciplinary approach, including surgery, radiation therapy and systemic therapy. The choice of treatment varies depending on the stage, tumour grade, histological type, expression of hormone receptors (HR; oestrogen and progesterone) and expression of human epidermal growth factor receptor 2 (HER2). These parameters are included in the St Gallen breast cancer guidelines, the National Institute of Health consensus and Adjuvant! Online to facilitate clinical decision-making [30, 31].

Surgery and radiation therapy are local therapies for managing primary disease, which is generally considered to decrease the risk of local recurrence. Surgery involves either lumpectomy (breast conservation surgery) or mastectomy when the tumour is >40mm in size. Radiation therapy is offered to patients who have undergone lumpectomy, have primary tumours >5 cm, skin/chest wall involvement or multiple lymph nodes involved (>4).

Systemic therapy (chemotherapy, anti-oestrogen/endocrine therapy and anti-HER2 therapy) is considered when there is a high risk of recurrence (e.g. when there is lymph node positivity, or other indicators of poor prognosis, see below). This is an important component of treatment for patients diagnosed with either locally advanced or metastatic breast cancer to increase the likelihood of cure and/or survival. This mode of treatment is usually given to patients postoperatively or after surgery (adjuvant setting). Some patients are prescribed systemic therapy

before surgery (neoadjuvant setting) in order to debulk the primary tumour to facilitate breast conservation surgery, and to facilitate cure.

Breast cancer management decisions are based on detailed histopathologic review and classification of the primary tumour. This involves assessment of prognostic and predictive factors (such as lymph node, tumour grade/stage, histological type, ER/PR and HER2 status) to select treatment regimens most appropriate for individual patients (*e.g.* tumours expressing HRs are most likely to respond to endocrine therapy (section 1.3.4). In addition to these, increasingly newer molecular techniques are used to guide clinical decision-making (see section 1.4.4).

1.3.1 Histological type

According to the latest World Health Organization (WHO, fourth edition) classification, there are more than 21 different histological types of breast cancer, which are characterised by differences in cellular morphology and growth patterns [32]. Most breast cancers (~75%) are classified as invasive carcinoma of 'no-special type' (IC NST), a large group of invasive breast cancers that do not demonstrate distinguishing morphological characteristics. IC NST is a very heterogeneous group and the diagnosis is one of exclusion. IC NST group of tumours is not only hugely variable in histologic presentation, but also in clinical outcome/prognosis/response to therapy. Thus histological typing is not particularly useful for guiding management for tumours in this category. The remaining tumour types are morphologically distinct and are categorised as 'special types', where ≥90% of the tumour exhibits distinguishing cytological and/or architectural differentiation [32, 33]. These include invasive lobular, tubular, mucinous and metaplastic carcinoma and carcinoma with medullary, neuroendocrine or apocrine features. Invasive lobular carcinomas (ILCs) account for about ~15% of all breast cancers while the other special types of breast cancers such as tubular, metaplastic, medullary, apocrine or mucinous etc., account for <2% of all invasive breast carcinomas [32]. Importantly, some of these breast cancer special types are associated with particular clinical behaviour. For example, tubular carcinomas relatively have better prognosis [34-37] whereas, metaplastic carcinomas show poor prognosis [38, 39]. Representative images of some morphological variants of invasive breast cancer are shown in Figure 1.3 [40] .

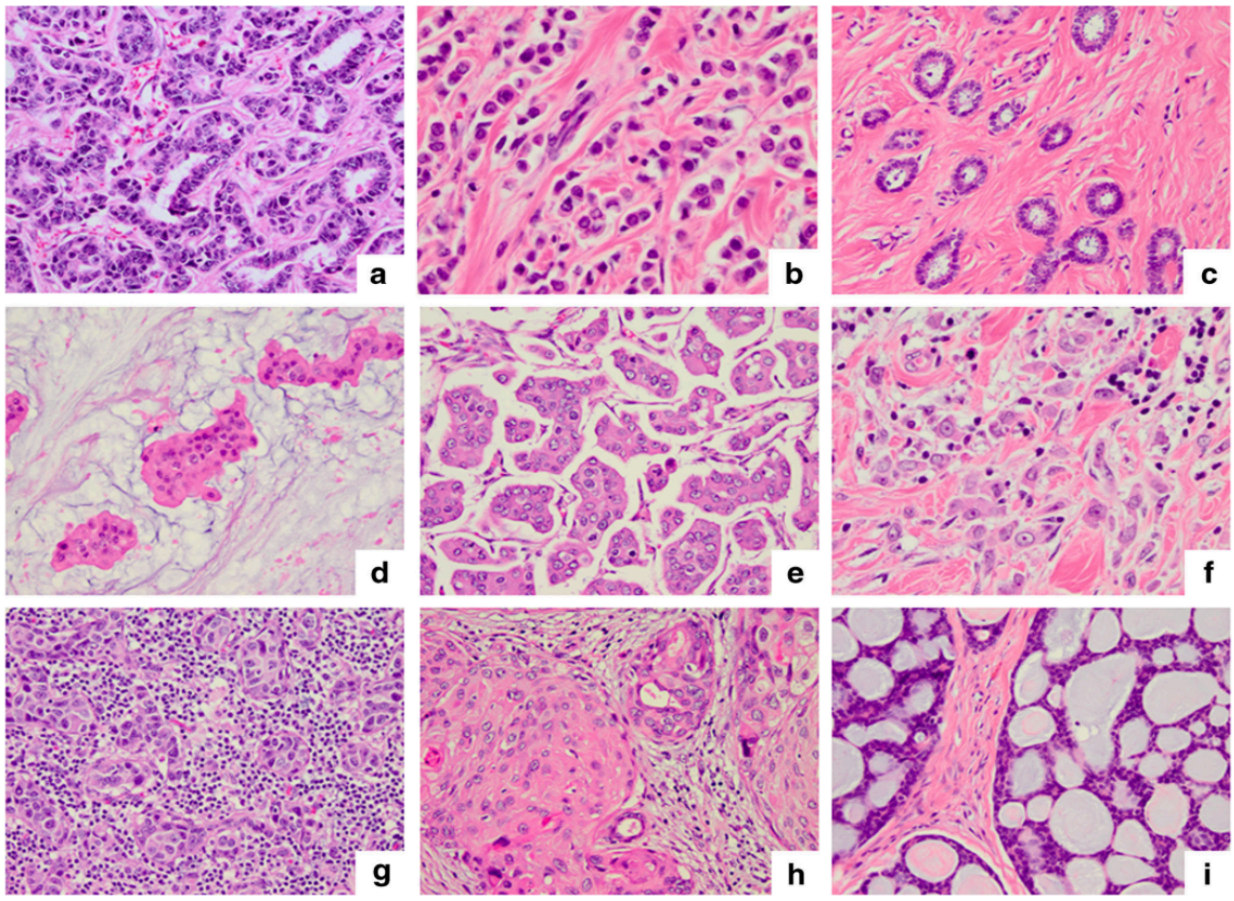


Figure 1.3: Representative images of some of the morphological types of breast cancer: (**a**) invasive carcinoma NST, (**b**) classic lobular carcinoma, (**c**) tubular carcinoma, (**d**) mucinous Carcinoma, (**e**) invasive micropapillary carcinoma; (**f**) pleomorphic lobular carcinoma, (**g**) medullary carcinoma; (**h**) metaplastic carcinoma, (**i**) adenoid cystic carcinoma. Figure adapted from [40]. Permission to use figure was obtained from the publisher through Copyright Clearance Center's RightsLink Service.

1.3.2 Histological grade

Histological grade is one of the main prognostic determinants used in routine practice and involves morphological assessment. The Nottingham Combined Histological Grading System evaluates three morphologic features scored in routine Haematoxylin and Eosin (H&E)-stained slides: the percentage of tubule formation, the degree of nuclear pleomorphism and the mitotic count. Each criteria is given a score between 1-3; the lowest score for a tumour being 3 and the highest being 9. Breast cancers are then classified as grade 1 or well differentiated (when the score is between 3 to 5), grade 2 or moderately differentiated (when the score is either 6 or 7) and grade 3 or poorly differentiated (with a score of 8 or 9, Figure 1.4 a-c) [41]. Histological grade is an essential component of breast cancer prognostication because stratification of patients according to this system has shown clear differences in disease-free survival (DFS) and overall survival (OS) (Figure 1.4 d and e) [42]. The relationship between grade and prognosis was well documented in a study of 1830 breast cancer patients; most of the cases were invasive carcinoma NST. The results showed 19% were grade I, 34% grade II and 47% grade III. The recurrence-free survival (RFS) and OS

were reduced to about 40% in patients with a poorly differentiated tumour (grade III) compared with those with well-differentiated tumours (grade I, >80% DFS).

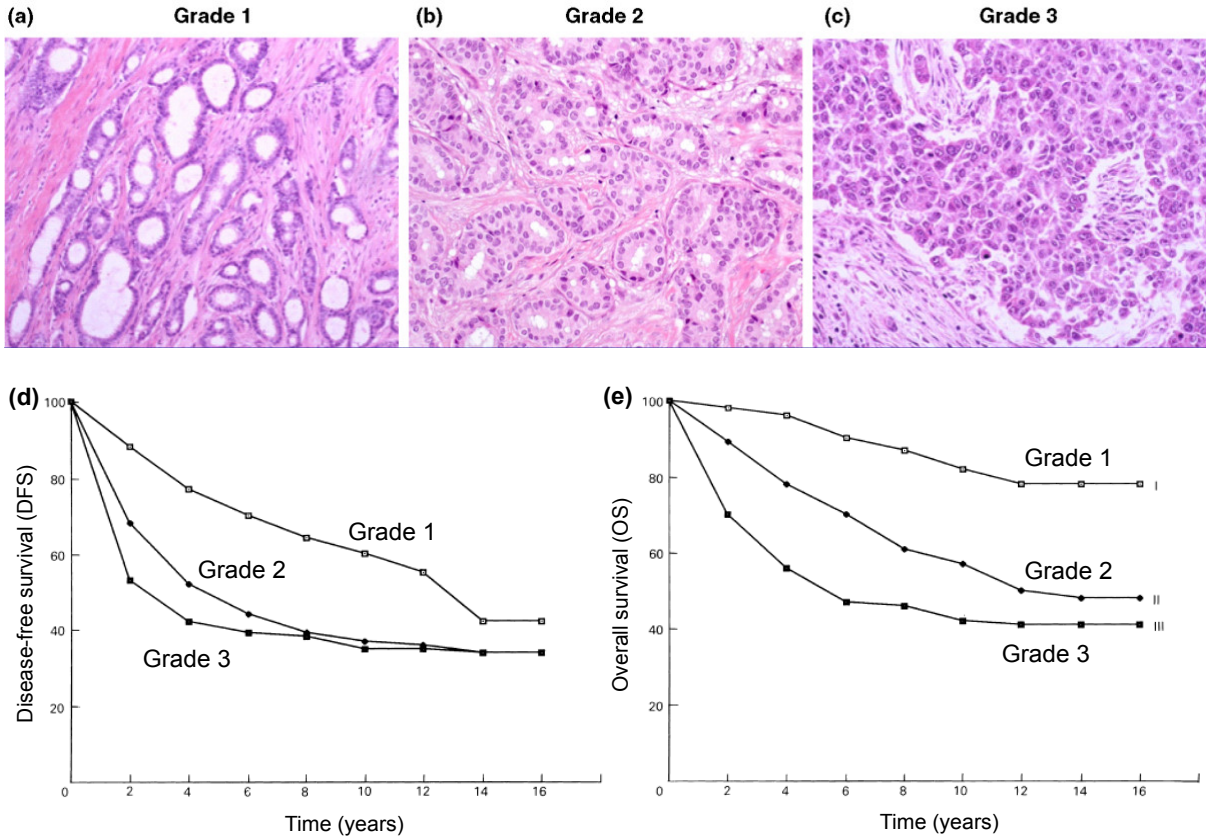


Figure 1.4: Histological grading of breast cancer according to Nottingham Grading System and its relationship with survival outcome. Images represent **(a)** well- **(b)** moderately-and **(c)** poorly differentiated IDC-NST **(d)** recurrence-free survival and **(e)** overall survival of 1830 patients with invasive breast carcinoma stratified according to histological grading (P<0.001). Figure from [41] and [42]. Permission to use figures was obtained from the publisher through Copyright Clearance Center's RightsLink Service.

1.3.3 Assessment of predictive tumour biomarkers

1.3.3.1 Hormone receptors (oestrogen and progesterone)

The ER and PR are nuclear transcription factors essential for normal mammary gland development and function. Ligand binding induces nuclear translocation of receptor complexes, which leads to the transcriptional regulation (activation or repression) of the target genes involved in cellular growth, differentiation and proliferation [43]. Expression of HRs in the normal breast epithelium is cyclic and fluctuates with the menstrual cycle [44].

About 80% of invasive breast cancers express ER [45, 46], which is significant because ER expression is a powerful predictive indicator for adjuvant endocrine therapy. For example, five years of adjuvant tamoxifen treatment in patients with HR+ tumours led to proportional reductions

in the risk of recurrence and mortality of 47% and 26%, respectively [47]. A collaborative meta-analysis of data from 20 clinical trials involving 21,457 HR+ breast cancer patients showed that treatment with tamoxifen halved the rate of recurrence during years 0-4 and reduced it by one-third during years 5-9, with an overall reduction of 39% [48]. Expression of the PR is strongly dependent on oestrogen, and therefore PR-positivity is interpreted clinically as a marker of functional ER signalling [49, 50]. Around 25% of ER+ breast cancers lack PR expression and are less responsive to endocrine therapy [49]. More recently published data has shown that PR is more than just a downstream marker of ER, it may re-program ER to different chromatin binding sites within breast cancer cells resulting in a unique gene expression programme that is associated with better clinical outcome [51].

The relationship between HR expression and disease progression has also been assessed in the treatment-naïve setting. The NSABP B-06 trial randomised women with early-stage breast cancer to mastectomy, lumpectomy alone, or lumpectomy followed by radiation therapy [52]. No adjuvant systemic therapy was administered. The 5-year DFS and OS rates were 74% and 92% for ER+ cases, and 66% and 82%, for ER-negative cases respectively [52]. Because ER and PR expression in invasive breast carcinomas has a proven prognostic and predictive utility, positive staining by immunohistochemistry (IHC) is used clinically to provide prognostic information and to identify tumours that may respond to endocrine therapy [53, 54]. The current recommended cut-offs for ER and PR positivity by the American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP) is nuclear staining in at least 1% of tumour cells [55].

1.3.3.2 Human epidermal growth factor 2 (HER2)

The human epidermal growth factor receptor 2 (HER2) is a receptor tyrosine kinase encoded by the *ERBB2* gene on chromosome 17q12. Over-expression of HER2 resulting from gene amplification drives the basic level of activity above a threshold stimulating tumour growth [56]. HER2 over-expression is observed in 18-20% of breast cancers and is associated with decreased survival [57]. The current recommended cut-off for HER2 positivity by the ASCO and CAP is 3+ by IHC (strong, complete membrane staining in at least >30% of homogeneous tumour cells) and > 6 gene copies by *in-situ* hybridization (ISH) [58].

Development of the anti-HER2 monoclonal antibody trastuzumab (Herceptin) marked a major breakthrough in breast cancer management in the early 90's [59]. Trastuzumab is the first line of treatment against $HER2^{+}$ breast cancers in their early-stage; and has shown to significantly reduce the risk of recurrence after surgery [60]. Trastuzumab inhibits HER2-mediated signalling *via* the

phosphatidylinositol 3-kinase (*PI3K*) and mitogen-activated protein kinase (*MAPK*) pathways, leading to an increase in apoptosis and decrease in cellular proliferation [61, 62]. When trastuzumab was used as a single agent, the response rate in 111 patients with HER2 amplification patients with 3+ IHC staining was 35% and the response rate for patients with 2+ cases was 0%; the response rates in patients with or without *ERBB2* amplification examined by FISH were 34% and 7%, respectively [63]. In another study of breast cancer treated with trastuzumab in combination with chemotherapy, the overall response rates ranged from 67-81% in patients with HER2 amplification and 41-46% in patients with normal expression of HER2 [64]. Randomised clinical trials have demonstrated that adjuvant chemotherapy with and without trastuzumab had an absolute difference in DFS of 12% at 3 years; and was associated with a 33% reduction in the risk of death [65]. In metastatic breast cancers that over-expressing HER2, the use of trastuzumab in combination with chemotherapy is associated with a longer time to disease progression (median, 7.4 vs. 4.6 months), higher rate of objective response (50% vs. 32%), longer duration of response (median, 9.1 vs. 6.1 months), lower rate of death at 1 year (22% vs. 33%), longer survival (median survival, 25.1 vs. 20.3 months), and 20% reduction in the risk of death compared with chemotherapy alone [66].

Lapatinib, a dual tyrosine kinase inhibitor that blocks both HER2 and Epidermal Growth Factor Receptor (EGFR) [67, 68] inhibits the growth of breast cancer cell lines that are $HER2^+$ in culture and in tumour xenografts [69, 70]. Lapatinib is used for the treatment of $HER2^+$ metastatic breast cancer that progresses after initial treatment with trastuzumab or taxane-derived chemotherapies. Combination therapy of lapatinib with capecitabine in patients with $HER2^+$ advanced breast cancer has been shown to reduce the risk of disease progression compared with capecitabine alone [71]. New agents are constantly being developed to improve the toxicity profiles and to reduce the development of resistance to trastuzumab. The use of pertuzumab, another recently developed humanised monoclonal antibody, targets specifically and has shown significant decrease of tumour growth in $HER2^+$ metastatic breast cancers [72]. The combination of pertuzumab, trastuzumab and docetaxel, as compared with placebo, trastuzumab and docetaxel, used as first-line treatment for $HER2^+$ metastatic breast cancer, markedly prolonged progression-free survival in the randomised, multinational, phase III CLEOPATRA (Clinical Evaluation of Pertuzumab and Trastuzumab) trial [73]. Several other clinical trials are currently examining the efficacy of pertuzumab in patients with advanced or metastatic $HER2^+$ breast cancer. One such example is the phase II NeoSphere trial (NCT00545688) [72]. Recently, an antibody drug conjugate, ado-trastuzumab emtansine (T-DM1) that combines the antitumor properties of the humanized anti-HER2 with maytansinoid, DM1 (a potent microtubule-disrupting agent) has been approved by FDA for treatment of patients with HER2 positive metastatic breast cancer [74]. HER2 expression in breast cancer strongly predicts the

response to HER2-targeted therapies and is therefore a good example of both prognostic and a predictive biomarker

1.3.3.3 Triple-negative and basal-like breast cancer (TN and BLBC)

Triple-negative breast cancer (TNBC) is defined by a lack ER, PR and HER2 expression by IHC, and account for 15-18% of all breast cancers [75, 76]. Patients with TNBC are generally younger (<50 years) and tumours are more frequently high-grade. TNBCs are often more aggressive than other breast cancer types [77] and are associated with higher risk of distant recurrence and death, particularly in the first three years after diagnosis. There are no targeted treatment guidelines for TNBC and clinical management is challenging. Generally speaking, many TNBCs initially respond to first-line adjuvant treatment with anthracycline and taxane-based chemotherapy, but there is a high risk of relapse [78, 79]. Recently, Poly-(ADP-ribose)-polymerase (PARP) inhibitors have shown favourable results in some TNBC patients [80, 81], however assessment of the long-term effects of PARP inhibitors is needed. Combination therapies (including ixabepilone and capecitabine) and other useful antiangiogenic agents such as tyrosine kinase and EGFR inhibitors have shown varying response rates but with little or no survival benefit. Other molecular-targeted agents (*e.g.* inhibitors of EGFR, c-kit, Raf/MEK/MAPK and mTOR) have shown limited clinical benefit (reviewed in [82, 83]).

The 'basal-like' subset of TNBCs is characterised by expression of markers normally found in the basal (myoepithelial) compartment of the breast. Classification requires expression of at least one of the basal cytokeratins (*e.g.* CK5/6, CK14 or CK17) and/or EGFR [84, 85]. Other basal markers are often also expressed (*e.g.* p63, P-cadherin, caveolins 1 and 2, nestin, αB crystallin and CD109) [85-87]. The basal-like breast cancer (BLBC) breast cancer subtype came to prominence following landmark breast cancer molecular subtyping studies conducted by Perou, Sorlie and colleagues around 15 years ago, which identified a BLBC cluster as one of five intrinsic molecular subtypes [88, 89] (see section 1.6.1). Although BLBC and basal-like TNBC (defined by gene expression profiling and IHC respectively) are sometimes used interchangeably, there is up to 30% discordance between the types [85, 90]. For instance, not all BLBCs lack expression of ER, PR, and HER2, but there is significant overlap (~75% of BLBCs are TN) [90-92].

TNBCs tend to metastasise to the lung and brain, and patients have significantly shorter survival after the first metastatic event compared with $HER2^+$ and ER^+ cancers (median 22 months, *vs* 30 and 63.5 months, respectively) [93]. This was supported by a study from our group, where basal-like tumours were classified based on CK14 expression using IHC in a cohort of grade III IC NST. Overall, CK14+ cases had better long-term survival, yet it was evident that there were subgroups of patients with either good or poor long-term outcomes (Figure 1.5A). Indeed, all CK14+ patients who developed metastatic disease died within 2 years (Figure 1.5B) [94]. Thus outcomes in this group are heterogeneous, and it is not yet possible to predict clinical behaviour in routine diagnostic practice. Current management strategies are failing to achieve disease control and this subgroup requires better prognostic indicators (*e.g.* tumour biomarkers) as well as new therapeutic targets.

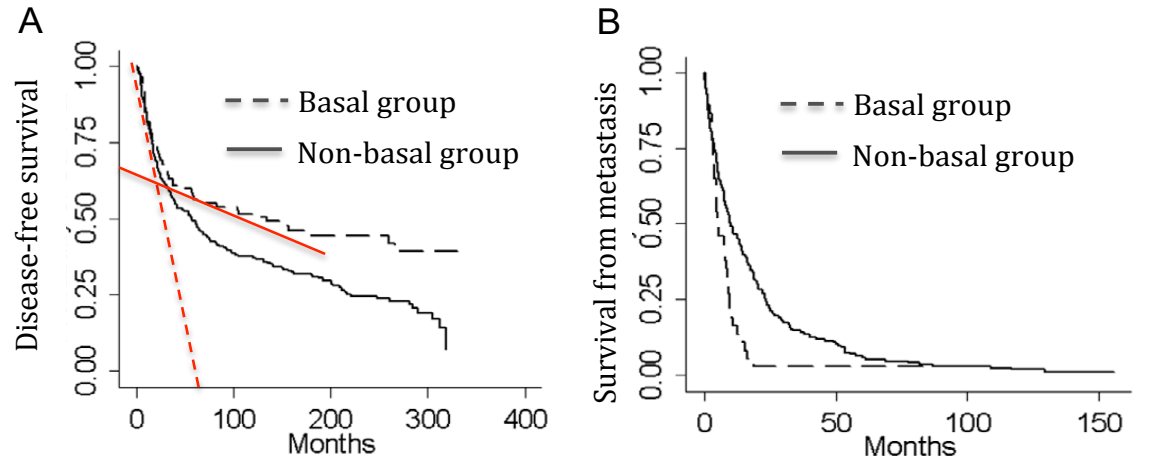


Figure 1.5: Long term survival in basal (dotted black line) versus non-basal tumour (solid black line). Kaplan-Meier survival plot for basal versus non-basal showing **(A)** disease-free survival and **(B)** survival from metastasis. The red dotted line represents patients with worse outcome and the red solid line represents patients with better outcome. Figure adapted from [94].

1.4 Molecular classification of breast cancer

The conventional method of breast cancer histological typing, grading and staging has been the foundation of outcome indicators. However, it has become obvious that not all breast cancers with the same type, grade or stage have similar underlying biology or clinical outcome [95, 96]. Although the expression of biomarkers ER/PR and HER2 offers information about how a tumour behaves clinically, the underlying biology of a tumour can only be resolved further by combining histopathologic factors with additional tissue-based molecular testing. In line with this idea, attempts have been made to classify breast cancers into different subgroups based on their molecular features in order to account for heterogeneity within the histopathologic classification. Two large breast cancer genomic datasets have substantially broadened our understanding of the molecular basis of breast cancer. The Cancer Genome Atlas (TCGA) research network has

characterised large numbers of human tumour samples to identify molecular aberrations at the DNA, RNA, protein and epigenetic levels [97]. Similarly, the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset contains clinical traits, expression, copy-number and single nucleotide polymorphism (SNP) genotype data from breast tumours collected from participants in the METABRIC trial [98].

1.4.1 Breast cancer intrinsic subtypes

cDNA microarray experiments performed by Perou and Sorlie showed that breast tumours could be classified into different molecular subtypes [88]. The original classification was further re-defined by the same group and others, which further contributed to characterisation of new molecular subtypes. Based on the similarities and differences in the gene expression profiles, breast tumours were mainly classified into five intrinsic or molecular subtypes (Stanford taxonomy, Figure 1.6A) called, luminal A, luminal B, $HER2^{+}$, basal-like and normal-like (Figure 1.6A) [89]. The luminal A subtype is the most frequent (24%-39%), followed by basal-like (17%-37%), luminal B (10%-18%), HER2 (4-10%) and normal-like (0-5%) [89, 99, 100]. Subsequent studies using larger cohorts have shown greater reproducibility (with >75% samples clustering within the same group) and also contributed to identifying new molecular subtypes [101-103]. Although the reproducibility of this classification was high, there were some discrepancies in more recent studies. By using three different classifiers, Weigelt et al., demonstrated that, only the basal-like subgroup was consistently classified as basal-like regardless of the method used. Classification of the remaining molecular subtypes (Luminal A, luminal B, HER2 and normal breast) varied considerably depending on the classifier used [104]. Also studies have shown that the normal-like subtype was not a robust category and represented an artefact of gene signatures related to stromal cells and adipose tissue (in the low cellularity tumour samples) [105, 106]. The proportion of unclassified tumours varies significantly across different platforms [89, 99, 100] suggesting some discrepancies.

This molecular classification of breast cancer is based on the combined and consistent expression of specific gene clusters. Luminal A and B tumours express the ER, PR and genes associated with ER pathway activation, such as *FOXA1, GATA3* and *CCND1* [89, 102]. Unlike luminal A, luminal B tumours have higher expression of proliferation-related genes (*e.g. CCNB1* and *MK167*), lower expression of hormone receptors (*ESR1/PGR*) and variable expression of *ERBB2* [89, 107]. The HER2 subtype is characterised by over-expression of *ERBB2* and other genes in the 17q12-21 amplicon (*e.g. GRB7*). BLBCs are characterised by expression of basal cytokeratins *(CKs 5,6,14), KIT*, and *FOXC1*, frequent *TP53* mutations and high proliferative activity [108]. BLBC overlaps with the group of basal-like TNBCs, which has been known to pathologists for many years, and this

overlap validates the molecular taxonomy to some extent. Importantly, classification based on intrinsic subtypes has shown predictive value due to the distinct clinical behaviour and overall survival of the different subtypes (Figure 1.6 B/C) [89, 109]. For example, luminal A tumours are associated with better survival, whereas BLBC and HER2 subgroups show the worst metastasis-free- and OS rates [110]. The intrinsic subtypes have been validated using more samples and newer, denser microarrays [89, 102, 111, 112], and have also been correlated with survival [113, 114] and treatment response [106, 115], which gives clinical context to the classification system.

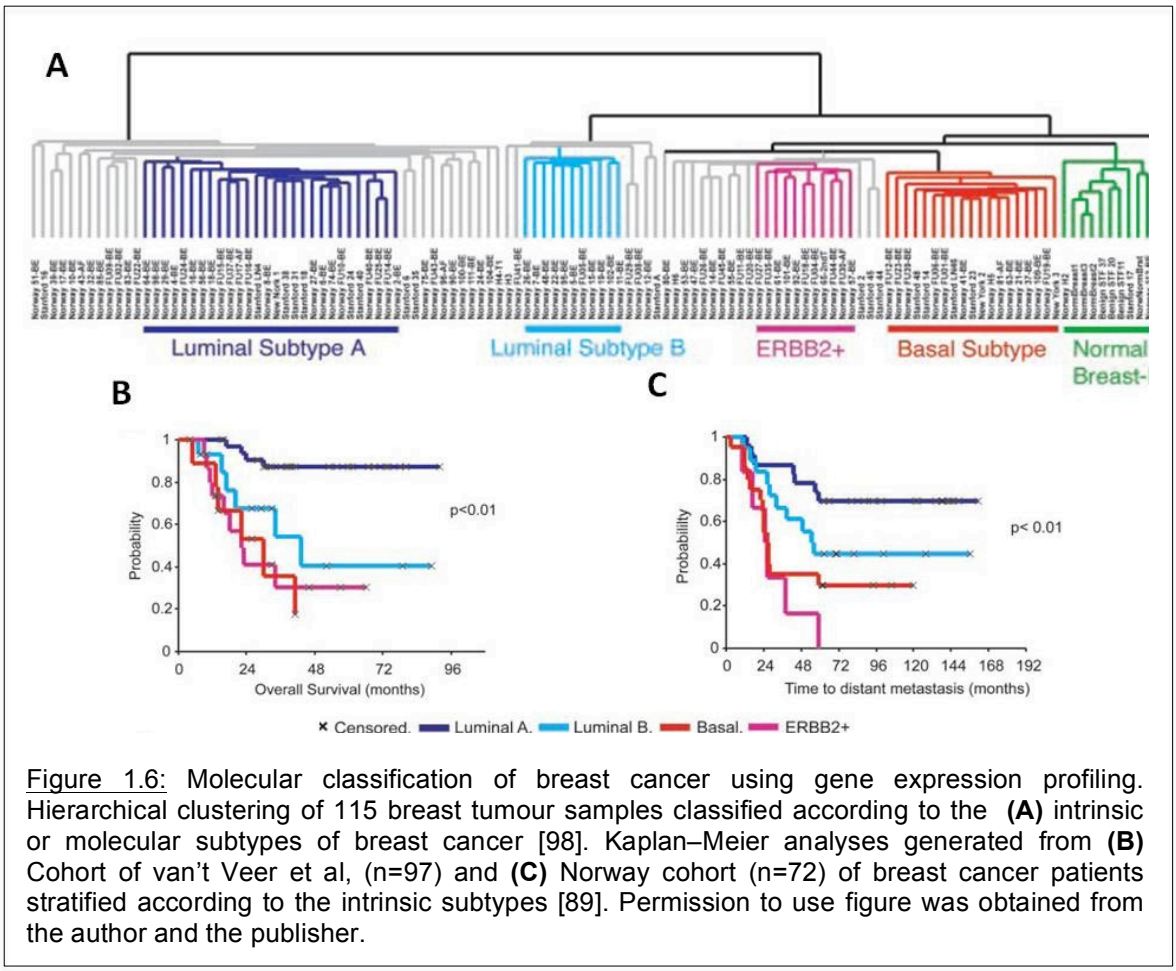


Figure 1.6: Molecular classification of breast cancer using gene expression profiling. Hierarchical clustering of 115 breast tumour samples classified according to the **(A)** intrinsic or molecular subtypes of breast cancer [98]. Kaplan–Meier analyses generated from **(B)** Cohort of van't Veer et al, (n=97) and **(C)** Norway cohort (n=72) of breast cancer patients stratified according to the intrinsic subtypes [89]. Permission to use figure was obtained from the author and the publisher.

1.4.2 Emerging molecular subtypes of breast cancer

With an increasing number of publicly available breast cancer datasets and emerging molecular profiling technologies, molecular classification of breast cancer is continually being refined. New subtypes have emerged through analysis of larger cohorts than the original Stanford study, including meta-analyses that account for technical disparities between array platforms. For example, 'claudin-low' and 'molecular apocrine' subtypes of breast cancer have been identified as distinct entities within the heterogeneous TN group. The Claudin-low subtype is characterised by low expression of cell adhesion genes (E-cadherin, Claudin 3, 4, 7 and Occludin) and high

expression of epithelial-mesenchymal transition (EMT) and stem cell ($CD44^{+}CD24^{-}$) related signatures and accounts for 7-14% of all invasive breast cancers [103]. Molecular apocrine tumours are enriched with androgen receptor (AR) expression (which regulates ER-responsive genes), calcium and ErbB signalling, lipid and fatty acid synthesis [101, 116] and accounts for about 10-15% of all invasive breast cancers [101, 105, 112].

Several studies have attempted to tease out the heterogeneity of BLBCs to better understand the biology of different subgroups and identify biomarkers within this group. Lehmann and colleagues identified seven TNBC subgroups based on gene ontologies and differential gene expression profiles. These include basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory, mesenchymal, mesenchymal stem-like, luminal androgen receptor, and an unstable cluster. BL1 and BL2 were enriched with cell cycle and DNA damage response gene signatures and were highly proliferative. The mesenchymal (M) and mesenchymal stem-like (MSL) groups were enriched for epithelial-mesenchymal transition (EMT) genes, whereas the immune-modulatory (IM) subtype was characterised by immune cell signalling features. The luminal androgen receptor (LAR) subtype was ER negative, but enriched in hormone-related pathways and found to express high levels of *AR* (androgen receptor) [117]. Survival analysis showed BL1 had better prognosis than BL2, LAR had the best OS rate and M had the worst [118]. The clinical relevance of these TNBC entities was investigated by determining pathological complete response (pCR) rates after neoadjuvant therapy. BL1 had the highest pCR (52%), whereas BL2 and LAR had the lowest (0% and 10%, respectively) [118]. This re-emphasises TNBC is a heterogeneous disease with unique molecular and clinical characteristics with different prognosis.

1.4.3 Integrated molecular classification of breast cancer

Curtis *et al.* recently presented an integrated analysis by combining gene expression and DNA copy-number alterations in ~2000 breast tumours (Figure 1.7) [98]. This integrated approach identified that >39% of genes were affected by copy-number alterations. By correlating *cis*-acting copy-number alterations and gene expression, this integrative approach identified 10 breast cancer subgroups with differences in clinical outcome (IntClust 1–10) [98]. This variation observed in clinical outcome across breast cancer subtypes reinforces molecular heterogeneity, as more tumours are being included in an unsupervised analysis, more clinically relevant and biologically informed subgroupings can be made.

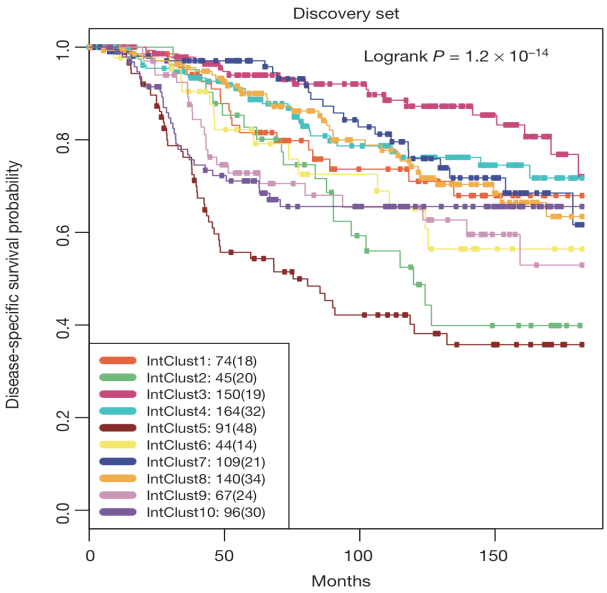


Figure 1.7: Kaplan–Meier analysis of disease-specific survival (truncated at 15 years) for the integrative subgroups of breast cancer. For each cluster, the number of samples at risk is indicated as well as the total number of deaths (in parentheses) [98]. Permission to use figures was obtained from the publisher through Copyright Clearance Center's RightsLink Service.

1.4.4 Multi-gene prognostic and predictive signatures

Many research and clinical groups have pursued the use of gene signature-based classifiers for prognostication and to facilitate treatment selection [119]. One example is **Mamma-Print®** (Agendia, Netherlands), a commercially available 70-gene test that provides prognostic information for patients with stage 1 or 2, node-negative invasive breast cancer of tumour size <5.0 cm. It provides an estimate of the risk of metastasis in early-stage breast cancer, and stratifies patients into two distinct groups (low vs high risk of distant recurrence) independently from ER status and any prior treatment [99]. The **76-gene prognostic signature** (VDX2; Veridex LLC, USA) predicts relapse in ER+ and ER- breast cancer [120]. The **Breast Cancer Index (**BCI; bioTheranostics) is based on the *HOXB13:IL17BR* gene ratio, and predicts recurrence in Tamoxifen-treated patients [121]. The **Genomic Grade Index** (GGI, 97-gene signature), based on expression of proliferation- and cell cycle-associated genes, can accurately classify grade-1 and -3 tumours, and critically, provides further stratification of grade-2 tumours into grade-1- and grade-3-like cases with corresponding clinical behaviour [122]. **Oncotype DX®** (Genomic Health, USA) [109] is a qRT-PCR assay of 21 genes associated with the ER pathway, proliferation, HER2 and invasion, and determines a risk of recurrence (ROR) score ranging from 0 to 100, which is an independent prognostic factor in Tamoxifen-treated patients with node-negative, ER+ breast cancer [109]. Finally, the **PAM50** signature (a 50-gene qRT-PCR assay) enables classification of breast tumours into intrinsic molecular subtypes [123-126]. It is commercially available from Nanostring as the **Prosigna™** assay. A ROR score is generated using an algorithm, which includes the tumour gene signature, tumour size, number of involved lymph nodes and its proliferative index. Similar to

Oncotype DX® and the BCI, the ROR score stratifies patients into low, intermediate and high-risk groups indicating the probability of disease recurrence within 10 years. The PAM50 intrinsic subtype classification currently holds superior to a six-marker immunohistochemical panel (ER, PR, HER2, Ki-67, CK5/6, EGFR; see section below) for both prognosis and the prediction of benefit from adjuvant tamoxifen [127].

1.4.5 Molecular subtyping based on surrogate IHC markers

Studies have shown IHC surrogate markers such as ER, PR, HER2, Ki67 and basal markers (CKs 5/6,14 and EGFR) can mimic molecular classification of breast cancer [85, 128-130] into different subtypes (Figure 1.8):

i) Luminal A ($HR^+/HER2^-$, Ki67 low);
ii) Luminal B ($HR^+/HER2^{-/+}$, Ki67 high);
iii) HER2 positive ($HR^-/HER2^+$);
iv) Triple Negative (TN; i.e. $ER/PR/HER2^-$); and
v) TN basal-like (i.e. TN with any positivity for CK5/6, CK14 or EGFR).

Cheang and colleagues showed that a combination of HER2 status and expression of the proliferation marker Ki67 (with a cut-off of 13.25% tumour cell positivity) could distinguish luminal A and B subtypes [84]. In this study, the luminal B (defined by gene expression analysis) and luminal $HER2^+$ (defined by IHC) breast cancers were associated with worse RFS and DFS relative to luminal A tumours. These results suggested that Ki67 labelling index may be a clinically valuable biomarker [84].

The non-luminal group can be further divided into $HER2^+$ and TN tumours based on HER2 expression status. Basal-like cancers can be identified by a lack of ER, PR and HER2 coupled with the expression of basal/myoeptihelial markers (*e.g.* EGFR, CKs 5/6 and 14) [85]. Breast cancer classification based on IHC surrogate markers has also showed a significant difference in Overall survival (OS) rates similar to gene expression studies. Luminal A subgroup had the best OS compared with Luminal B and basal-like groups, and that the HER2 group had the worst OS [129, 130]. The six-biomarker profile seems to be the most useful panel for resembling the molecular subtype of breast cancers as determined by gene expression profiling.

