Prognostic and Predictive Biomarkers in Breast Cancer

Shristi Bhattarai

Follow this and additional works at: https://scholarworks.gsu.edu/biology_diss

Recommended Citation
doi: https://doi.org/10.57709/20361228

This Dissertation is brought to you for free and open access by the Department of Biology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Biology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
ABSTRACT

Breast cancer (BC) is a heterogeneous disease consisting of distinct subtypes that vary in prognosis. Routine diagnosis is limited to the assessment of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), there are just a few others that have been clinically validated to guide chemotherapy and medicolegal decisions for patients with BC. Although androgen receptor (AR) has recently emerged as a promising predictive and prognostic biomarker for BC, especially triple negative breast cancers (TNBCs), there is still an unmet need to risk-stratify risk BC patients and predict response to therapy. Thus, we hypothesize that a biomarker-guided deeper stratification of patients will improve prognostication; aid in tailored therapy and decision making in medicolegal cases.
My research has primarily focused on evaluating biomarkers that can determine \textit{in vivo} tumor growth rate, predict response to neoadjuvant in BC patients, and risk-stratify TNBC patients using a combination of \textit{in silico} analysis, \textit{in vitro} assays and RNA-sequencing. Our clinically relevant growth rate model derived from Ki67, histological tumor size and mitotic index, stratified tumors into fast-growing versus slow-growing tumor subgroups, wherein patients with fast-growing tumors experienced poorer BC-specific survival. Evaluation of different biomarkers to predict pCR in BC patients revealed that HER2+ and TNBC subtypes had higher pCR rates compared with the luminal subtype. ER and PR negativity, HER2 positivity, Nottingham grade 3, increased TLI and SLI, high mitotic count and Ki67 score correlated significantly with pCR. Evaluating AR status shows population-specific patterns of association with patients’ overall survival after controlling for age, grade, population, and chemotherapy. My study validates the striking association of AR loss with worse clinical outcome. The collective data offers compelling evidence to support misregulation of oncogenic Wnt/\(\beta\)-catenin in AR negative scenario.

Collectively, my work has revealed a prognostic model that can predict the \textit{in vivo} breast tumor growth rate and offers several useful application; identified immunohistochemical and clinicopathological biomarkers that are independent predictors of neoadjuvant chemotherapy; stratify risk in TNBC patients based on AR status; and uncovered molecular pathways that can optimize targeted therapy to combat TNBCs that lack AR.

INDEX WORDS: Breast Cancer, In-vivo growth rate, Neoadjuvant chemotherapy, Androgen Receptor, Wnt signaling pathway, Prognosis, Predictive, Biomarkers
PROGNOSTIC AND PREDICTIVE BIOMARKERS IN BREAST CANCER

by

SHRISEI BHATTARAI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020
PROGNOSTIC AND PREDICTIVE BIOMARKERS IN BREAST CANCER

by

SHRISTI BHATTARAI

Committee Chair: Ritu Aneja
Committee: Emad Rakha
Zhi-Ren Liu

Electronic Version Approved:

Office of Graduate Services
College of Arts and Sciences
Georgia State University
December 2020
DEDICATION

This dissertation is dedicated to the memory of my father in-law. Although he was one of my biggest inspiration to pursue my doctoral degree, he was unable to see my graduation. This is for him who courageously fought his battle with cancer. I would also like to dedicate this work to my loving parents and my husband without whose never-failing sympathy and encouragement I would not have been able to endure to this path.
ACKNOWLEDGEMENTS

I would like to acknowledge and thank Dr. Ritu Aneja who is an excellent scientific mentor and role model. Her support in all ups and downs in my life was instrumental in my decision to continue my doctoral studies. I would also like to extend my heartfelt gratitude to Dr. Padmashree CG Rida for her vital encouragement, motivation and crucial contribution. I wish to thank my thesis committee Dr. Emad Rakha and Dr. Zhi Ren Liu for their mentorship and guidance in developing my research project. I also wish to thank Dr. Karuna Mittal and Dr. Sergey Klimov who were exceptional colleagues to me throughout my graduate studies. I am also indebted to Chakraborty Garlapati and Shriya for their extensive support during my graduate studies and being a gem friend throughout. I also wish to thank my former and current teammates: Deepika, Brittney, Riri, Lihan, Gaurav, Anjello, and all the graduate students in Aneja lab.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................................................... V