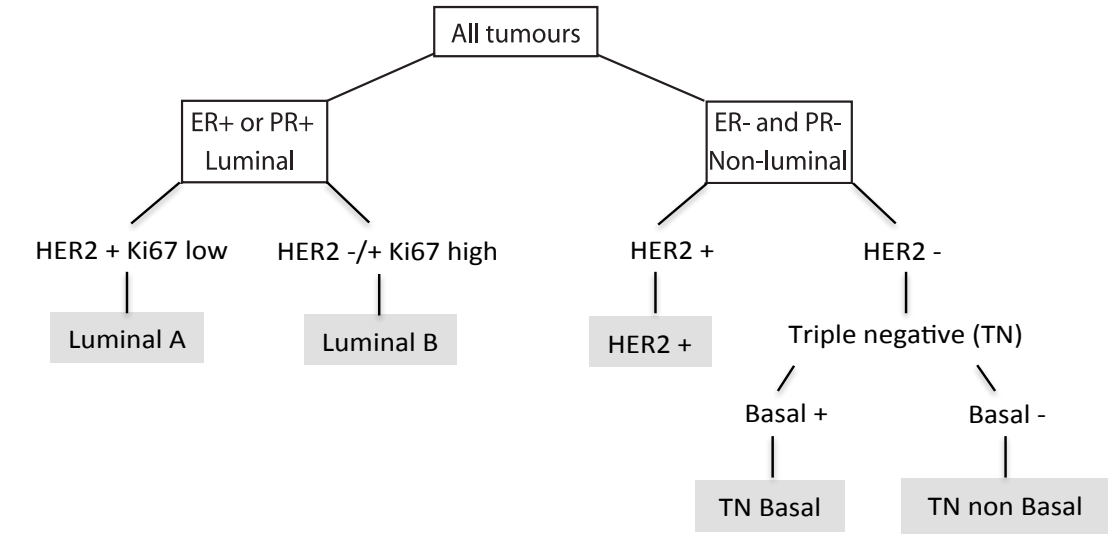


Figure 1.8: Classification of breast cancer subtype according to IHC marker panel.

Breast cancer classification based on gene expression profiling is only selectively used in the clinic because of the high cost and complexities associated with the assay and data analysis. Translation of molecular signatures into the clinic has been limited due to this reason. High resolution molecular profiling has been enormously helpful in understanding the biology of breast cancer, though it is likely that prognostication will rely on surrogate IHC markers for some time because of the technical and logistical challenges associated with gene signature analysis in a diagnostic setting.

1.5 Possible therapeutic targets in TNBC/BLBC

There is now fairly extensive evidence that inappropriate regulation of developmental genes can contribute to carcinogenesis. For example, silencing of tumour suppressor genes (*BRCA1, CDH1, HOXA5, CASP8 etc)* by hypermethylation shows epigenetic deregulation in breast cancer [131-133]. Similarly, developmental functions have been identified for many genes initially identified for their roles in cancer (*e.g.* Wnt, Notch, Pax and Hox genes [134-136]), further highlighting the relationships between these physiologic processes.

SOX (Sry-related HMG box) gene expression has been investigated comprehensively in human cancer because of its critical role in development and cell fate specification [137, 138]. Overexpression and amplification of *SOX2*, 4 and 9 have been reported in various types of cancers including breast cancer [137]. Recent studies have shown high expression of transcription factor SOX10 in a subset of TN and BLBCs [139, 140]. Supporting this, a pilot study from our research group has shown high SOX10 expression in a subset of TNBCs in a small cohort invasive breast cancers [141]. Little is known about role of SOX10 in breast cancer. However, in melanomas,

studies have shown SOX10 plays a role in tumour progression and metastasis using cell line and mouse models ([142-144] discussed in Chapter 4). Exploring the role of SOX10 protein in breast cancer cohort and inhibiting their oncogenic properties may be an approach for treatment. Similarly, **Cancer Testis** (CT) antigens are another good example given their expression is aberrantly activated in various malignancies [145]. Because of their tumour-selective expression, CT antigens have been recently exploited for therapeutic benefit (section 1.5.2).

1.5.1 SOX10

SOX family of genes function by activating or repressing transcription in a tissue-specific manner, and play an important role in cell fate specification and lineage differentiation during development [146, 147]. Key developmental processes regulated by SOX proteins include sex determination, neurogenesis, neural crest development, skeletogenesis and hematopoiesis [148, 149]. SOX genes were identified based on the presence of a homologous DNA-binding domain called the HMG-box (high mobility group) [150, 151]. So far, around 20 SOX proteins have been identified in mice and in humans, and are classified on the basis of HMG box domain similarities (SoxA to SoxE) [152, 153].

SOX10 belongs to the SoxE group (SOX8, 9 & 10) of genes [154] and is an important regulator at multiple steps of early neural crest development [155]. *SOX10* is expressed during the specification of migrating neural crest cells and its expression persists in neurons and melanocytes [155]. During melanocyte differentiation, SOX10 acts synergistically with *MITF* (Micropthalmia-associated transcription factor) and controls the expression of Dct/TRP2 (dopachrome tautomerase), an enzyme essential for melanin synthesis [156]. In the peripheral nervous system (PNS), SOX10 regulates the expression of the Schwann cell-specific myelin genes *P0* (protein zero), *MBP* (myelin basic protein), and *PLP* (proteolipid protein) [157, 158]. It also regulates expression of *connexin-32* and *connexin-47* in the PNS and is involved in the process of gliogenesis [155]. In the autonomic nervous system, SOX10 regulates the expression of *MASH1* and *PHOX2B,* two key transcription factors necessary for sympathetic nerve fate. In the enteric nervous system (ENS), SOX10 interacts with PAX3 and is involved in activation of c-Ret receptor tyrosine kinase pathway [155, 159]. It also synchronises normal ENS development in association with endothelin-3 by directly regulating its receptor, *EDNRB* (Figure 1.9) [155]. During brain development, *SOX10* not only controls glial specification and regulates the transcription of the *HER3 (ERBB3*), a receptor tyrosine kinase that is involved in activation of the *PI3K* and *MAPK* signalling pathways [160, 161].

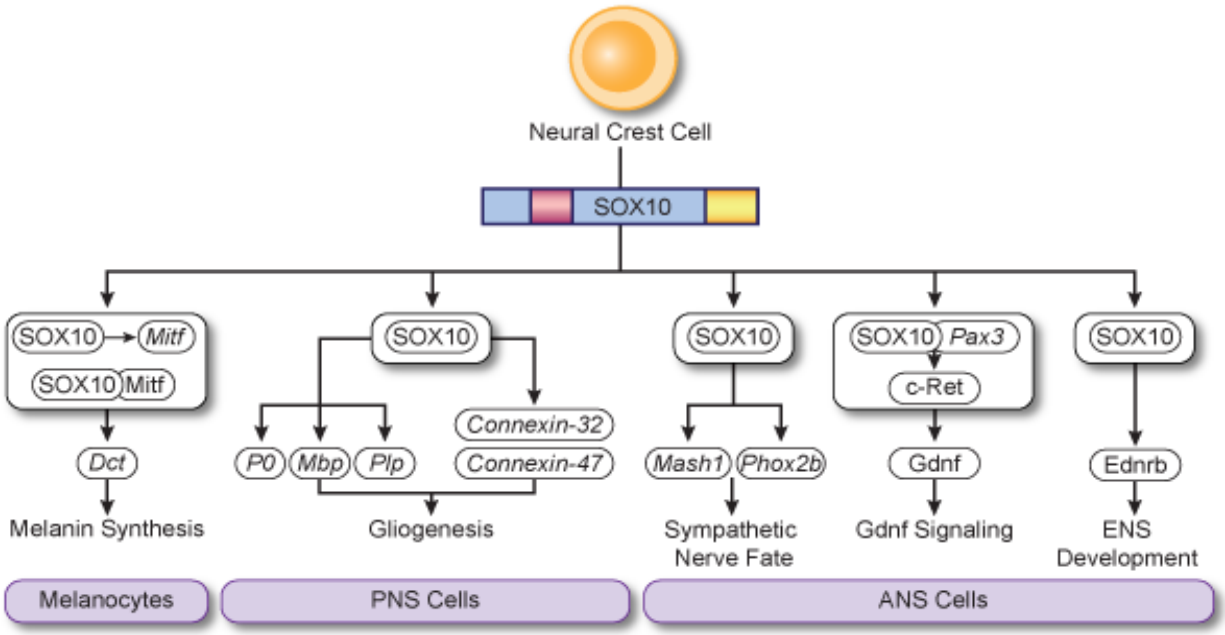


Figure 1.9: Role of SOX10 in neural crest (NC) cell development. SOX10 is first expressed during neural crest (NC) specification in premigratory, early delaminating and migrating NC cells. SOX10 is involved in the development of melanocytes. During gliogenesis, SOX10 is involved in the formation of peripheral nervous system (PNS) cells such schwann cells and glia. In autonomic nervous system (ANS), SOX10 controls other transcriptional factors (*Mash1* and *Phox2b*) that are responsible for sympathetic nerve fate and is involved in the development of enteric nervous system (ENS). Arrows indicate positive regulation by SOX10. Mitf: Micropthalmia-associated transcription factor; Dct: dopachrome tautomerase; P0: protein zero gene; Mbp: myelin basic protein; Plp: proteolipid protein; Gdnf: glial cell derived neurotrophic factor; Ednrb: endothelin receptor-β. Illustration by Diantha La Vine. MUTAGENETIX (TM), B. Beutler and colleagues, Center for the Genetics of Host Defense, UT Southwestern, Dallas, TX.

The SOX10 gene is located on chromosome 22q13.1 and spans ~15kb in length [162]. The predicted protein has 466 amino acids, contains 5 exons, and has a mass of ~50 kDa [162, 163]. Five *SOX10* splice variants have been reported so far (Ensembl, #ENSG00000100146). Two longest isoforms have identical coding sequences that specify full-length, functional protein, but differ by ~100nt within the 3' untranslated region. The other three variants are very small (124-1095nt), lack the N-terminal portion of full-length SOX10, and currently their function(s) are unknown. Full-length SOX10 has two functional domains: a high mobility group-DNA binding domain (HMG) and transactivation domain (TAD) [164]. It is an active nuclear-cytoplasmic shuttle protein, though is mostly detected in the nucleus and contains two nuclear localisation signals (NLS). It also has a functional Rev-type nuclear export signal (NES) within its DNA binding domain [164]. Mutational inactivation of NES affects shuttling of SOX10 from the nucleus to the cytoplasm and this decreases transactivation of endogenous target genes [165]. SOX10 can bind to its target DNA either as a monomer or as a dimer [166]. The dimeric binding cannot be obtained with the isolated HMG domain and requires the presence of specific 40 amino acid residues

immediately preceding the highly conserved HMG domain of SOX10 [166]. These residues cooperate with the HMG domain in a specific manner within the first two α-helices of the HMG domain and change the overall conformation of DNA-bound dimer [167]. Several SOX10 target genes have been identified, including *Mitf*, *P0*, *Cx32*, *Oct6/Scip,* and *Erbb3* that are involved in formation of Schwann cells melanocyte differentiation [161, 166, 168-170]. In addition, recent ChIP-Seq data has identified novel target genes of SOX10 activity, including *EGR2*, Sh3tc2, *Lgi4* and *Cldn19*, which are required for formation of peripheral myelin [171].

SOX10 expression has been reported in several malignancies including glial derived tumours [172], melanoma [144], digestive carcinoma [173] and breast carcinoma [139, 140]. In human gliomas, high SOX10 expression inversely correlates with tumour grade [174]. SOX10 is expressed in 87-97% of melanomas and so it has been proposed that IHC-based detection could be used in a clinical diagnostic setting [175-177]. Conversely, *SOX10* is heavily methylated and not expressed in digestive cancers [173, 178, 179]. In breast cancer, SOX10 is expressed in only a subset of TNBC and BLBCs [139-141] and its clinical implications is currently unknown.

1.5.2 Cancer-testis (CT) antigens

Cancer-testis (CT) antigens are a group of protein antigens that are predominantly expressed only in the germ cells of adult testis. These antigens are usually down regulated in somatic adult tissues but are aberrantly activated in various malignancies [145]. About 153 CT genes have been identified to date and are recorded in the CT database (www.cta.lncc.br/) [180, 181]. CT antigens encoded on chromosome X are referred to as CT-X antigens and are distinguished from those that are located on other chromosomes referred as non-CT-X antigens [145, 180, 181]. Although roles for CT antigens in transcription, translation, chromosomal recombination and signalling have been proposed, their physiological functions remain poorly understood [145, 180-182]. CT-X antigen expression varies significantly between tumour types, being more frequent in breast, melanomas, bladder, lung, ovarian, hepatocellular carcinomas ('CT-rich') and uncommon in renal, colon and gastric malignancies ('CT-poor') [182]. CT-X expression is common in high-grade, advanced stage tumours, and is associated with poorer outcome [183-187]. Given their tumour-selective expression, CT-X antigens are conceptually attractive therapeutic targets, particularly as targets for cancer vaccines [187]. A range of tumour antigen genes such as *MAGE* [188], *BAGE* [189], *GAGE1* [190], *SCP1* [191] and, *NY-ESO-1* [192] have been identified. In cancer patients, MAGE-A [193] and NY-ESO-1 were the first human CT-antigens to promote spontaneous cytotoxic T cell responses [194, 195] – they are amongst the most developed in terms of clinical translation. Recent studies have shown an association between expression of the CT-X antigens MAGE-A and NY-ESO-1, and

TN/basal-like phenotypes in breast cancer [196-200] (as well as brain metastases in a study from our group [197]). Since there are currently no molecular-targeted therapy options for TNBC, and these cancers often carry a worse prognosis, it has been proposed that CT-X immunotherapy could provide therapeutic benefit for some TNBC patients [201]. Clinical trials of vaccines targeting MAGE-A and NY-ESO-1 are underway in melanoma, and cancers of the lung and ovary [145, 202-207].

1.6 Project hypothesis and aims

TNBC is a genetically, histologically and clinically diverse disease. A subset of patients in this group has a very poor outcome, and currently the therapeutic options available are very limited. Recent studies have shown high expression of transcription factor SOX10 and CT-X antigens (MAGE-A and NY-ESO-1) in a subset of TN and BLBCs [139, 140]. Supporting this, a pilot study from our lab has shown high SOX10 expression in a subset of TNBCs in a small cohort invasive breast cancers [141]. Although published data show that SOX10, MAGE-A and NY-ESO-1 expression is associated with TNBC, their clinical implication is currently unknown. We hypothesise that SOX10 and CT-X antigens will have prognostic value in this disease, and could be viable therapeutic targets.

A study with the following aims was developed to address this hypothesis and generally improve our understanding of TNBC biology:

1. Investigate the expression and prognostic significance of SOX10 in breast cancer.
2. Investigate functions of SOX10 in the normal breast and experimental models of breast cancer.
3. Investigate the prognostic value of MAGE-A and NY-ESO-1 CT-X antigens in TNBC.

Chapter II

Materials and Methods

Chapter II: Materials and Methods

2.1 Clinical samples and ethics

This study included five clinical sample cohorts:

i. The *Queensland Follow-Up (QFU) cohort*: formalin-fixed, paraffin embedded (FFPE) breast tumour samples from 449 patients who underwent surgical breast cancer resection at the Royal Brisbane Women's Hospital (RBWH, Herston, Queensland) between 1987 and 1994. Familial breast cancer was not consistently identified and recorded at this time in Queensland, and thus we estimate the cohort comprises ~90% sporadic and ~10% of cases associated with inherited susceptibility [208, 209].
ii. A cohort of 105 FFPE *metaplastic carcinomas* (whole sections) was obtained from Pathology Queensland.
iii. FFPE TMA sections containing *BRCA1* (*n*=139), *BRCA2* (*n*=114) and *BRCAX* (*n*=353) cases, obtained from kConFab (Kathleen Cuningham Foundation Consortium for research into Familial Breast cancer). Mutation status was confirmed by denaturing high performance liquid chromatography (dHPLC) and sequencing [210]. The criteria for classifying all *BRCA1* and *BRCA2* variants by diagnostic or research laboratories are posted on kConFab's website (http://www.kconfab.org/progress/mutations.asp).
iv. A cohort of 18 FFPE reduction mammoplasty samples, obtained through collaboration with Dr William Cockburn, his team at the Wesley Tissue Bank (Wesley Hospital, Auchenflower, Queensland). Age, parity and menopausal status of these patients were unknown. Thirty-precent cases showed fibrocystic change and 10% presented with columnar cell lesions.
v. A cohort of 32 FFPE tumour-associated normal samples was obtained from patients who were part of the QFU cohort.

The use of clinical samples in this study was carried out in accordance with our governing human research ethics committees at the RBWH (2005/022; last amendment 14/5/14) and the University of Queensland (2005000785, last amendment 28/05/15, expires 31/12/16).

2.2 Histology review and clinical follow-up

2.2.1 QFU and metaplastic cohort

Four micron thick sections were cut from each tumour block and stained with haematoxylin and eosin (H&E) using an automatic staining machine (Tissue Tek DRS, Sakura). H&E slides for all donor tumour blocks were reviewed and tumour areas were annotated by a surgical pathologist. Clinical data from patient such as date of birth, age, date of diagnosis, death data, date of last follow-up and cause of death, histological grade, tumour size, presence or absence of axillary lymph

node, number of positive and negative nodes *etc* were obtained from the pathology reports (Dept. of Pathology, Royal Brisbane Women's Hospital (RBWH), Herston, Queensland), Queensland Cancer Registry (QCR), Auslab and Paris reports. Internal review of clinical features such as histological type, grade, invasive type, lymphovascular invasion, lymphatic infiltration, tumour border etc. was performed by Prof. Sunil Lakhani as described in the formal pathology report (Appendix 2.1). A detailed database containing biomarker status (used clinically and conceptually), histopathologic and clinical variables were built using information collected from different sources.

2.2.2 kConFab cohort

Date of birth, age, date of diagnosis, death data, date of last follow-up and cause of death for *BRCA1, 2* and *X* cases was obtained from original pathology reports (Cancer Council, Victoria). Hormone receptor, HER2 and basal marker (CK5/6, 14 and EGFR) status for all cases were obtained from kConFab and their collaborators. A database containing date of birth, date of diagnosis, death date and biomarker status were built using information collected from kConFab and their collaborators.

2.3 Construction of tissue microarrays

For the QFU cohort, tissue microarrays (TMAs) were constructed according to pre-designed maps (Appendix 2.2) using a tissue arrayer (model MTA-I, Beecher Instruments Inc., Sun Prairie, WI, USA). An H&E-stained section was selected from each case, and tumour-rich regions were identified by histopathological review, preferentially on the periphery of the tumour where the tissue tends to be best preserved and to avoid necrosis, selecting areas most representative of overall tumour histopathology. Two 0.6 mm biopsy cores per sample were arrayed into recipient paraffin blocks at 1 mm grid spacing. All arrays were built in duplicate. Unmatched, histologically normal tissue from breast, liver, testis, and pancreas were used as controls and for map orientation. After construction, blocks were heated to 55 °C for 10 min, then cooled to room temperature to solidify the donor tissue cores within the recipient block. Arrays were sectioned at 4 μm onto slides (SUPERFROST PLUS) for downstream analyses.

2.4 Immunohistochemistry (IHC)

2.4.1 Sectioning, deparaffinisation and antigen retrieval

Tissue sections were incubated at 37 °C overnight to facilitate full adhesion to glass slides, and were subsequently deparaffinised to expose the tissues for immunostaining: xylene (three x 3 min), 100% alcohol (two x 2 min), 90 % alcohol for 2 min, 70 % alcohol for 2 min and MilliQ water (two x 2 min). Heat induced epitope retrieval (HIER), also called as antigen retrieval was selected based

on optimisation tests for individual antibodies (Table 2). Antigen retrieval causes reversal or unfolding of proteins within the tissue exposing the epitope, enabling antigen binding. Tissue slides were put into a slide rack in a container with either of the following antigen retrieval buffers:

- Sodium citrate: 1 in 10 dilution of 10X stock solution of 0.01 M sodium citrate (pH 6.0).
- EDTA: 1 in 10 dilution of 10X stock solution of 0.001M EDTA (pH 8.8).
- Chymotrypsin (1% in 10mM HCl)

Antigen retrieval was performed in a decloaker at 125 °C for 5 min (sodium citrate) or 105 °C for 15 min (EDTA), and then 90 °C for 10 sec. After this, the slides were left to cool for 20 min at room temperature (RT) and washed with TBS buffer.

Washes: All washes (3 x 2 min, each) were performed using 1X TBS (Tris Buffered Saline: NaCl, Tris, HCL; 1.0N; ph 7.4). One litre of 10X TBS was prepared mixing 24 g Tris base (50mM), 88 g NaCl (150mM) and 900 ml of distilled water. To prepare a 1X solution, 10X TBS was diluted in distilled water.

2.4.2 Immunostaining and chromogenic detection

Staining was performed according to the manufacturer's instructions using MACH 1 Universal HRP-Polymer detection kit (Biocare Medical, Cat # M1U539 G, L10). The slides were incubated with 30% hydrogen peroxide solution for 10 min in order to block the endogenous peroxidase activity, followed by TBS wash. The non-specific background staining was blocked using a reagent called "background sniper" for 15 min at room temperature. This was followed by incubation with the primary antibody for 60 min. After removal of the primary antibody with TBS washes (3 x 2 min), incubation with the appropriate secondary antibodies (HRP-Polymer) was performed for 30 to 45 min at room temperature.

MACH 1 Universal HRP-Polymer Detection kit (Biocare Medical, Cat # M1U539 G, L10). It was performed as per the manufacture's instructions using reagents supplied in the kit, except for the primary antibody and the TBS washes. However, peroxidase blockage was performed using the reagent "Background Sniper" for 15 min at RT. This was followed by incubation with the primary antibody diluted in TBS buffer. 1-200 µL of diluted antibody was added to each slide depending on tissue section sizes. The dilutions and incubation times were pre-optimised using appropriate control tissues (Table 2.1). After incubation with the primary antibody, tissue sections were washed with TBS (three x 2 min) and the secondary antibody (HRP-Polymer) and incubated for 30 min at room temperature. The poly-HRP is used to remove unconjugated streptavidin molecules that reduce signal intensity by competing for binding sites with horseradish peroxidase (HRP)-

conjugates molecules. This allows the detection of low levels of the target molecule by decreasing the background staining. Finally the slides were washed (TBS: three x 2 min) followed by chromogenic staining.

DAB chromogenic staining was prepared using VECTASTAIN ABC or MACH 1 Universal HRP-Polymer Detection kit. Depending on the size of the tissue section, 100-200 μL of this working solution was added to each section and incubated for 10 min. Tissue slides were then rinsed with deionised water and counterstained using an automatic staining machine (Tissue Tek DRS, Sakura), based on a very light counterstaining program as follows: sections were incubated with haematoxylin for 30 s followed by washes in tap water (two x 3 min) and then 90% alcohol for 1 min, 100% alcohol (two x 2 min) and Xylene (two x 2 min). The slides were then coverslipped using an automatic glass coverslipper (LJUNG TECH by INSTRUMEC, GCS 600) with a standard water-based mounting media.

Table 2.1: Antibodies, staining methods and positivity thresholds used for immunohistochemistry in this study.

Antigen	Ab clone	Catalogue no.	Sp.	Supplier	Dilution	Ag Ret.	Cellular Localisation of staining scored?	Positivity threshold or parameters scored
ER	6F11	NCL-L-ER-6F11	M	Novocastra	1/100	C	Nuc	> 1%
PR	1A6	NCL-L-PGR-312	M	Novocastra	1/200	C	Nuc	> 1%
HER2	CB11	K5204	R	Dako	1/200	C	Memb	3+ by IHC and >6 by CISH
CK5/6	D5/16B4	MAB-1620	M	Chemicon	1/400	C	Memb, cyto	Any positivity
CK14	LL002	NCL-LL002	M	Novocastra	1/40	C	Memb, cyto	Any positivity
EGFR	31G7	280005	M	Invitrogen	1/100	E	Memb	Any positivity
Ki-67	MIB-1	M7240	M	Dako	1/200	C	Nuc	> 20%
AR	AR441	M3562	M	Dako	1/50	C	Nuc	>1 %
c-kit	-	A4502	R	Dako	1/800	E	Memb	Any positivity
SOX10	sc-17342	N-20	G	Santa Cruz	1/100	C	Nuc	% and intensity (1-3+)
p53	DO7	M7001	M	Dako	1/200	C	Nuc	% and intensity (1-3+)
Vimentin	V9	M0725	M	Dako	1/400	C	Nuc	Any positivity
MAGE-A	6C1	sc-20034	M	Santa Cruz	1/500	C	Nuc, cyto	% and intensity (1-3+)
NY-ESO-1	E978	sc-53869	M	Santa Cruz	1/30	E	Nuc, cyto	% and intensity (1-3+)

Abbreviations: Ab- antibody; Sp- species, M-mouse, R-rabbit, G-goat; .Ag Ret- antigen retrieval, C-citrate, E-EDTA; Memb, plasma membrane, Nuc, nucleus, Cyto- cytoplasm.

2.5 IHC biomarker scoring

Stained slides were scanned using the Aperio ScanScope T2 digital scanning system at 20x magnification. Images were subjected to array image segmentation using the TMA module of Spectrum software (Aperio). Cores were extracted as high-resolution JPEG images with file names

used as metadata to link them to corresponding positions in the array. Image files were then imported to iPhoto (iPhoto11, v. 9.2.1), and scored by two observers independently.

2.6 Chromogenic *in situ* hybridisation (CISH)

HER2 typing on QFU cohort was obtained from Pathology Queensland, RBWH laboratory and scored by Prof. Sunil Lakhani, according to clinical diagnostic criteria (CISH score ≥ 6).

2.7 Kaplan Meier survival analysis and chi square tests for association

Statistical analysis was performed using PRISM Software (version 6.0c). The association of SOX10 expression and different clinicopathologic parameters was evaluated using chi square and Fisher's exact tests. Survival associations were identified by Kaplan-Meier analysis using log-rank test to assess significance. Breast cancer specific survival (BCSS) was defined as the time from primary breast cancer diagnosis to death due to breast cancer. For cases with locally recurrent disease (second/multiple primary breast cancers) the first event was used. Death due to other causes was censored. *P*-values of < 0.05 were considered significant. Multivariate analysis was performed by Dr Jodi Saunus using MelCalc software (v12.7).

2.8 Immunofluorescence (IF) on FFPE sections

Immunofluorescence was performed for SOX10 and CK8/18 using FFPE sections of normal breast. Four μm tissue sections were cut from paraffin blocks using a microtome (Lieca RM2135). The slides were placed at 37 °C overnight followed by deparaffinisation as described above. Citrate buffer was used for antigen retrieval and the sections were processed as for IHC (section 2.4.2). Anti-SOX10 (goat polyclonal) and CK8/18 (mouse monoclonal) primary antibodies (Santa Cruz Biotechnology, Cat # sc-68853 and Novocastra, Cat # NCL-L- 5D3) diluted to 1:50 and 1:100 in TBS, respectively, were incubated on tissue sections for 1 h at room temperature. Slides were then washed in TBS buffer (three x 2 min) followed by incubation with secondary antibodies: rabbit anti-goat IgG (Alexa Fluor 594, Invitrogen, Cat #A11008; 1:500 in TBS) and goat anti-mouse IgG1 (Alexa Fluor 594, Invitrogen, Cat #A21125; 1:400 in TBS) for 30 min in the dark. The slides were then stained with SUDAN Black dye for 20 min in dark (Sigma, Cat # S-2380) to block auto-fluorescence, washed with TBS-Tween (0.1%) for 30 min and washed again with TBS for 10 min at room temperature. After washing, slides were mounted using Vectashield (Vecta Laboratories, H1000) with DAPI (Sigma, Cat # D9542), coverslipped and sealed using nail polish, and stored in the dark at 4 °C. Slides were imaged with a Carl Zeiss MicroImaging system and analysed using Axio Vision LE version 4.8.2 (PerkinElmer, Glen Waverley, VIC, Australia).

2.9 Cell culture

Breast cancer cell lines used for screening experiments were cultured according to the recommendations from American Type Culture Collection (ATCC) and authenticated by STR profiling (*GenePrint* 10 System, Promega Cat # B9510); Queensland Institute of Medical Research (QIMR) Berghofer cell line authentication service. The list of cell lines used in this study and their intrinsic subtypes are presented in the table 2.2.

Table 2.2: List of cells lines used in this study classified according to their intrinsic subtypes.

#	Cell line	Intrinsic subtype	#	Cell line	Intrinsic subtype
1	Bre80 Tert	B-UC	21	184B5	CL-UC
2	MDA-MB-468	Basal A	22	T47D	Luminal
3	PMC 42ET	Basal A	23	MCF7	Luminal
4	HCC70	Basal A	24	ZR751	Luminal
5	HCC1937	Basal A	25	MDA-175-VII	Luminal
6	SVCT	Basal A	26	KPL-1	Luminal
7	HCC1954	Basal A	27	UACC812	Luminal
8	HCC1143	Basal A	28	MDA-MB-330	Luminal
9	BT-20	Basal A	29	SKBR3	Luminal ($HER2^+$)
10	8701BC	BB	30	BT474	Luminal ($HER2^+$)
11	SUM159	Basal B/Claudin low	31	MDA-MB-453	Luminal ($HER2^+$)
12	MDA-MB-231	Basal B/Claudin low	32	NIH 3T3	Non breast
13	MDA-MB-436	Basal B/Claudin low	33	HaCaT	Non breast
14	Hs578T	Basal B/Claudin low	34	293 T	Non breast
15	HBL100	Basal B/Claudin low	35	HeLa	Non breast
16	BT549	Basal B/Claudin low	36	A549	Non breast
17	184A1	Basal B/Claudin low	37	D22	Melanoma
18	MCF10A	Basal B/Claudin low	38	D05	Melanoma
19	RHB1	Basal B/Claudin low	39	D41	Melanoma
20	MDA-MB-435	Basal B/Claudin low			

2.10 Generation of MDA-MB-435 cells with stable, shRNA-mediated knock-down of SOX10

2.10.1 Transformation of pLKO.1 into competent bacterial host cells

The pLKO.1 is a transfer vector backbone suitable for cloning and expressing shRNA sequence. pLKO.1 can be introduced into cells via direct transfection, or can be converted into lentiviral particles for successive infection of a target cell line. pLKO.1 vector has a puromycin resistance marker encoded within that allows convenient stable selection. For transformation, LB media was prepared by adding 10g Trypton, 10g NaCl and 5g of yeast to 1000 mL distilled water. LB agar was prepared by adding 15 g agar to 1000 mL LB media. LB agar (1.5%) plates were prepared by pouring preheated LB broth containing Ampicillin onto culture plates. Round bottom BD falcon polypropylene tube (14 mL) was pre-chilled on ice. Fifty µL XL1-Blue competent *E. Coli* cells

were aliquotted into the falcon tube, to which 50 ng pLKO.1 plasmid DNA was added, mixed gently and incubated on ice for 20 min. The tubes were subjected to heat-pulse in a water bath at 42 °C for 45 s and transferred on ice for 2 min. Approximately 200 µL of transformation mixture was plated onto LB agar plates and incubated at 37 °C, overnight.

2.10.2 Plasmid preparation

A homogeneous starter culture of 5mL LB medium was prepared by inoculating a single colony from a freshly streaked plate. The mixture was incubate at 37 °C for approximately 8 h while shaking at 250–300 rpm. Fifty mL culture was prepared from the starter culture and incubated at 37 °C overnight. Cells were harvested from an overnight culture by centrifuging the culture media at 5000 x *g* for 10 min. Plasmid isolation was performed using the GenElute HP Plasmid Midiprep Kit (Cat #NA0200, Sigma) according to the manufacturers instructions. The concentration of DNA was estimated by measuring absorbance (optical density) at 260 nm (A_{260}) using NanoDrop 1000 spectrophotometer. The read outs are based on linear relationship between absorbance and concentration of an absorbing species (Beer-Lambert's Law) using the formula, Concentration (µg/ml) = (A_{260} reading-A_{320} reading) x dilution factor x 50µg/ml. The ratio between 260 and 280 nm was used to assess purity of DNA. An average ratio of 1.8-2.0 was considered good quality DNA.

2.10.3 shRNA retroviral transduction

A set of pre-validated pLKO.1 (puromycin-resistance) constructs encoding three different *SOX10*-targeted shRNAs, and a non-target shRNA control pLKO.1-puro construct (NTNC; encodes a hairpin that does not target any known human gene) were purchased from Sigma as glycerol stocks of host *E. Coli* suspensions (TRCN0000018984, TRCN0000018987, TRCN0000018988 and SHC002, respectively). Plasmid DNA was isolated from overnight bacterial cultures as described above (section 2.10.2).

Lentiviral particles encoding each of the shRNAs were produced by triple transient transfection of HEK-293T (human embryonic kidney) packaging cells with one of the four transfer plasmids (pLKO.1-puro; 2 µg), together with two companion plasmids encoding lentiviral packaging and replication elements (2 µg pHR'8.2ΔR and 0.25 µg pCMV-VSV-G, which encode viral packaging and envelope elements; donated by Dr Wei Shi, QIMR Berghofer) in OptiMEM media (Cat #31985070, Invitrogen). Packaging cells were transfected at ~60% confluence in 100 mm dishes according to the manufacturer's instructions. Following completion of transfection the next morning, media was replaced with 5 mL target cell (MDA-MB-435) media that evening for overnight viral production. Viral supernatants were collected on the following two days, and filtered

(0.45 µm 25 mm filters, Acrodisc Syringe Filters, Cat #4614, Pall Corporation) before target cell transduction to ensure no cross-contamination with packaging cells. MDA-MB-435 target cells were seeded at $3.1\times10^4/cm^2$ in 6-well plates, then after 24-48 h (at ~50% confluence), cells were infected with filtered viral supernatants, supplemented with 1 mg/mL polybrene (Cat #H9268, Sigma) for 24 h. Stably transduced cells were then selected with 1 µg/mL puromycin (Cat #P9620, Sigma) for two weeks to eliminate uninfected cells. SOX10 protein levels were then investigated by Western analysis.

2.11 Western analysis

2.11.1 Sample preparation and quantification

Cells were briefly washed in PBS and the pellet was resuspended in 150 µl RIPA buffer: 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% deoxycholic acid, and fresh protease inhibitors (1 mM PMSF (phenyl-methylsulfonylfluoride) with protease inhibitor cocktail (Thermo Scientific, Cat #78440; 10 µg/ml leupeptin, 10 µg/ml aprotinin, and 1 µM pepstatin)). After incubation at 4 °C for 30 min, the lysates were subjected to sonication for 10 s (10 times) and centrifuged at 13000 g for 30 min at 4 °C. Protein concentration was determined using Bradford Protein Assay (Biorad) using Bovine Serum Albumin concentration standards (BSA; Sigma: 0, 50, 100, 200 and 500 µg/mL). Equal volumes of protein lysates (diluted in PBS to within the quantitative range of the assay) were analysed in 96-well plates in triplicate. The protein assay reagent was diluted 1:5 and then mixed with the standards and the samples and incubated for 10 min at RT, before optical density measurement (595 nm). Protein concentration was determined by fitting the absorbance data to the standard curve obtained from the BSA samples (Microsoft Excel).

2.11.2 Western transfer and immunoblotting

Protein lysates (30 mg) were denatured in sample buffer (cat #161-0747, BIO-RAD) with reducing agent (2-mercaptoethanol, cat #21985-023, Thermo Scientific) for 5 min at 95 °C, and subjected to Sodium Dodecyl Sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) on gradient gels (4-15%; cat # NP0322BOX, Life Technologies). The separated proteins were transferred to PVDF membrane (Cat #IPVH00010, Merck-Millipore) for 60 min at 4 °C. The membrane was air-dried and then non-specific binding sites were blocked using 5% milk in TBS-T (20 mM Tris•HCl, 137 mM NaCl and 0.05% Tween 20, pH 7.4) for 60 min. After blocking, membranes were incubated with primary antibodies (SOX10, Santa-Cruz Cat # sc-17342, 1:500 dilution; β-actin, Cell Signalling, cat # 8H10D10, 1:5000 dilution) in TBS with 1% milk overnight at 4 °C. The blots were then incubated with an appropriate secondary antibody (HRP-conjugated anti-goat or anti-mouse) at 1:5000 in TBS-T for 60 min at room temperature. Bound antibodies were detected by chemiluminescence (ECL, Cat #34087, Thermo Scientific), and signals developed with a Konica

Minolta SRX-101A tabletop processor.

2.12 *In vitro* tumourigenicity assays

2.12.1 Proliferation assay

Cells were seeded at 5×10^4 cells/well density in a 6-well plate. The IncuCyte real-time imaging system (Essen Bioscience) was used to record proliferation under normal conditions every 6 h over a period of 5 d. Measurement of cell growth is based on area or confluence. Three technical replicates were assayed for each sample and two biological replicates were performed on different days.

2.12.2 Scratch wound assay

Cells were seeded in triplicate (3×10^4/well) in 96-well Essen ImageLock plates and cultured in regular conditions overnight. Monolayers were then scratched using a 96-pin WoundMaker (IncuCyte, Essen Bioscience) to create precise and reproducible wounds in all wells simultaneously. IncuCyte image analysis software was set to scan the experiment every 2 h for 72 h. Three technical replicates were assayed for each sample and two biological replicates were performed on different days.

2.12.3 Colony formation assay

Cells were seeded at low density (500 cells/well) in a 6-well plate in triplicate and maintained in complete media (DMEM supplemented with 10% foetal bovine serum and 1% penicillin/streptomycin) for 14 d at 37 °C in a standard CO_2 incubator. The colonies were fixed in methanol for 15 min and stained with crystal violet (0.01% v/w) for 20 min. Colonies (size threshold of >50 cells) were counted by microscopy. Three technical replicates were assayed for each sample and four biological replicates were performed on different days.

2.12.4 Tumoursphere formation assay

Cells were seeded at very low densities (500 cells/well) in a 6-well plate in triplicate and maintained in NSA media [DMEM F12 containing recombinant human epidermal growth factor (EGF; Sigma; 20 ng/mL), recombinant human basic fibroblast growth factor (bFGF; R&D Systems; 10 ng/mL), heparin (Sigma; 4mg/mL), human or mouse proliferation supplement (NeuroCultH; Stem Cell Technologies; 10%), bovine serum albumin (BSA; Sigma; 0.15%), and penicillin G-streptomycin solution (Gibco; 1%)] for 7 d at 37 °C in a standard CO_2 incubator. Tumourspheres (discriminated from cell clumps using a size threshold of >50 cells) were counted manually (microscopic examination with the 10 x objective). Three technical replicates were assayed for each sample and two biological replicates were performed on different days.

2.13 *In vivo* xenograft experiments

2.13.1 Animal Housing and ethics

Four-week-old female NOD/SCID mice were purchased from Animal Recourse Centre (ARC). Mice were housed at the QIMR Berghofer Animal Facility in the optiMICE® caging system (Centennial, Colorado) on a 12 h light/dark cycle at 25 °C. Animals were handled under sterilised airflow conditions. Animal work was approved by the QIMR Berghofer Animal Ethics Committee and performed by Mariska Miranda and Fares Al-Ejeh in accordance with their guidelines.

2.13.2 Mammary fat pad xenografts

NOD/SCID mice were allowed to acclimatise to their surroundings for 7 days then used at 5 weeks of age for xenograft experiments. Mice were given anaesthesia before injecting the cells. For injections, $5x10^6$ MDA-MB-435 cells were resuspended in PBS with 50% matrigel (BD, Biosciences, Bedford, USA) to a final volume of 50 uL and injected into the right 4th mammary fat pad. Tumour growth was measured twice weekly for a period of six weeks by experienced experimentalists (MM/FA-E) using manual calliper measurement. To calculate tumour volume the following formula was used: tumour volume = (B/2)*S2 where B = largest tumour measurement and S = the smallest, based on two-dimensional calliper measurements. Statistical analysis for *in vitro* and *in vivo* functional assays was performed using PRISM Software (version 6.0c) p value $\leq$ 0.05 considered statistically significant. Mice were euthanased by cervical dislocation when the tumours reached $600mm^3$ in size. Tumours were fixed in paraformaldehyde (PFA) for 24 h, washed with 70% ethanol and stored in ethanol at 4 °C. Five mice per group were used and the experiment was performed only once due to time constrain. This experiment will is currently being repeated at QIMR Berghofer Animal Facility.

2.14 *In-silico* analysis of published datasets

2.14.1 Analysis of *SOX10* expression, methylation and copy-number in published datasets

Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) gene expression and copy-number data were obtained from the European Genome-Phenome Archive (n=1992; accession number EGAS00000000083). This included 209 cases classified as basal-like, 218 as claudin-low, 224 as HER2, 700 as Luminal A, 475 as Luminal B and 148 as normal-like (according to the PAM50 classifier [98]). A second, independent *SOX10* dataset, comprising RNA-sequencing (Illumina), copy-number (Affymetrix SNP 6.0 array) and methylation (Illumina 450k) was extracted from The Cancer Genome Atlas (TCGA) provisional (as they are more updated compared to published) breast cancer dataset (1104 tumours) using the cBio Cancer Genomics Portal [97].

Methylation value included was the average of probes across the gene and not just main promoter data. PAM50 classification results of all samples were obtained from the TCGA group, distributed as follows: Basal-like (n=187), Her2 (n=80), Luminal A (n=549) Luminal B (n=213), and Normal-like (n=38).

In order to understand the impact of *SOX10* methylation on its expression, we performed spearman correlation analysis of log^2-transformed *SOX10* expression data and methylation data (n=737) in basal-like versus other breast cancers. A non-parametric test was used because the datasets were found to be non-normally distributed using the D-Agostino/Pearson omnibus and Shapiro-Wilk normality tests (GraphPad Prism, v6.0f). SOX10 is highly and consistently expressed in melanoma, and was hypomethylated compared to breast cancer. We applied a 90^{th} percentile methylation threshold that captured the majority of the low-expression melanoma cases (11/13), and found this also segregated BLBC from non-BLBC. This threshold value (0.68) captured ~75% of melanoma cases with bottom 5^{th} percentile expression. Melanoma data derived from TCGA (n=~450) was used for comparison as good model because *SOX10* is over-expressed in ~95% cases. Similarly, in order to understand whether SOX10 expression is regulated in part by copy-number, spearman correlation analysis of log2 transformed RNA-seq mRNA expression and log ratio (cbs) smoothed copy-number data for SOX10 was performed. We used spearmen correlation analysis as cbs smoothed copy-number data was not normally distributed. "R" statistical software was used for all statistical tests [211].

2.14.2 Normal breast epithelial compartment analysis

To investigate SOX10 mRNA expression in the normal breast, we made use of publicly available gene expression array profiling data derived from fluorescence-activated cell sorting (FACS)-based enrichment of particular epithelial compartments: mature luminal cells, luminal progenitor cells, myoepithelial/mammary stem cells (MaSC) and a stromal compartment [24]. Gene expression array data were extracted from the Gene Expression Omnibus (GEO: GSE16997), quartile-normalised and assessed for quality by Mr Samir Lal using arrayQualityMetrics tool where individual array quality was determined by MA plot, spatial effects, density plots and heatmaps (as described in [212]). Probes were mapped using the illuminaHumanv3.db annotation package. For cases where multiple probes mapped to a single gene, probes with the highest expression across all samples were used. Analysis of variance (ANOVA) tests were conducted using R statistical software [211] to investigate expression *SOX10* and a panel of genes associated with mammary epithelial differentiation (*CK14*, *CK5*, *KIT*, *ELF5* and *CK18*) between the compartments. $p<0.05$ was considered significant.

2.14.3 Gene set enrichment analysis (GSEA)

Gene signature list from Lim *et al.*[24] (section 2.15.2) was divided into positive and negative fold-change subsets. mRNA expression data for luminal progenitor genes (RNA-seq) from the TCGA cohort was extracted from cBioportal. Data were transformed using voom tool (which transforms RNA-seq data to micro array format) prior to GSEA analysis [213]. For mRNA expression data, the relative expression of an individual gene in a reference population (both normal and tumour) is computed. The reference population is all samples that are diploid for the gene (by default for mRNA), or normal samples (when specified). The returned value from cBioportal indicates the number of standard deviations away from the mean of expression in the reference population (z-score). This measure is useful to determine whether a gene is up- or down-regulated relative to the normal samples or all other tumour samples. *SOX10* RNAseq z-scores were extracted from the cBioportal (TCGA provisional dataset; n=1104 cases). The cohort was separated into 'SOX10-low' and 'SOX10-high' groups using z-score cut-off of 2 as it separated the two groups. A total of 1000 permutations based on the SOX10-high *vs* -low phenotype were used to calculate a false discovery rate using multiple hypothesis testing. GSEA results were shown using a normalised enrichment score (NES), which is the value assigned to each gene-set after normalisation across all analysed gene sets. The NES was determined as: actual enrichment score (ES)/mean (ES against all permutations in the dataset). $p<0.05$ was considered significant.

2.15 SOX10 expression analyses

2.15.1 RNA extraction and quantification

Isolation of total RNA from cell pellets was performed using TRIzol (Invitrogen) according to the manufacturer's protocol. Briefly, cells were collected after trypsinisation and centrifuged at 1000 rpm for 10 min. 0.2 mL chloroform/1 mL TRIzol was added to the cell pellets for homogenisation. The samples were vigorously shaken for 15 s, incubated at RT for 3 min and then centrifuged at 12,000×*g* for 15 min at 4 °C. The mixture was separated into three phases: upper aqueous, interphase and a lower phenol-chloroform phase. The aqueous phase (containing RNA) was removed and transferred to a new tube. For RNA precipitation, 0.5 mL of 100% isopropanol was added to the tube; the samples were incubated at room temperature (RT) for 10 min and then centrifuged at 12,000×g for 10 min at 4 °C. RNA pellets were washed with 1 mL of 75% ethanol and centrifuged at 7500×g for 5 min at 4 °C, then air-dried for 30 min at RT and dissolved in 20 µL nuclease-free water. RNA samples were qualified and quantified by qubit/bioanalyser. RNA integrity numbers were found to be between 9.4 and 10.

2.15.2 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

SuperScript III Reverse Transcriptase (Invitrogen) was used for reverse transcription of 1μg RNA from each sample, according to the manufacturer's instructions. Briefly, 1 μL (50 mM) oligo (dT), 1 μL (10 mM) dNTP mix and 1 μg of RNA was added to nuclease-free water to make up the total volume 13 μL. The mix was heated to 65°C for 5 min and then incubated on ice for 1 min. 4 μL 5X First- Strand Buffer, 1 μL 0.1 M DTT, 1 μL SuperScript III RT and 1 μL RNaseOUT Recombinant RNase Inhibitor was added to each tube and the mixtures were incubated at 25 °C for 5 min followed by 50°C for 1 h. The reaction was inactivated by heating at 70 °C for 15 min.

Primers were designed using pBLAST (http://www.ncbi.nlm.nih.gov/tools/primerblast/index). (SOX10- CTTCATGGTGTGGGCTCAGG and TCACTTTCGTTCAGCAGCCT) qRT-PCR was performed on a StepOnePlus instrument (Applied Biosystems). The PCR assay was performed in a final volume of 20 μL containing 1x SYBR® Green SuperMix (Invitrogen, Cat #11733-038) , 2 μL of cDNA samples and 10 pmol each of forward and reverse primers. PCR cycles (40 total) consisted of 15 s denaturation at 95 °C and a combined annealing/extension step for 30 s at 60 °C. Melt curve analysis was applied to validate that primers produced single PCR products. Hypoxanthine phosphoribosyltransferase1 (*HPRT1*) was amplified as a loading control, and the fold change in expression of each target mRNA relative to *HPRT1* was calculated using the comparative ΔΔCt method [214].

2.15.3 Differential gene expression analysis of SOX10-depleted MDA-MB-435 cells

Whole genome gene expression array analysis of non-target shRNA (NTC)- and shSOX10-infected MDA-MB-435 cells (n=3 independently infected, pooled clones each) was performed on the Illumina HT-12 v4 platform and analysed using the beadarray package [215] Arrays were run at the Ramaciotti Centre for Genomics (University of New South Wales) and the differential gene expression analyses were performed in collaboration with Mr Samir Lal. BASH [216] program was used to correct for any defects (spatial artefacts) on the array. Briefly, the bead-level data were checked for quality (including log transformation) and only probes with annotations were selected for further analysis. arrayQualityMetrics [212] was used for array comparisons to identify potential outliers across all arrays in the dataset by visual inspection of the arrays through sample correlation heatmaps and boxplots. Normalised gene expression data was then used for multidimensional scaling (MDS) which highlighted differences in expression profiles between samples (NTC and shSOX10) and differential expression analysis was conducted using the 'limma' package [217]. Differential gene expression analysis was performed on multiple genes (>10,000), which increases our chances of false positivity therefore, p-values were adjusted (as $p<0.01$ was not significant) using the Benjamini-Hotchberg method as it controls for false discovery rate.

Chapter III

SOX10 as a prognostic marker in triple-negative and basal-like breast cancer

Chapter III: SOX10 as a prognostic marker in triple-negative and basal-like breast cancer

3.1 Background

Triple negative breast cancers (TNBCs) are a diverse and a heterogeneous group of tumours that are characterised by lack of hormone receptors (oestrogen and progesterone) and amplification of HER2 growth factor [218]. TNBC are typically high grade, aggressive tumours, seen in young women and most harbour a basal-like phenotype [77, 219]. Gene expression studies have recently identified six distinct TNBC subtypes, each of which displays a unique biology [117]. The prognosis for patients diagnosed with TNBC is variable. A subset has very poor outcome, with higher risk of relapse including development of lung and brain metastases [220, 221]. Current clinicopathologic markers, including grade, tumour size and lymph node status, do not fully explain this clinical behaviour, as TNBC subtype is already enriched for high-grade tumours and are lymph node negative. This separation of TNBCs in overall breast cancer-related survival is particularly marked for the 'basal-like' subset of TNBCs (TN-BLBC) [94], where the survival rates in not change considerably after five years post-diagnosis [222].

Management of TNBC/BLBCs is limited and challenging as they have an overall poor outcome compared to other subgroups and they lack therapeutic targets in addition to standard therapy (unlike ER^+ and $HER2^+$ breast cancers). Despite the heterogeneity (exemplified by molecular profiling (section 1.6) and non-linear survival curves (Figure 1.5), they're all managed the same way, using surgery, radio- and chemotherapy (although some refinement is possible by considering the grade and stage). There is a major requirement to better understand the molecular basis of TNBC development – it is anticipated that this will help identify novel drug targets and facilitate the development diagnostic tests and new prognostic indicators for this very heterogeneous disease. Current management strategies are failing to achieve disease control, and therefore this subgroup needs better early prognostic indicators (*e.g.* tumour biomarkers) and novel therapeutic targets. Recent meta-analysis study identified *TTK*, a gene associated with chromosomal instability (CIN) that was deregulated in TNBC [222]. Molecular inhibition of *TTK* supports the idea of targeting chromosomal segregation in tumours with a high CIN phenotype as a therapeutic strategy. This study demonstrated that IHC assessment of TTK provided better characterization and understanding for the involvement of CIN to tumour aggressiveness and prognosis [222]. Biomarkers that can be assessed by IHC using procedures and infrastructure that are already established in molecular diagnostic labs and that are associated with a high degree of accuracy and precision are needed as that could facilitate better diagnostic tests.

SOX proteins are a family of transcriptional factors that have an Sry-related HMG DNA-binding domain [223]. During development, SOX genes function as transcriptional activators and are often associated with regulation of fate specification and cellular differentiation. SOX10 has been studied extensively in melanomas and in glial-derived tumours. In human gliomas, high SOX10 expression inversely correlates with tumour grade, and is higher in low grade than in high-grade gliomas [174]. IHC staining of SOX10 is now being used in pathology to support the preliminary diagnosis of melanoma [175-177]. Initial interest in SOX10 as a biomarker of interest in breast cancer came from gene expression data from our lab, showing high expression of *SOX10* mRNA in TNBC compared to non-TNBCs, albeit the data were obtained from a very small cohort [141]. Validation of this finding in an independent cohort of invasive breast cancer (n=37) showed a bimodal *SOX10* mRNA distribution within TNBC (n=21) with ~60% tumours (13/21) exhibited significantly higher levels of *SOX10* compared with luminal (n=9) and HER2$^+$ tumours (n=7) (Figure 3.1). Recently published studies have reported SOX10 protein expression in a small cohort of invasive breast carcinomas and that its expression was enriched in TNBC and BLBCs [139, 140].

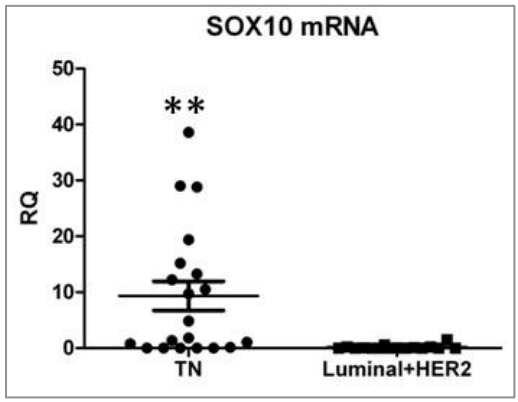


Figure 3.1: *SOX10* mRNA expression levels in Triple negative (TN) vs Luminal/HER2 amplified samples (non-TN breast cancer). Relative quantification (RQ) data normalised to the endogenous control (*RLP13A*) and to reference samples (normal breast). Logarithmic scale is displayed, **: p<0.001 (Mann Whitney U test).

3.2 Hypothesis and aims

Although our preliminary and published data suggest that SOX10 expression correlates with triple-negative and basal-like phenotypes, the relationships with clinical outcomes are unknown, and the role of SOX10 in breast cancer has not been fully characterised. Given the aforementioned expression pattern in breast cancer, we hypothesised that:

1. SOX10 may be important in the biology of a subset of TNBC; and
2. That its expression might stratify outcome in TN-BLBC, which is known to comprises subgroups with good and bad outcome subgroups.

To test these hypotheses, we designed a study with the following aims:

i) To investigate *SOX10* mRNA expression, and any relationships with molecular subtypes and survival, using published datasets (METABRIC and TCGA)

ii) To investigate the genomic mechanisms regulating *SOX10* mRNA expression in breast cancer using published whole-genome expression, methylation and copy-number data

iii) To generate a breast cancer tissue resource comprising up to 500 cases, with accompanying clinical follow-up data and TMAs for high-throughput biomarker analyses

iv) To investigate the relationships between SOX10 protein expression, breast cancer clinic-pathological variables and overall survival in a large cohort of invasive breast cancers.

3.3 Results

3.3.1 *SOX10* expression and regulation in published genomic datasets

3.3.1.1 Pan-Cancer analysis of *SOX10* mRNA expression

The Pan-Cancer project initiative by The Cancer Genome Atlas (TCGA) research network has profiled a large number of tumour samples of different cancer types [224]. When *SOX10* expression was investigated across different cancer types, we noticed that *SOX10* levels were high in melanoma and glioblastoma/glioma (LGG, GBMLGG, GBM), followed by neuroendocrine tumours and breast cancer (BRCA) in the most variable of cancers (Figure 3.2). The expression pattern in breast cancer in particular seems to be heterogeneous and is represented by wide error bars (Figure 3.2).

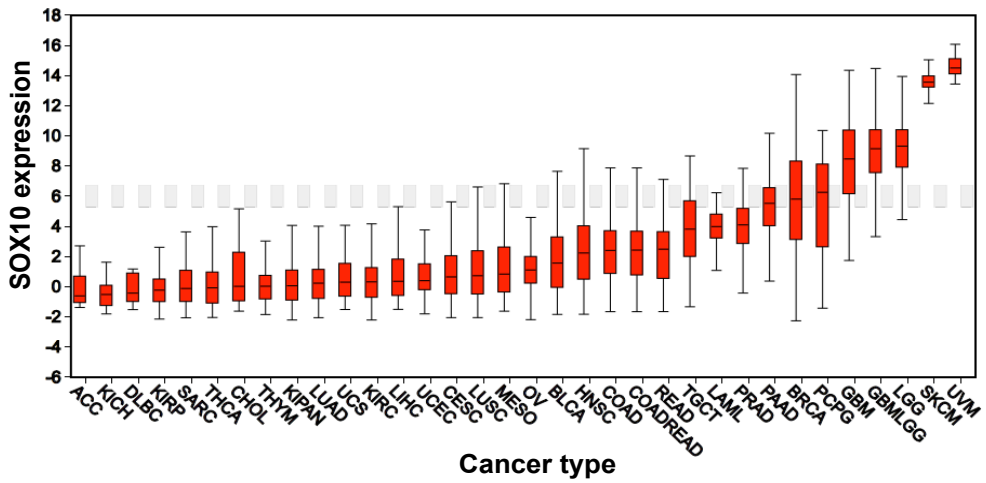


Figure 3.2: Box plot represents *SOX10* gene expression in different cancer types. Expression is highest in melanoma (SKCM and UVM), followed by glial derived tumours (PCPG, GBM,GBMLGG, LGG). In breast cancer, SOX10 expression shows a wider spread compared with other types of cancer. *(Abbreviations: **ACC**, Adrenocortical carcinoma; **KICH**, Kidney Chromophobe; **DLBC**, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; **KIRP**, Kidney renal papillary cell carcinoma; **SARC**, Sarcoma; **THCA**, Thyroid carcinoma; **CHOL**, Cholangiocarcinoma; **THYM**, Thymoma; **KIPAN**, Pan-kidney (KICH+KIRC+KIRP); **LUAD**, Lung adenocarcinoma; **UCS**, Uterine Carcinosarcoma; **KIRC**, Kidney renal clear cell carcinoma; **LIHC**, Liver hepatocellular carcinoma; **UCEC**, Uterine Corpus Endometrial Carcinoma; **CESC**, Cervical squamous cell carcinoma and endocervical adenocarcinoma; **LUSC**, Lung squamous cell carcinoma; **MESO**, Mesothelioma; **OV**, Ovarian serous cystadenocarcinoma; **BLCA**, Bladder Urothelial Carcinoma; **HNSC**, Head and Neck squamous cell carcinoma; **COAD**, Colon adenocarcinoma; **COAD/READ**, Colon and rectal adenocarcinoma combined; **READ**, Rectum adenocarcinoma; **TGCT**, Testicular Germ Cell Tumours; **LAML**, Acute Myeloid Leukemia; **PRAD**, Prostate adenocarcinoma; **PAAD**, Pancreatic adenocarcinoma; **BRCA**, Breast invasive carcinoma; **PCPG**, Pheochromocytoma and Paraganglioma; **GBM**, Glioblastoma multiforme; **GBMLGG**, Glioblastoma multiforme and Brain Lower Grade Glioma combined; **LGG**, Brain Lower Grade Glioma; **SKCM**, Skin Cutaneous Melanoma; UVM, Uveal Melanoma).*

3.3.1.2 *SOX10* mRNA expression in breast cancer molecular subtypes

To assess *SOX10* mRNA levels across different breast cancer subtypes, breast cancer datasets from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and The Cancer Genome Atlas (TCGA) were investigated [97, 98]. These datasets comprise clinical information, global gene expression, copy-number variation and methylation profiles of breast tumours collected as part of various clinical trials, and the data are readily available online.

Using molecular subtype classifications applied by the respective study groups, analysis of the subtype distribution of *SOX10* mRNA expression showed that it is enriched in basal-like tumours in both datasets (Figure 3.3), although it is bimodally distributed. A large proportion of claudin-low molecular subtype have a basal-like phenotype, and are characterised by increased expression of genes associated with the EMT phenotype, and also are enriched for stem cell-like gene expression signatures. Therefore, in this study, we have also investigated *SOX10* expression in claudin-low subtype. We noticed high *SOX10* expression and bimodal distribution patterns in the claudin-low subtype, which was not surprising given that a larger proportion of these tumours have basal-like phenotype. *SOX10* expression was also high in some Luminal A and normal-like breast cancers.

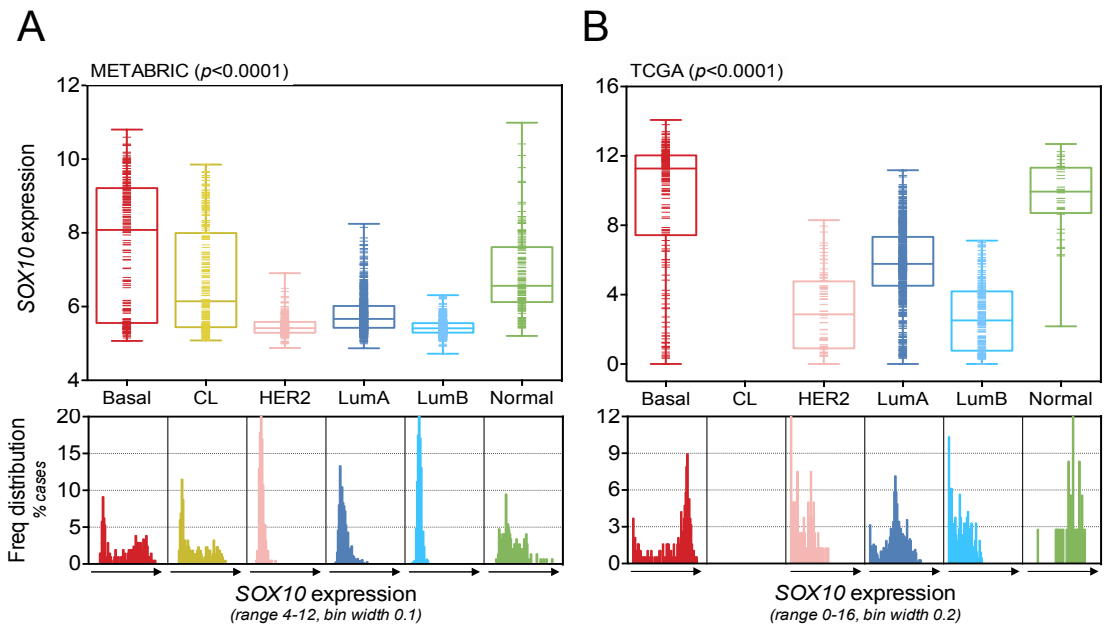


Figure 3.3: *SOX10* mRNA expression across six molecular subtypes of breast cancer classified according to the PAM50 classifier in **(A)** TCGA (n=1104) and **(B)** METABRIC (n=1992) dataset. Boxplot shows enrichment for *SOX10* expression in basal-like subtype in both TCGA and METABRIC dataset (top panel). Frequency distribution plot (bottom panel) shows bimodal distribution pattern for *SOX10* in basal-like (both METABRIC and TCGA) and claudin-low subtype of breast cancer (METABRIC only). TCGA had only eight cases defined as claudin-low and so was excluded for analysis. Boxplots shows the minimum, first quartile, median, third quartile and maximum for each subtype. Significance was assessed using Kruskal-Wallis test (p<0.0001).

SOX10 expression in integrative clusters (IntClust) of breast cancer (classified by a combination of gene expression and copy-number variations) showed high levels of expression in the IntClust-10 subgroup and was bimodal in distribution (Figure 3.4). Given that the IntClust-10 group is thought to be a basal-like subgroup [98], this emphasises the enrichment of *SOX10* is enriched in the basal-like subtype regardless of classification system used.

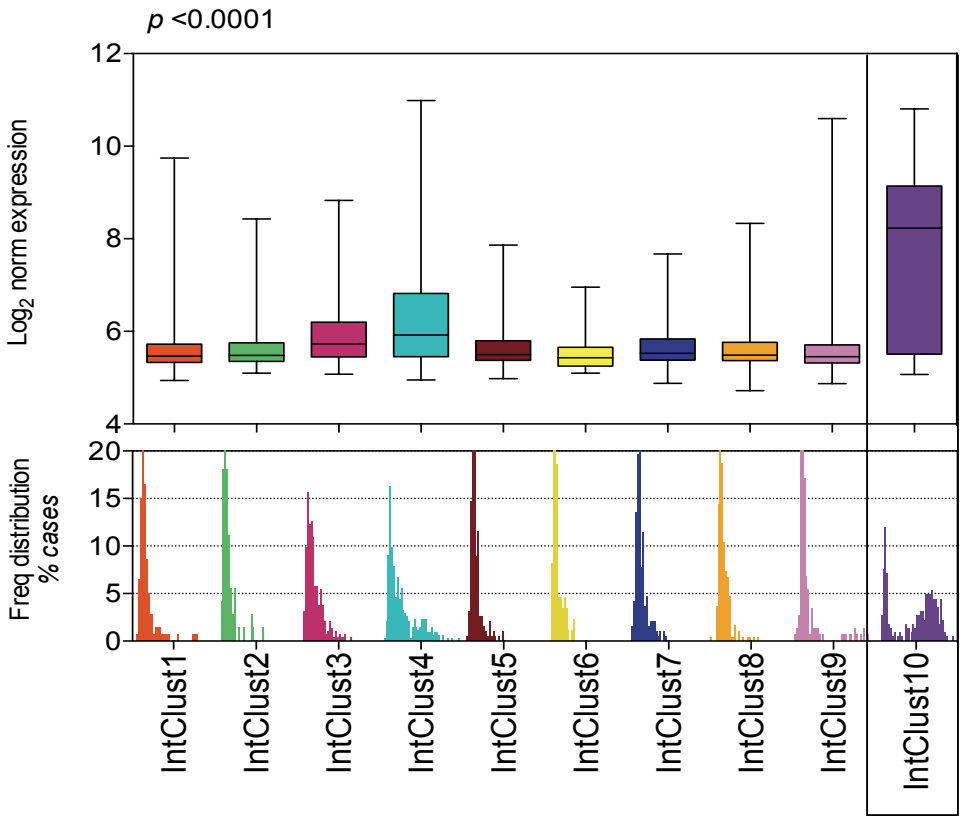


Figure 3.4: *SOX10* mRNA expression in integrative molecular classification of breast cancer as defined by copy-number and gene expression profiles. Box plot shows enrichment for SOX10 expression within integrative cluster 10. The frequency distribution plot (bottom panel) shows bimodal distribution pattern for SOX10 in cluster 10 subgroup. Boxplot shows the minimum, first quartile, median, third quartile and maximum for each IntClust cohort. Significance was assessed using Kruskal-Wallis test (p<0.0001).

3.3.1.3 *SOX10* copy-number (CN) variations in breast cancer

Since *SOX10* expression was enriched in the basal-like subtype of breast cancer, we wanted to investigate the mechanisms that were driving its high expression. We know that SOX10 is regulated by epigenetic mechanisms during neural crest development and melanocyte differentiation therefore; we investigated whether genomic alterations such as copy-number changes and methylation are involved in regulating SOX10 expression in breast cancer. We used matching data from melanoma (n=~450, from TCGA) as a reference, since SOX10 is highly expressed in

melanoma (87-97% of cases). *SOX10* copy-number gains/amplifications were more frequent in basal-like cancers compared to non-BLBC (χ^2 p=0.0002, Figure 3.5A). There are CN aberrations affecting SOX10 and these do effect expression, but overall there are only few cases with amplification, and across the whole cohort this generates only a weak association with expression (r=0.438 vs r=0.068) (Figure 3.5B). The pattern of correlation was visually similar when compared to melanoma cases (Figure 3.5C). Figure 3.5 B and C represents a simple linear correlation analysis that shows a linear relationship between CN and expression.

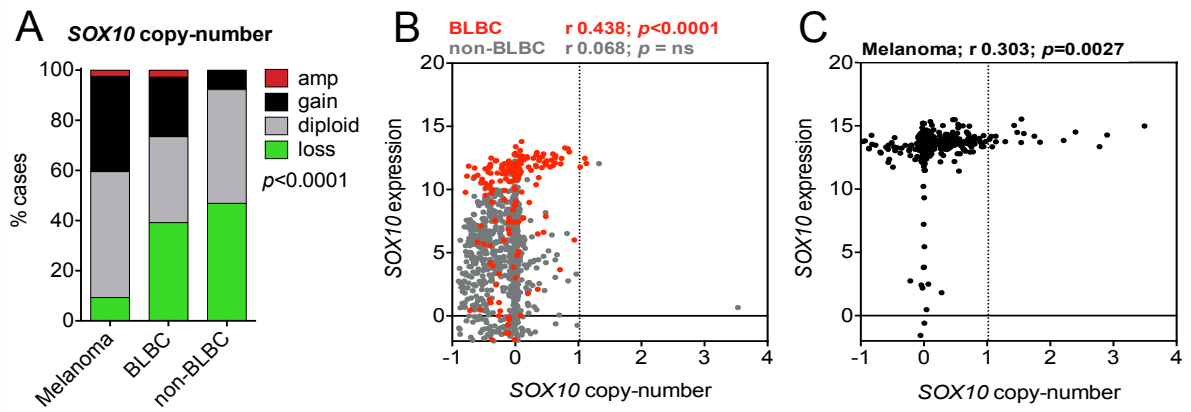


Figure 3.5: Copy-number variation at the SOX10 locus and its association with gene expression **(A)** Bar plot depicting the percentage of samples with copy-number changes in breast cancer and melanoma. Scatter plot showing the correlation between *SOX10* gene expression and copy-number changes in **(B)** breast cancer and **(C)** melanoma. Melanoma cases (TCGA, n=~450) were used as a reference for comparison and to define thresholds. X-axis in B and C reflects raw CN data (log2 SNP array data). Significance was assessed using Chi-square and Spearman's correlation test.

3.3.1.4 *SOX10* genomic methylation in breast cancer

Methylation data across all probes with the strongest negative association between methylation signal and gene expression were included in this correlation analysis. The results showed that a subgroup of basal-like breast cancers (~65%) demonstrated a high expression compared with non-BLBC (χ^2 p<0.0001, Figure 3.6A). SOX10 is highly and consistently expressed in melanoma, and was hypomethylated compared to breast cancer. We applied a 90th percentile methylation threshold that captured the majority of the low-expression melanoma cases (11/13), and found this also segregated BLBC (r=-0.766, p<0.0001) from non-BLBC (r=-0.822, p<0.0001, Figure 3.6B). This strong correlation between SOX10 expression and methylation status in breast cancer suggests that this gene is tightly regulated by methylation in breast cancer and that hypomethylation characterizes BLBCs with high expression. The pattern of methylation-gene expression correlation in BLBC (red dots) was similar when compared with that for melanoma cases (black dots, Figure 3.6C).

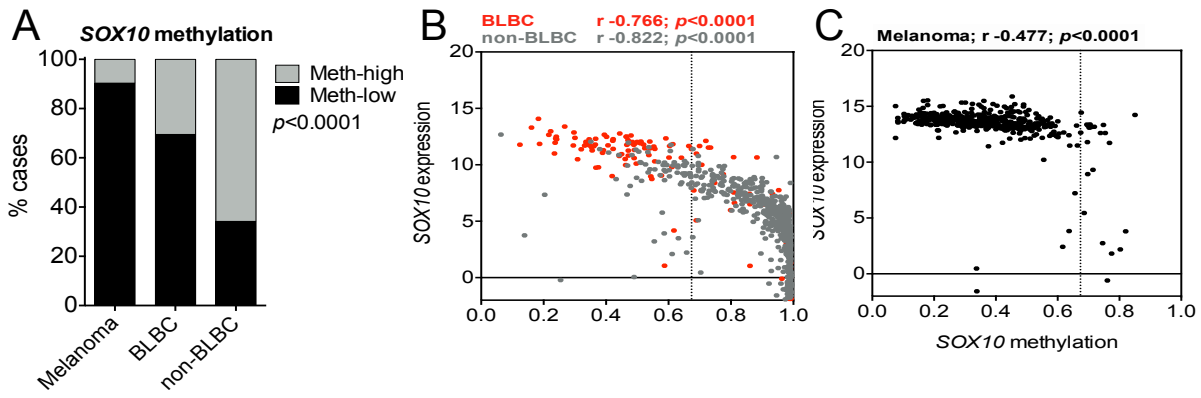


Figure 3.6: Methylation status at the SOX10 locus and its association with gene expression **(A)** Bar plot depicting the percentage of samples with high/low methylation levels in breast cancer and melanoma. Scatter plot shows correlation between *SOX10* gene expression and methylation changes in **(B)** breast cancer and **(C)** melanoma. Melanoma cases (TCGA, n=~450) were used as a reference for comparison and to define thresholds. High and low methylation levels were defined based on cut-off value of 0.68 as this cut-off separated high/low methylation levels in melanoma samples. The same cut-off was then superimposed on breast cancer samples, which separated BLBCs from non-basals. Dotted line indicates the 90th percentiles of *SOX10* methylation. Significance was assessed using Chi-square and Spearman's correlation test.

3.3.1.5 Relationships between *SOX10* mRNA expression and breast cancer survival

There was no significant relationship between SOX10 mRNA expression and breast cancer survival at five years in METABRIC and TCGA datasets. However, survival after 5 years showed a trend towards poor prognosis in *SOX10* positive cases (Figure 3.7 A and B). Since METABRIC and TCGA datasets represents total mRNA levels, it is important to investigate whether there is a correlation between SOX10 mRNA and protein levels and whether SOX10 protein expression correlates with poor survival.

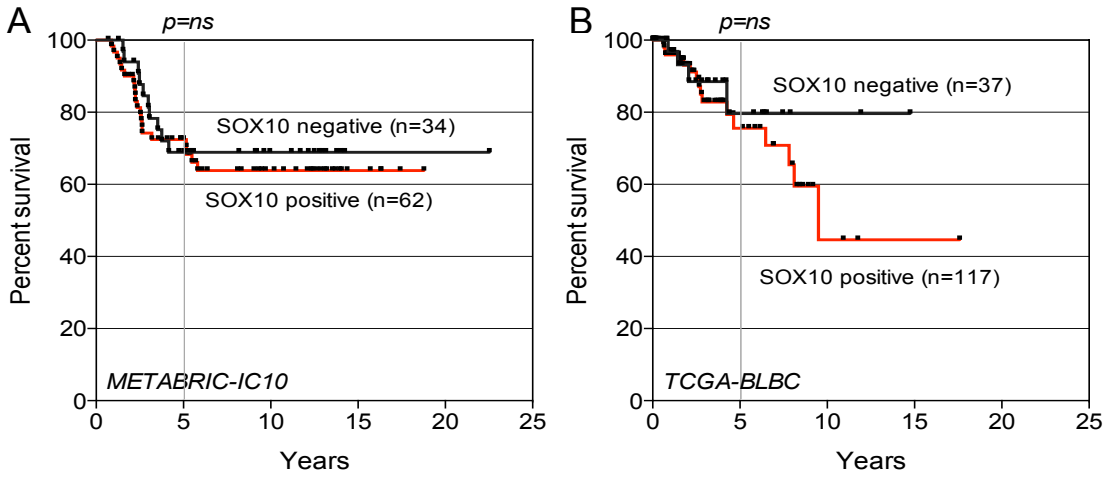


Figure 3.7: Kaplan-Meier survival analysis comparing *SOX10* positive and negative cases in basal-like subtype of breast cancer in **(A)** METABRIC and **(B)** TCGA datasets.

3.3.2 Construction of the QFU breast cancer resource

Preliminary data from the group, and also data from publicly available breast cancer genomic datasets suggested that *SOX10* hypomethylation (and copy-number gains in some cases) contributes to the relatively high expression of *SOX10* mRNA in BLBC. To investigate whether this corresponds to SOX10 protein expression, we assembled a large cohort of invasive breast cancer samples with long-term clinical follow-up data, and constructed tissue microarrays (TMAs) comprising duplicate samples of each tumour for high-throughput IHC analysis.

Of the 500 cases identified and retrieved for the study, 449 satisfied the inclusion criteria: survival and clinical information accessible via the Queensland Cancer Registry, and diagnostic blocks conducive to sampling of areas with sufficient tumour cellularity (minimum 20%). We performed a comprehensive clinicopathological review of the cohort, involving: (i) collection of survival data, patient age, lymph node status and tumour size information from clinical pathology reports; (ii) independent histological assessment of H&E-stained whole sections (Prof Lakhani); and (iii) expression of prognostic biomarkers by IHC, in order to classify cases as $HER2^+$, ER^+ (high *vs* low proliferation), TN and basal-like. The median age of patients in the cohort was 60 years. The median follow-up and follow-up of patients still alive were 13.2 and 25 years, respectively. The histopathological features of the cohort are summarised in Table 2 (Appendix 3.1).

3.3.2.1 Validation of the QFU cohort: investigation of known prognostic indicators

We validated the QFU resource by performing univariate survival analysis of the prognostic indicators used currently in diagnostic practice: grade, lymph node status, ER/PR and HER2 status. The trends observed were as expected, with high grade, lymph node-positivity, HR negativity and HER2 positivity (by ISH) all associated with poor outcome (Figure 3.8). Stratification of breast cancer cases into different prognostic subgroups based on expression of immunohistochemistry markers ER, PR, HER2, Ki67, EGFR and CK 5/6 and 14 is represented in the form of a flow chart (see Appendix 3.2). Kalpan-Meier survival analysis of QFU cohort showed that the cohort separated into different breast cancer subtypes as expected (see Appendix 3.3). Ki67 threshold was determined by analyzing different expression cutoff values such as 10% (low) and 20% (high, percentage represents tumour cell positivity). Twenty per cent cutoff value showed a difference in breast cancer survival outcome compared to 10% (Appendix 3.4), and thus was used for stratification of breast tumours in this study. Comparison of the clinicopathological variables within different breast cancer subtypes is presented in Appendix 3.5. This analysis demonstrated that the QFU resource were robust and reliable for analysis of new candidate biomarkers. The baseline data

generated were also then available for future correlation and subgroup-specific analyses (*e.g.* analysis of SOX10 prognostic significance specifically in TN-BLBC; see section 3.3.2.4 below).

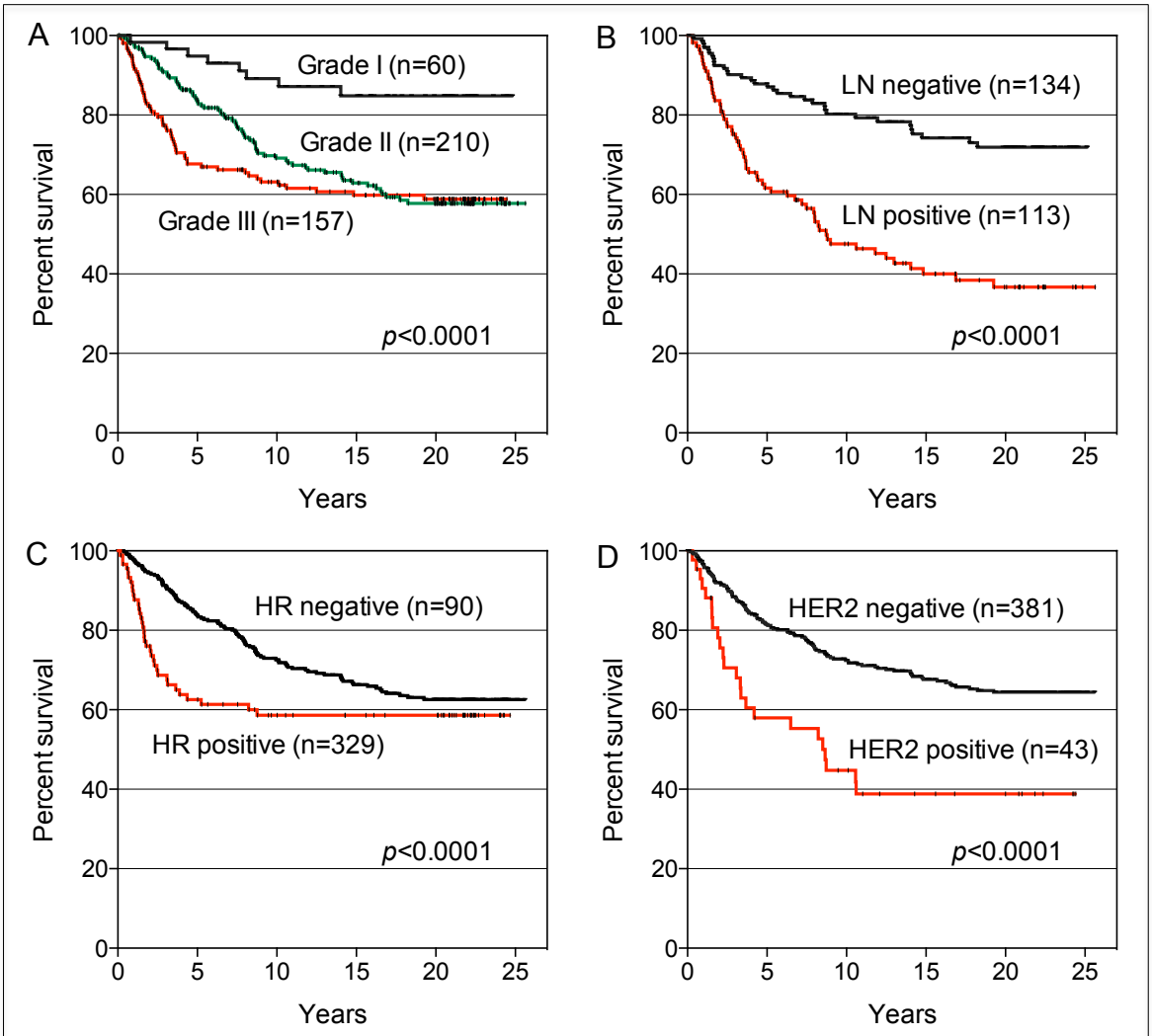


Figure 3.8: Validation of QFU cohort according to expected outcomes for different prognostic variables. Kalpan-Meier survival analysis according to **(A)** Grade and **(B)** Lymph node status **(C)** ER Status and **(D)** HER2 status. Significance was assessed using Gehan-Breslow-Wilcoxon test.