LIST OF TABLES ........................................................................................................................................................ XII

LIST OF FIGURES ........................................................................................................................................................ XIII

LIST OF ABBREVIATIONS ......................................................................................................................................... XV

1 INTRODUCTION ........................................................................................................................................................ 1

1.1 Breast cancer: an overview .................................................................................................................................... 1

1.1.1 Epidemiology and risk factors ........................................................................................................................... 1

1.1.2 Screening and diagnosis ................................................................................................................................... 2

1.1.3 Treatment .......................................................................................................................................................... 3

1.2 Biomarkers in breast cancer ................................................................................................................................ 5

1.3 Tumor growth rate ............................................................................................................................................... 7

1.4 Molecular subtypes ................................................................................................................................................ 7

1.5 Triple-negative breast cancer ................................................................................................................................... 9

1.6 Androgen receptor signaling ................................................................................................................................ 12

1.6.1 AR: nuclear hormone receptors .......................................................................................................................... 12

1.6.2 Mechanism of action ....................................................................................................................................... 13

1.6.3 AR expression in TNBC ................................................................................................................................... 13

1.6.4 AR expression in different TNBC subtypes ....................................................................................................... 14

1.7 Canonical Wnt signaling pathway .......................................................................................................................... 15
2 MACHINE LEARNING-BASED PREDICTION OF BREAST CANCER

2.1 Abstract ..................................................................................................................29

2.2 Introduction ............................................................................................................29

2.3 Methods ..................................................................................................................31

2.3.1 Study cohort .......................................................................................................31

2.3.2 Calculating tumor volumes and growth rates ......................................................32

2.3.3 Selecting optimal tumor volume, growth rate combination and development of
SM-INVIGOR ..............................................................................................................32

2.3.4 Assessing and scoring immunohistochemical staining ......................................33

2.3.5 Development of the machine learning-based surrogate model (Surr-INVIGOR)35

2.3.6 Validation of Surr-INVIGOR ............................................................................35

2.3.7 Statistical analysis ..............................................................................................36

2.4 Results ....................................................................................................................36

2.4.1 Clinicopathological and molecular features of cases in the study cohort ..........36

2.4.2 Development of SM-INVIGOR, a significant predictor of breast cancer-specific
survival .........................................................................................................................36

2.4.3 Development of a clinically-relevant surrogate model (Surr-INVIGOR) for in-vivo
growth rate prediction ...............................................................................................38
2.4.4 Validation of Surr-INVIGOR in an independent BC case series demonstrates its robust prognostic value

2.4.5 Surr-INVIGOR can be used to determine tumor age at diagnosis in a subset of breast tumors

2.5 Discussion

2.6 References

3 BIOMARKERS PREDICTING PATHOLOGIC COMPLETE RESPONSE TO NEOADJUVANT CHEMOTHERAPY IN BREAST CANCER

3.1 Abstract

3.2 Introduction

3.3 Methods

3.3.1 Patient Selection and Clinicopathologic Characteristics

3.3.2 Pathologic Evaluation

3.3.3 Breast Cancer Classification and Definition of pCR

3.3.4 Statistical Analysis

3.4 Results

3.4.1 Clinicopathologic Features of the Cases

3.4.2 HER2+ and TNBCs have higher pCR Rate than Luminal Breast Cancer

3.4.3 High TLI and SLI are significantly associated with pCR rate in HER2+ breast cancers and TNBCs
3.4.4  ER and PR expression are associated with decreased pCR in the Overall Cohort and HER2+ Breast Cancers ................................................................. 63

3.4.5  High mitotic count and Ki67 score and Nottingham histologic Grade 3 are significantly associated with pCR rate in the overall cohort .................... 63

3.4.6  High correlation between stromal and intratumoral lymphocytic infiltration..... 63

3.4.7  Luminal type breast cancers have low pCR rate and no parameter is strongly associated with pCR ................................................................. 64

3.5  Discussion ........................................................................................................... 64

3.6  References ......................................................................................................... 67