3.3.2.2 SOX10 expression in sporadic breast cancer

The intensity of nuclear expression was recorded as either negative (n=363), weakly positive (n=7) or strongly positive (n=30, Figure 3.9). The percentage of tumour nuclei positivity was 80-100%. Similarly, SOX10 staining in the cytoplasm was recorded as either negative (n=191), weakly positive (n=199), moderately positive (n=37) or strongly positive (n=10) as shown in Figure 3.10. Approximately 73% (27/37) of SOX10 nuclear positive cases showed weak cytoplasmic staining, about 13% (5/37) showed moderate to strong cytoplasmic staining and the remaining 14% showed no staining in the cytoplasm.

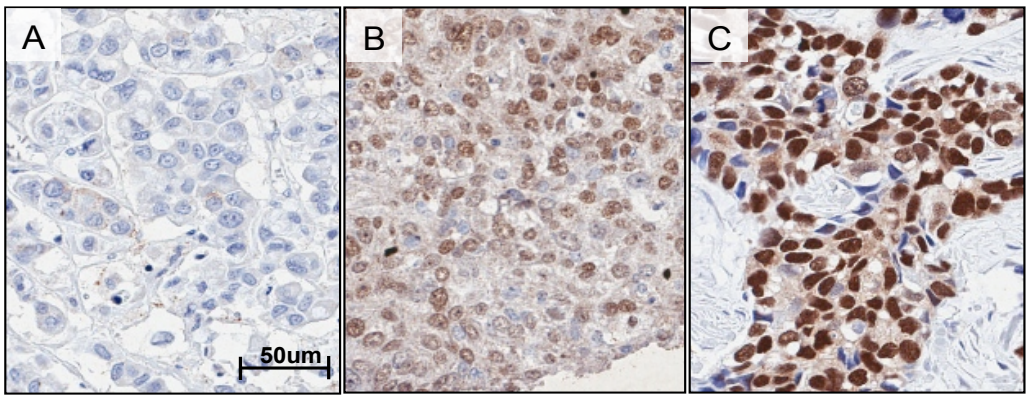


Figure 3.9: Representative images of SOX10 nuclear staining on breast tumour cases at 10X magnification showing **(A)** completely negative **(B)** weakly positive and **(C)** strong positive.

SOX10 nuclear and cytoplasmic expression across different subtypes of breast cancer is shown in Figure 3.11. The results showed 50% (27/54) of TN-BLBC, 25% (4/16) TN non-basal, 10% (5/49) Luminal B and <1% (2/266) Luminal A cancers were positive for SOX10 nuclear staining. Chi-square analysis showed a positive association between SOX10 nuclear expression and TN-BLBC (Figure 3.11A, $p<0.0001$). However, between 60-70% of all cases in each subtype showed SOX10 cytoplasmic staining was not associated with any of the breast cancer subtypes (Figure 3.11B).

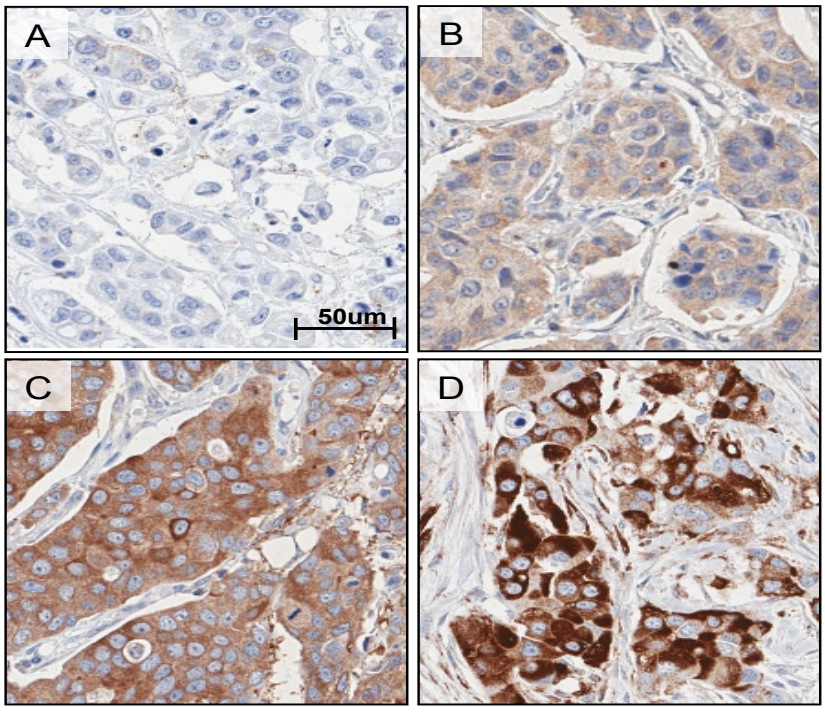


Figure 3.10: Representative images of SOX10 cytoplasmic staining on breast tumour cases at 10X magnification showing **(A)** completely negative **(B)** weakly positive (**C)** moderately positive and **(D)** strong positive.

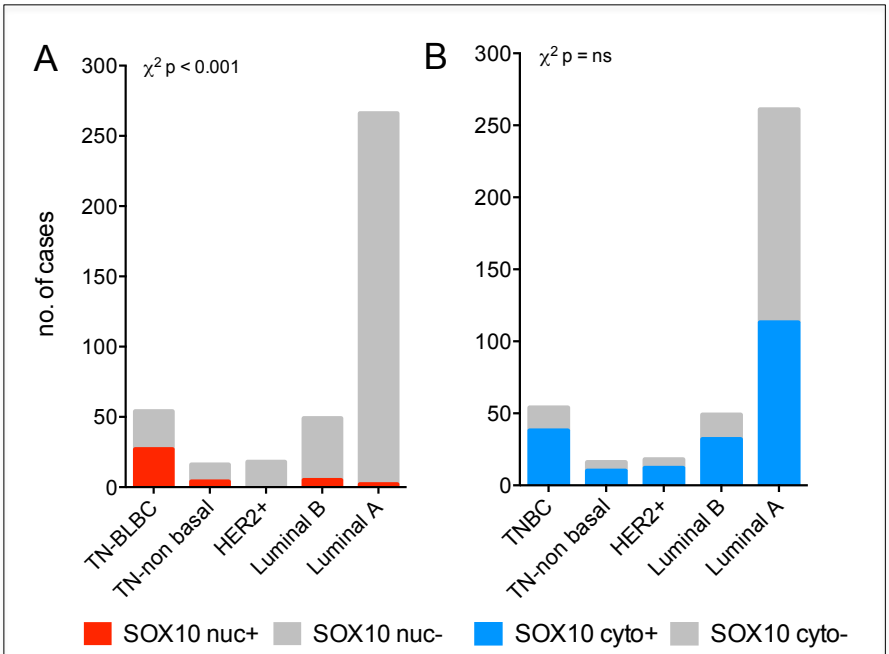


Figure 3.11: Frequency of SOX10 expression in the **(A)** nucleus and **(B)** cytoplasm across different subtypes of breast cancer. Significance was assessed using Chi-square test (p<0.0001).

3.3.2.3 Relationships between SOX10 expression and clinicopathologic variables

Chi-square analysis showed a strong positive association between SOX10 nuclear expression and histopathologic parameters (*e.g.* grade and mitotic score, $p<0.0001$) within the whole cohort. This was not surprising because SOX10 expression marks a subset of BLBCs, which are high-grade tumours to begin with. Among the different biomarkers tested, SOX10 expression showed a positive correlation with Ki67, p53, Vimentin and c-kit and an inverse correlation with the androgen receptor (AR), $p<0.0001$ for each) (Appendix 3.6). However, there was no association with age, tumour size, lymph node status and lymphovascular invasion.

Comparison of the clinicopathological variables based on SOX10 status within TN and TN-BLBC is presented in Table 3 (Appendix 3.7). Mitotic score showed a positive correlation with SOX10 in the TN group ($p=0.0467$). Among the different biomarkers examined, SOX10 expression correlated positively with c-kit and vimentin within the TN group ($p=0.0068$ and 0.007 respectively). Androgen receptor showed an inverse correlation with SOX10 expression within the TN and TN-BLBC groups ($p=0.0010$ and $p=0.0177$ respectively).

3.3.2.4 Relationship between SOX10 expression and breast cancer-related survival

Kaplan-Meier analysis of the QFU cohort (n=426 with informative staining) showed SOX10 cytoplasmic positivity was not associated with survival (Figure 3.12C). By contrast, nuclear positivity was associated with a 40% decrease in the disease specific survival at five years (Figure

3.11A, $p<0.0001$; hazard ratio (HR): 3.66 +95% confidence interval (CI): 1.78-7.51. Given that most of the nuclear staining was in a subset of TN/BLBCs, we analysed survival specifically in this subgroup and found a similar result (Figure 3.12B; $p=0.0188$; HR: 3.20 +95% CI: 1.28-8.03). SOX10 nuclear expression did not stratify outcome in the other subgroups; however it was observed infrequently and the case numbers were not sufficient for robust statistical analysis.

In a univariate model, SOX10 nuclear expression in TN/BLBC showed strong prognostic significance at 5 years post-diagnosis with a difference in the survival rate was similar to that of grade (GI vs GIII) and lymph node status (log-rank test), which suggests that the SOX10 nuclear expression may have independent prognostic value. Our TN/BLBC cohort was of insufficient size to investigate this in a multivariate model. Whether SOX10 nuclear expression should be considered as a useful marker in the diagnostic setting needs to be confirmed in studies with a larger cohort, and is therefore a priority for future work along with validating the data in an independent cohort.

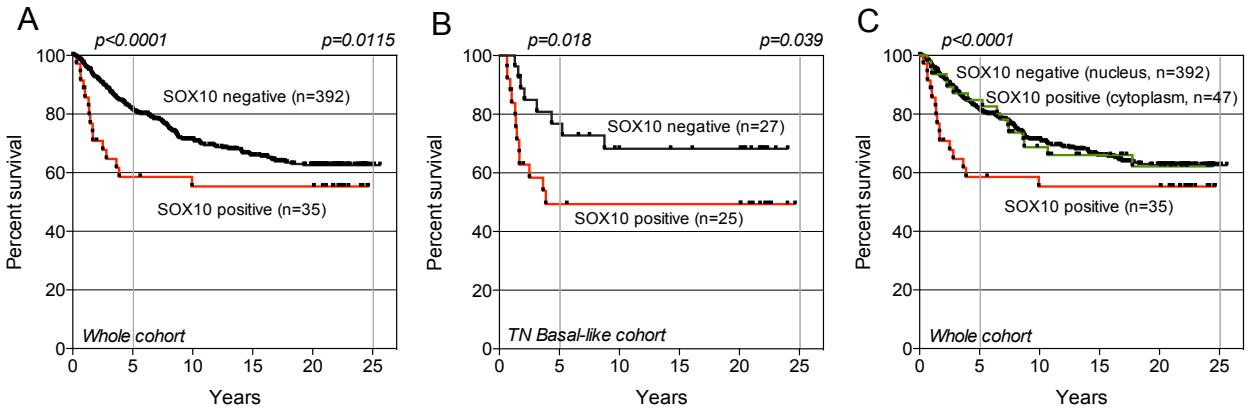


Figure 3.12: Kaplan-Meier survival analysis of SOX10-negative and–positive subgroups within **(A)** the whole cohort (QFU) **(B)** TN-BLBC and **(C)** survival analysis comparing SOX10 positive staining in the nucleus and cytoplasm. *p* values are represenated at 5 and 25 years respectively. Significance was assessed using Gehan-Breslow-Wilcoxon test.

3.3.3 SOX10 expression in breast cancer special types enriched for basal-like phenotypes

Immunohistochemistry based study has shown ~91% (59/65) of metaplastic breast carcinomas have basal-like immunophenotype as defined by Nielsen et al [85] regardless of type of metaplastic element [225]. These tumours express several other markers usually found in BLBCs such as vimentin [226, 227], caveolin 1 [228] and EGFR [229, 230]. More recently, microarray based gene expression analysis of 20 metaplastic breast cancers revealed that 95% of all cases have basal-like molecular subtype [231].

Similarly, immunohistochemistry profiles and morphological characteristics of sporadic basal-like breast cancers shows striking similarities to that of tumours arising in *BRCA1* mutations carriers [94, 232, 233]. Using overlapping but not identical definitions, it has been demonstrated that a large proportion of *BRCA1* mutations carriers have a basal-like phenotype [89, 232, 234]. Lakhani et al., has shown that ~70% of *BRCA1* tumours express basal cytokeratins (CK5/6 and/or CK14) are negative for ER [234]. BRCA1 tumours have shown to segregate with sporadic basal-like breast cancers consistently in hierarchical clustering analysis using gene expression profiling data [89]. There is also evidence that metaplastic carcinomas harbour BRCA1 dysfunction where *BRCA1* gene promoter is being methylated in about 60% of these cases [235]. Therefore, keeping these observations in mind, a natural extension of our work was to see whether SOX10 expression has prognostic significance in metaplastic and familial breast cancers as a large proportion of these tumours have basal-like phenotype.

3.3.3.1 Metaplastic breast cancer

Metaplastic carcinomas comprise a group of typically high-grade tumours, which have both epithelial and mesenchymal elements and often have basal-like phenotype. They account for 4-5% of all breast cancers, and are often associated with aggressive clinical behaviour and poor outcome. SOX10 expression has been previously investigated in a small cohort of metaplastic carcinomas, with 6/13 (46%) of which showed high SOX10 expression. We therefore investigated SOX10 expression in a larger, independent cohort of metaplastic carcinomas (n=105); whole sections rather than TMA were used (to ensure that we considered intrinsic morphologic heterogeneity). Overall, 48% of cases (51/104) showed SOX10 nuclear positivity. Unlike sporadic breast cancers, SOX10 expression in was not only restricted to TNBCs, it was seen in ~50% cases regardless of which subtype (Figure 3.13A). SOX10 staining intensities ranged from weak (3/51), to moderate (30/51) and strong (18/51) (Figure 3.13B-D). The percentage of tumour cell positivity was 20-100% in each group. Most positive cases were classified as mixed metaplastic type (42/51) the other subtypes also included squamous cell spindle cell and chondroid cell carcinoma (9/51). SOX10 expression within the whole cohort showed a trend towards better outcome in Kaplan-Meier survival analysis. This was not a strong correlation, as the data did not reach statistical significance (p=0.06). However, within only mixed metaplastic subtype, SOX10 expression showed a significant correlation towards better outcome (p=0.04, Figure 3.12F).

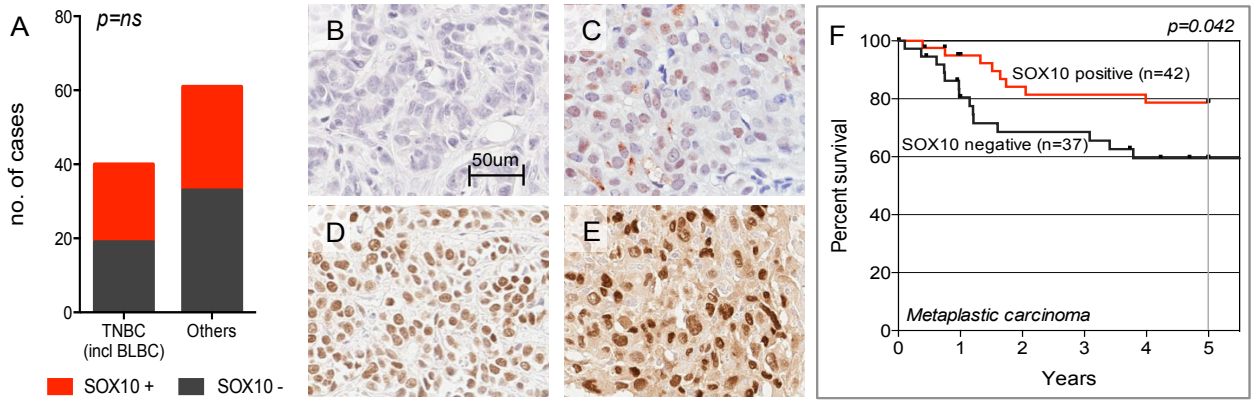


Figure 3.13: SOX10 nuclear expression in metaplastic carcinomas. **(A)** frequency of SOX10 expression in TN group versus others. Light microscopy images showing **(B)** negative **(C)** weak positive **(D)** moderately positive and **(E)** strong positive; **(F)** Kaplan-Meier survival analysis comparing SOX10 positive and negative staining. Significance was assessed using Chi-square and Gehan-Breslow-Wilcoxon test.

3.3.3.2 Familial breast cancers

The relationships between tumour SOX10 expression, the basal-like phenotype and survival are unknown in familial breast cancers. Since tumours arising in germline *BRCA1* mutation carriers are enriched with basal-like phenotypes, we examined whether SOX10 also has prognostic significance in a large cohort of familial breast cancers with *BRCA1* germline mutations (n=139). About 71% *BRCA1* cases were basal-like based on expression of basal markers (CK5/6 and/or CK14 and/or EGFR) and about 35% were TN basal-like. Our results showed within *BRCA1* cohort, 46% (64/139) were SOX10 positive. We also assessed the relationships between SOX10 expression and breast cancer-specific survival in *BRCA2* (n=114) and *BRCAX* cases (n=353) for comparison. Within these groups 46% *BRCA2* and 69% *BRCAX* cases were basal-like; about 11% *BRCA2* and 9% *BRCAX* cases were TN basal-like. SOX10 expression in *BRCA2* and *BRCAX* groups were 10.5% (12/114) and 9% (33/353) respectively.

SOX10 expression in familial cancers was not associated with triple negative phenotype and was similar to metaplastic cancers with no difference in the frequency of expression between TNBC (~50%) and other subtypes (~60%, Figure 3.14A). Interestingly, SOX10 expression correlated with better outcome in *BRCA1* cases (Figure 3.14B) in contrast to sporadic cancers. There was no difference in breast cancer-specific survival in *BRCA2* cases (Figure 3.14C), though the small size of the SOX10 positive subgroup prevented robust statistical analysis. In *BRCAX* cases, which are not enriched for any particular subtype, SOX10 expression correlated with poor outcome at five years post-diagnosis. Interestingly, four out of six cases responsible for the early decline in mortality were basal-like TNBCs. This effect was lost when observed across 25 years (Figure 3.14D).

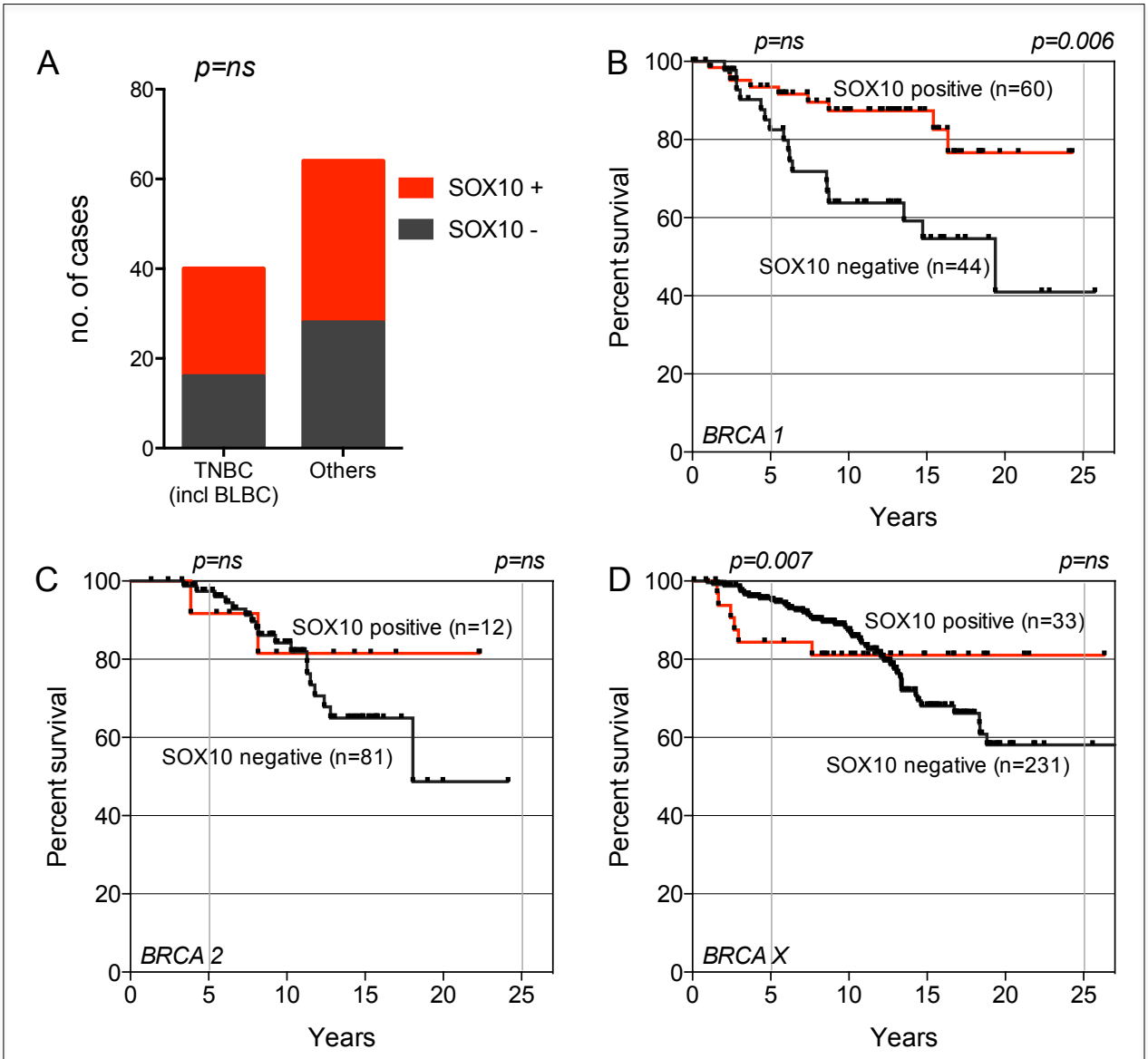


Figure 3.14: SOX10 nuclear expression in familial brest cancer. **(A)** frequency of SOX10 expression in TN group versus others. Kaplan-Meier survival analysis comparing SOX10 positive and negative staining in **(B)** *BRCA1*+, **(C)** *BRCA2*+ **(D)** *BRCAX* cases. *p* values are represenated at 5 and 25 years respectively. Significance was assessed using Chi-square and Gehan-Breslow-Wilcoxon test.

As SOX10 expression correlated with better outcome in *BRCA1* cases, we wanted to investigate whether this difference existed in the TNBC subset. Similarly, we wanted to know whether SOX10 expression in *BRCA2* and *BRCAX* cases showed any difference in survival within TN and basal-like TN subsets of these cohorts. The results showed no significant difference in survival at 5 years in both *BRCA1* and *BRCAX* groups (Figure 3.15) within TN and basal-like TN subsets. We could not investigate this in *BRCA2* cases, as the sample size was small.

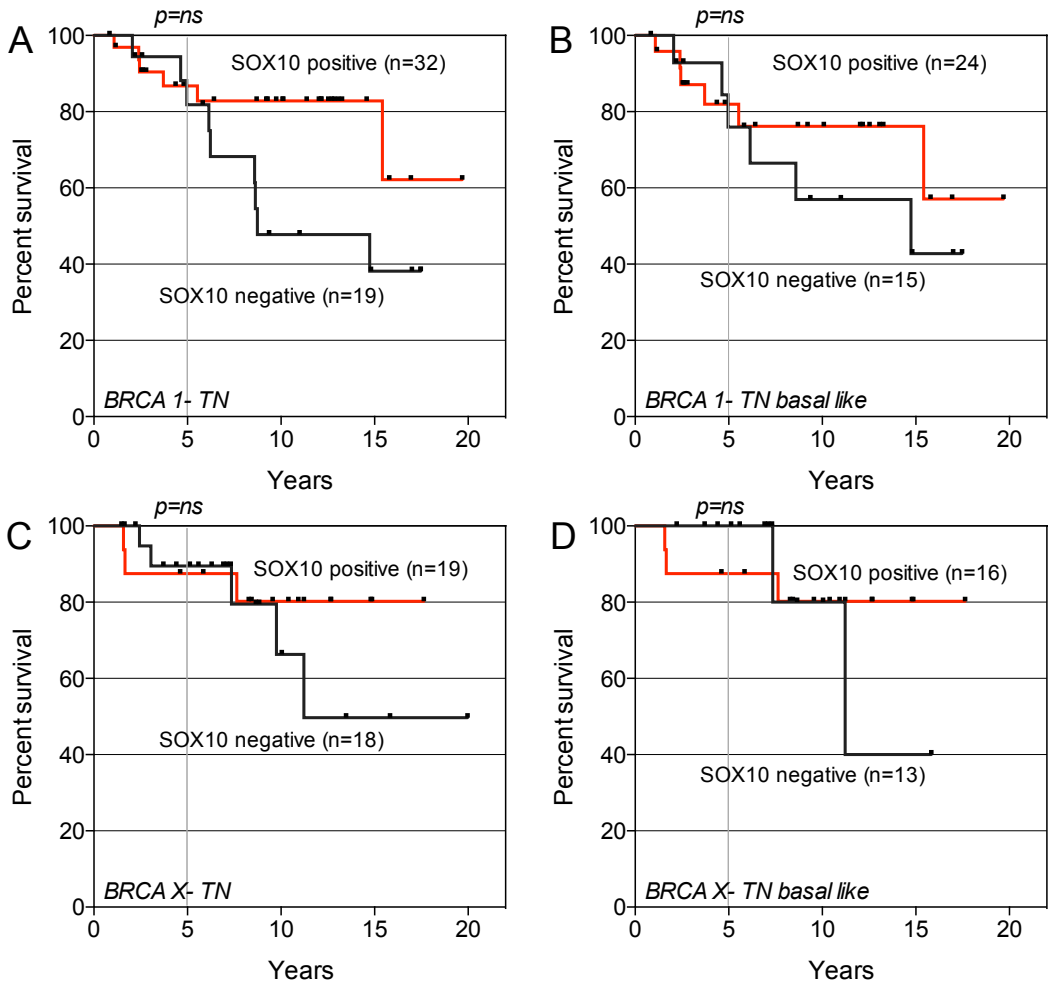


Figure 3.15: Kaplan-Meier survival analysis comparing SOX10 positive and negative cases within triple negative (TN) and TN basal-like subgroups in familial breast cancer **(A-B)** *BRCA1+* (top panel) and **(C-D)** *BRCAX* cases (bottom panel). Significance was assessed using Gehan-Breslow-Wilcoxon test.

3.4 Discussion

Breast cancer is a complex heterogeneous disease that encompasses different tumour entities with different biological features and clinical outcomes. New therapeutic options are particularly needed for patients with breast tumours that are TN, as there are currently no targeted therapies available, and these tumours can be associated with aggressive clinical behaviour. Identifying new biomarkers may hold the key to a better biological understanding of the development of these tumours, and may also improve the clinical management through more accurate prognostication and/or as companion diagnostics for new targeted therapies.

In this study, analysis of published breast cancer genomic datasets showed that high *SOX10* mRNA expression was enriched in basal-like breast cancer subtype and was inversely correlated with methylation leading to increased expression, suggesting epigenetic regulation is important in breast cancer. There is clear evidence that DNA methylation profiles for BLBCs are different to other types of breast cancer [236, 237]. BLBC have lower methylation levels compared to Luminal or HER2 [238, 239]. *SOX10* may be one of the differentially methylated genes in breast cancer, which would explain its high expression levels in BLBC compared to other subtypes.

We also found that the nuclear expression of SOX10 is strongly associated with poor outcome in TN/BLBC. This observation needs to be validated in an independent cohort in order to strengthen our hypothesis for SOX10 as a potential breast cancer biomarker. This is in process through collaboration with the Nottingham group (Prof. Ian Ellis). The use of SOX10 monoclonal antibody (Santa Cruz, Clone N20) as a diagnostic marker for melanoma has been reported in different studies [175, 176]. Our data therefore may raise an interesting possibility regarding potential utility of SOX10 assessment to guide management decisions in TN/BLBC. Only a subset of TNBCs have poor survival outcome and yet at this point all TNBCs are treated with standard adjuvant therapies. There is a subset within TNBC with better come post 5 years of diagnosis that are currently not identified or separated in routine diagnostic practice who may not require standard therapy. In this chapter, we have shown TNBCs can be characterised as “good” and “bad” outcome groups based on SOX10 expression using IHC and this stratification may help in better management of these patients in future. SOX10 positive patients can therefore be classified as high-risk group (as they have the worst survival outcome within 5 years of diagnosis) and this group may benefit from intense surveillance and risk reducing interventions in future. Having said this, identifying new TNBC biomarkers is not just enough given that there is no therapy that targets SOX10. This study is leading to further understanding of functional role of SOX10/downstream signalling pathways regulated by SOX10, which would help identify targets for new therapeutic regimens in future.

The reason for different relationships observed between SOX10 and survival in the different TN cohorts in this study may be for the following reasons. Metaplastic carcinomas are enriched with mesenchymal and stem cell-like characteristics, and are often inherently aggressive in terms of clinical behaviour. This is distinct from sporadic, triple-negative IC NST, which comprise both aggressive and treatment-responsive cases with very good outcomes. TN tumours arising in *BRCA1* mutation carriers are characterised by DNA repair deficiency and genome instability, and overall have a poor outcome, but lower mortality rate in the first 5 years after diagnosis compared with sporadic TN cases [240], which tend to metastasise rapidly if they progress. Thus although they have in common a TN, basal-like phenotype, there are still clear differences in the biology underlying their development. Recent published study showed that sporadic and *BRCA1* BLBC have independent miRNA expression profiles supporting the underlying difference [241]. There is a possibility that these differing biological backgrounds might affect SOX10 function, which may explain the different relationships with clinical correlates observed in the different cohorts. Alternatively, SOX10 expression may be an inconsequential correlate in metaplastic and *BRCA1* mutation carriers. Therefore, further functional experiments are required to fully role of SOX10 in breast cancer (see chapter IV).

Chapter IV

Evaluating the functional role of SOX10 in breast cancer and normal breast

Chapter IV: Evaluating the functional role of SOX10 in breast cancer and normal breast

4.1 Background

Preliminary and published data show that SOX10 is strongly associated with aggressive clinical behaviour in sporadic breast cancers when expressed and localised in the nucleus, suggesting this could be a causal association, since SOX10 is a transcription factor. This tends to occur in a subset of triple-negative, basal-like tumours. However, little is known about SOX10 function in the breast, in either normal or neoplastic contexts.

SOX10 expression is initiated in the pre-migratory neural crest cells during development [242]. Its expression controls multipotency, survival, proliferation and differentiation of neural crest cells into melanocytes, as well as specialised cells of peripheral and autonomous nervous system (as described in Chapter 3) [155, 243, 244]. During melanocyte differentiation, SOX10 expression is restricted to the bipotent melanoblast glia progenitor cells and is completely absent in the differentiated melanocyte suggesting an inverse association with differentiation (Figure 4.1) [156]. SOX10 in adult mature tissues is restricted to a few cell types, expressed mainly in mature glial lineages from the peripheral and central nervous system and melanocytes (reviewed in [155]). Less common cell types include the basal cells of salivary glands and a subset of smooth muscle cells of the pancreas and prostate [175, 245]. The expression SOX10 in the normal breast was originally reported to be restricted to the basal compartment [175]. A recent study reported SOX10 expression in both luminal and basal compartment [140]. There is evidence from the literature that SOX10 is essential for the survival of notch4-immortalised mouse mammary epithelial cells [246], which is consistent with its role in regulating differentiation, although this was in melanocyte development.

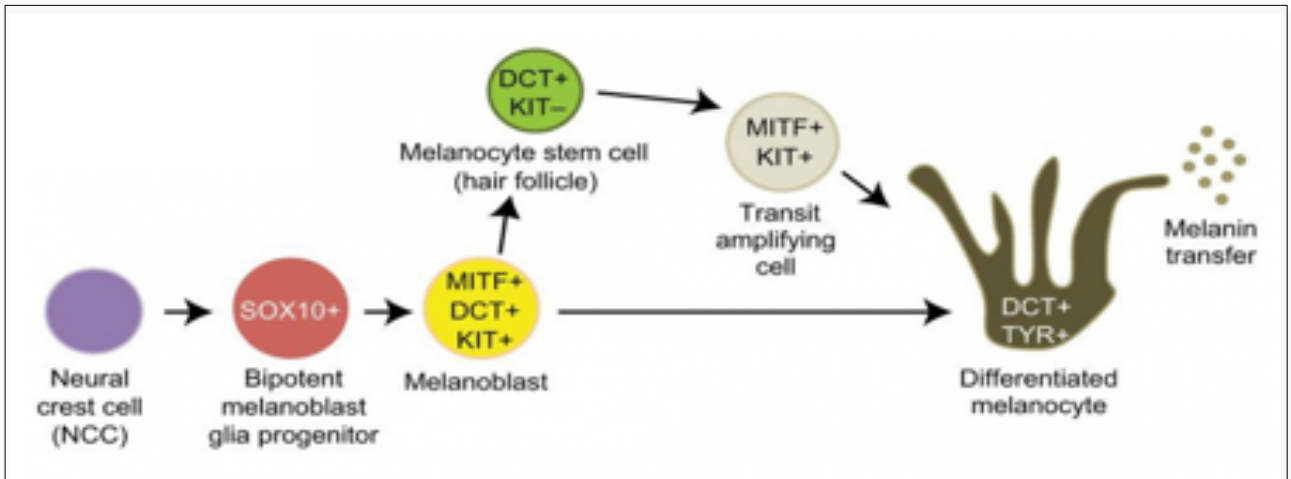


Figure 4.1: Differentiation of neural crest cell (NCC) into melanocyte [156]. Permission to use figure was obtained from the publisher through Copyright Clearance Center.

SOX10 knockdown in *Xenopus* model showed loss of NC precursors, decreased expression of Slug and Foxd3, enlargement of non-NC domains, increased apoptosis and decreased proliferation. *In vivo* and *In vitro*, SOX10 knockdown blocks the formation of melanocytes and ganglia [247]. *SOX10* homozygosity in mice is associated with defects of the entire PNS and motor neurons [166]. *SOX10* haploinsuffiency causes aganglionosis of the colon and pigmentation defects whereas, homozygous deletion of SOX10 in mice leads to embryonic lethality [248]. In humans, heterozygous mutations in SOX10 gene are associated with Waardenburg-Hirschprung disease (colonic aganglionosis, hearing defects, pigmentation abnormalities and myelinopathies [249]).

Little is known about functional role of SOX10 in cancer development, studies on functional aspects of SOX10 have been largely in melanomas. Low frequencies of intragenic mutations in the SOX10 gene have indicated that maintenance of SOX10 wild type function is essential for melanoma formation [142, 250]. Role of SOX10 in melanomagenesis has been demonstrated using different mouse models [142, 144]. Cronin and co-workers have shown that loss of SOX10 causes cell cycle arrest at G1 phase and demonstrated that SOX10 expression is fundamental for melanomagenesis *in vivo,* low SOX10 expression reduced tumour formation in melanoma mouse model [142, 144]. Another study showed SOX10 controls migration of B16F10 melanoma cells through multiple regulatory target genes. This study showed that a cascade of gene expression initiated by SOX10 and subsequently facilitated in part by MITF (microphthalmia-associated transcript factor) represented an important regulatory axis of migration and metastasis in a subset of melanoma cases [251]. Additionally, Graf and co-workers investigated role of SOX10 in different human melanoma cell lines and showed SOX10 inhibition revealed a critical role in melanoma cell invasion and survival [143].

4.2 Hypotheses and aims

Melanocytes, glia and mammary epithelia all derive from a common ectodermal origin, thus it is possible that they might retain some common developmental features. Indeed, high expression of SOX10 is a feature of neoplastic transformation in cancers of all three tissues (possibly due to epigenetic dysregulation; Figure 3.2), and inappropriate reactivation of developmental/embryonic pathways is known to be a mechanism used in cancer to achieve critical growth advantages [252]. Given that SOX10 expression is inversely associated with differentiation in the melanocyte lineage, we hypothesise that it is associated with an undifferentiated phenotype in the normal breast. Since high SOX10 expression in basal-like breast cancer (BLBC) is associated with poor outcome, we propose that it is causally associated with aggressive/tumourigenic cell behaviour.

To address these hypotheses, we aim to:

i) Investigate SOX10 protein expression in normal breast tissue *in situ*.

ii) Investigate *SOX10* mRNA expression in published gene expression data from fluorescence-activated cell sorting (FACS)-derived normal breast cell compartments.

iii) Investigate SOX10 expression in a panel of breast and melanoma cell lines.

iv) Investigate the effect of short hairpin (sh) RNA-mediated *SOX10* knockdown on proliferation, migration, invasion and colony formation by breast cancer cells *in vitro*.

v) Investigate the effect of *SOX10* shRNA on tumour growth and metastasis *in vivo*.

4.3 Results

4.3.1 SOX10 expression in the normal breast

4.3.1.1 *In situ* analysis of SOX10 expression in the normal breast

A report from Nonaka et al., suggested that SOX10 expression is restricted to the myoepithelium [175]. However, a more recent study reported its expression in myoepithelial and some luminal cells [140]. To clarify this discrepancy, we investigated SOX10 expression by IHC analysis of 11 reduction mammoplasty specimens using commercially available antibody from Santa Cruz.

The results showed SOX10 nuclear expression in normal breast was heterogeneous in both ducts and lobules. Six of 11 cases (55%) showed both myoepithelial and luminal cell staining in the lobules of normal breast sections. Four cases (36%) showed combinations of staining within the same section i.e, some lobules stained only myoepithelial cells (Figure 4.2 A) and the others stained both myoepithelial and luminal cell (Figure 4.2 B). Only 1 case (9%) showed SOX10 staining in only myoepithelial cells.

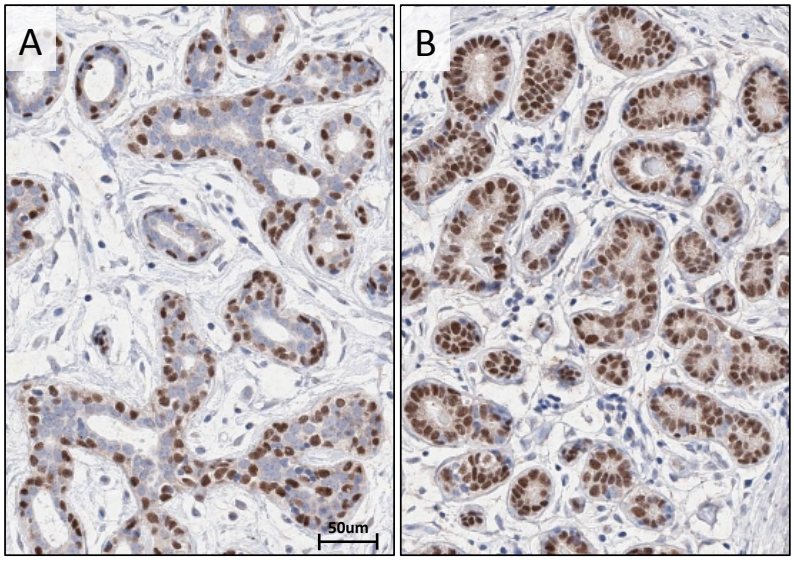


Figure 4.2: Representative images of SOX10 staining in normal breast sections (10X magnification) in **(A)** basal compartment (myoepithelial cell) and **(B)** luminal and basal compartment (myoepithelial/luminal cell) within the same case.

Next, we investigated if there was any difference in expression between ducts and lobular units. The percentage of luminal cells positive for SOX10 was calculated in ducts and lobules of 20 normal breast cases. Within the ducts, 70% cases were homogeneously positive and 20% expressed SOX10 heterogeneously (~40-90% luminal-positive). Ten per cent of cases (Q695 and Q725) were excluded due to a lack of ductal structures (Figure 4.3A). Within lobules, luminal expression of SOX10 was mixed. Of the cases containing lobular units (80%), ~30% were homogeneously positive, and ~50% cases showed mixed staining; with either 100% (positive), or 0% (negative) or 10-80 % (heterogeneous). Twenty per cent of cases (Q556, Q689, Q725 and Q778) were excluded due to a lack of ductal structures (Figure 4.3 B). When looked at the difference in the percentage of SOX10 positive luminal cells, lobules showed higher percentage compared to the ducts (Figure 4.3 C, p=0.0168).