4  PROGNOSTIC ROLE OF ANDROGEN RECEPTOR IN TRIPLE NEGATIVE BREAST CANCER: A MULTI-INSTITUTIONAL ............................................ 74

4.1  Abstract ............................................................................................................. 74

4.2  Introduction ....................................................................................................... 74

4.3  Methods ............................................................................................................ 76

4.3.1  Study Cohorts and Samples ....................................................................... 76

4.3.2  Immunohistochemistry (IHC) ................................................................. 77

4.3.3  Assessment of IHC Staining ..................................................................... 77

4.3.4  Statistical Analyses ...................................................................................... 78

4.4  Results and Discussion ..................................................................................... 79

4.5  References ......................................................................................................... 82
UPREGULATION OF WNT/β-CATENIN SIGNALING UNDERLIES AGGRESSIVE DISEASE COURSE IN QUADRUPLE NEGATIVE BREAST CANCER

5.1 Abstract ......................................................................................................................91

5.2 Introduction ................................................................................................................92

5.3 Methods .......................................................................................................................94

5.3.1 Study Cohorts and Samples ......................................................................................94

5.3.2 Immunohistochemistry (IHC) ..................................................................................95

5.3.3 Assessment of IHC staining .....................................................................................95

5.3.4 Cell culture ..............................................................................................................95

5.3.5 AR knockdown and overexpression ..........................................................................96

5.3.6 Western blotting and antibodies ..............................................................................96

5.3.7 Cell proliferation assay ..........................................................................................97

5.3.8 Invasion assay .......................................................................................................97

5.3.9 Wound healing assay ............................................................................................98

5.3.10 Wound healing assay ..........................................................................................98

5.3.11 RNA isolation and quantitative RT PCR ..............................................................98

5.3.12 Statistical analysis of experimental data ..............................................................99

5.3.13 Statistical analysis of clinical data .........................................................................99

5.3.14 Insilico data analysis ............................................................................................99
5.4 Results........................................................................................................................................100

5.4.1 Lack of AR in TNBC patients is associated with poor overall survival.............100

5.4.2 In silico analysis reveals upregulation of Wnt signaling in QNBCs .................101

5.4.3 Upregulation of Wnt/β-catenin is associated with loss of AR expression ........102

5.4.4 Aberrant localization of membrane β-catenin staining is associated with poor overall survival........................................................................................................................................104

5.5 Discussion ....................................................................................................................................104

5.6 References....................................................................................................................................107

6 CONCLUSIONS..................................................................................................................................123

REFERENCES .......................................................................................................................................131
LIST OF TABLES

Table 2.6.1 Clinicopathological characteristics ..............................................................53
Table 2.6.2 Table showing the mathematical representation of the growth rates tested.....54
Table 2.6.3 Details of antibodies used and dilutions.........................................................55
Table 2.6.4 The list of surrogate markers to develop Surr-INVIGOR..........................56
Table 2.6.5 Table with the chosen hyperparameters for the optimal KNN model........56
Table 2.6.6 Table showing model fit statistics (Akaike information criterion) for various combinations of growth rate functions and tumor volumes...............................57

Table 3.6.1 Clinicopathological characteristics of study cohort................................71
Table 3.6.2 Univariate logistic analysis of pCR in the total cohort and within subtypes. ....72
Table 3.6.3 Multivariate Logistic Regression Analysis for pCR in the Overall Cohort and HER2+ and TNBC Subtypes.................................................................73

Table 4.5.1 Table representing number of cases and sample type evaluated for AR expression in global TNBC cohort and diverse institutions from US......................87
Table 4.5.2 Patient demographic and tumor data stratified by cohort; n (%). .................88
Table 4.5.3 Table showing overall survival of AR positive TNBC at different thresholds of AR. .......................................................................................................................89
Table 4.5.4 Multivariate analysis to reveal population-specific differences between cohorts. .........................................................................................................................90

Table 5.6.1 Clinicopathological variables of the study cohort.........................................114
Table 5.6.2 Multivariate analysis after adjusting the confounding variables..................115
LIST OF FIGURES

Figure 2.1 Schematic depicting sequences of steps in our study leading to the calculation of SM-INVIGOR and the development of Surr-INVIGOR that predicts in-vivo tumor growth rate in BC. ...........................................................48

Figure 2.2 Prognostic significance of SM-INVIGOR...........................................................49

Figure 2.3 Tumor Growth Rate group (fast versus slow) classification accuracies obtained using different machine learning algorithms and features (included sequentially as indicated on the x-axis). ........................................................................50

Figure 2.4 Scatterplot representing the optimal predicted in-vivo growth rate (through regression using surrogate markers) versus the SM-INVIGOR from the study cohort ........................................................................................................51

Figure 2.5 Plot representing the duration before diagnosis that patients in the slow growth subgroup have had a tumor volume >0 mm3 (black lines)........................................52

Figure 4.1 Androgen receptor (AR) expression in the seven cohorts of our multi-institutional study. ..........................................................................................................................85

Figure 4.2 Prognostic significance of the androgen receptor (AR) in different study cohorts. ..........................................................................................................................86

Figure 5.1 AR expression and its association with overall survival.............................116

Figure 5.2 QNBCs are associated with higher proliferation, invasion and migration......117

Figure 5.3 Loss of AR is associated with higher proliferation, invasion and migration.....118

Figure 5.4 Upregulation of Wnt/β-catenin signaling pathway in QNBCs..................119

Figure 5.5 Upregulated of Wnt/β-catenin is associated with loss of AR..................120

Figure 5.6 Aberrant localization of β-catenin is associated with QNBC patients.........121
Figure 5.7 Loss of membrane β-catenin expression is associated with poor survival among QNBCs.
LIST OF ABBREVIATIONS

A
AR: Androgen receptor

B
BC: breast cancer
BCSCs: breast cancer stem cell
BCSS: breast cancer specific survival
BRCA1 breast cancer type-1 susceptibility protein
BRCA2 breast cancer type-2 susceptibility protein

D
DAB: 3,3'-diaminobenzidine
DBD: DNA-binding domain
DCIS: Ductal carcinoma in situ
DFS: disease-free survival
DMFS: Distant metastasis-free survival

E
EGF: Epidermal growth factor
EGFR: epidermal growth factor receptor
EMT: epithelial-mesenchymal transition
EpCAM: Epithelial cell adhesion molecule
ER: Oestrogen receptor
ER+: ER positive

F
FISH fluorescence in situ hybridisation

FZD: frizzled receptors

G

GSK-3β: glycogen synthase kinase 3 beta

H

H&E: haematoxylin and eosin

H2O: Water

HER2: human epidermal growth factor receptor-2

HIF: hypoxia-inducible factor

HR: hazard ratio

HRP: Horse-radish peroxide

H-score: Histological score

I

IBC: Invasive breast cancer

IHC: Immunohistochemistry

ILC: Invasive lobular carcinoma

L

LBD: ligand-binding domain

LN: lymph nodes

LVI: lymphovascular invasion

M

mg/ml: milligram/milliliter

MHC1 major Histocompatibility complex class I
N
NACT: Neoadjuvant chemotherapy
NPI: Nottingham Prognostic Index
P
PAM50: prediction analysis of microarray of 50 genes
PBS: phosphate-buffered saline
pCR: pathological complete response
PR: Progesterone receptor
R
RA: retinoic acid
RARs: retinoid receptors
S
SIL: Stromal infiltrating lymphocytes
T
TBS: Tris Buffered Saline
TGF-β: Transforming growth factor beta
TILs: tumor infiltrating lymphocytes
TMA: tissue microarray
TNBC: Triple negative breast cancer
V
V: Voltage
1 INTRODUCTION

1.1 Breast cancer: an overview

1.1.1 Epidemiology and risk factors

Breast cancer is the most frequently diagnosed and a common type of cancer in women (24.2%) with an estimated number of more than 2 million new cases worldwide. According to the World Health Organization (WHO), breast cancer is the leading global cause of cancer-related death among women (15%), with approximately 627,000 breast cancer-related deaths annually worldwide (1). The higher incidence rates of breast cancer are in North America, Western Europe, Oceania, and Argentina (1). This may be attributed to genetic predisposition, a high prevalence of environmental risk factors for breast cancer, and/or overdiagnosis due to screening for the disease, as many of the countries have implemented national screening programs (2). In the United States (US) alone, approximately 268,600 new cases of invasive breast cancer were diagnosed, and 41,760 women died from this disease in 2019 (3).