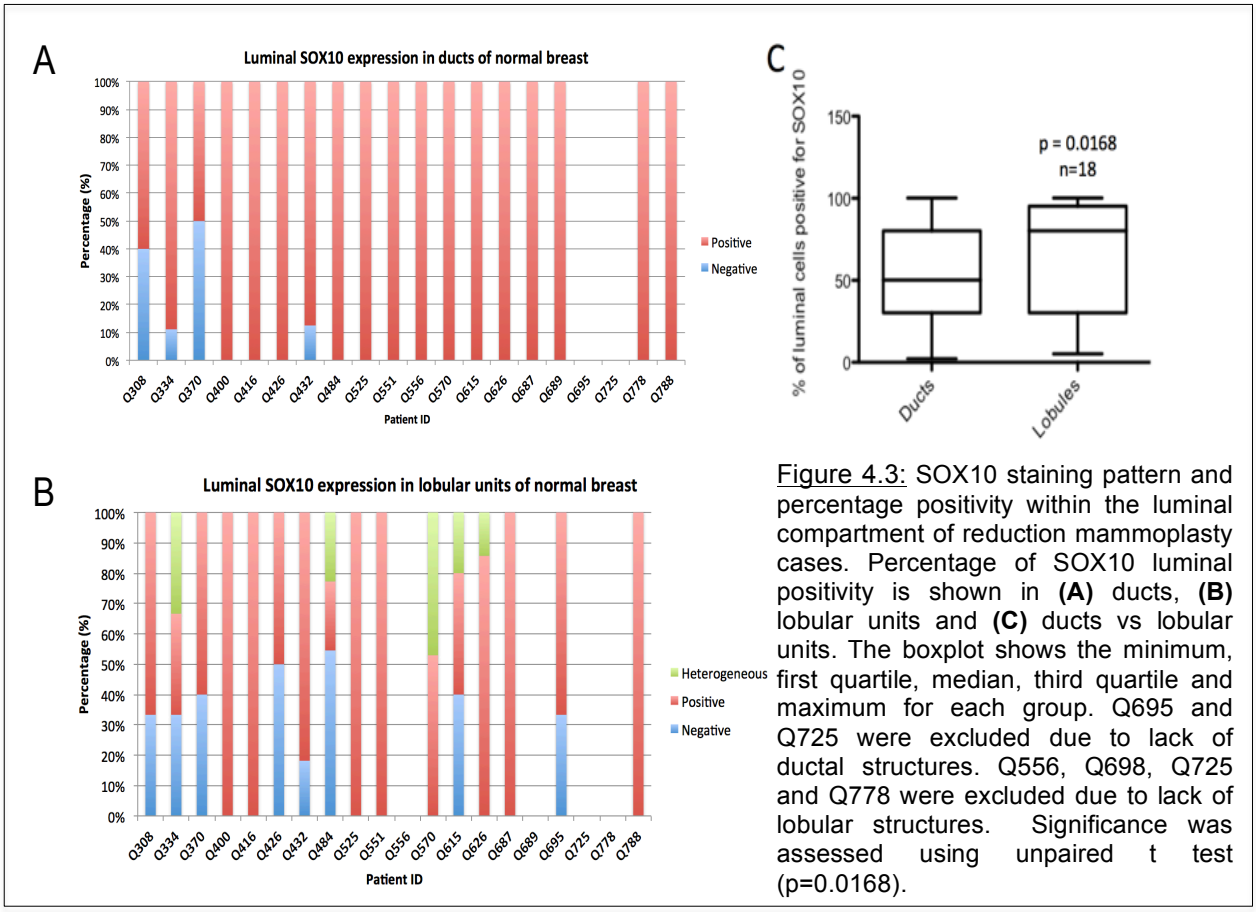


Figure 4.3: SOX10 staining pattern and percentage positivity within the luminal compartment of reduction mammoplasty cases. Percentage of SOX10 luminal positivity is shown in **(A)** ducts, **(B)** lobular units and **(C)** ducts vs lobular units. The boxplot shows the minimum, first quartile, median, third quartile and maximum for each group. Q695 and Q725 were excluded due to lack of ductal structures. Q556, Q698, Q725 and Q778 were excluded due to lack of lobular structures. Significance was assessed using unpaired t test (p=0.0168).

4.3.1.2 *SOX10* expression in mammary epithelial compartments

Tissue and cell type-specific expression of genes can often provide clues about their functions, as can an understanding which other genes are co-expressed. In 2009, Lim, Visvader and colleagues published a landmark paper describing a method for generating single cell suspensions from reduction mammoplasty tissues that were enriched for functionally distinct compartments of the normal breast [23, 24]. The method used fluorescence-activated cell sorting (FACS) to segregate

cells expressing different combinations of CD49f and EpCAM, characteristic of the myoepithelium and bipotent progenitor cells (CD49f^{+}/EpCAM^{-}), mature luminal (CD49f^{-}/EpCAM^{+}), luminal progenitor (CD49f^{+}/EpCAM^{+}) and stromal (CD49f^{-}/EpCAM^{-}) cells. Importantly, the authors also performed gene expression profiling analysis on the four compartments from three separate patients, and these data are now publically available.

SOX10 was moderately expressed in the Myo/MaSc compartment, consistent with our IHC data (Figure 4.2 A). Within the luminal compartment, where protein expression was heterogeneous by IHC, *SOX10* mRNA levels were significantly higher in luminal progenitor compared with mature luminal cells (Figure 4.4A), with expression levels similar to known luminal progenitor markers *KIT* [24] and *ELF5* [253] (Figure 4.4). Together these data suggest that *SOX10* could be involved in regulating epithelial differentiation in the breast.

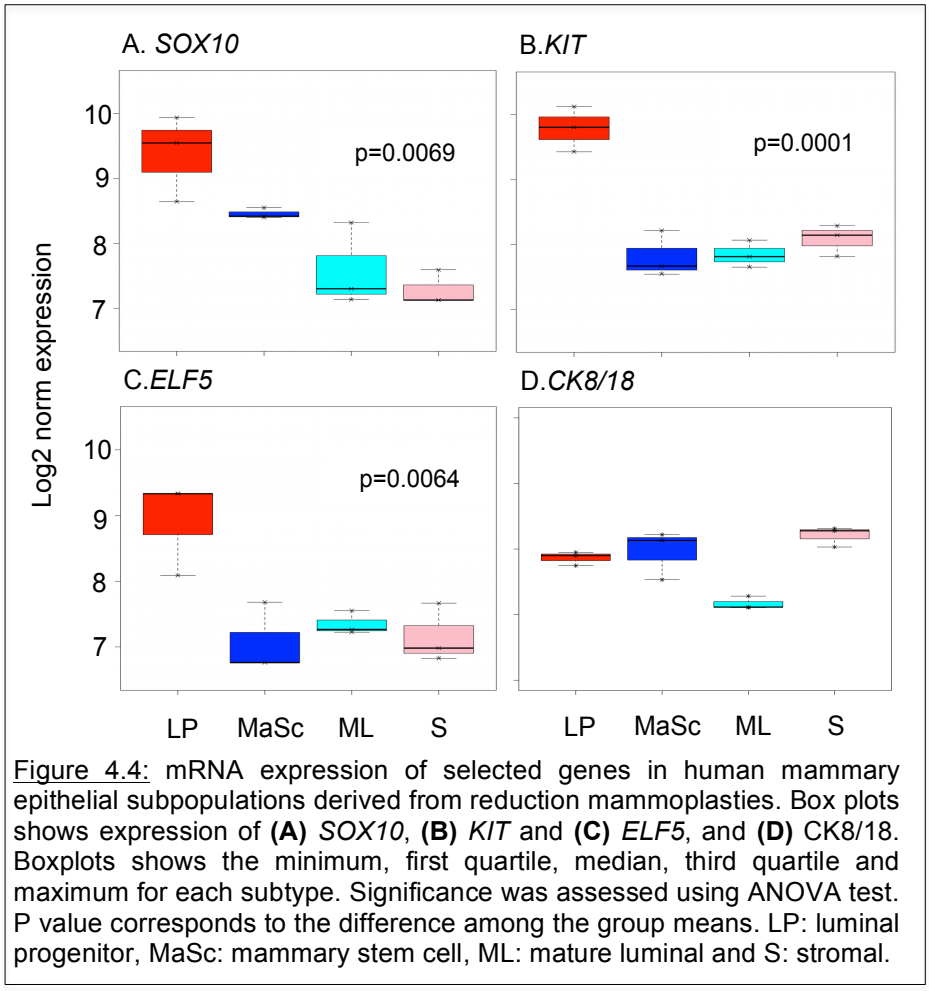


Figure 4.4: mRNA expression of selected genes in human mammary epithelial subpopulations derived from reduction mammoplasties. Box plots shows expression of **(A)** *SOX10*, **(B)** *KIT* and **(C)** *ELF5*, and **(D)** CK8/18. Boxplots shows the minimum, first quartile, median, third quartile and maximum for each subtype. Significance was assessed using ANOVA test. P value corresponds to the difference among the group means. LP: luminal progenitor, MaSc: mammary stem cell, ML: mature luminal and S: stromal.

Since SOX10 is associated with neural crest stem cell activity in development and in certain derivatives in the adult (*e.g.* melanocyte differentiation), we hypothesised that it may also be associated with stem/progenitor cell activity in the breast. The observation that SOX10 is highly expressed in luminal epithelia is consistent with this hypothesis, as the frequency of progenitor cells

expressed in the luminal compartment (specially in the lobules) is high during different stages of development (*e.g.* at puberty and during pregnancy) [254, 255].

In order to investigate the involvement of SOX10 in luminal differentiation, we analysed the relationship between SOX10, c-kit (luminal progenitor marker), ER (marker of mature epithelial cells) and Ki67 (marker of cell proliferation) by performing IHC analysis on serial sections of normal breast (11 of the most epithelial-rich samples from the cohort used in section 4.3.1.1). SOX10 expression in luminal cells showed a direct correlation with c-kit and an inverse correlation with ER (Figure 4.5) on a cell-by-cell level in both ducts and lobules. There was no relationship with Ki67 expression.

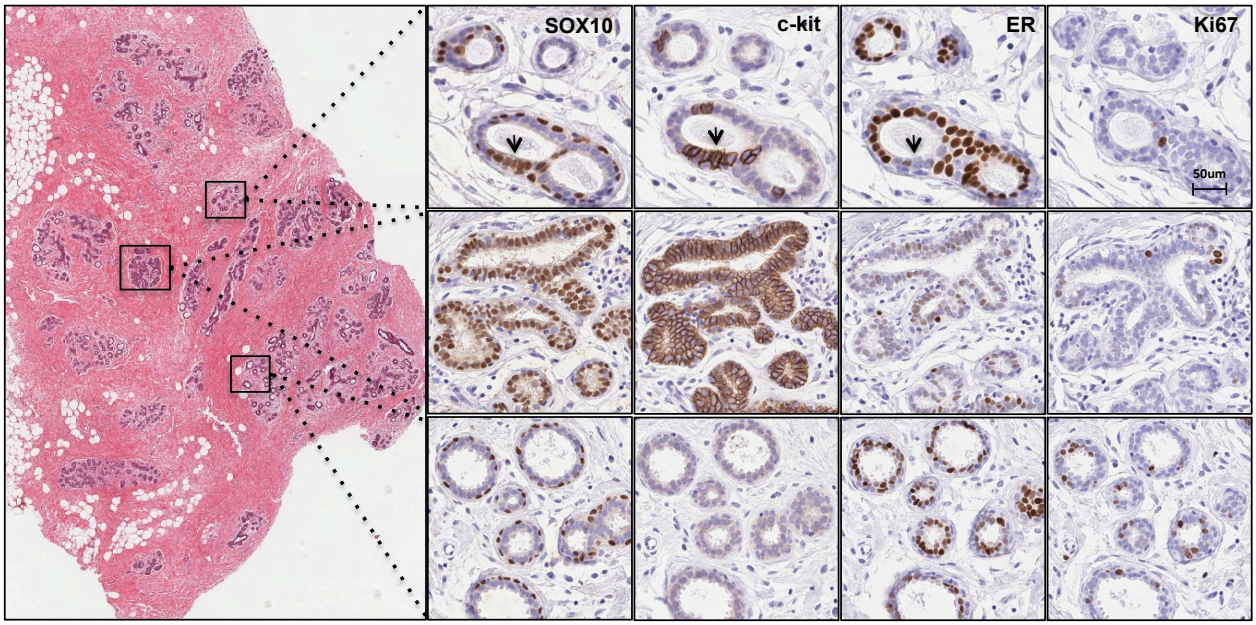


Figure 4.5: Light microscopy images of normal breast sections showing SOX10, c-kit, ER and Ki67 staining on serial sections (20X). Arrows indicate SOX10 positive luminal cells have a direct correlation with c-kit and an inverse correlation with ER. Images were scanned and analysed using Aperio software.

We also analysed SOX10, c-kit, ER and Ki67 staining on serial sections of normal breast associated with triple-negative breast tumours (n=19; described in chapter 3). Among these, 11 cases were SOX10 positive and the remaining were SOX10 negative. Compared to reduction mammoplasty specimens, similar staining patterns of SOX10, c-kit and ER were observed in the tumour-adjacent normal tissue (data not shown). However, we did not find any correlation with proliferation marker Ki67. These data further suggest that SOX10 maybe involved in regulating mammary epithelial differentiation.

We then analysed the relationship between expression of SOX10 and CK8/18, a marker of mature luminal cells [256] (Figure 4.6). For this analysis we used immunofluorescence to enable

multiplexing of the antibodies (dual analysis in the same cells; Figure 4.6 A and B), as well as IHC, using a high dilution of CK8/18 to make the analysis semi-quantitative (Figure 4.6 C and D). In lobules with a myoepithelial-restricted pattern of SOX10 expression (no luminal staining), the luminal compartment was CK8/18-positive (Figure 4.6Aii); whereas CK8/18 was weakly expressed in lobules with SOX10-positive luminal cells (Figure 4.6Bii).

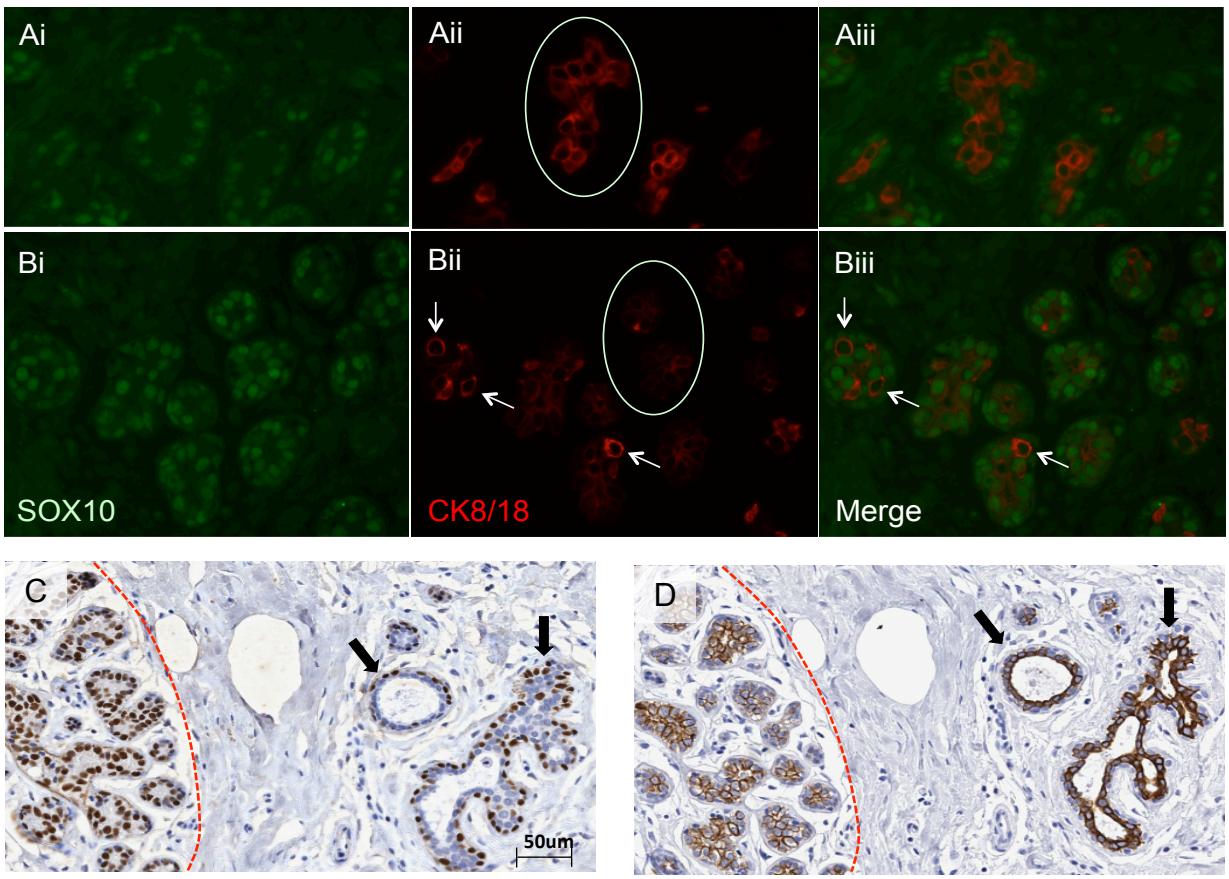


Figure 4.6: Immunofluorescent staining on FFPE sections **(A)** SOX10 staining (green) in myepithelial cells and CK8/18 staining (red) in mature luminal cells. **(B)** SOX10 staining in both myepithelial and luminal cells. Arrows indicate SOX10 negative CK8/18 postive cells. Circles highlight strong and weak CK8/18 in lobular units. Light microscopy images of **(C)** SOX10 nuclear staining and **(D)** CK8/18 cytoplasmic staning. Dotted red lines marks a lobular unit and arrows shows neighbouring duct.

Individual cells in SOX10 heterogeneous lobules also exhibited this inverse relationship (Figure 4.6Biii, arrows). IHC analysis showed SOX10-negative luminal cells (Figure 4.6 D) express higher levels of CK8/18 compared to SOX10-positive luminal cells (Figure 4.6 C) with weak staining of CK8/18. Arrows in the figure indicate presence of adjacent duct where SOX10 is negative in the luminal cells, while area marked in red indicates a lobule with SOX10-positive luminal cells. Based on these data we hypothesised that SOX10 may be involved in regulation of terminal differentiation in the luminal epithelial compartment of the normal breast.

4.3.2 *SOX10* expression in breast cancer cell lines.

In order to select an appropriate cell line model into investigate functional role of SOX10 in breast cancer, we looked at *SOX10* expression levels across a panel of breast cancer cell lines. In-house screening showed only 2/30 cell lines, which were basal-like (HCC1143 and MDA-MB-435) showed detectable levels of *SOX10* (Figure 4.7A) suggesting *SOX10* could be silenced and/or a SOX10-positive cell subpopulation could be lost during adherent *in vitro* selection. Melanoma cell lines were used as a control in our in-house screening experiment as we know that SOX10 is expressed in most melanomas (up to 97%). Cell culture conditions that promote de-differentiation (growth factor deprivation and suspension format) did not induce SOX10 expression in breast cancer cell lines (Figure 4.7C). This suggests that *in vitro* cell culture conditions promoting selection of stem/progenitor cells did not induce SOX10 expression in a panel of basal-like breast cancer cell lines.

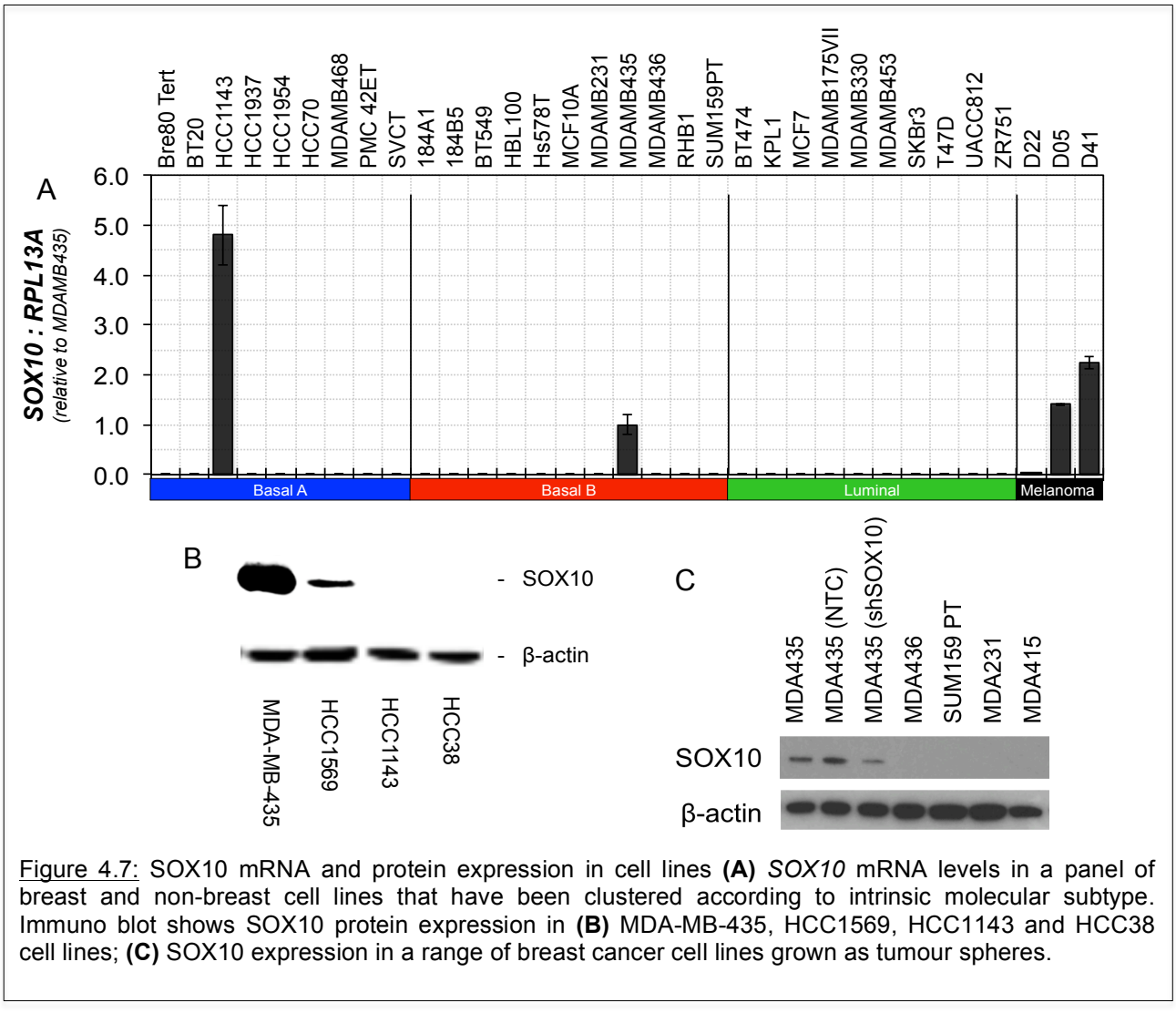


Figure 4.7: SOX10 mRNA and protein expression in cell lines **(A)** *SOX10* mRNA levels in a panel of breast and non-breast cell lines that have been clustered according to intrinsic molecular subtype. Immuno blot shows SOX10 protein expression in **(B)** MDA-MB-435, HCC1569, HCC1143 and HCC38 cell lines; **(C)** SOX10 expression in a range of breast cancer cell lines grown as tumour spheres.

Although in-house screening detected SOX10 mRNA in MDA-MB-435 and HCC1143 basal-like cell lines, analysis of protein expression by IHC (on cell-line TMA) and Western blot showed only MDA-MB-435 was positive for SOX10 (Figure 4.7B). We also tested the basal-like cell lines HCC1569 and HCC38 by Western analysis, since a published report showed SOX10 expression in these lines [140], but we could only detect SOX10 in HCC1569 (Figure 4.7B). Therefore, we used MDA-MB-435 and HCC1569 cell line as models to investigate the functional role of SOX10 in breast cancer.

4.3.3 shRNA-mediated knockdown of *SOX10* using lentiviral transduction

We used lentiviral transduction to establish stable shRNA-mediated knockdown of SOX10 in HCC1569 and MDA-MB-435 cells. We were able to generate two stable MDA-MB-435 sublines with very efficient knockdown (SOX10 protein virtually undetectable by Western blot), and one sub-line with partial knockdown (Figure 4.8A; red and black respectively). HCC1569 cells showed changes in cell morphology after viral transduction from cobble stone-like (Figure 4.8B) to mesenchymal/spindle-like (Figure 4.8C), and ultimately failed to proliferate and survive post-transduction. Therefore, we were not able to perform western blot, as there was not enough cells that survived for protein extraction. These results suggest that SOX10 is essential for cellular proliferation, growth and survival in HCC1569 cells. There was no difference observed in cellular morphology for MDA-MB-435 post transfection (hence, images not shown).

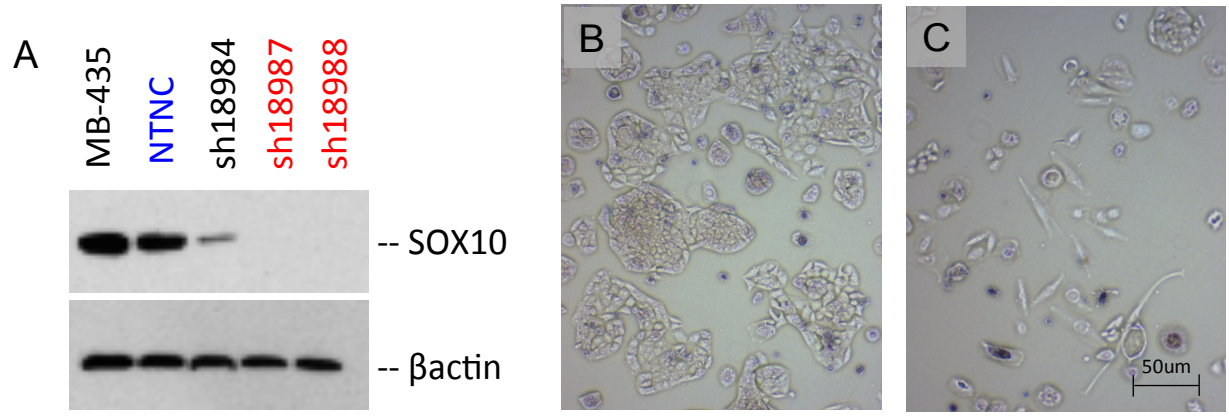


Figure 4.8: SOX10 knockdown in MDA-MB-435 and HCC1569 cell lines using lentiviral transfection system **(A)** Immuno bolt shows SOX10 expression in MDA-MB435 cell line and effective knockdown in shRNA transduced lines, NTNC was used as a non-target negative control. Change in cell morphology of HCC1569 cells from **(B)** cobble stone-like to **(C)** mesenchymal/spindle like.

4.3.3.1 Effect of SOX10 knockdown *in vitro* in MDA-MB-435 cell line

We next investigated whether loss of SOX10 had any effect on cell proliferation, migration and invasion. To assessing proliferation rate, cell growth was measured at regular time intervals based

on the area of confluence. To assess cell migration and invasion, the relative wound density (wound area expressed relative to the cell density outside the wound area) at was measured at regular time intervals. The results showed SOX10 knockdown did not alter the rate of proliferation (Figure 4.9A) or migratory behaviour (Figure 4.9B) compared to the non-target negative control (NTC/NTNC) (Figure 4.9).

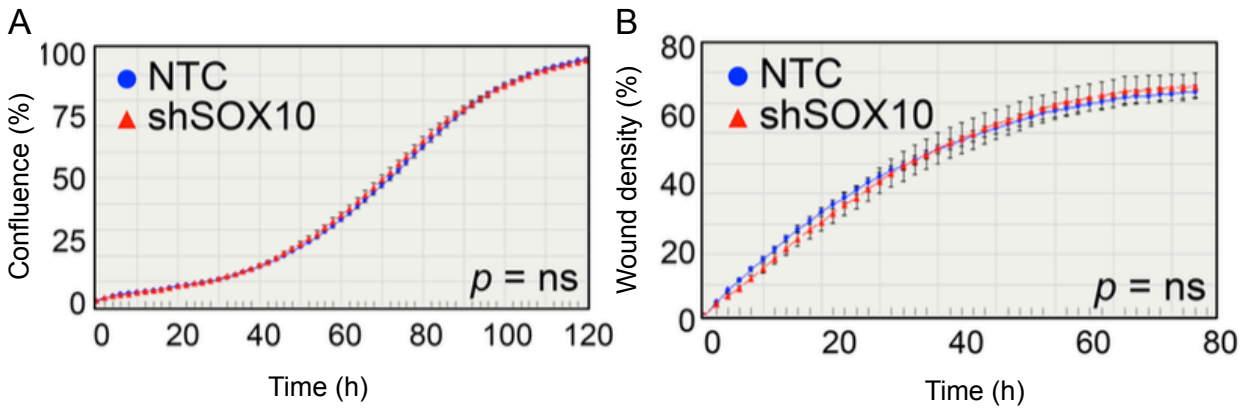


Figure 4.9: Effect of SOX10 knockdown on **(A)** proliferation and **(B)** migration between NTNC and shSOX10 in MDA-MB-435 cell line. Error bars represents standard error of mean (SEM). Significance was assessed using unpaired 't' test.

Given that SOX10 expression is inversely associated with differentiation in the melanocyte lineage and possibly in mammary epithelia (as seen in section 4.3.1.2), we investigated whether SOX10 knockdown affected colony and sphere forming capacities (as they mark an undifferentiated state) under normal culture. There was a significant decrease in the number of colonies formed in the knockdown line compared to the control line (Figure 4.10A, p=0.028). However, we did not see any difference in tumour sphere formation (Figure 4.10B).

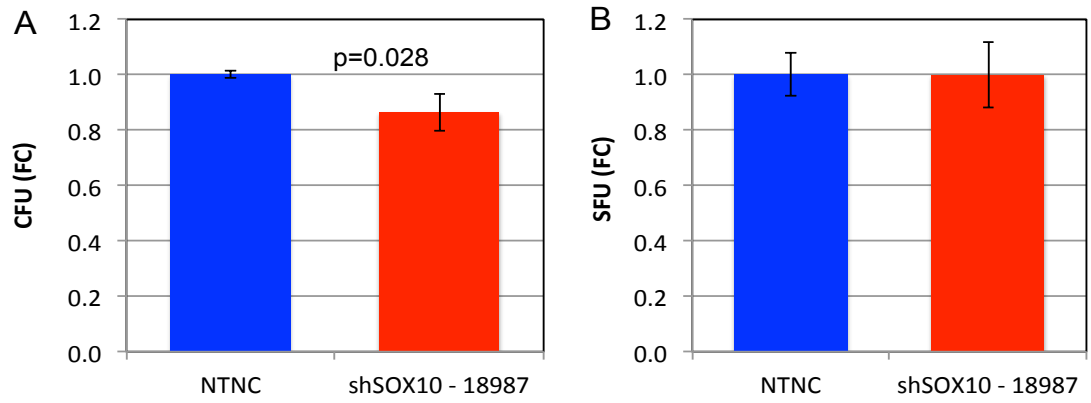


Figure 4.10: Effect of SOX10 knockdown on **(A)** colony formation and **(B)** sphere formation in MDA-MB-435 cell line. The number of colonies/spheres were normalised to NTNC. CFU: colony forming units and SFU: sphere-forming units. Error bars represents standard error of mean (SEM). Significance was assessed using unpaired 't' test.

4.3.3.2 Effect of SOX10 knockdown *in vivo*

Analysis of the relationship between SOX10 expression and outcome in breast cancer showed that high expression in basal-like breast tumours is associated with significantly poorer survival (chapter 3). Although, loss of SOX10 showed no effect on cell proliferation and migration *in vitro*, we wanted to investigate SOX10 loss in a physiological context using an *in vivo* xenograft model. Cells stably expressing the *SOX10* shRNA construct or the control were injected into the mammary fat pads of BALB/c *nu/nu* (immunocompromised) mice; five mice were used in each group. The mice were monitored for 30 days, and xenografts were regularly measured with vernier calipers. Tumour growth was calculated from calliper measurement of tumour's longest (a) and shortest (b) diameters using the formula, tumour volume (mm^3) = a/2 x b^2. The results showed that *SOX10* knockdown significantly impaired primary tumour growth (Figure 4.11, p=0.0013). These data suggest SOX10 expression has a significant impact on cell proliferation promoting tumour growth *in vivo*. The results showed that *SOX10* knockdown significantly impaired primary tumour growth (Figure 4.11, p=0.0013). These data suggest SOX10 expression has a significant impact on cell proliferation promoting tumour growth *in vivo*.

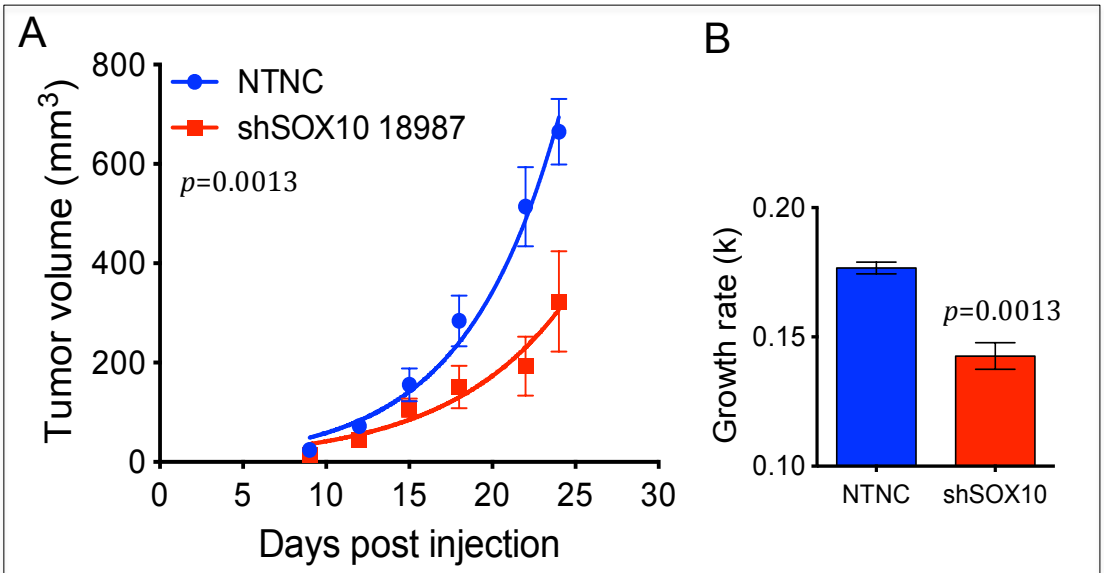


Figure 4.11: MDA-MB-435 cell line xenografts. (A) Effect of SOX10 knockdown on primary tumour growth and (B) Bar graph shows fold change in growth rate. Error bars represents standard error of mean (SEM). Significance was assessed using unpaired 't' test with Welch's corrections.

4.3.4 Differential gene expression analysis of MDA-MB-435 basal-like breast cancer cells with and without SOX10 silencing

In order to investigate possible mechanisms underlying the aggressive phenotype of SOX10 positive breast cancers, we performed comparative gene expression analysis of MDA-MB-435 cells

with (NTNC) and without SOX10 (shSOX10), using gene expression array profiling. This identified a total of ~257 differentially expressed genes (120 induced and 137 suppressed, with at least two-fold change in expression and a corrected p value of <0.05). Strikingly, the 15/20 differentially expressed probes (highlighted in blue) were highly enriched for genes in the MHC-II antigen presentation pathway (Table 4.1, 15 probes corresponding to 8 genes).

All these genes were significantly induced in shSOX10 cells (which were not seen in the wild type cells). Consistent with this, gene ontology analysis of the full list of differentially expressed genes showed that enrichment for the immune response antigen presentation by MHC class II, Cell adhesion glycoconjugates, immune response MIF mediated glucocorticoid regulation and immune response oncostatin M signalling via MAPK in human cells, immune response antigen presentation and inflammation IL4 signalling. Collectively these data implicate SOX10 in suppression of antigen presentation in MDA-MB-435 cells, raising the possibility that this may also occur in BLBC. Functional validation of this idea was outside the scope of this study, and further studies are required to fully understand the link between SOX10 and immunogenicity in BLBC.

PROBES	SYMBOL	GENENAME	logFC	AveExpr	t	P.Value	adj.P.Val
ILMN_2157441	HLA-DRA	MHC, class II, DR alpha	6.66	10.33	69.34	1.67E-16	8.05E-12
ILMN_1792455	TMEM158	transmembrane protein 158	5.56	10.68	60	8.86E-16	2.13E-11
ILMN_1689655	HLA-DRA	MHC, class II, DR alpha	5.02	10.77	50.33	6.70E-15	1.07E-10
ILMN_1761733	HLA-DMB	MHC, class II, DM beta	4.66	8.83	48.4	1.05E-14	1.27E-10
ILMN_3228688	HLA-DRB1	MHC, class II, DR beta 1	5.62	9.28	45.72	2.03E-14	1.94E-10
ILMN_3228688	HLA-DRB4	MHC, class II, DR beta 4	5.62	9.28	45.72	2.03E-14	1.94E-10
ILMN_3228688	HLA-DRB3	MHC, class II, DR beta 3	5.62	9.28	45.72	2.03E-14	1.94E-10
ILMN_1736567	CD74	MHC, class II invariant chain	4.77	8.66	45.01	2.42E-14	1.94E-10
ILMN_1772218	HLA-DPA1	MHC, class II, DP alpha 1	5.48	8.78	43.76	3.35E-14	2.30E-10
ILMN_2379644	CD74	MHC, class II invariant chain	4.2	8.04	42.86	4.26E-14	2.56E-10
ILMN_2066066	HLA-DRB6	MHC, class II, DR beta 6	5.03	8.81	41.36	6.41E-14	3.39E-10
ILMN_2066060	HLA-DRB6	MHC, class II, DR beta 6	4.75	8.68	41.02	7.04E-14	3.39E-10
ILMN_3243714	HLA-DRB1	MHC, class II, DR beta 1	4.17	8.19	40.35	8.51E-14	3.72E-10
ILMN_3243714	HLA-DRB3	MHC, class II, DR beta 3	4.17	8.19	40.35	8.51E-14	3.72E-10
ILMN_1703178	SCG2	secretogranin II	4.43	9.34	39.82	9.91E-14	3.97E-10
ILMN_1725139	CA9	carbonic anhydrase IX	3.69	8.36	39.4	1.12E-13	4.14E-10
ILMN_1813704	CEMIP	cell migration inducing protein	3.57	11.07	36.98	2.31E-13	7.54E-10
ILMN_1695311	HLA-DMA	MHC, class II, DM alpha	3.95	9.98	36.72	2.51E-13	7.54E-10
ILMN_1762899	EGR1	early growth response 1	-3.49	9.78	-36.79	2.45E-13	7.54E-10
ILMN_1752592	HLA-DRB4	MHC, class II, DR beta 4	3.77	9.52	36.22	2.94E-13	8.31E-10

Table 4.1: Differentially expressed probes in shSOX10 and non-target negative (NTNC) sublines of MDA-MB-435 breast cancer cells *in vitro* (e.g. mean of n=3 shSOX10 or NTNC samples were compared for each probe).

4.4 Discussion

Although, SOX10 expression is reported in several malignancies including melanoma, glioma, digestive cancers and breast carcinomas, very little is known about its functional role in cancers. SOX10 role has been broadly investigated in melanomas and studies using mice models have demonstrated that loss of SOX10 causes cell cycle arrest at G1 phase that SOX10 expression is fundamental for melanomagenesis [142, 144]. There is evidence that SOX10 controls invasion, migration in melanoma cells lines and is involved in metastasis in melanoma cases [251].