Breast cancer is considered to be a genetic disorder. There are many factors that may contribute to the development of breast cancer, encompassing both reproductive and lifestyle-related factors. Germline mutations reflect the hereditary predisposition to breast cancer (4). Different external and lifestyle factors like age, gender, menarche, late menopause, nulliparity, and late age at first childbirth, use of oral contraceptives, and hormone replacement therapy may heighten the risk of developing breast cancer (5-9). Several lifestyle factors such as obesity, lack of exercise, excessive alcohol consumption, tobacco smoking, and a high dietary fat intake have been associated with the development of breast cancer (10-19). About 25% of hereditary breast cancers are associated with mutations in highly penetrant genes including BRCA1, BRCA2, TP53,
PTEN, STK11, and CDK1 (20). These mutations are associated with a greater than 50% lifetime risk of developing breast cancer.

1.1.2 Screening and diagnosis

Breast cancer is diagnosed and treated in a multidisciplinary approach, involving specialists from surgical oncology, medical oncology, radiation oncology, pathology, and radiology. This multidisciplinary approach has resulted in a significant decrease in breast cancer mortality compared to what it was a decade ago. Mammogram screening is commonly used for the detection of breast cancer; however, in instances where cancer is not detected via mammogram screening, patients are still assessed for symptoms. The most common symptom is the presence of a lump in the breast; other symptoms may include swelling, skin irritation, pain, redness, and nipple discharge or retraction (21).

During mammogram screening, low doses of X-rays are used to examine the breast for possible cancerous masses. Mammogram screenings aim to detect breast cancers at an earlier stage of the disease compared to symptom-based detection (22), to increase the chance of cure. The success of breast screening lies in the timely detection of cancer by mammography. False-negative mammography is one of the principal reasons for delayed diagnosis of breast cancer (23-26). Even though some studies report high sensitivity (>90%) for diagnostic mammography, such results have not proven universal (27). Among many factors, age appears to be one of the important factors underlying false-negative results because the high radiographic density of breasts in young women makes cancer detection difficult (28). Mammograms are generally capable of detecting tumors as small as 2 mm in diameter, which equates to a tumor that consists of approximately $10^7$ cells and whose mass has doubled about 23 times (29). Even though early detection is often reported to improve breast cancer outcomes, the impact of mammography screening on breast cancer mortality
is still debated (30). No significant effect on breast cancer mortality rates has been observed solely based on mammogram screening (31); however, an independent panel in the United Kingdom has estimated, based on available published data, a 20% survival improvement among women that undergo mammography screening between 50–70 years of age (30). Controversial survival benefit is often juxtaposed with potential harms such as overdiagnosis and overtreatment, but the possibility of early breast cancer diagnosis in asymptomatic women with low tumor burden, which can potentially lead to de-escalation of cancer therapy, should not be overlooked (32, 33).

1.1.3 Treatment

Systemic treatment includes chemotherapy, endocrine therapy, and/or human epidermal growth factor 2 receptor-targeted therapy and immunotherapy. Recommendations for the use of systemic therapy are based on the individual patient's risk and the balance between absolute benefit and toxicity. Overall, chemotherapy regimens based on anthracyclines and taxanes reduce breast cancer mortality by about one-third (34, 35). Chemotherapy can be applied in the neoadjuvant or adjuvant setting.

1.1.3.1 Neoadjuvant chemotherapy

Neoadjuvant chemotherapy, also known as primary or preoperative chemotherapy, consists of chemotherapy delivered prior to local treatment (surgery). Neoadjuvant chemotherapy is an alternative to adjuvant chemotherapy for both early and locally advanced breast cancer and results in an equivalent disease outcome in terms of disease-free and overall survival (36, 37). This type of chemotherapy has become the standard of care in patients with locally advanced breast cancer or borderline irresectable breast cancer, as more effective drugs have become available. Moreover, this treatment is being developed to extend it to patients with less advanced or resectable breast cancer. For patients with early stages of breast cancer, downstaging of the primary tumor as a
result of neoadjuvant chemotherapy may facilitate breast conserving therapy; this treatment may also result in a possible downstaging of the axilla, thus obviating the need for axillary treatment in some patients (38). In a trial setting, a neoadjuvant approach is more attractive because a more rapid outcome is available with fewer patients, compared with an adjuvant approach. Moreover, neoadjuvant chemotherapy has shown to eradicate nodal disease in 20–40% of the patients (39). Performing a sentinel node biopsy post-neoadjuvant chemotherapy might be an attractive strategy to take maximum benefit of its effect on nodal downstaging and may potentiate axilla-conserving treatment. Predicting which patients will achieve pathologic complete response (pCR) to neoadjuvant chemotherapy is important because this approach is not without risk. For instance, although neoadjuvant chemotherapy can prolong disease-free and overall survival, it may increase the rate of ipsilateral tumor recurrence compared with adjuvant therapy. In addition, delaying surgery may decrease survival. Therefore, in Chapter 3, we conducted a comprehensive evaluation of tumor morphology and biomarker status and correlated these parameters with pCR rate in a neoadjuvant chemotherapy setting.

1.1.3.2 Adjuvant chemotherapy

Historically, adjuvant chemotherapy was administered to patients with node-positive breast cancer. However, the development of different predictive and prognostic biomarkers has led to a more complex assessment of relapse risk and chemo-/endocrine-sensitivity prior to treatment decisions on an individual level. Data from a large meta-analysis of 100,000 women revealed a one-third reduction in breast cancer mortality after adjuvant chemotherapy (35). Initially, the combination of cyclophosphamide, methotrexate, and 5-fluorouracil resulted in a one-fourth reduction in breast cancer mortality at 20 years post-treatment, compared to that with no adjuvant chemotherapy (40) treatment. Addition of anthracyclines decreased the relative risk for breast
cancer mortality by 21% (absolute risk reduction at 10 years post-treatment, 6.5%) (35), compared to that with no adjuvant chemotherapy. There is a strong biological rationale for intensifying chemotherapy regimens. Increasing the dose or shortening the treatment interval (dose-dense chemotherapy) can potentiate treatment effect and produce better cancer-related outcomes (41, 42). Moreover, adjuvant chemotherapy for breast cancer can result in short-term toxicity and particularly, anthracycline-based regimens tend to cause more short-term toxicity than non-anthracycline regimens (43). The most commonly described symptoms of short-term toxicity are fatigue, nausea and emesis, alopecia, myelosuppression, and mucositis. The reported frequency of nausea has been varying; however, more recently, with the widespread adaptation of primary prophylaxis with serotonin antagonists (such as ondansetron and palonosetron) and the neurokinin-1 receptor antagonist (aprepitant) (44), the frequency of nausea is expected to be lower. Specific chemotherapy regimens such as taxanes can also cause peripheral neuropathy (in up to 22% of cases) and myalgia (43, 45).

1.2 Biomarkers in breast cancer

Due to increased life expectancy and changes in lifestyle, the incidences of many cancers, among them breast cancer, have slowly increased. Complete prevention or eradication of all cancers will probably not be possible. Instead, early detection and further improving treatments will become increasingly important. The treatment of breast cancer has drastically improved in the past years. Today, most women who are diagnosed with breast cancer have a good prognosis. However, due to the high incidence of breast cancer, many women still die of the disease.