In breast cancer, we noticed a subset of basal-like breast cancer show high SOX10 expression and are linked to poor survival outcome. As there is limited data on SOX10 and its function in breast cancer and normal breast, we investigated its functional role in malignant and normal cells. Analysis of in-house screening experiments demonstrated infrequent expression of SOX10 in breast cancer cell lines, suggesting *SOX10* is silenced and/or a SOX10-positive cell subpopulation is lost during adherent *in vitro* selection. Another possible reason for difference in SOX10 expression in primary breast cancer cell lines could be due to the fact that the majority of cell line samples are derived from metastases. *In vitro* cell culture conditions promoting selection of stem/progenitor cells (suspension format and/or growth factor-deprivation) did not induce SOX10 expression in a panel of basal-like cell lines. shRNA-mediated knockdown of SOX10 in basal-like MDA-MB-435 breast cancer cells did not alter their proliferation rate, migratory behaviour or sphere formation *in vitro*, though did reduce colony formation on plastic (p=0.028). However, *in vivo* tumourigenicity assays showed that SOX10 knockdown in mice reduced tumour formation suggesting high SOX10 expression promotes tumour growth in mice.

TNBC are highly mutagenic [257] and are characterised by an increased number of tumour infiltrating lymphocytes, which can facilitate an immune response [258]. Cells with greater influx of tumour infiltrating lymphocytes respond better to neoadjuvant therapy compared to less immunogenic tumours [259]. Immunotherapy strategies in treatment of breast cancer has shown some hope in recent studies, for example, two new related classes of drugs called anti-PD-1s and anti-PD-L1s, also called as ‘immune checkpoint blockers’ have shown some promising results in breast cancer. PD-1 inhibitor, pembrolizumab in a phase Ib KEYNOTE-012 trial demonstrated an overall response rate of 18.5% in patients with PD-L1–positive TNBC (unpublished, San Antonio Breast cancer Symposium 2015). In the current study, we found that differential gene expression analysis using MDA-MB-435 cell line showed enrichment for immune related genes post SOX10 knockdown suggesting a link between SOX10 and immunogenicity. This suggests that targeting SOX10 may improve survival outcome in a subset of TNBC/BLBC.

In normal breast, SOX10 expression was initially described to be restricted to the myoepithelial compartment [175]. However, a more recent study by Ivanov et al., reported SOX10 expression in myoepithelial and in some luminal cells [140]. Our analysis in both normal and tumour-associated normal breast tissue, agreed with the findings of Smith and colleagues that showed both myoepithelial cells and luminal epithelial cells were positive for SOX10 in normal ducts and lobular units. Based on IHC and IF results, we noticed SOX10 expression is heterogeneous within normal breast structures. Regardless of ducts and lobules, SOX10 expression was not only restricted to myoepithelial cells but is also expressed in luminal epithelial cells. In this chapter, *in silico* and *in situ* analysis confirmed a direct correlation of SOX10 expression with luminal progenitor marker c-kit. Complementary to our findings, a recent study has reported that SOX10 is specifically expressed in mammary cells exhibiting the highest levels of stem/progenitor activity [260]. These data, together with our data on clonogenicity and knowledge of SOX10's role in progenitor function in other tissues (*e.g.* melanocytes), we hypothesise that SOX10 could be involved in progenitor function in basal-like breast cancers. Further functional studies are required to fully understand the role of SOX10 in breast cancer and normal breast (see chapter 6 for future directions).

Chapter V

Analysis of prognostic significance of MAGA-A and NYESO-1 cancer testis antigens in breast cancer

Chapter V: Analysis of prognostic significance of MAGE-A and NY-ESO-1 cancer testis antigens in breast cancer

5.1 Background

Melanoma-associated antigen A (MAGE-A) was identified in melanoma cells as a testis tumour antigen gene [193]. To date, ~60 *MAGE* genes have been discovered, which all share a central conserved region called the MAGE homology domain (MHD), and are subdivided into groups I and II based on their protein expression patterns [261]. The MAGE-A subfamily includes MAGE-A1 to A15 and all known genes of this family are located on chromosome Xq28. Each gene is ~4.5 kb, has three exons and encodes a protein of 309-319 amino acids (with the exception of MAGE-A10, which is slightly larger at 369 amino acids) [262].

Cancer testis antigen 1B (CTAG1B), also known as, NY-ESO-1 is another prominent CT antigens located on the X-chromosome. It spans about 1.6 kb, has 3 exons and encodes a protein of 180 amino acids [263]. NY-ESO-1 is expressed in variety of human cancers [263] but in the normal tissues it is primarily limited to the testis and ovaries. NY-ESO-1 is thought to be one of the most immunogenic CT antigens, and is capable of inducing spontaneous humoral immunity in patients whose tumors express this antigen [200, 264, 265]. Because of this property, NY-ESO-1 has been an attractive candidate for immunotherapy.

Amongst the studies evaluating MAGE-A and NY-ESO-1 expression in breast cancer, there have been contradicting reports on the expression frequency in breast cancer, with MAGE-A-positivity ranging from 20-74% and NY-ESO-1 from 2-40% [196-200]. One possible explanation could be variation in the size and composition of the tumour cohorts (for example, the cohort used in [197] was enriched for ER-negative and basal-like tumours). Other factors that varied among these studies were the use of different antibodies and their antigen retrieval techniques (reviewed in [201]).

At the molecular level, MAGE proteins promote the viability of malignant melanoma cell lines by suppressing apoptosis [266, 267]. Several MAGE-A proteins (1-3 and 6) inhibit p53 transactivation by recruiting histone deactylase 3 (HDAC3) to sites of p53 interaction at the promoters of target genes, leading to resistance to chemotherapies that promote apoptosis? (*e.g.* etoposide) [266]. Mechanistically, MAGE-A suppresses the p53 transcriptional program during tumour development by interacting with its DNA-binding domain, blocking its association with cognate sites in chromatin. Silencing MAGE-A induces p53-responsive genes in a p53-dependent manner and stimulates interaction of p53 with the p21, MDM2, and PUMA promoters [268].

CT-X antigen expression in breast cancer was interrogated by meta-analysis of a large massively parallel sequencing dataset (nine studies, 1259 breast tumours and 51 breast cancer cell lines) [197]. Expression of some MAGE-A genes was associated with ER-negativity (MAGE A3, A6 and A9), and others were specifically associated with the basal-like subgroup (A1, A4-6, A12 and NY-ESO-1-1), and these relationships were validated at the protein level by TMA-based IHC analysis of two separate breast cancer cohorts. Other large cohort studies have also confirmed these associations [269, 270]. Since there are currently no molecular-targeted therapy options for TNBC, and these cancers often carry a worse prognosis, it has been proposed that CT-X immunotherapy could provide therapeutic benefit for some TNBC patients [201]. Although published data show that MAGE-A expression is associated with TNBC, it is currently unknown whether it stratifies outcome in this group. A recent study assessing MAGE-A expression in breast cancer by qRT-PCR reported a trend towards poorer survival in MAGE-A 3/6-high, hormone receptor-negative cases, however the association was not statistically significant [271].

In ER-positive breast cancer, gene expression profiles of ER^+ Tamoxifen resistant (TR) breast tumour-derived lines identified and functionally validated MAGE-A2 as a novel TR-associated gene [272]. MAGE-A2 localises to the nucleus and forms complexes with p53 and ERα in ER positive Tamoxifen-resistant cells, resulting in repression of the p53 pathway but increased ER-dependent signalling [272]. Expression of MAGE-A2 in 144 ER^+ TR breast cancers was assessed by IHC. Thirty-five showed high expression of MAGE-A2 and positive staining was significantly associated with reduction in overall survival in this cohort [272], suggesting a link between expression of MAGE-A proteins and failure of Tamoxifen therapy.

5.2 Hypothesis and aims

The main focus of this thesis has been investigating novel prognostic biomarkers within hormone receptor negative breast cancers. Published data suggest MAGE-A and NY-ESO family proteins show an association with high-grade triple negative breast cancers. MAGE-A interacts directly with p53 *in vitro*, resulting in repression of the p53 pathway whereas, very little is known about NY-ESO. Others members in our lab have examined the relationship between MAGEA and NY-ESO-1 in breast cancers, and identified an association of MAGE-A with p53 expression. Although, mRNA expression of MAGE-A in ER negative and protein expression MAGE-A in ER positive breast cancers are linked with poor outcome, it is currently not known whether MAGE-A and NY-ESO-1 protein expression is related to clinical outcomes within triple negative breast cancer. Therefore, we hypothesize that i) MAGE-A and/or NY-ESO-1 expression can stratify human breast cancers into

clinically meaningful groups and that ii) MAGE-A family proteins could be potential therapeutic target in triple negative breast cancers. To address this, we aim to

i) Investigate MAGE-A and NY-ESO-1 expression in a larger cohort of invasive breast cancers (*n*=449) with long-term follow-up (25 years, QFU cohort) in order to correlate with clinical outcome

ii) Associate MAGE-A and NY-ESO-1 expression with biomarkers used either clinically or conceptually

iii) Investigate prognostic significance of MAGE-A and NY-ESO-1 in breast cancer cohort

5.3 Results

5.3.1 MAGEA and NY-ESO-1 expression in breast cancer

Using commercially available antibodies, we optimised IHC staining conditions using a small panel of normal tissues, and selected conditions that produced the strongest signal in testis compared to other normal tissues (*e.g.* intestinal, lymphoid and epithelia). The MAGE-A antibody used is known/predicted to bind to A1-4, A6, A10 and A21 MAGE-A proteins. When staining pattern was investigated in QFU tumour cohort, MAGE-A was characterised by localisation to the nucleus and cytoplasm of tumour epithelial cells. Among positive cases, 33% stained exclusively in the cytoplasm, 33% were exclusively nuclear and 34% exhibited both nuclear and cytoplasmic staining. The intensity of nuclear expression was recorded as either negative (n=364), weakly positive (1+; n=2), moderate (2+; n=7) or strongly positive (3+; n=22) (Figure 5.1). The intensity of cytoplasmic expression was recorded separately as either negative (n=364), weakly positive (1+; n=7), moderate (2+; n=5) or strongly positive (3+; n=3).

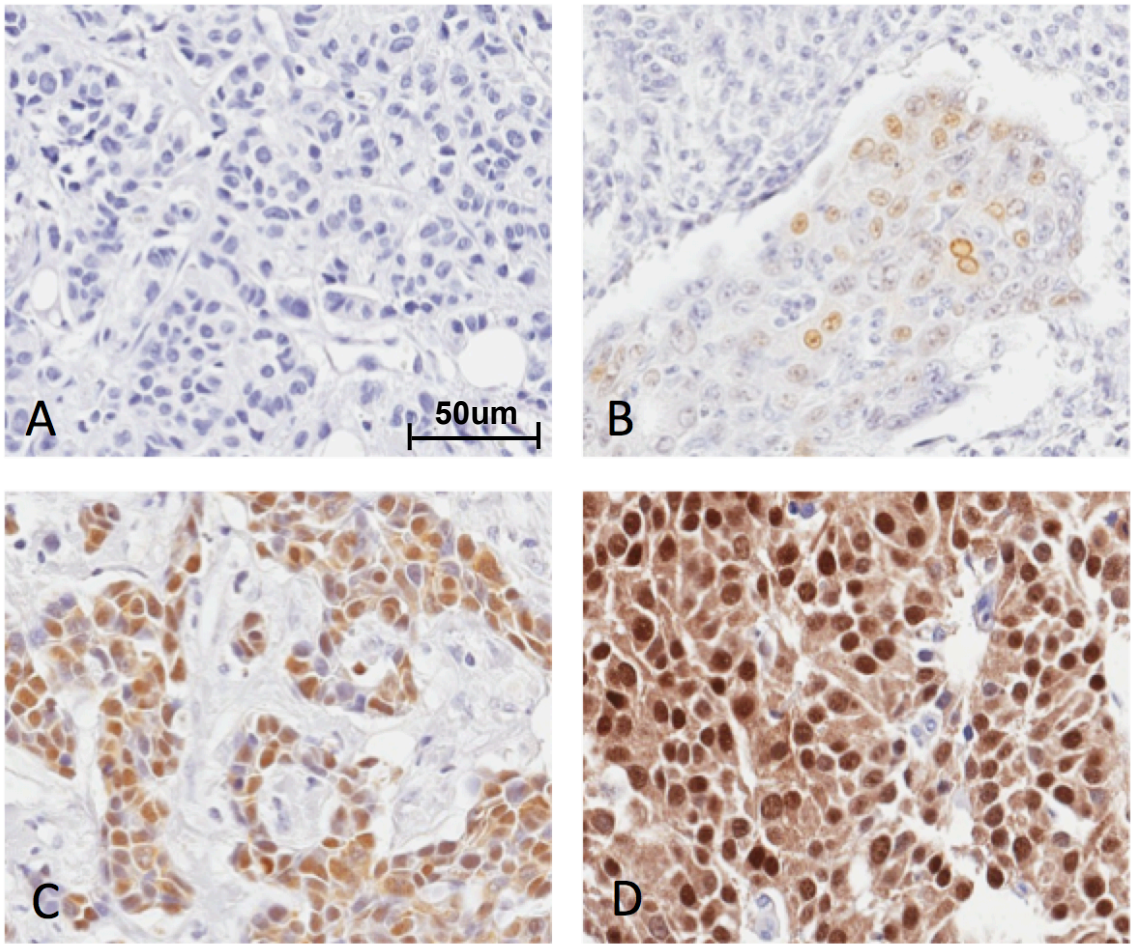


Figure 5.1: Representative images of MAGE-A staining on breast tumour cases in the QFU cohort showing **(A)** negative **(B)** weak positive **(C)** moderately positive and **(D)** strong positive staining.

NY-ESO-1 staining was also characterised by localisation to the nucleus and cytoplasm of tumour epithelial cells in the QFU tumour cohort. Only 14/403 cases were positive for NY-ESO-1, of which two were exclusively cytoplasmic and 12/14 exhibited both nuclear and cytoplasmic staining. The intensity of expression was recorded as either negative (n=391), weakly positive (1+; n=4), or strongly positive (2+; n=8) (Figure 5.2). The percentage of tumour nuclei positive was categorised as 1 (<33 % tumour cell nuclei stained), 2 (33-66%) or 3 (>66%). Fifty precent cases (6/12) were in category 3, 33% (4/12) were in category 2 and 17% (2/12) were in category 1.

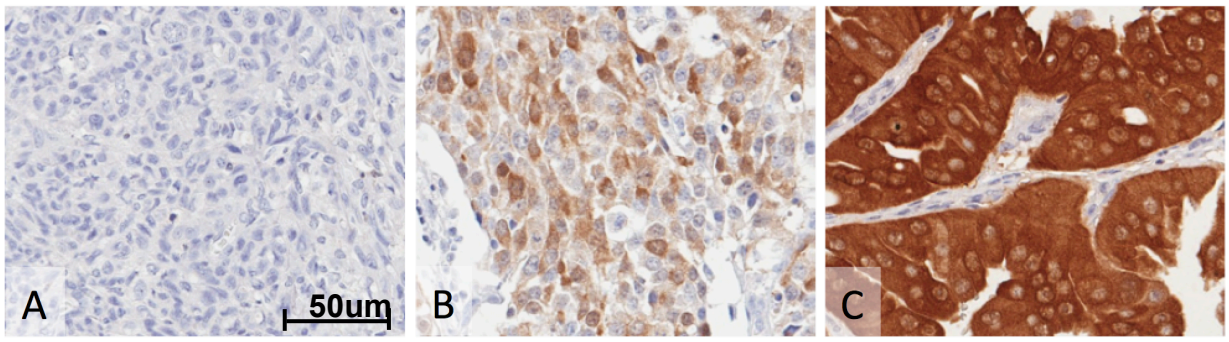


Figure 5.2: Representative images of NY-ESO-1 staining on breast tumour cases in the QFU cohort showing **(A)** negative **(B)** weak positive and **(C)** strong positive staining.

5.3.2 Distribution of MAGE-A and NY-ESO-1 expression across breast cancer subtypes

Chi-square analysis showed a positive association between MAGE-A and NY-ESO-1 expression with the triple-negative phenotype (p<0.0001; Figure 5.3), which is consistent with published data for MAGE-A and NY-ESO-1 [201, 270, 273]. MAGE-A positivity was observed in ~34% (18/53) TNBCs with basal-like features, 27% (4/15) of non-basal TNBCs, 4% (9/241) of $HR^{+/}HER^{-/+}$ ($Ki67^{low}$) and 8% (4/48) $HR^{+}/HER2^{-}$ ($Ki67^{high}$) cases. NY-ESO-1 positivity was restricted to ~21% (11/52) of basal-like TNBCs.

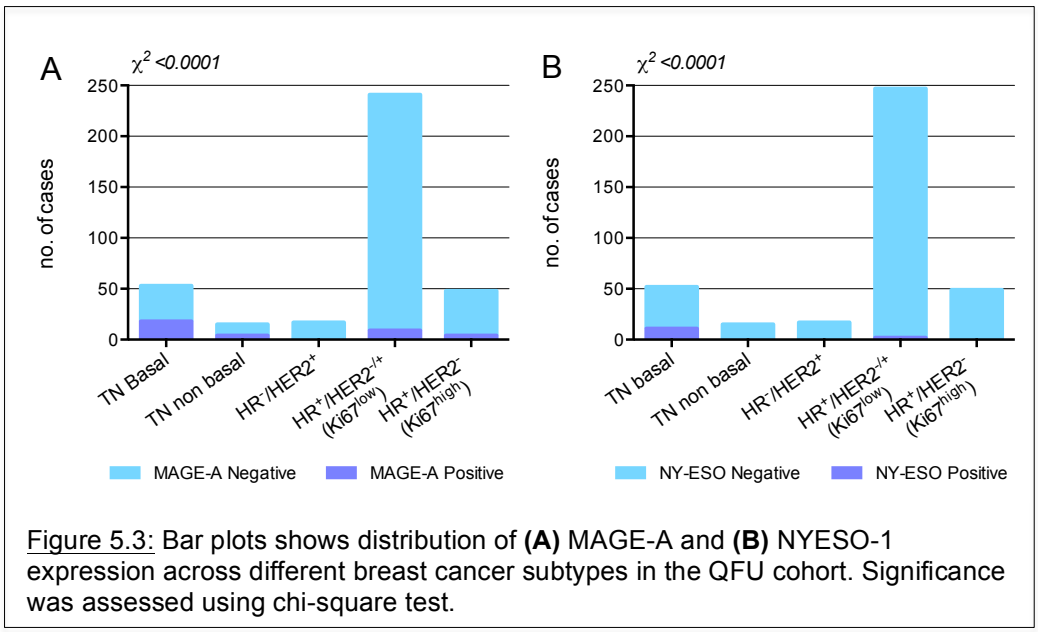


Figure 5.3: Bar plots shows distribution of **(A)** MAGE-A and **(B)** NYESO-1 expression across different breast cancer subtypes in the QFU cohort. Significance was assessed using chi-square test.

5.3.3 Comparison of MAGE-A expression with breast cancer clinicopathologic variables

Across all invasive breast cancers, regardless of molecular subtype, Chi square analysis showed a strong positive association between MAGE-A nuclear expression and markers of poor prognosis (grade, mitotic score, HR status, triple-negative status and Ki67 expression). Among different

biomarkers tested, MAGE-A expression showed a positive correlation with basal markers (CK5/6, CK14 and EGFR), p53 expression, c-kit, Vimentin, and an inverse correlation with androgen receptor (AR), (Figure 5.4 and 5.5, and Appendix 5.1). There was no association with age, tumour size, LN or LVI (lymph node status, lympho-vascular invasion). Within TNBCs, there was no association between MAGE-A nuclear expression and known clinicopathologic variables, though interestingly, there was an association with the luminal progenitor marker, c-kit (Appendix 5.2).

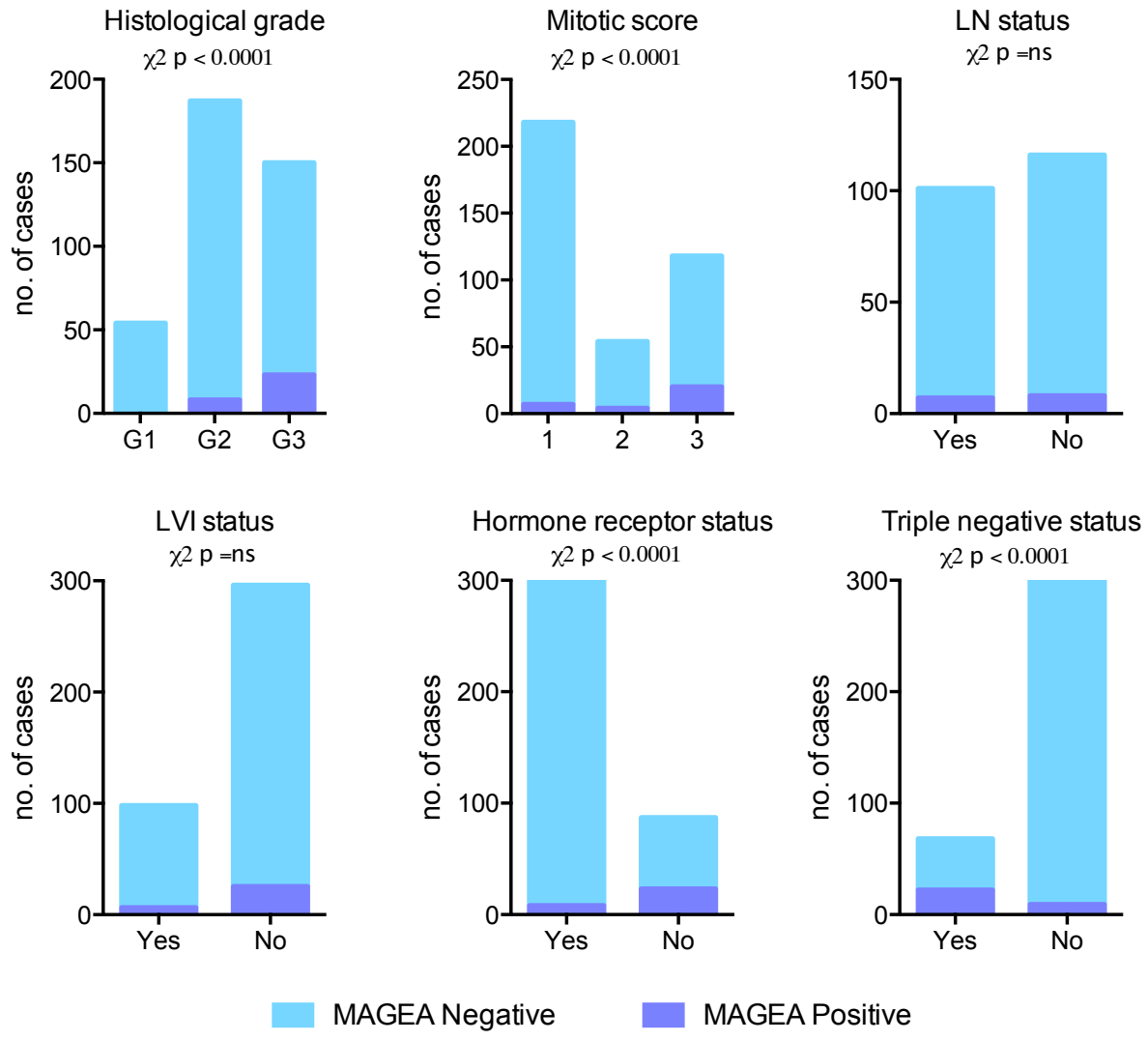


Figure 5.4: Association between MAGE-A expression with clinical and histopathological indicators in breast cancer. Significance was assessed using chi-square test.

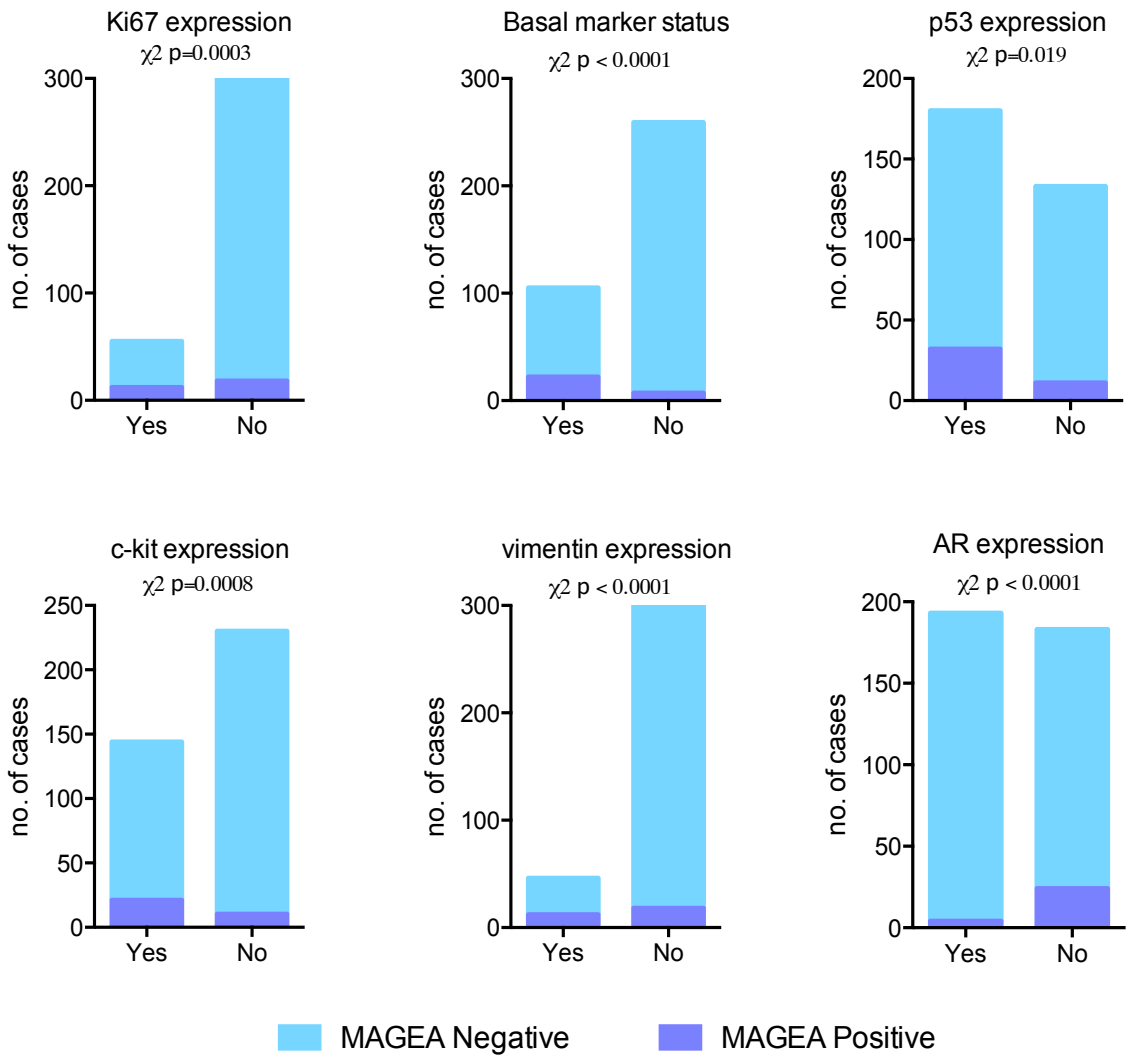


Figure 5.5: Association between MAGE-A expression with biomarkers such as Ki67, basal markers (CK5/6, CK14, EGFR), p53, c-kit, vimentin and androgen receptor (AR). Significance was assessed using chi-square test.

5.3.4 Prognostic significance of MAGE-A in breast cancer

Kaplan-Meier analysis for MAGE-A nuclear-positive cases within the whole cohort (n=426) showed a 40% decrease in disease-specific survival at 5 years (Figure 5.6A, p=0.0035). However there was no stratification *within* TNBCs and basal-like TN (Figure 5.6B and C), suggesting that in published reports where MAGE-A family genes was suggested to be a prognostic indicator in breast cancer based on its mRNA expression [271], it could be simply acting as a surrogate marker for poor outcome TNBCs. Also, the number of TN samples may not be sufficient to determine if MAGE-A can alter disease outcome. There was no relationship with survival in TNBC when we analysed nuclear and cytoplasmic expression separately (data not shown).

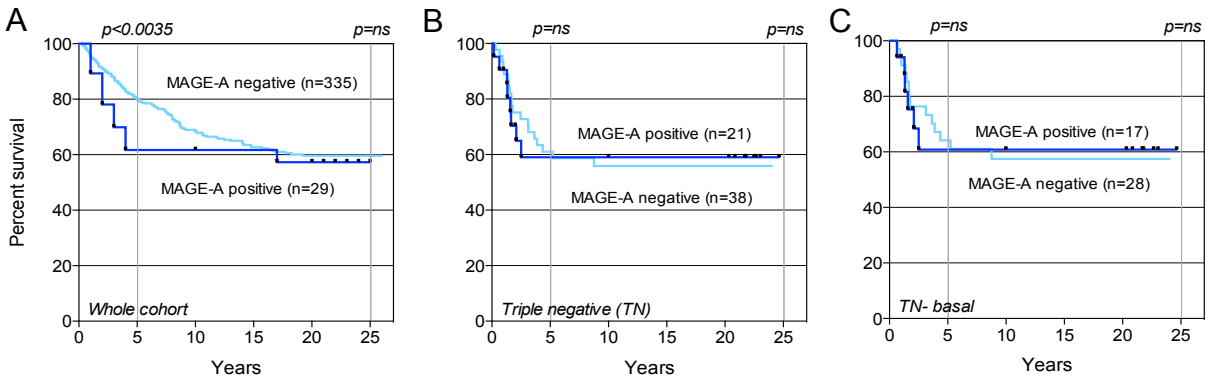


Figure 5.6: Kaplan-Meier survival analysis of MAGE-A negative and positive cases within **(A)** whole cohort **(B)** triple negative and **(C)** triple negative (TN) basal-like breast cancers. Significance was assessed using Gehan-Breslow-Wilcoxon test.

5.3.5 Characteristics of NY-ESO-1 positive tumours and correlation with survival outcome

NY-ESO-1 expression in this study was observed in a small group of basal-like TNBCs (14/403 cases; ~3%). The clinical characteristics of NY-ESO-1 positive tumours are presented in Appendix 5.3. All positive cases were either grade 2 or 3, 68% were above the age of 50 years, 27% showed high Ki67 expression (>20%) and almost all cases were positive for MAGE-A and p53. Analysis of the relationship between NY-ESO-1 expression and survival showed a trend towards poorer outcome in the whole cohort (QFU) and the TNBC subgroup, which was interesting. However, these relationships were not statistically significant as the numbers were too small for any statistical power. These findings should however be followed up in a larger breast cancer cohort in future (Figure 5.8).

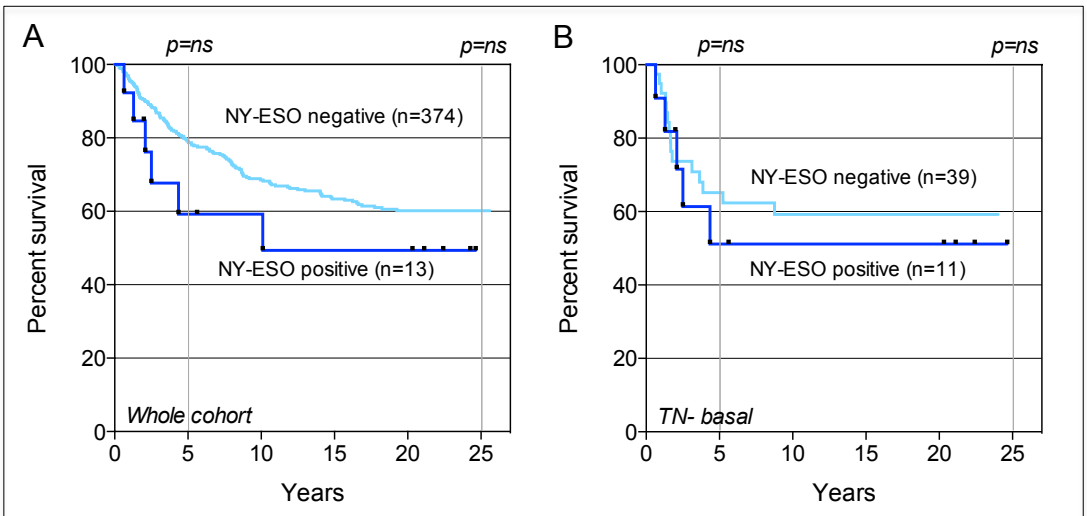


Figure 5.7: Kaplan-Meier survival analysis of NYESO-1 negative and positive cases within **(A)** whole cohort and **(B)** triple negative (TN) basal-like breast cancers. Significance was assessed using Gehan-Breslow-Wilcoxon test.

5.4 Discussion

Triple negative breast cancers (TNBC) constitute a heterogeneous group of tumours. The clinical outcome of patients with this subgroup of breast cancer are often linked to poor outcome compared to those with hormone receptor positive breast cancer that generally have better outcome. There are currently no molecular targets available for treatment of TNBC and unfortunately chemotherapy is the only standard treatment given to these patients. Anti-cancer vaccines may help in the management of these patients if a suitable target antigen/s and patient population can be identified. Cancer testis (CT) antigens are a group of protein antigen that are predominantly expressed in human germ line cells and become activated frequently in different cancer types. Numerous CT antigens have proved to be useful markers and promising therapeutic targets for cancer vaccines.

A study from our group has previously reported that expression of CT-X antigens, *MAGE-A* and *NY-ESO-1* was observed in ER^- breast cancer cell lines and primary breast cancers using gene expression analysis [197]. These results were in concordance with protein expression of MAGE-A and NY-ESO-1 in ER^- breast cancers, and showed a trend for their co-expression with basal cell markers [197]. Although, several published studies have shown expression of MAGE-A and NY-ESO-1 in breast cancer, results from these studies have shown contradicting results with MAGE-A positive rate ranging from 20-74 % and NY-ESO-1 positive rate ranging from 2-40 %. Subsequent studies with large cohorts have shown members of MAGE-A family and NYESO-1 expression to be correlating with high grade TNBC [269, 270]. Matkovic and colleagues have reported expression of MAGE-A and NY-ESO-1 in medullary breast carcinomas and the results from this study showed, 33% of cases were positive for MAGE-A and positive cases were associated with poor prognosis (p=0.004). Whereas, NY-ESO-1 expression in the same cohort was reported in 22% cases and also showed a trend towards poor prognosis (p=0.007) [274]. Currently, there is only one study that has reported most rigorous investigation of NY-ESO-1 expression in a larger cohort of TNBC (n=168) and showed 16% of triple negative breast cancers (TNBC) were positive for NY-ESO-1. There was no association with survival outcome between positive and negative cases in this study; on the other hand, prognostic significance of MAGE-A protein expression in TNBC has not been investigated so far. In this study, we have investigated frequency of MAGE-A and NY-ESO-1 expression in a large cohort of invasive breast cancers and also investigated prognostic significance of MAGE-A and NY-ESO-1 expression specifically in TNBC.

Results from this study showed, MAGE-A nuclear expression strongly correlated with markers of poor prognosis such as grade, mitotic score, hormone receptor (HR) status, triple negative status, and Ki67 expression within a large cohort of invasive breast cancers. Across all invasive breast

cancers, MAGE-A were predominantly expressed in a subset of TN basal-like (48%) and TN non-basal breast cancers (33%). Although, Kaplan-Meier analysis for MAGE-A nuclear positive cases within the whole cohort (n=426) showed a 40% decrease in the disease specific survival at early stages within 5 years (p=0.0035), there was no difference in survival when specifically looked at TN and TB-basal group in this cohort. NYESO-1 expression on the other hand was observed in 2% of all invasive breast cancers and was restricted to a subset of TN-basal like breast cancers (21%). Kaplan-Meier analysis of NY-ESO-1 positive expression showed a trend towards poor outcome within the whole cohort and TN-basal group but the results were not significant in this cohort. In conclusion, these results suggests that it is still unclear whether cancer antigen expression particularly MAGE-A and NY-ESO-1 contributes to tumourigenesis, as the sample size in TN group may not be sufficient to determine if MAGE-A can alter disease outcome. Further investigations are required on larger cohort to fully understand the role of MAGE-A and NY-ESO-1 in breast cancer.

Chapter VI

General discussion

Chapter VI: General discussion and future directions

Breast cancer is the most common cancer in Australian women and is the second leading cause of cancer related death worldwide. The incidence of breast cancer in Australia is increasing however; the mortality rate has declined for the last three decades as a result of a collection of measures including improvements in screening, local management and the development of targeted therapies for hormone receptor positive and HER2 positive disease [275].

Breast cancer is a complex heterogeneous disease encompassing different tumour entities with different biological features and clinical outcomes. Management decisions are based on clinic-pathologic parameters including age, tumour size, histological grade, clinical stage, and expression of ER, PR, and HER2. Treatment of breast cancer requires a multidisciplinary approach, including surgery, radiation and systemic therapy and may vary depending on the stage, histological type. Women with ER and/or PR positive tumours have better overall survival and DFS at 5 years compared to women with ER negative tumours. The presence of ER and/or PR is a powerful predictive factor for the likelihood of benefit from adjuvant endocrine therapy. Similarly, studies have shown that HER2 overexpression predicts response to trastuzumab. Therefore, the presence of ER, PR and HER2 has proven to be both prognostic and predictive of therapeutic response in invasive breast cancer. In contrast, management of patients with TNBC has been challenging because of the lack of therapeutic targets. Cytotoxic chemotherapy (anthracycline and taxane) is currently the only standard treatment options for treating this particular group. Although PARP inhibitors have shown favourable results in some TNBC patients with germline *BRCA* mutations [80, 81], assessment of its long-term effects is needed. Other therapeutic targets (*e.g.* inhibitors of EGFR, c-kit, Raf/MEK/MAPK and mTOR) have been developed, but their clinical benefits are limited (reviewed in [82, 83]).

There is significant overlap between TN and BLBC; studies have shown that around 75% of TN cancers have basal-like phenotype. A subset of TNBC/BLBCs patients have a high risk of relapse including development of lung and brain metastases and poor outcome within five years after diagnosis. Long-term survival rates for the remaining TN/BLBC patients are actually very good [77, 94, 222]. Current clinicopathologic markers, including grade, tumour size and lymph node status, do not fully explain this dichotomous clinical behaviour. Recent molecular studies have highlighted there is significant heterogeneity even within TN/BLBCs and these tumours show different pCR rates (after neoadjuvant therapy) and clinical outcomes [97, 98, 117]. Therefore, it would be favourable to develop methods to predict the risk of relapse earlier in the course of disease

to inform clinical management and to provide an appropriate level of surveillance for metastatic disease.

The primary aim of this study was to therefore to identify new prognostic and/or predictive biomarkers within TN/BLBCs to help improve the management of this group of patients in the future. The main focus was on SOX10, but we also investigated the CT-X antigens MAGE-A and NY-ESO-1. Some of the key findings from our immunohistochemical analysis of SOX10 in a cohort of 449 sporadic breast cancers with long term follow up are reiterated below:

i) the vast majority of breast tumours are uniformly negative for SOX10 protein, which appears to be related to hypermethylation of the *SOX10* gene;
ii) positive staining of SOX10 protein, depicted as nuclear localisation, was detected in 8% of all tumours and correlated with a TN/BLBC phenotype. Importantly the staining is easy to interpret, with 80-100% of tumour cells exhibiting strong nuclear positivity.
iii) SOX10 expression correlated with various clinical and biological features related to poor prognosis (including grade, mitotic score, hormone receptor and Ki67 status).
iv) Kaplan-Meier survival analysis demonstrated SOX10 nuclear positivity is associated with poor outcome in all breast cancers and importantly identifies a poor outcome subgroup of TN/BLBC.

This exciting data suggests SOX10 might be useful prognostic marker for TN/BL breast cancer. There are, however a number factors that need to be considered in order to move this idea forward. Firstly, our cohort was of insufficient size to enable a multivariate analysis to determine whether SOX10 represents an independent prognostic marker. Secondly, the SOX10 positive TN/BLBC group remains heterogeneous, with respect to outcome. Approximately half the patient group died within 5 years post diagnosis whereas the other half experienced a good long-term outcome. Again our cohort was too small to try and tease out whether this dichotomy could be related to specific clinic-pathological or biological factors. Thirdly, these observations need to be validated in an independent cohort to strengthen the role of SOX10 as a potential breast cancer biomarker. To address these issues we have initiated a collaboration with the Nottingham Breast Cancer group (Prof. Ian Ellis, Emad Rakha, Andrew Green), who have developed a very large breast cancer cohort with long term clinical follow up on TMAs and have published widely using this resource to investigate the expression of a large number of biomarkers in breast cancer [276-278]. Analysis of SOX10 using this additional cohort will occur in due course. This may also lead to a better understanding of the biology of SOX10 positive tumours given the large volume of biomarker data already available for tumours in this cohort.

Meta-analysis of gene expression profiling data also confirmed the observation that SOX10 correlates with the TN/BL phenotype, indicating there is probably good agreement between mRNA and protein for this molecule. It will therefore also be interesting to determine whether SOX10 mRNA correlates with a particular TN/BLBC subgroup, as defined by Lehmann and colleagues [117, 118] (see section 1.4.2). In that study, six distinctive subgroups were defined based on gene expression profiles (*e.g.* BL1, BL2, mesenchymal, mesenchymal stem cell, immunomodulatory, luminal androgen receptor) and these were shown to have different prognosis and response to chemotherapy. Determining whether SOX10 identifies a specific subgroup may help us further understand the biological role of SOX10 in breast cancer and whether it can be developed as a robust biomarker of treatment response and outcome.

We have undertaken a preliminary analysis to begin to understand the functional role of SOX10 in breast cancer. Our *in-silico, in vitro* and *in vivo* analysis demonstrated that,

i) SOX10 is rarely expressed in breast cancer cell lines. Only ~15% had detectable levels of *SOX10* mRNA levels across a panel of breast cancer cell lines. SOX10 protein expression was evident only in the MDA-MB-435 basal-like breast cancer cell line.
ii) *In vitro* cell culture conditions promoting selection of stem/progenitor cells did not induce SOX10 expression in a panel of SOX10-negative basal-like cell lines.
iii) shRNA-mediated knockdown of SOX10 in MDA-MB-435 breast cancer cells did not alter their proliferation rate, migratory behavior or sphere formation *in vitro*, though did reduce colony formation on plastic.
iv) *In vivo* tumourigenicity assays using MDA-MB-435 cells injected into the mouse mammary fat showed that SOX10 knockdown reduced primary tumour formation.
v) Differential gene expression profiling of shRNA mediated SOX10 knockdown vs. normal MDA-MB-435 cells showed enrichment for immune related genes.

Although we found some interesting results using *in vitro* and *in vivo* functional assays, we were limited to using a single cell line model (MDA-MB-435), as SOX10 was not frequently expressed in breast cancer cell lines. Therefore, we will have to complement our findings with additional experiments in order to further tease out the role of SOX10 in breast cancer and support our hypothesis that SOX10 overexpression contributes to an aggressive phenotype. To address this, future experiments need to focus on SOX10 over expression in other basal-like breast cancer cell lines (*e.g.* MDA-MB-468 and MDA-MB-231). *In vivo* analysis suggests SOX10 may be playing an important role in tumour growth, and this may be the reason for its aggressive phenotype in basal-

like breast cancers. However, due to time limitations we were unable to investigate the effect of SOX10 knockdown on spontaneous metastasis in this study. This will be completed in due course by analysing the expression of markers such as trichrome, Ki67, ApopTag and SOX10 using immunohistochemistry on sections from resected tumours and/or organs. This will give insight as to whether SOX10 has a role in distant metastasis. Alternately, this may be achieved by transducing shSOX10 with luciferase and injecting the transduced control and knockdown lines into the mice. The mice would then be monitored for the formation of spontaneous metastasis using bioluminescence imaging.

A really interesting finding that needs a significant level of additional work was an association between SOX10 and immune-related gene expression. Differential gene expression analysis of SOX10 control and knockdown lines showed that loss of SOX10 in MDA-MB-435 cells is mediating/regulating an immune-related gene expression program that may have an effect *in vivo* on the immune response. This observation needs to be further validated using experimental models that can help determine the effect of SOX10 knockdown on immune system. In hindsight, use of the NOD-SCID mice model to investigate tumour progression in this study may not have been an appropriate model to study the effects of the immune response because these mice are immuno-compromised. However, the 4T1 metastatic mouse model may be a more appropriate model to study the possible link between SOX10, the immune system and tumour growth/metastasis in future. There are many advantages of using this model; firstly, it is a syngeneic xenograft model with a luciferase-expressing clone, it is easily transplantable, highly tumourigenic and highly metastatic [279]. This makes it easier to study the effect of SOX10 overexpression and knockdown on spontaneous metastasis using bioluminescence imaging. Secondly, this model closely resembles the human breast cancers (particularly the basal-like cancer) as they have a propensity to metastasise to the lungs, liver, brain and bone. This would be interesting as we know that SOX10 expression marks a subset of aggressive basal-like breast cancers and investigating loss/gain of SOX10 in a model like this would further give us some insight into its aggressive tumourigenic behavior.

We have also investigated SOX10 in the context of normal breast biology and showed, overall, a possible association with luminal progenitor cell populations. The following were key findings:

i) Immunohistochemistry and immunoflouresence analysis of SOX10 in normal epithelia from reduction mammoplasty samples and in normal adjacent to tumour demonstrated that SOX10

expression was either restricted to the myoepithelial cell compartment or was found in both myoepithelial and luminal epithelial cell compartments.

ii) We found a striking correlation between SOX10 and the luminal progenitor marker c-kit, and an inverse association with markers of fully differentiated luminal epithelia, ER and CK 8/18.

iii) An *in-silico* analysis showed that *SOX10* mRNA expression was highest in a FACS sorted luminal progenitor cell population, was moderately expressed in mammary stem cell-enriched (MaSC) populations, and was lowest in mature luminal cells.

Based on this association and in light of what is known about SOX10 in melanocytes where it controls a differentiation program, we hypothesise therefore that SOX10 may be regulating mammary epithelial lineage differentiation in the breast. We know that there is an increasing requirement for stem/progenitor cell population at various stages of mammary gland development (*e.g.* during puberty, pregnancy and lactation) that undergoes rapid cell proliferation and differentiation. SOX10 may play an important part in this developmental program and that further investigations are required to test this hypothesis. This could be explored further by performing mammary transplant assays, which can be achieved by isolating normal mammary epithelial cell populations from reduction mammoplasty cases (based on their cell surface markers) using FACS analysis [24]. Isolated cells can be separated as mature luminal, luminal progenitor and MaSc and maintained as individual cultures. shRNA-mediated SOX10 knockdown can be performed on these different cell populations, transduced cells can then be injected into the mice to examine ductal elongation and morphogenesis. The effect of SOX10 knockdown on mammary gland development could then be assessed by whole mount analysis. A similar analysis has been done to study the role of ID4 in breast cancer development [280].

How SOX10 exerts its possible effects on differentiation is presumably through its role as a transcription factor, although this is still poorly characterised. Only a handful of target genes for SOX10 are known so far and these include *MITF, P0, c-Ret, EGR2* and *EDNRB* [161, 166, 168-170]. These gene targets are known to play an important role in melanocyte development and differentiation of Schwann cells. However, we are currently limited by our understanding of biological networks that SOX10 could be involved in regulating in normal breast and in breast cancer. Recent advances in the techniques of chromatin immunoprecipitation (ChIP) combined with DNA sequencing (ChIP-Seq) would enable one to create a global map of specific DNA-protein interactions in a given cell type and can facilitate identification of gene –gene/gene –protein interactions [281]. ChiP involves crosslinking DNA-binding proteins to DNA by treatment of the cells with formaldehyde and preparation of chromatin by enzyme digestion or sonication.

Immunoprecipitation of the cross-linked chromatin could then be performed using SOX10 antibody (as we know the SOX10 antibody used in this study is of good quality based on IHC and western blot results); followed by next generation sequencing of the captured DNA sequences. This analysis would allow for the identification of potential SOX10 biding sites throughout the genome. Thus, identification of downstream target genes for SOX10 will further help understand the functional role of SOX10 in normal breast and in breast cancer.

There is evidence from the literature that breast cancer subtypes arise from different cell populations of the mammary cell hierarchy. Molecular profiling of basal-like cancers suggests that luminal progenitors are the cell of origin for BLBCs [24]. Mammary development studies have led to greater understandings into breast cancer biology, and it is now clear that transcription factors such as *GATA3* and *BRCA1* controlling luminal epithelial cell lineage are powerful breast tumour suppressors that are frequently mutated in malignancy [282-284]. Through cancer genomics, *in vitro* assays and experimental xenograft models, we have demonstrated that the transcription factor SOX10 is a luminal progenitor marker and that its expression is maintained in a subset of basal-like cancers and contribute to poor outcomes. This suggests that SOX10 could be one of the breast cancer genes that could be important in both normal mammary development and in breast cancer.

As mentioned earlier, the main focus of this study was to identify prognostic biomarkers in TN/BLBs. In line with work on SOX10, we extended our investigations on two X-chromosome cancer testis (CT-X) antigens, MAGE-A and NY-ESO-1:

i) MAGE-A expression identified a group of high grade, highly proliferative TN/BL tumours with poor outcome within the whole cohort. MAGE-A expression was not prognostic, however, within TN/BL tumours subgroup.
ii) NY-ESO-1 identified a very small group of high-grade patients, which are also MAGE-A positive and showed a trend towards poor outcome.

Overall, MAGE-A and NY-ESO-1 do not offer much value as prognostic markers in this cohort, but it is possible that with a larger sample size of TN/BLBCs one might see an effect for these makers. Further investigations of NY-ESO-1 are required to fully understand the role of MAGE-A and NY-ESO-1 in breast cancer.

In conclusion, this study is the first to assess the prognostic significance of SOX10, MAGE-A and NY-ESO-1 in breast cancer. All three proteins are associated with tumours with TN/BL phenotype, and SOX10 shows promise as a biomarker of prognostic significance. We have shown that SOX10

is a luminal progenitor marker whose expression is maintained in a subset of basal-like cancers with poor outcome. This is an observational study and there is a requirement for a more mechanistic approach in the future in order to fully understand the role of SOX10 in normal breast and in breast cancer. Some of the future experiments discussed in this section may further increase our knowledge on the functional role of SOX10.

References

Reference

1. *Australian Institute of Health and Welfare & Australasian Association of Cancer Registries*, in *Breast cancer in Australia: an overview*, C.s.n. 74, Editor. 2012: Cnberra: AIHW.
2. *Australian Institute of Health and Welfare & Cancer Australia. Breast cancer in Australia: an overview*. 2012: Canberra: AIHW.
3. Early Breast Cancer Trialists' Collaborative, G., *Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.* Lancet, 2005. **365**(9472): p. 1687-717.
4. Mego, M., S.A. Mani, and M. Cristofanilli, *Molecular mechanisms of metastasis in breast cancer--clinical applications.* Nat Rev Clin Oncol, 2010. **7**(12): p. 693-701.
5. Beenken, S.W.a.K.I.B., Eds., *Inherited Cancer Syndromes: Current Clinical Management*. Breast Cancer Genetics. Vol. New York, Springer. 2004.
6. Martin, A.M. and B.L. Weber, *Genetic and hormonal risk factors in breast cancer.* J Natl Cancer Inst, 2000. **92**(14): p. 1126-35.
7. Mavaddat, N., et al., *Genetic susceptibility to breast cancer.* Mol Oncol, 2010. **4**(3): p. 174-91.
8. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
9. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21.* Science, 1990. **250**(4988): p. 1684-9.
10. Margaritte, P., et al., *Linkage of familial breast cancer to chromosome 17q21 may not be restricted to early-onset disease.* Am J Hum Genet, 1992. **50**(6): p. 1231-4.
11. Narod, S.A., et al., *Familial breast-ovarian cancer locus on chromosome 17q12-q23.* Lancet, 1991. **338**(8759): p. 82-3.
12. Solomon, E. and D.H. Ledbetter, *Report of the committee on the genetic constitution of chromosome 17.* Cytogenet Cell Genet, 1990. **55**(1-4): p. 198-215.
13. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.
14. Mallon, E., et al., *The basic pathology of human breast cancer.* J Mammary Gland Biol Neoplasia, 2000. **5**(2): p. 139-63.
15. Cotran, R.a., *Pathologic basis of disease*. 8th Edition ed. 2010.
16. Brinton, L.A., et al., *Menstrual factors and risk of breast cancer.* Cancer Invest, 1988. **6**(3): p. 245-54.
17. McPherson, K., C.M. Steel, and J.M. Dixon, *ABC of breast diseases. Breast cancer--epidemiology, risk factors and genetics.* BMJ, 1994. **309**(6960): p. 1003-6.
18. Tirona, M.T., R. Sehgal, and O. Ballester, *Prevention of breast cancer (part I): epidemiology, risk factors, and risk assessment tools.* Cancer Invest, 2010. **28**(7): p. 743-50.
19. Hsieh, C.C., et al., *Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study.* Int J Cancer, 1990. **46**(5): p. 796-800.
20. Feuer, E.J., et al., *The lifetime risk of developing breast cancer.* J Natl Cancer Inst, 1993. **85**(11): p. 892-7.
21. Bradley, C.J., C.W. Given, and C. Roberts, *Race, socioeconomic status, and breast cancer treatment and survival.* J Natl Cancer Inst, 2002. **94**(7): p. 490-6.
22. Daniel, C.W. and G.H. Smith, *The mammary gland: a model for development.* J Mammary Gland Biol Neoplasia, 1999. **4**(1): p. 3-8.
23. Visvader, J.E., *Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.* Genes Dev, 2009. **23**(22): p. 2563-77.
24. Lim, E., et al., *Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.* Nat Med, 2009. **15**(8): p. 907-13.
25. Prat, A. and C.M. Perou, *Mammary development meets cancer genomics.* Nat Med, 2009. **15**(8): p. 842-4.
26. O'Connell, P., et al., *Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci.* J Natl Cancer Inst, 1998. **90**(9): p. 697-703.

27. Simpson, P.T., et al., *Columnar cell lesions of the breast: the missing link in breast cancer progression? A morphological and molecular analysis.* Am J Surg Pathol, 2005. **29**(6): p. 734-46.
28. Hussain, M. and G.H. Cunnick, *Management of lobular carcinoma in-situ and atypical lobular hyperplasia of the breast--a review.* Eur J Surg Oncol, 2011. **37**(4): p. 279-89.
29. Meisner, A.L., M.H. Fekrazad, and M.E. Royce, *Breast disease: benign and malignant.* Med Clin North Am, 2008. **92**(5): p. 1115-41, x.
30. Goldhirsch, A., et al., *Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009.* Ann Oncol, 2009. **20**(8): p. 1319-29.
31. Mook, S., et al., *The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 and 70 years of age.* Ann Oncol, 2010. **21**(4): p. 717-22.
32. Lakhani, S.R., Ellis I.O., Schnitt S.J., Tan P.H., van de Vijver M.J, *WHO Classification of Tumours of the Breast*. 2012.
33. Weigelt, B., F.C. Geyer, and J.S. Reis-Filho, *Histological types of breast cancer: how special are they?* Mol Oncol, 2010. **4**(3): p. 192-208.
34. Fisher, E.R., et al., *Fifteen-year prognostic discriminants for invasive breast carcinoma: National Surgical Adjuvant Breast and Bowel Project Protocol-06.* Cancer, 2001. **91**(8 Suppl): p. 1679-87.
35. Kitchen, P.R., et al., *Tubular carcinoma of the breast: prognosis and response to adjuvant systemic therapy.* ANZ J Surg, 2001. **71**(1): p. 27-31.
36. Northridge, M.E., et al., *The importance of histologic type on breast cancer survival.* J Clin Epidemiol, 1997. **50**(3): p. 283-90.
37. Reed, W., et al., *The prognostic value of p53 and c-erb B-2 immunostaining is overrated for patients with lymph node negative breast carcinoma: a multivariate analysis of prognostic factors in 613 patients with a follow-up of 14-30 years.* Cancer, 2000. **88**(4): p. 804-13.
38. Galea, M.H., et al., *The Nottingham Prognostic Index in primary breast cancer.* Breast Cancer Res Treat, 1992. **22**(3): p. 207-19.
39. Luini, A., et al., *Metaplastic carcinoma of the breast, an unusual disease with worse prognosis: the experience of the European Institute of Oncology and review of the literature.* Breast Cancer Res Treat, 2007. **101**(3): p. 349-53.
40. Vuong, D., et al., *Molecular classification of breast cancer.* Virchows Arch, 2014. **465**(1): p. 1-14.
41. 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.
42. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.* Histopathology, 1991. **19**(5): p. 403-10.
43. Diaz, L.K. and N. Sneige, *Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy.* Adv Anat Pathol, 2005. **12**(1): p. 10-9.
44. Pike, M.C., et al., *Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk.* Epidemiol Rev, 1993. **15**(1): p. 17-35.
45. Elledge, R.M., et al., *Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study.* Int J Cancer, 2000. **89**(2): p. 111-7.
46. Hammond, M.E., et al., *American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer.* J Clin Oncol, 2010. **28**(16): p. 2784-95.
47. *Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group.* Lancet, 1998. **351**(9114): p. 1451-67.
48. Early Breast Cancer Trialists' Collaborative, G., et al., *Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials.* Lancet, 2011. **378**(9793): p. 771-84.
49. Osborne, C.K., et al., *Endocrine responsiveness: understanding how progesterone receptor can be used to select endocrine therapy.* Breast, 2005. **14**(6): p. 458-65.

50. Sommer, S. and S.A. Fuqua, *Estrogen receptor and breast cancer.* Semin Cancer Biol, 2001. **11**(5): p. 339-52.
51. Mohammed, H., et al., *Progesterone receptor modulates ERalpha action in breast cancer.* Nature, 2015. **523**(7560): p. 313-7.
52. Fisher, B., et al., *Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study.* J Natl Cancer Inst, 1998. **90**(18): p. 1371-88.
53. Jordan, V.C. and B.W. O'Malley, *Selective estrogen-receptor modulators and antihormonal resistance in breast cancer.* J Clin Oncol, 2007. **25**(36): p. 5815-24.
54. Patel, R.R., C.G. Sharma, and V.C. Jordan, *Optimizing the antihormonal treatment and prevention of breast cancer.* Breast Cancer, 2007. **14**(2): p. 113-22.
55. *Pathologists' Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer.* Breast Care (Basel), 2010. **5**(3): p. 185-187.
56. Akiyama, T., et al., *The product of the human c-erbB-2 gene: a 185-kilodalton glycoprotein with tyrosine kinase activity.* Science, 1986. **232**(4758): p. 1644-6.
57. Slamon, D.J., et al., *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.* Science, 1987. **235**(4785): p. 177-82.
58. Wolff, A.C., et al., *Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update.* J Clin Oncol, 2013. **31**(31): p. 3997-4013.
59. Carter, P., et al., *Humanization of an anti-p185HER2 antibody for human cancer therapy.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4285-9.
60. Moja, L., et al., *Trastuzumab containing regimens for early breast cancer.* Cochrane Database Syst Rev, 2012. **4**: p. CD006243.
61. Baselga, J., *Targeting the phosphoinositide-3 (PI3) kinase pathway in breast cancer.* Oncologist, 2011. **16 Suppl 1**: p. 12-9.
62. Serra, V., et al., *PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer.* Oncogene, 2011. **30**(22): p. 2547-57.
63. Vogel, C.L., et al., *Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.* J Clin Oncol, 2002. **20**(3): p. 719-26.
64. Seidman, A.D., et al., *Weekly trastuzumab and paclitaxel therapy for metastatic breast cancer with analysis of efficacy by HER2 immunophenotype and gene amplification.* J Clin Oncol, 2001. **19**(10): p. 2587-95.
65. Romond, E.H., et al., *Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer.* N Engl J Med, 2005. **353**(16): p. 1673-84.
66. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* N Engl J Med, 2001. **344**(11): p. 783-92.
67. Baselga, J., *Targeting tyrosine kinases in cancer: the second wave.* Science, 2006. **312**(5777): p. 1175-8.
68. Wood, E.R., et al., *A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells.* Cancer Res, 2004. **64**(18): p. 6652-9.
69. Chu, I., et al., *The dual ErbB1/ErbB2 inhibitor, lapatinib (GW572016), cooperates with tamoxifen to inhibit both cell proliferation- and estrogen-dependent gene expression in antiestrogen-resistant breast cancer.* Cancer Res, 2005. **65**(1): p. 18-25.
70. Konecny, G.E., et al., *Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells.* Cancer Res, 2006. **66**(3): p. 1630-9.
71. Geyer, C.E., et al., *Lapatinib plus capecitabine for HER2-positive advanced breast cancer.* N Engl J Med, 2006. **355**(26): p. 2733-43.
72. Gianni, L., et al., *Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial.* Lancet Oncol, 2012. **13**(1): p. 25-32.

73. Keating, G.M., *Pertuzumab: in the first-line treatment of HER2-positive metastatic breast cancer.* Drugs, 2012. **72**(3): p. 353-60.
74. Lambert, J.M. and R.V. Chari, *Ado-trastuzumab Emtansine (T-DM1): an antibody-drug conjugate (ADC) for HER2-positive breast cancer.* J Med Chem, 2014. **57**(16): p. 6949-64.
75. Rakha, E.A., et al., *Invasive lobular carcinoma of the breast: response to hormonal therapy and outcomes.* Eur J Cancer, 2008. **44**(1): p. 73-83.
76. Reis-Filho, J.S. and A.N. Tutt, *Triple negative tumours: a critical review.* Histopathology, 2008. **52**(1): p. 108-18.
77. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clin Cancer Res, 2007. **13**(15 Pt 1): p. 4429-34.
78. Sparano, J.A., et al., *Randomized phase III trial of ixabepilone plus capecitabine versus capecitabine in patients with metastatic breast cancer previously treated with an anthracycline and a taxane.* J Clin Oncol, 2010. **28**(20): p. 3256-63.
79. Colleoni, M., et al., *Classical cyclophosphamide, methotrexate, and fluorouracil chemotherapy is more effective in triple-negative, node-negative breast cancer: results from two randomized trials of adjuvant chemoendocrine therapy for node-negative breast cancer.* J Clin Oncol, 2010. **28**(18): p. 2966-73.
80. Fong, P.C., et al., *Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers.* N Engl J Med, 2009. **361**(2): p. 123-34.
81. O'Shaughnessy, J., et al., *Iniparib plus chemotherapy in metastatic triple-negative breast cancer.* N Engl J Med, 2011. **364**(3): p. 205-14.
82. Perou, C.M., *Molecular stratification of triple-negative breast cancers.* Oncologist, 2011. **16 Suppl 1**: p. 61-70.
83. Chacon, R.D. and M.V. Costanzo, *Triple-negative breast cancer.* Breast Cancer Res, 2010. **12 Suppl 2**: p. S3.
84. Cheang, M.C., et al., *Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer.* J Natl Cancer Inst, 2009. **101**(10): p. 736-50.
85. Nielsen, T.O., et al., *Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma.* Clin Cancer Res, 2004. **10**(16): p. 5367-74.
86. Bhargava, R., et al., *Prevalence, morphologic features and proliferation indices of breast carcinoma molecular classes using immunohistochemical surrogate markers.* Int J Clin Exp Pathol, 2009. **2**(5): p. 444-55.
87. Badve, S., et al., *Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists.* Mod Pathol, 2011. **24**(2): p. 157-67.
88. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
89. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
90. Bertucci, F., et al., *How basal are triple-negative breast cancers?* Int J Cancer, 2008. **123**(1): p. 236-40.
91. Rody, A., et al., *A clinically relevant gene signature in triple negative and basal-like breast cancer.* Breast Cancer Res, 2011. **13**(5): p. R97.
92. de Ronde, J.J., et al., *Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response.* Breast Cancer Res Treat, 2010. **119**(1): p. 119-26.
93. Heitz, F., et al., *Triple-negative and HER2-overexpressing breast cancers exhibit an elevated risk and an earlier occurrence of cerebral metastases.* Eur J Cancer, 2009. **45**(16): p. 2792-8.
94. Fulford, L.G., et al., *Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival.* Breast Cancer Res, 2007. **9**(1): p. R4.
95. Foulkes, W.D., et al., *Tumor size is an unreliable predictor of prognosis in basal-like breast cancers and does not correlate closely with lymph node status.* Breast Cancer Res Treat, 2009. **117**(1): p. 199-204.
96. Foulkes, W.D., J.S. Reis-Filho, and S.A. Narod, *Tumor size and survival in breast cancer--a reappraisal.* Nat Rev Clin Oncol, 2010. **7**(6): p. 348-53.

97. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
98. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.
99. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
100. West, M., et al., *Predicting the clinical status of human breast cancer by using gene expression profiles.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11462-7.
101. Farmer, P., et al., *Identification of molecular apocrine breast tumours by microarray analysis.* Oncogene, 2005. **24**(29): p. 4660-71.
102. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC Genomics, 2006. **7**: p. 96.
103. Prat, A., et al., *Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer.* Breast Cancer Res, 2010. **12**(5): p. R68.
104. Weigelt, B., et al., *Breast cancer molecular profiling with single sample predictors: a retrospective analysis.* Lancet Oncol, 2010. **11**(4): p. 339-49.
105. Guedj, M., et al., *A refined molecular taxonomy of breast cancer.* Oncogene, 2011.
106. Rouzier, R., et al., *Breast cancer molecular subtypes respond differently to preoperative chemotherapy.* Clin Cancer Res, 2005. **11**(16): p. 5678-85.
107. Tran, B. and P.L. Bedard, *Luminal-B breast cancer and novel therapeutic targets.* Breast Cancer Res, 2011. **13**(6): p. 221.
108. Badve, S., et al., *Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2011. **24**(2): p. 157-67.
109. Paik, S., et al., *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med, 2004. **351**(27): p. 2817-26.
110. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
111. Sotiriou, C., et al., *Breast cancer classification and prognosis based on gene expression profiles from a population-based study.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(18): p. 10393-8.
112. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* Journal of Clinical Oncology, 2009. **27**(8): p. 1160-7.
113. Harrell, J.C., et al., *Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse.* Breast Cancer Research and Treatment, 2011.
114. Kennecke, H., et al., *Metastatic behavior of breast cancer subtypes.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2010. **28**(20): p. 3271-7.
115. Korde, L.A., et al., *Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer.* Breast Cancer Research and Treatment, 2010. **119**(3): p. 685-99.
116. Doane, A.S., et al., *An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen.* Oncogene, 2006. **25**(28): p. 3994-4008.
117. Lehmann, B.D., et al., *Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.* J Clin Invest, 2011. **121**(7): p. 2750-67.
118. Masuda, H., et al., *Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes.* Clin Cancer Res, 2013. **19**(19): p. 5533-40.
119. Ross, J.S., et al., *Commercialized multigene predictors of clinical outcome for breast cancer.* Oncologist, 2008. **13**(5): p. 477-93.
120. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. **365**(9460): p. 671-9.
121. Ma, X.J., et al., *The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer.* J Clin Oncol, 2006. **24**(28): p. 4611-9.

122. Sotiriou, C., et al., *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* J Natl Cancer Inst, 2006. **98**(4): p. 262-72.
123. Nielsen, T.O., et al., *A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer.* Clin Cancer Res, 2010. **16**(21): p. 5222-32.
124. Nielsen, T., et al., *Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens.* BMC Cancer, 2014. **14**: p. 177.
125. Dowsett, M., et al., *Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy.* J Clin Oncol, 2013. **31**(22): p. 2783-90.
126. Dowsett, M., et al., *Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study.* J Clin Oncol, 2010. **28**(11): p. 1829-34.
127. Chia, S.K., et al., *A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen.* Clin Cancer Res, 2012. **18**(16): p. 4465-72.
128. Verma, S., et al., *Immunohistochemical characterization of molecular subtypes of invasive breast cancer: a study from North India.* APMIS, 2012. **120**(12): p. 1008-19.
129. Blows, F.M., et al., *Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies.* PLoS Med, 2010. **7**(5): p. e1000279.
130. Vallejos, C.S., et al., *Breast cancer classification according to immunohistochemistry markers: subtypes and association with clinicopathologic variables in a peruvian hospital database.* Clin Breast Cancer, 2010. **10**(4): p. 294-300.
131. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors.* J Natl Cancer Inst, 2000. **92**(7): p. 564-9.
132. Grady, W.M., et al., *Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer.* Nat Genet, 2000. **26**(1): p. 16-7.
133. Raman, V., et al., *Compromised HOXA5 function can limit p53 expression in human breast tumours.* Nature, 2000. **405**(6789): p. 974-8.
134. Classon, M. and E. Harlow, *The retinoblastoma tumour suppressor in development and cancer.* Nat Rev Cancer, 2002. **2**(12): p. 910-7.
135. Lustig, B. and J. Behrens, *The Wnt signaling pathway and its role in tumor development.* J Cancer Res Clin Oncol, 2003. **129**(4): p. 199-221.
136. Wicking, C., I. Smyth, and A. Bale, *The hedgehog signalling pathway in tumorigenesis and development.* Oncogene, 1999. **18**(55): p. 7844-51.
137. Castillo, S.D. and M. Sanchez-Cespedes, *The SOX family of genes in cancer development: biological relevance and opportunities for therapy.* Expert Opin Ther Targets, 2012. **16**(9): p. 903-19.
138. Dong, C., D. Wilhelm, and P. Koopman, *Sox genes and cancer.* Cytogenet Genome Res, 2004. **105**(2-4): p. 442-7.
139. Cimino-Mathews, A., et al., *Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas.* Hum Pathol, 2012.
140. Ivanov, S.V., et al., *Diagnostic SOX10 gene signatures in salivary adenoid cystic and breast basal-like carcinomas.* Br J Cancer, 2013. **109**(2): p. 444-51.
141. Vargas, A.C., et al., *Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression.* Breast Cancer Res Treat, 2012. **135**(1): p. 153-65.
142. Cronin, J.C., et al., *SOX10 ablation arrests cell cycle, induces senescence, and suppresses melanomagenesis.* Cancer Res, 2013. **73**(18): p. 5709-18.
143. Graf, S.A., et al., *SOX10 promotes melanoma cell invasion by regulating melanoma inhibitory activity.* J Invest Dermatol, 2014. **134**(8): p. 2212-20.
144. Shakhova, O., et al., *Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma.* Nat Cell Biol, 2012. **14**(8): p. 882-90.
145. Simpson, A.J., et al., *Cancer/testis antigens, gametogenesis and cancer.* Nat Rev Cancer, 2005. **5**(8): p. 615-25.

146. Sarkar, A. and K. Hochedlinger, *The sox family of transcription factors: versatile regulators of stem and progenitor cell fate.* Cell Stem Cell, 2013. **12**(1): p. 15-30.
147. Kamachi, Y. and H. Kondoh, *Sox proteins: regulators of cell fate specification and differentiation.* Development, 2013. **140**(20): p. 4129-44.
148. Pevny, L.H. and R. Lovell-Badge, *Sox genes find their feet.* Curr Opin Genet Dev, 1997. **7**(3): p. 338-44.
149. Wegner, M., *From head to toes: the multiple facets of Sox proteins.* Nucleic Acids Res, 1999. **27**(6): p. 1409-20.
150. Jantzen, H.M., et al., *Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins.* Nature, 1990. **344**(6269): p. 830-6.
151. Kelly, M., et al., *Four mating-type genes control sexual differentiation in the fission yeast.* EMBO J, 1988. **7**(5): p. 1537-47.
152. Bowles, J., G. Schepers, and P. Koopman, *Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators.* Dev Biol, 2000. **227**(2): p. 239-55.
153. Schepers, G.E., R.D. Teasdale, and P. Koopman, *Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families.* Dev Cell, 2002. **3**(2): p. 167-70.
154. Wright, E.M., B. Snopek, and P. Koopman, *Seven new members of the Sox gene family expressed during mouse development.* Nucleic Acids Res, 1993. **21**(3): p. 744.
155. Sauka-Spengler, T. and M. Bronner-Fraser, *A gene regulatory network orchestrates neural crest formation.* Nat Rev Mol Cell Biol, 2008. **9**(7): p. 557-68.
156. Mort, R.L., I.J. Jackson, and E.E. Patton, *The melanocyte lineage in development and disease.* Development, 2015. **142**(4): p. 620-32.
157. Peirano, R.I., et al., *Protein zero gene expression is regulated by the glial transcription factor Sox10.* Mol Cell Biol, 2000. **20**(9): p. 3198-209.
158. Stolt, C.C., et al., *Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10.* Genes Dev, 2002. **16**(2): p. 165-70.
159. Schlierf, B., et al., *The high-mobility group transcription factor Sox10 interacts with the N-myc-interacting protein Nmi.* J Mol Biol, 2005. **353**(5): p. 1033-42.
160. Buac, K., et al., *A Sox10 expression screen identifies an amino acid essential for Erbb3 function.* PLoS Genet, 2008. **4**(9): p. e1000177.
161. Prasad, M.K., et al., *SOX10 directly modulates ERBB3 transcription via an intronic neural crest enhancer.* BMC Dev Biol, 2011. **11**: p. 40.
162. Pusch, C., et al., *The SOX10/Sox10 gene from human and mouse: sequence, expression, and transactivation by the encoded HMG domain transcription factor.* Hum Genet, 1998. **103**(2): p. 115-23.
163. Kuhlbrodt, K., et al., *Sox10, a novel transcriptional modulator in glial cells.* J Neurosci, 1998. **18**(1): p. 237-50.
164. Yusuf, D., et al., *The transcription factor encyclopedia.* Genome Biol, 2012. **13**(3): p. R24.
165. Rehberg, S., et al., *Sox10 is an active nucleocytoplasmic shuttle protein, and shuttling is crucial for Sox10-mediated transactivation.* Mol Cell Biol, 2002. **22**(16): p. 5826-34.
166. Peirano, R.I. and M. Wegner, *The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences.* Nucleic Acids Res, 2000. **28**(16): p. 3047-55.
167. Schlierf, B., et al., *Cooperative binding of Sox10 to DNA: requirements and consequences.* Nucleic Acids Res, 2002. **30**(24): p. 5509-16.
168. Widlund, H.R. and D.E. Fisher, *Microphthalamia-associated transcription factor: a critical regulator of pigment cell development and survival.* Oncogene, 2003. **22**(20): p. 3035-41.
169. Bondurand, N., et al., *Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome.* Hum Mol Genet, 2000. **9**(13): p. 1907-17.
170. Ghislain, J. and P. Charnay, *Control of myelination in Schwann cells: a Krox20 cis-regulatory element integrates Oct6, Brn2 and Sox10 activities.* EMBO Rep, 2006. **7**(1): p. 52-8.
171. Srinivasan, R., et al., *Genome-wide analysis of EGR2/SOX10 binding in myelinating peripheral nerve.* Nucleic Acids Res, 2012. **40**(14): p. 6449-60.

172. Ferletta, M., et al., *Sox10 has a broad expression pattern in gliomas and enhances platelet-derived growth factor-B--induced gliomagenesis.* Mol Cancer Res, 2007. **5**(9): p. 891-7.
173. Tong, X., et al., *SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/beta-catenin pathway.* Oncotarget, 2014. **5**(21): p. 10571-83.
174. Schlierf, B., et al., *Expression of SoxE and SoxD genes in human gliomas.* Neuropathol Appl Neurobiol, 2007. **33**(6): p. 621-30.
175. Nonaka, D., L. Chiriboga, and B.P. Rubin, *Sox10: a pan-schwannian and melanocytic marker.* Am J Surg Pathol, 2008. **32**(9): p. 1291-8.
176. Karamchandani, J.R., et al., *Sox10 and S100 in the diagnosis of soft-tissue neoplasms.* Appl Immunohistochem Mol Morphol, 2012. **20**(5): p. 445-50.
177. Shin, J., et al., *Sox10 is expressed in primary melanocytic neoplasms of various histologies but not in fibrohistiocytic proliferations and histiocytoses.* J Am Acad Dermatol, 2012. **67**(4): p. 717-26.
178. Ying, J., et al., *WNT5A exhibits tumor-suppressive activity through antagonizing the Wnt/beta-catenin signaling, and is frequently methylated in colorectal cancer.* Clin Cancer Res, 2008. **14**(1): p. 55-61.
179. Wang, J., et al., *Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing.* BMC Genomics, 2013. **14**: p. 11.
180. Almeida, L.G., et al., *CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens.* Nucleic Acids Res, 2009. **37**(Database issue): p. D816-9.
181. Hofmann, O., et al., *Genome-wide analysis of cancer/testis gene expression.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20422-7.
182. Scanlan, M.J., A.J. Simpson, and L.J. Old, *The cancer/testis genes: review, standardization, and commentary.* Cancer Immun, 2004. **4**: p. 1.
183. Gure, A.O., et al., *Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer.* Clin Cancer Res, 2005. **11**(22): p. 8055-62.
184. Velazquez, E.F., et al., *Expression of the cancer/testis antigen NY-ESO-1 in primary and metastatic malignant melanoma (MM)--correlation with prognostic factors.* Cancer Immun, 2007. **7**: p. 11.
185. Napoletano, C., et al., *MAGE-A and NY-ESO-1 expression in cervical cancer: prognostic factors and effects of chemotherapy.* Am J Obstet Gynecol, 2008. **198**(1): p. 99 e1-7.
186. Andrade, V.C., et al., *Prognostic impact of cancer/testis antigen expression in advanced stage multiple myeloma patients.* Cancer Immun, 2008. **8**: p. 2.
187. Scanlan, M.J., et al., *Cancer/testis antigens: an expanding family of targets for cancer immunotherapy.* Immunol Rev, 2002. **188**: p. 22-32.
188. Chomez, P., et al., *An overview of the MAGE gene family with the identification of all human members of the family.* Cancer Res, 2001. **61**(14): p. 5544-51.
189. Boel, P., et al., *BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes.* Immunity, 1995. **2**(2): p. 167-75.
190. De Backer, O., et al., *Characterization of the GAGE genes that are expressed in various human cancers and in normal testis.* Cancer Res, 1999. **59**(13): p. 3157-65.
191. Tureci, O., et al., *Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5211-6.
192. Chen, Y.T., et al., *A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1914-8.
193. van der Bruggen, P., et al., *A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma.* J Immunol, 2007. **178**(5): p. 2617-21.
194. Gaugler, B., et al., *Human gene MAGE-3 codes for an antigen recognized on a melanoma by autologous cytolytic T lymphocytes.* J Exp Med, 1994. **179**(3): p. 921-30.
195. Jager, E., et al., *Humoral immune responses of cancer patients against "Cancer-Testis" antigen NY-ESO-1: correlation with clinical events.* Int J Cancer, 1999. **84**(5): p. 506-10.
196. Bandic, D., et al., *Expression and possible prognostic role of MAGE-A4, NY-ESO-1, and HER-2 antigens in women with relapsing invasive ductal breast cancer: retrospective immunohistochemical study.* Croat Med J, 2006. **47**(1): p. 32-41.

197. Grigoriadis, A., et al., *CT-X antigen expression in human breast cancer.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13493-8.
198. Mischo, A., et al., *Prospective study on the expression of cancer testis genes and antibody responses in 100 consecutive patients with primary breast cancer.* Int J Cancer, 2006. **118**(3): p. 696-703.
199. Sugita, Y., et al., *NY-ESO-1 expression and immunogenicity in malignant and benign breast tumors.* Cancer Res, 2004. **64**(6): p. 2199-204.
200. Theurillat, J.P., et al., *NY-ESO-1 protein expression in primary breast carcinoma and metastases: correlation with CD8+ T-cell and CD79a+ plasmacytic/B-cell infiltration.* Int J Cancer, 2007. **120**(11): p. 2411-7.
201. Chen, Y.T., et al., *Multiple cancer/testis antigens are preferentially expressed in hormone-receptor negative and high-grade breast cancers.* PLoS One, 2011. **6**(3): p. e17876.
202. Atanackovic, D., et al., *Booster vaccination of cancer patients with MAGE-A3 protein reveals long-term immunological memory or tolerance depending on priming.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1650-5.
203. Bender, A., et al., *LUD 00-009: phase 1 study of intensive course immunization with NY-ESO-1 peptides in HLA-A2 positive patients with NY-ESO-1-expressing cancer.* Cancer Immun, 2007. **7**: p. 16.
204. Davis, I.D., et al., *Recombinant NY-ESO-1 protein with ISCOMATRIX adjuvant induces broad integrated antibody and CD4(+) and CD8(+) T cell responses in humans.* Proc Natl Acad Sci U S A, 2004. **101**(29): p. 10697-702.
205. Jager, E., et al., *Recombinant vaccinia/fowlpox NY-ESO-1 vaccines induce both humoral and cellular NY-ESO-1-specific immune responses in cancer patients.* Proc Natl Acad Sci U S A, 2006. **103**(39): p. 14453-8.
206. Odunsi, K., et al., *Vaccination with an NY-ESO-1 peptide of HLA class I/II specificities induces integrated humoral and T cell responses in ovarian cancer.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12837-42.
207. van Baren, N., et al., *Tumoral and immunologic response after vaccination of melanoma patients with an ALVAC virus encoding MAGE antigens recognized by T cells.* J Clin Oncol, 2005. **23**(35): p. 9008-21.
208. Coughlin, S.S., M.J. Khoury, and K.K. Steinberg, *BRCA1 and BRCA2 gene mutations and risk of breast cancer. Public health perspectives.* Am J Prev Med, 1999. **16**(2): p. 91-8.
209. Nathanson, K.L., R. Wooster, and B.L. Weber, *Breast cancer genetics: what we know and what we need.* Nat Med, 2001. **7**(5): p. 552-6.
210. Mann, G.J., et al., *Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource.* Breast Cancer Res, 2006. **8**(1): p. R12.
211. *R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* 2014.
212. Kauffmann, A., R. Gentleman, and W. Huber, *arrayQualityMetrics—a bioconductor package for quality assessment of microarray data.* Bioinformatics, 2009. **25**(3): p. 415-416.
213. Law, C., et al., *voom: precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biology, 2014. **15**(2): p. R29.
214. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
215. Dunning, M.J., et al., *beadarray: R classes and methods for Illumina bead-based data.* Bioinformatics, 2007. **23**(16): p. 2183-2184.
216. Cairns, J.M., et al., *BASH: a tool for managing BeadArray spatial artefacts.* Bioinformatics, 2008. **24**(24): p. 2921-2922.
217. Smyth, G.K., *limma: Linear Models for Microarray Data*, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, R. Gentleman, et al., Editors. 2005, Springer New York. p. 397-420.
218. Carey, L., et al., *Triple-negative breast cancer: disease entity or title of convenience?* Nat Rev Clin Oncol, 2010. **7**(12): p. 683-92.
219. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* JAMA, 2006. **295**(21): p. 2492-502.
220. Carey, L.A., et al., *The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes.* Clin Cancer Res, 2007. **13**(8): p. 2329-34.