Identification of novel biomarkers and understanding the biological processes of breast cancer development are important early steps in improving survival. To date, only a handful of useful breast cancer biomarkers that offer actionable information on the patient’s disease course
are currently recommended by the American Society of Clinical Oncology (ASCO). The ASCO guidelines recommend routine testing for the estrogen receptor (ER) and progesterone receptor (PR), along with the ErbB family member human epidermal growth factor receptor, to aid breast cancer treatment. Overexpression of ER and PR indicates that the patient will benefit from endocrine therapy. The ERBB2 proto-oncogene encodes the receptor tyrosine kinase ErbB-2, also known as HER2, and is amplified and/or overexpressed in approximately 15–20% of breast cancers (46, 47). HER2 overexpression prognosticates increased tumor aggressiveness and a higher incidence of recurrence. Furthermore, HER2 overexpression is a predictive factor for response to targeted therapy using the monoclonal antibody trastuzumab (48). Ki-67 is a cellular marker for proliferation that identifies ER-positive breast cancer patients who would benefit from adjuvant chemotherapy (49). Patients whose tumors are negative for all three receptors (triple-negative breast cancers, TNBCs) also have a less favorable prognosis than those with hormone receptor-positive breast cancers. Unlike HER2+ tumors, there are no approved targeted therapies for TNBCs, so the only available systemic treatment is cytotoxic chemotherapy with its often-devastating side effects. Several biomarkers and genomic tests are available that may predict the risk of recurrence and guide the selection of adjuvant therapy for patients with hormone receptor-positive early-stage invasive breast cancer. These biomarkers include urokinase-type plasminogen activator, plasminogen activator inhibitor type 1, Oncotype DX, PAM50, EndoPredict, and Breast Cancer Index (50). However, for HER2+ and TNBC patients, these tests are contraindicated because there is insufficient evidence to suggest that they are clinically useful. Thus, there is an unmet need for biomarkers of clinical utility for these high-risk patients. Any new biomarker needs to contribute clinically useful information beyond what is already provided by the current clinical and histopathological markers.
1.3 Tumor growth rate

Tumor response to therapy may also be studied by analyzing the effect of therapy on the natural growth of the tumor. Modeling the natural (therapy-naive) growth of tumors is not only valuable for the study of tumor progression, but will also support the optimization of screening programs, prognostication (50), optimal scheduling of chemotherapy (51) and radiation therapy, and assessment of tumor spread (number and size distribution of metastases, including micrometastases) (52, 53). The ability to mathematically predict tumor growth has always garnered much interest since the early days of cancer research. Many models have been proposed, but there is still no consensus on the growth patterns exhibited by solid tumors (54). It is important to have an accurate model of tumor growth for evaluating screening strategies (55), optimizing radiation treatment protocols (56, 57), and making decisions about patient treatment (58, 59). However, determining the rate of breast cancer growth in vivo, which, in turn, can predict prognosis, has remained elusive.

In Chapter 2, I discuss the need for a predictive growth model and describe the model that I developed, which predicts the rate of in vivo tumor growth using a unique study cohort of breast cancer patients who had two serial mammograms wherein the tumor, visible in the diagnostic mammogram, was missed in the first screen.

1.4 Molecular subtypes

Although the current WHO classification of breast cancer relies on tumor morphology, molecular analyses of tumors have caused a paradigm shift. Microarray-based gene expression profiling has shown that breast cancer encompasses a collection of different diseases with unique patterns of gene expression. Distinct molecular patterns in breast cancer samples described by Perou et al. paved the way for a more tailored, tumor biology-guided classification and
management of both early and metastatic breast cancer. The different gene expression patterns led to 4 clusters: basal-like, HER2+, normal breast-like, and luminal/epithelial (60). Given the global heterogeneity of access to gene expression analysis, immunohistochemistry is currently used as a surrogate test to distinguish the different molecular subtypes. Based on that method, five distinct intrinsic subtypes have been identified (61-63):

1. Luminal A: Represents the most common molecular subtype of breast cancer (60–70%) and is characterized by low proliferation measured by Ki67, strong ER and PR expression, and are usually HER2-negative (HER2−).

2. Luminal B-like HER2−: Accounts for about 10–20% of breast cancers and is characterized by ER expression, PR expression, high tumor cell proliferation according to Ki67, and lack of HER2 expression.

3. Luminal B-like HER2+: Characterized by the same biomarkers as above but with additional expression of HER2.

4. HER2-enriched (non-luminal): Presents with high tumor cell proliferation and HER2-expression but lacks ER and PR expression. All HER2-positive cancers, regardless of ER/PR status, account for about 15% of breast cancers.

5. TNBC: Lack both endocrine receptors (ER and PR) and HER2 expression and is commonly accompanied by high tumor cell proliferation.