221. Liedtke, C., et al., *Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer.* J Clin Oncol, 2008. **26**(8): p. 1275-81.
222. Al-Ejeh, F., et al., *Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer.* Oncogenesis, 2014. **3**: p. e100.
223. Gubbay, J., et al., *A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.* Nature, 1990. **346**(6281): p. 245-50.
224. Cancer Genome Atlas Research, N., et al., *The Cancer Genome Atlas Pan-Cancer analysis project.* Nat Genet, 2013. **45**(10): p. 1113-20.
225. Reis-Filho, J.S., et al., *Metaplastic breast carcinomas are basal-like tumours.* Histopathology, 2006. **49**(1): p. 10-21.
226. Leibl, S. and F. Moinfar, *Metaplastic breast carcinomas are negative for Her-2 but frequently express EGFR (Her-1): potential relevance to adjuvant treatment with EGFR tyrosine kinase inhibitors?* J Clin Pathol, 2005. **58**(7): p. 700-4.
227. Reis-Filho, J.S., et al., *Novel and classic myoepithelial/stem cell markers in metaplastic carcinomas of the breast.* Appl Immunohistochem Mol Morphol, 2003. **11**(1): p. 1-8.
228. Savage, K., et al., *Caveolin 1 is overexpressed and amplified in a subset of basal-like and metaplastic breast carcinomas: a morphologic, ultrastructural, immunohistochemical, and in situ hybridization analysis.* Clin Cancer Res, 2007. **13**(1): p. 90-101.
229. Gilbert, J.A., et al., *Molecular analysis of metaplastic breast carcinoma: high EGFR copy number via aneusomy.* Mol Cancer Ther, 2008. **7**(4): p. 944-51.
230. Reis-Filho, J.S., et al., *EGFR amplification and lack of activating mutations in metaplastic breast carcinomas.* J Pathol, 2006. **209**(4): p. 445-53.
231. Weigelt, B., B. Kreike, and J.S. Reis-Filho, *Metaplastic breast carcinomas are basal-like breast cancers: a genomic profiling analysis.* Breast Cancer Res Treat, 2009. **117**(2): p. 273-80.
232. Foulkes, W.D., et al., *Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer.* J Natl Cancer Inst, 2003. **95**(19): p. 1482-5.
233. Honrado, E., et al., *Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations.* Oncogene, 2006. **25**(43): p. 5837-45.
234. Lakhani, S.R., et al., *Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype.* Clin Cancer Res, 2005. **11**(14): p. 5175-80.
235. Turner, N.C., et al., *BRCA1 dysfunction in sporadic basal-like breast cancer.* Oncogene, 2007. **26**(14): p. 2126-32.
236. Bediaga, N.G., et al., *DNA methylation epigenotypes in breast cancer molecular subtypes.* Breast Cancer Res, 2010. **12**(5): p. R77.
237. Kamalakaran, S., et al., *DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables.* Mol Oncol, 2011. **5**(1): p. 77-92.
238. Lee, J.S., et al., *Basal-like breast cancer displays distinct patterns of promoter methylation.* Cancer Biol Ther, 2010. **9**(12): p. 1017-24.
239. Holm, K., et al., *Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns.* Breast Cancer Res, 2010. **12**(3): p. R36.
240. Lee, L.J., et al., *Clinical outcome of triple negative breast cancer in BRCA1 mutation carriers and noncarriers.* Cancer, 2011. **117**(14): p. 3093-100.
241. Yan, M., et al., *Comparative microRNA profiling of sporadic and BRCA1 associated basal-like breast cancers.* BMC Cancer, 2015. **15**: p. 506.
242. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.
243. Kelsh, R.N., *Sorting out Sox10 functions in neural crest development.* Bioessays, 2006. **28**(8): p. 788-98.
244. Heanue, T.A. and V. Pachnis, *Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies.* Nat Rev Neurosci, 2007. **8**(6): p. 466-79.

245. Miettinen, M., et al., *Sox10-A Marker for Not Only Schwannian and Melanocytic Neoplasms But Also Myoepithelial Cell Tumors of Soft Tissue: A Systematic Analysis of 5134 Tumors.* Am J Surg Pathol, 2015.
246. Zhu, Y.T., et al., *Peroxisome-proliferator-activated receptor-binding protein (PBP) is essential for the growth of active Notch4-immortalized mammary epithelial cells by activating SOX10 expression.* Biochem J, 2010. **425**(2): p. 435-44.
247. Honore, S.M., M.J. Aybar, and R. Mayor, *Sox10 is required for the early development of the prospective neural crest in Xenopus embryos.* Dev Biol, 2003. **260**(1): p. 79-96.
248. Herbarth, B., et al., *Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5161-5.
249. Pingault, V., et al., *SOX10 mutations in patients with Waardenburg-Hirschsprung disease.* Nat Genet, 1998. **18**(2): p. 171-3.
250. Cronin, J.C., et al., *Frequent mutations in the MITF pathway in melanoma.* Pigment Cell Melanoma Res, 2009. **22**(4): p. 435-44.
251. Seong, I., et al., *Sox10 controls migration of B16F10 melanoma cells through multiple regulatory target genes.* PLoS One, 2012. **7**(2): p. e31477.
252. Kelleher, F.C., D. Fennelly, and M. Rafferty, *Common critical pathways in embryogenesis and cancer.* Acta Oncol, 2006. **45**(4): p. 375-88.
253. Chakrabarti, R., et al., *Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling.* Stem Cells, 2012. **30**(7): p. 1496-508.
254. Bai, L. and L.R. Rohrschneider, *s-SHIP promoter expression marks activated stem cells in developing mouse mammary tissue.* Genes Dev, 2010. **24**(17): p. 1882-92.
255. Williams, J.M. and C.W. Daniel, *Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis.* Dev Biol, 1983. **97**(2): p. 274-90.
256. Cimpean, A.M., et al., *Relevance of the immunohistochemical expression of cytokeratin 8/18 for the diagnosis and classification of breast cancer.* Rom J Morphol Embryol, 2008. **49**(4): p. 479-83.
257. Watson, I.R., et al., *Emerging patterns of somatic mutations in cancer.* Nat Rev Genet, 2013. **14**(10): p. 703-18.
258. Teschendorff, A.E., et al., *An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.* Genome Biol, 2007. **8**(8): p. R157.
259. Denkert, C., et al., *Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.* J Clin Oncol, 2010. **28**(1): p. 105-13.
260. Dravis, C., et al., *Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells.* Cell Rep, 2015. **12**(12): p. 2035-48.
261. Ohman Forslund, K. and K. Nordqvist, *The melanoma antigen genes--any clues to their functions in normal tissues?* Exp Cell Res, 2001. **265**(2): p. 185-94.
262. Vatolin, S., et al., *Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes.* Cancer Res, 2005. **65**(17): p. 7751-62.
263. Nicholaou, T., et al., *Directions in the immune targeting of cancer: lessons learned from the cancer-testis Ag NY-ESO-1.* Immunol Cell Biol, 2006. **84**(3): p. 303-17.
264. Jager, E., et al., *Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes.* J Exp Med, 1998. **187**(2): p. 265-70.
265. Valmori, D., et al., *Naturally occurring human lymphocyte antigen-A2 restricted CD8+ T-cell response to the cancer testis antigen NY-ESO-1 in melanoma patients.* Cancer Res, 2000. **60**(16): p. 4499-506.
266. Monte, M., et al., *MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11160-5.
267. Yang, B., et al., *MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines.* Cancer Res, 2007. **67**(20): p. 9954-62.

268. Marcar, L., et al., *Mage-A cancer/testis antigens inhibit p53 function by blocking its interaction with chromatin.* Cancer Res, 2010. **70**(24): p. 10362-70.
269. Badovinac Crnjevic, T., et al., *High expression of MAGE-A10 cancer-testis antigen in triple-negative breast cancer.* Med Oncol, 2012. **29**(3): p. 1586-91.
270. Curigliano, G., et al., *Cancer-testis antigen expression in triple-negative breast cancer.* Ann Oncol, 2011. **22**(1): p. 98-103.
271. Ayyoub, M., et al., *Expression of MAGE-A3/6 in primary breast cancer is associated with hormone receptor negative status, high histologic grade, and poor survival.* J Immunother, 2014. **37**(2): p. 73-6.
272. Wong, P.P., et al., *Identification of MAGEA antigens as causal players in the development of tamoxifen-resistant breast cancer.* Oncogene, 2014. **33**(37): p. 4579-88.
273. Ademuyiwa, F.O., et al., *NY-ESO-1 cancer testis antigen demonstrates high immunogenicity in triple negative breast cancer.* PLoS One, 2012. **7**(6): p. e38783.
274. Matkovic, B., et al., *Expression of MAGE-A and NY-ESO-1 cancer/testis antigens in medullary breast cancer: retrospective immunohistochemical study.* Croat Med J, 2011. **52**(2): p. 171-7.
275. Jemal, A., et al., *Cancer statistics, 2010.* CA Cancer J Clin, 2010. **60**(5): p. 277-300.
276. Abdel-Fatah, T.M., et al., *HAGE (DDX43) is a biomarker for poor prognosis and a predictor of chemotherapy response in breast cancer.* Br J Cancer, 2014. **110**(10): p. 2450-61.
277. Jerjees, D.A., et al., *ERK1/2 is related to oestrogen receptor and predicts outcome in hormone-treated breast cancer.* Breast Cancer Res Treat, 2014. **147**(1): p. 25-37.
278. 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.
279. Pulaski, B.A. and S. Ostrand-Rosenberg, *Mouse 4T1 breast tumor model.* Curr Protoc Immunol, 2001. **Chapter 20**: p. Unit 20 2.
280. Junankar, S., et al., *ID4 controls mammary stem cells and marks breast cancers with a stem cell-like phenotype.* Nat Commun, 2015. **6**: p. 6548.
281. Farnham, P.J., *Insights from genomic profiling of transcription factors.* Nat Rev Genet, 2009. **10**(9): p. 605-16.
282. Asselin-Labat, M.L., et al., *Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation.* Nat Cell Biol, 2007. **9**(2): p. 201-9.
283. Kouros-Mehr, H., et al., *GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland.* Cell, 2006. **127**(5): p. 1041-55.
284. Liu, S., et al., *BRCA1 regulates human mammary stem/progenitor cell fate.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1680-5.

Appendices

Appendix 2.1: Proforma used for pathology review.

Case No:

DCIS: ☐ Gd1 ☐ Gd2 ☐ Gd3 ☐

LCIS: ☐ PLCIS ☐ LCIS variant ☐ .....specify

Invasive: Ductal NOS ☐ Lobular ☐ Lobular Variant ☐ .....specify

Tubular/Crib ☐ Mucinous ☐ Medullary ☐

Papillary ☐ Metaplastic ☐ Mixed ☐.....specify

Other ☐ .....specify

GRADE : Tubule Score ☐ Pleo Score ☐ Mitotic Score ☐

Total Mitoses/10hpf.....

Gd1 ☐ Gd2 ☐ Gd3 ☐

Lympho-Vascular Permeation ☐ Yes ☐ No

Lymph Infiltrate None ☐ Mild ☐ Mod ☐ Severe ☐

Tumour Border Pushing ☐ >50% ☐ <50% ☐ Infiltrative ☐

Central scar/fibrosis Yes ☐ No ☐

Spindle/Sq metaplasia Yes ☐ No ☐

Appendix 2.2: Example of a TMA map.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1					L											L
2		L					L									L
3		L			L				L				L			
4									L				L			
5		L	L	L	L				L						L	L
6					L	L	L				L		L		L	L
7		L														L
8		L		L							L			L		L
9	L					L	L							L		
10				L								L	L	L	L	
11	L	L	L	L	L	P	P									

Controls used in this TMA: L= Liver, *P*= Placenta.

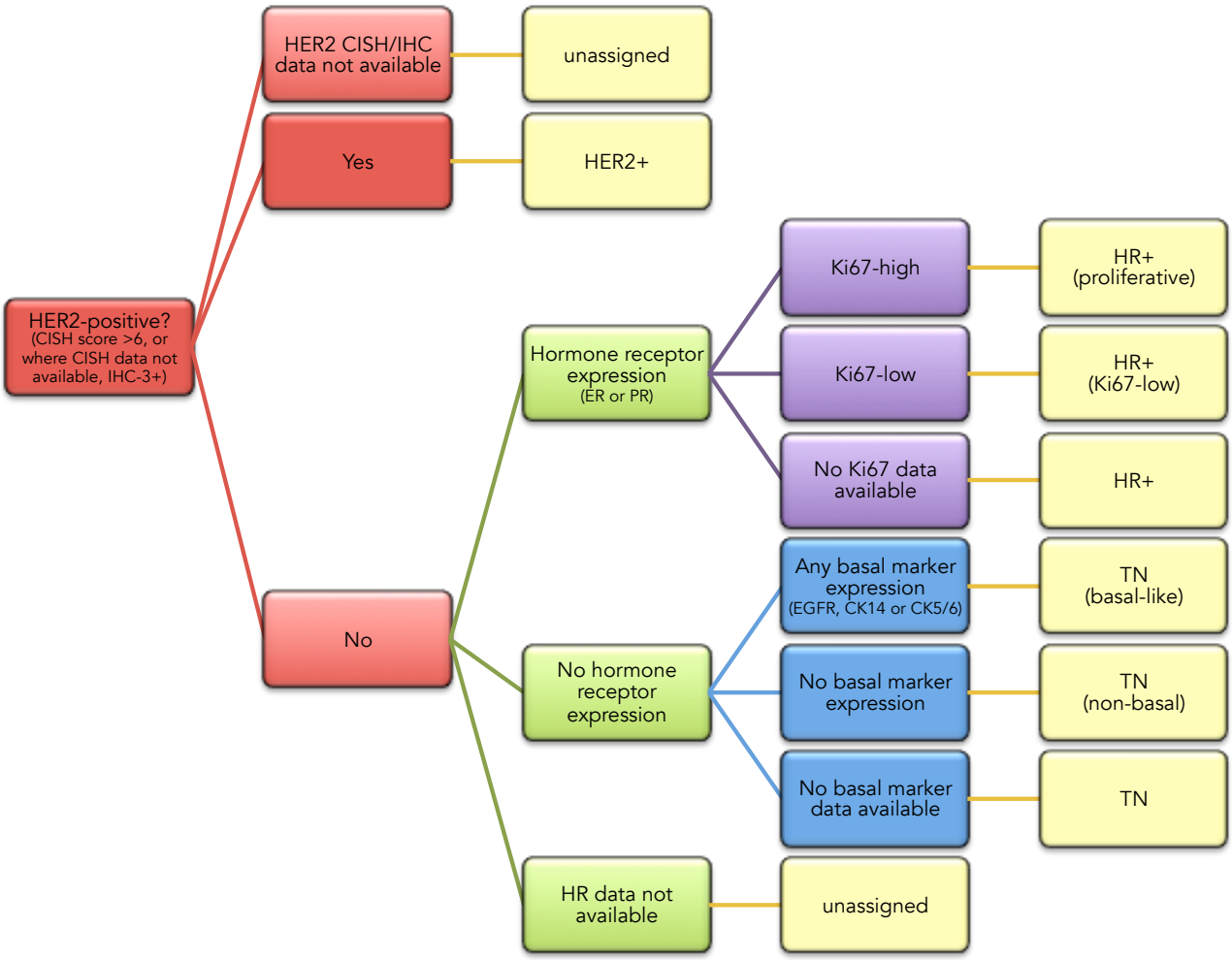
Appendix 3.1: Histopathologic features of 449 breast cancer cases used in this study.

Histopathologic variables	n	%
Tumour grade		
1	63	14
2	220	49
3	161	35.9
Unknown	5	1.1
Total	449	
Lymph node status		
Positive	116	25.8
Negative	135	30.1
Unknown	198	44.1
Total	449	
Tumour size		
<2	222	49.4
2 to 5	131	29.2
>5	18	4
Unknown	78	17.4
Total	449	
Histological type		
Ductal NOS	256	57
Lobular	34	7.6
Lobular variant	30	6.7
Mixed	77	17.1
Tubular/Crib	10	2.2
Mucinous	3	0.7
Mataplastic	18	4
Medullary	2	0.4
Papillary	2	2.4
Others	11	2.4
Unknown	6	1.3
Total	449	

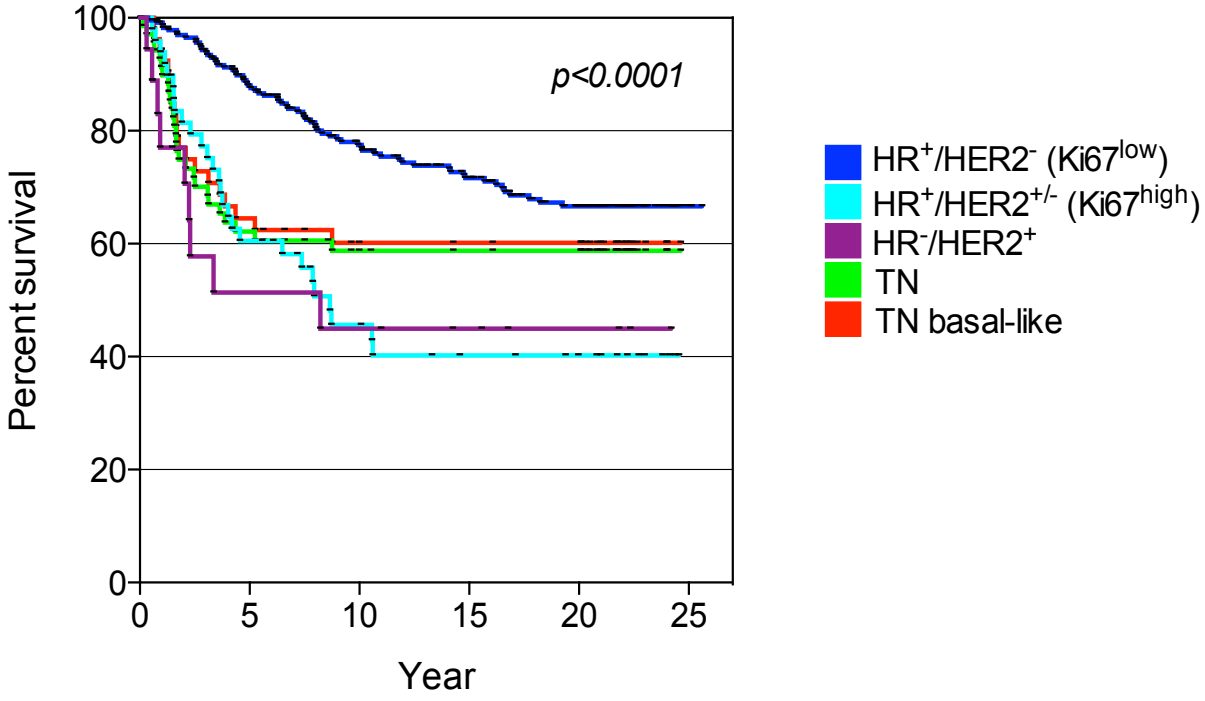
Appendix 3.1 continued: Histopathologic features of 449 breast cancer cases used in this study.

Histo pathological variables	n	%
Immuno-marker positivity		
ER	334	77.1
Unknown	16	
PR	267	62.5
Unknown	22	
HER2 (IHC 3+ only)	72	17
HER2 (IHC 3+ & ISH >6)	38	9.2
Unknown	37	
CK5/6	43	10.6
Unknown	44	
CK14	29	7.4
Unknown	55	
EGFR	84	21.4
Unknown	57	
Ki67 > 10%	145	36.6
Ki67 > 20%	55	13.9
Unknown	53	

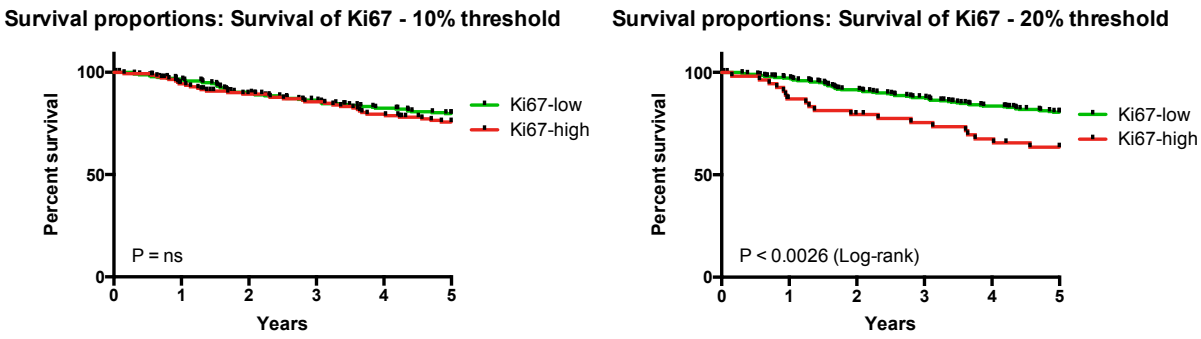
Appendix 3.2: Flow chart representing stratification of breast cancer cases into different prognostic subgroups based on expression of immunohistochemistry markers ER, PR, HER2, Ki67, EGFR and CK 5/6 and 14.



Appendix 3.3: Kalpan-Meier survival analysis of QFU cohort according to breast cancer subgroups defined by IHC surrogate marker panel.



Appendix 3.4: Kalpan-Meier survival analysis of QFU cohort based on expression of Ki67. Ten per cent threshold showed no difference in survival whereas 20% threshold showed a significant difference in survival outcome.



Appendix 3.5: Comparison of clinicopathologic variables within different breast cancer subtypes.

Parameters	Total	$HR^+/HER2^-$ (low Ki67)		$HR^+/HER2^{-/+}$ (high Ki67)		$HR^-/HER2^+$		Triple negative (TN)		TN-basal like	
		n	%	n	%	n	%	n	%	n	%
Number	386	248	64.2	49	12.7	18	4.7	71.0	18.4	54.0	14.0
Tumour grade											
1	52	51	98.1	0	0.0	0	0.0	1.0	1.9	0.0	0.0
2	177	142	80.2	20	11.3	2	1.1	13.0	7.3	9.0	5.1
3	154	53	34.4	29	18.8	15	9.7	57.0	37.0	46.0	29.9
Lymph node status											
Positive	103	61	59.2	15	14.6	11	10.7	16.0	15.5	11.0	10.7
Negative	114	71	62.3	16	14.0	2	1.8	25.0	21.9	21.0	18.4
unknown	169	116	68.6	18	10.7	5	3.0	30.0	17.8	23.0	13.6
Tumour size											
<2	199	144	72.4	21	10.6	7	3.5	27.0	13.6	21.0	10.6
2 to 5	114	62	54.4	14	12.3	8	7.0	30.0	26.3	22.0	19.3
>5	17	7	41.2	2	11.8	1	5.9	7.0	41.2	6.0	35.3
unknown	48	35	72.9	10	20.8	2	4.2	1.0	2.1	6.0	12.5
Histological type											
Ductal NOS	230	147	63.9	31	13.5	10	4.3	42.0	18.3	34.0	14.8
Lobular	41	32	78.0	5	12.2	2	4.9	2.0	4.9	0.0	0.0
Mixed	77	39	50.6	8	10.4	3	3.9	17.0	22.1	13.0	16.9
Tubular/Crib	9	9	100.0	0	0.0	0	0.0	0.0	0.0	0.0	0.0
Mucinous	3	2	66.7	1	33.3	0	0.0	0.0	0.0	0.0	0.0
Mataplastic	17	5	29.4	2	11.8	2	11.8	8.0	47.1	7.0	41.2
Medullary	2	2	100.0	0	0.0	0	0.0	0.0	0.0	0.0	0.0
Papillary	2	2	100.0	0	0.0	0	0.0	0.0	0.0	0.0	0.0
others	10	8	80.0	2	20.0	0	0.0	0.0	0.0	1.0	10.0
Immuno marker positivity											
ER (>1%)	297	248	83.5	49	16.5	0	0.0	0.0	0.0	0.0	0.0
PR (>1%)	243	209	86.0	34	14.0	0	0.0	0.0	0.0	0.0	0.0
HER2 (3+, >6 by ISH)	41	0	0.0	23	56.1	18	43.9	0.0	0.0	0.0	0.0
CK5/6	42	5	11.9	5	11.9	2	4.8	30.0	71.4	30.0	71.4
CK14	26	6	23.1	2	7.7	1	3.8	17.0	65.4	17.0	65.4
EGFR	81	21	25.9	9	11.1	10	12.3	41.0	50.6	41.0	50.6
Ki67 > 20%	52	0	0.0	26	50.0	3	5.8	23.0	44.2	19.0	36.5
SOX10	37	2	5.4	4	10.8	0	0.0	31.0	83.8	27.0	73.0
p53 (3+)	57	14	24.6	10	17.5	7	12.3	26.0	45.6	20.0	35.1
AR	331	241	72.8	41	12.4	17	5.1	32.0	9.7	19.0	5.7
c-kit	154	90	58.4	27	17.5	5	3.2	32.0	20.8	29.0	18.8

Appendix 3.6: Comparison of clinicopathologic variables based on SOX10 status in QFU cohort.

Invasive breast carcinomas

Variables		Total	n		%		P-value[#]
			SOX10 positive	SOX10 negative	SOX10 positive	SOX10 negative	
Clinical parameters							
Age	<40	41	6	35	14.6	85.4	ns
	40-50	90	11	79	12.2	87.8	
	51-60	86	7	79	8.1	91.9	
	>60	205	11	194	5.4	94.6	
Tumour size	<2 cm	171	14	157	8.2	91.8	ns
	2-5 cm	163	15	148	9.2	90.8	
	>5 cm	33	4	29	12.1	87.9	
	n	367					
Lymph node status	Positive	130	8	122	6.2	93.8	ns
	Negative	119	12	107	10.1	89.9	
	n	249					
LVI	Positive	103	9	94	8.7	91.3	ns
	Negative	338	28	310	8.3	91.7	
	n	441					
Histopathologic parameters							
Histological type	Ductal NOS	252	23	229	9.1	90.9	ns
	Lobular	64	0	64	0.0	100.0	
	Mixed type	76	11	65	14.5	85.5	
	Metaplastic	18	2	16	11.1	88.9	
	others	29	1	28	3.4	96.6	
	n	439					
Grade	I	61	0	61	0.0	100.0	**<0.0001**
	II	218	7	211	3.2	96.8	
	III	160	30	130	18.8	81.3	
	n	439					
Mitotic score	1	256	4	252	1.6	98.4	**<0.0001**
	2	58	2	56	3.4	96.6	
	3	124	31	93	25.0	75.0	
	n	438					

Appendix 3.6 continued: Comparison of clinicopathologic variables based on SOX10 status in QFU cohort.

Invasive breast carcinomas

Variables		Total	n		%		P-value[#]
			SOX10 positive	SOX10 negative	SOX10 positive	SOX10 negative	
Biomarkers							
Ki67 expression (20% threshold)	High	55	15	40	27.3	72.7	**<0.0001**
	Low	338	22	316	6.5	93.5	
	n	393					
HR Status (ER and/or PR)	Yes	340	6	334	1.8	98.2	**<0.0001**
	No	91	31	60	34.1	65.9	
	n	431					
HER2 status	Yes	40	0	40	0.0	100.0	**0.0338**
	No	352	36	316	10.2	89.8	
	n	392					
Triple negative (TN) status	Yes	71	31	40	43.7	56.3	**<0.0001**
	No	357	6	351	1.7	98.3	
	n	403					
Basal markers (CK5/6, CK14, EGFR)	Yes	110	32	78	29.1	70.9	**<0.0001**
	No	280	3	277	1.1	98.9	
	n						
TN-Basal status	Yes	49	22	27	44.9	55.1	**<0.0001**
	No	375	10	365	2.7	97.3	
	n	424					
p53 expression	Positive	57	14	43	24.6	75.4	**<0.0001**
	Negative	267	15	252	5.6	94.4	
	n	324					
AR expression	Positive	362	12	350	3.3	96.7	**<0.0001**
	Negative	42	24	18	57.1	42.9	
	n	404					
Vimentin expression	Positive	46	24	22	52.2	47.8	**<0.0001**
	Negative	357	12	345	3.4	96.6	
	n	403					
c-kit expression	Positive	162	26	136	16.0	84.0	**<0.0001**
	Negative	249	10	239	4.0	96.0	
	n	411					

[#]*Chi-square test*

Appendix 3.7: Comparison of clinicopathologic variables based on SOX10 status within triple negative and basal-like breast cancers

Variables		Triple negative (TN)				TN basal-like			
		total	**SOX10 positive**	**SOX10 negative**	**P-value[#]**	**total**	**SOX10 positive**	**SOX10 negative**	**P-value[#]**
Clinical parameters									
Age	<40	2	1	1	ns	6	4	2	ns
	40-50	2	0	2		13	8	5	
	51-60	5	2	3		8	2	6	
	>60	5	0	5		25	11	14	
Tumour size	<2 cm	26	10	16	ns	16	7	9	ns
	2-5 cm	30	14	16		27	13	14	
	>5 cm	7	4	3		6	4	2	
	n	63				49			
Lymph node status	Positive	16	8	8	ns	11	6	5	ns
	Negative	25	11	14		21	10	11	
	n	41				32			
LVI	Positive	15	6	9	ns	15	6	9	ns
	Negative	55	25	30		42	21	21	
	n	70				57			
Histopathologic parameters									
Histological type	Ductal NOS	41	19	22	ns	33	16	17	ns
	Lobular	2	0	2		0	0	0	
	Mixed type	18	10	8		14	10	4	
	Metaplastic	8	1	7		7	1	6	
	others	1	1	0		1	1	0	
	n	70				55			
Grade	I	1	0	1	ns	0	0	0	ns
	II	13	5	8		9	5	4	
	III	56	26	30		46	23	23	
	n	70				55			
Mitotic score	1	12	4	8	**0.0467**	7	4	3	ns
	2	9	1	8		8	1	7	
	3	49	26	23		39	22	17	
	n	70				54			

Appendix 3.7 continued: Comparison of clinicopathologic variables based on SOX10 status within triple negative and basal-like breast cancers

Variables		Triple negative (TN)				TN basal-like			
		total	**SOX10 positive**	**SOX10 negative**	**P-value[#]**	**total**	**SOX10 positive**	**SOX10 negative**	**P-value[#]**
Biomarkers									
Ki67 expression (20% threshold)	High	23	11	12		19	9	10	
	Low	43	20	23	ns	32	18	14	ns
	n	66				51			
p53 expression	Positive (3+)	25	11	14		19	10	9	
	Negative	31	14	17	ns	23	12	11	ns
	n	56				42			
AR expression	Positive	32	8	24		19	6	13	
	Negative	32	22	10	**0.001**	29	20	9	**0.0177**
	n	64				48			
Vimentin expression	Positive	31	20	11		27	16	11	
	Negative	35	10	24	**0.007**	23	10	13	ns
	n	66				50			
c-kit expression	Positive	32	20	12		29	18	11	
	Negative	36	10	26	**0.0068**	23	8	15	ns
	n	68				52			

[#]*Chi-square test*

Appendix 5.1: Comparison of clinicopathologic variables based on MAGE-A status in QFU cohort.

Invasive Breast Carcinoma

Variables		Total	n		%		P-value[#]
			MAGE-A positive	MAGE-A negative	MAGE-A positive	MAGE-A negative	
Clinical parameters							
Age	<50	108	10	98	9.3	90.7	ns
	>50	269	18	251	6.7	93.3	
Tumour size	<2 cm	160	9	151	5.6	94.4	ns
	2-5 cm	157	16	141	10.2	89.8	
	>5 cm	17	3	14	17.6	82.4	
	n	334					
Lymph node status	Positive	101	7	94	6.9	93.1	ns
	Negative	116	8	108	6.9	93.1	
	n	217					
LVI	Positive	98	6	92	6.1	93.9	ns
	Negative	296	25	271	8.4	91.6	
	n	394					
Histopathologic parameters							
Histological type	Ductal NOS	232	16	216	6.9	93.1	0.0244
	Lobular	43	1	42	2.3	97.7	
	Mixed type	73	8	65	11.0	89.0	
	Metaplastic	14	4	10	28.6	71.4	
	others	26	2	24	7.7	92.3	
	n	388					
Grade	I	54	0	54	0.0	100.0	< 0.0001
	II	187	8	179	4.3	95.7	
	III	150	23	127	15.3	84.7	
	n	391					
Mitotic score	1	218	7	211	3.2	96.8	< 0.0001
	2	54	4	50	7.4	92.6	
	3	118	20	98	16.9	83.1	
	n	390					

Appendix 5.1 continued: Comparison of clinicopathologic variables based on MAGE-A status in QFU cohort.

Invasive Breast Carcinoma

Variables		Total	n		%		P-value[#]
			MAGE-A positive	MAGE-A negative	MAGE-A positive	MAGE-A negative	
Biomarkers							
Ki67 expression (20% threshold)	High	55	12	43	21.8	78.2	0.0003
	Low	324	18	306	5.6	94.4	
	n	379					
HR Status (ER and/or PR)	Yes	305	8	297	2.6	97.4	< 0.0001
	No	87	23	64	26.4	73.6	
	n	392					
HER2 status	Yes	68	3	65	4.4	95.6	ns
	No	309	27	282	8.7	91.3	
	n	377					
Triple negative (TN) status	Yes	68	22	46	32.4	67.6	< 0.0001
	No	323	9	314	2.8	97.2	
	n	391					
TN-Basal status	Yes	53	18	35	34.0	66.0	< 0.0001
	No	320	11	309	3.4	96.6	
	n	373					
Basal markers (CK5/6, CK14, EGFR)	Yes	105	22	83	21.0	79.0	< 0.0001
	No	259	7	252	2.7	97.3	
	n	364					
p53 expression	Positive (3+)	180	32	148	17.8	82.2	0.0195
	Negative	133	11	122	8.3	91.7	
	n	313					
AR expression	Positive	193	4	189	AR	97.9	< 0.0001
	Negative	183	24	159	13.1	86.9	
	n	376					
Vimentin expression	Positive	46	12	34	26.1	73.9	< 0.0001
	Negative	344	18	326	5.2	94.8	
	n	390					
c-kit expression	Positive	144	21	123	14.6	85.4	0.0008
	Negative	230	10	220	4.3	95.7	
	n	374					
NY-ESO-1 expression	Positive	12	9	3	75.0	25.0	< 0.0001
	Negative	381	23	358	6.0	94.0	
	n	393					

[#]*Chi-square test*

Appendix 5.2: Comparison of clinicopathologic variables based on MAGE-A status within triple negative and basal-like breast cancers

Variables		Triple negative (TN)				TN basal-like			
		total	**MAGEA positive**	**MAGEA negative**	**P-value[#]**	**total**	**MAGEA positive**	**MAGEA negative**	**P-value[#]**
Clinical parameters									
Age	≤50	22	8	14		18	6	12	
	≥50	46	14	32	ns	35	12	23	ns
	n	68				53			
Tumour size	<2 cm	18	7	11		14	4	10	
	2-5 cm	35	11	24	ns	26	10	16	ns
	>5 cm	9	3	6		8	3	5	
	n	62				48			
Lymph node status	Positive	15	5	10		10	3	7	
	Negative	25	6	19	ns	21	6	15	ns
	n	40				31			
LVI	Positive	5	10	15		12	4	8	
	Negative	17	36	53	ns	41	14	27	ns
	n	22				53			
Histopathologic parameters									
Histological type	Ductal NOS	35	12	23		32	9	23	
	Lobular	2	0	2		0	0	0	
	Mixed type	15	7	8	ns	14	6	8	ns
	Metaplastic	7	3	4		7	3	4	
	others	0	0	0		0	0	0	
	n	59				53			
Grade	I	1	0	1		0	0	0	
	II	13	3	10	ns	9	2	7	ns
	III	54	19	35		44	16	28	
	n	68				53			

Appendix 5.2 continued: Comparison of clinicopathologic variables based on MAGE-A status within triple negative and basal-like breast cancers

Variables		Triple negative (TN)				TN basal-like			
		total	**MAGEA positive**	**MAGEA negative**	**P-value[#]**	**total**	**MAGEA positive**	**MAGEA negative**	**P-value[#]**
Biomarkers									
Ki67 expression (20% threshold)	High	23	10	13		19	8	11	
	Low	42	11	31	ns	31	9	22	ns
	n	65				50			
p53 expression	Positive (3+)	25	10	15		19	8	11	
	Negative	31	9	22	ns	23	7	16	ns
	n	56				42			
AR expression	Positive	31	8	23		19	5	14	
	Negative	32	13	19	ns	29	12	17	ns
	n	63				48			
Vimentin expression	Positive	31	9	22		27	7	20	
	Negative	34	12	22	ns	23	10	13	ns
	n	65				50			
c-kit expression	Positive	31	16	15		28	14	14	
	Negative	35	6	29	0.0041	23	4	19	0.0201
	n	66				51			
NY-ESO-1 expression	Positive	9	7	2		9	7	2	
	Negative	57	15	42	0.0048	42	11	31	0.006
	n	66				51			

[#]*Chi-square test*

Appendix 5.3: Characteristics of NY-ESO-1 positive tumours in QFU cohort.

#	Age	Size (mm)	Grade	Invasive type	LFP	Subtype	Tumour border	NY-ESO -1 (%)	NY-ESO-1 Intensity	NY-ESO- 1 Location
1	48	15	3	IC NST	alive	TNBC	infiltrative	3	weak	Cy
2	53	15	3	IC NST	dead	TNBC	infiltrative	2	weak	Cy
3	49	35	3	Mixed	dead	TNBC	infiltrative	2	weak	Nu
4	46	15	3	IC NST	dead	TNBC	infiltrative	2	weak	Nu/Cy
5	50	12	3	IC NST	alive	ER/PR+HER2-	infiltrative	1	weak	Nu/Cy
6	54	25	3	IC NST	alive	TNBC	Pushing <50%	3	weak	Nu/Cy
7	75	25	2	IC NST	dead	TNBC	infiltrative	3	srong	Nu/Cy
8	72	25	2	Mixed	alive	TNBC	infiltrative	3	strong	Nu/Cy
9	uk	uk	2	Special type	dead	uk	Pushing >50%	3	strong	Nu/Cy
10	55	70	3	Mixed	alive	TNBC	infiltrative	3	strong	Nu/Cy
11	73	180	3	IC NST	alive	TNBC	Pushing <50%	2	strong	Nu/Cy
12	72	40	3	Mixed	alive	TNBC	infiltrative	3	strong	Nu/Cy
13	58	11	3	IC NST	alive	TNBC	infiltrative	2	strong	Nu/Cy
14	uk	uk	2	IC NST	uk	ER/PR+HER2-	infiltrative	1	weak	Nu/Cy

LFU- Last follow up, Cy- cytoplasm and Nu- nucleus