These intrinsic subtypes have been reproduced across independent gene expression datasets (64-66), and they have demonstrated differences in behavior and prognosis (65, 67, 68). The luminal subclass is characterized by ER expression, genes related to the ER pathway, and high expression of luminal cytokeratin (CK7, CK8, CK18, and CK19). The luminal A subgroup expresses higher levels of ER-related genes and lower levels of proliferation-related genes than
those of luminal B tumors (68). The HER2 subclass is associated with the amplification of the HER2 gene. The basal-like subgroup lacks expression of ER and HER2-related genes. These tumors express basal cytokeratins (CK 5, 6, 14, 15, and 17), and demonstrate high proliferative activity (69). The luminal A and B, HER2-enriched, and basal-like subtypes were also confirmed by The Cancer Genome Atlas (TCGA) research network, although significant heterogeneity within these subclasses was underscored (11, 70). In addition, three mainly ER-negative subtypes have been proposed, comprising the interferon-rich (64), molecular apocrine (71, 72), and claudin-low subtypes (73, 74). Further refinement of the triple-negative subclass (75, 76) and the HER2-enriched breast cancers have also been suggested (4, 77, 78). In 2012, a large global gene study (METABRIC) identified 10 distinct disease subgroups which further subclassify both ER-positive and ER-negative tumors (79). Interestingly, unique genomic portraits of the morphologically defined lobular carcinomas have recently been demonstrated (80-82). Still, this is only the beginning of cancer genomics, and the definitive molecular classification of breast carcinomas is yet to be established.

1.5 **Triple-negative breast cancer**

TNBC, which is defined by the lack of all three markers (ER, PR and HER2) used today to personalize treatment, represents about 15% of all breast cancers but is a highly malignant subtype, with earlier age of onset, high risk of metastasis, and unfavorable clinical prognosis. In the absence of an available targeted therapy, TNBC patients have a worse prognosis than patients with other subtypes, with reduced overall survival and greater risk of recurrence (83).

Given the nature of heterogeneity, a collection of studies has profiled the distinct genetic landscape and therapeutic response of TNBC. However, as a diagnosis of exclusion, tumors in TNBC are not necessarily a single disease, and, in fact, demonstrate great diversity in
histopathologic features and genetic profiles. A TCGA study of 510 exome-sequenced breast tumors found that in a majority of TNBC tumors, point mutations were most commonly detected in the TP53 gene—80% of study patients harbored at least one mutation in that gene. The gene coding for phosphatidylinositol-4,5-biphosphate 3-7 kinase catalytic subunit alpha (PIK3CA) was found to be the second most commonly mutated gene with 8% of patients harboring mutations in that gene. Most other somatic mutations were scattered among a multitude of genes at low frequency (84). In another study, Lehman et al analyzed the transcriptomic profiles of 587 TNBC cases and demonstrated that TNBC consists of seven subtypes and displays a heterogeneous biology with differing responses to various therapies (76). By measuring a total of 2,188 genes and consensus clustering, they recognized seven molecular subtypes, namely basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M), mesenchymal stem-like (MSL), immunomodulatory (IM), luminal androgen receptor (LAR), and unstable. Each subtype represents a unique genetic background and driver signaling pathway. For example, the BL1 subtype represents increased cell cycle and DNA damage response gene signature, while BL2 involves high growth factor signaling. M represents gene enrichment in cell motility, growth, and differentiation, which is partially similar to MSL, but with low expression of proliferative genes. IM is associated with immune cell processes, whereas LAR is associated with elevated androgen signaling. Lehmann and colleagues recently refined their sub classification into four tumor-specific subtypes (BL1, BL2, M, and LAR), after recognizing the influence of infiltrating lymphocytes and tumor-associated stromal cells on IM and MSL subtypes (85). TNBC stratification using mRNA and DNA profiling include the four subtypes classified by Burstein et al: basal-like immune-suppressed (BLIS), basal-like immune-activated (BLIA), mesenchymal (MES), and LAR (75). The diversity in TNBC subtypes and their notable overlapping features not only reaffirm the significant heterogeneity in this disease
but also highlight the need for a more comprehensive and optimized designation of TNBC molecular subtypes to eventually translate into clinical settings.

Clinically, TNBC patients have a wide range of outcomes and chemotherapy responses. Although TNBC has a worse prognosis than other types of breast cancer, patients who achieve pCR after chemotherapy (defined by the complete absence of residual tumor tissue on pathologic examination) show similar progression-free survival to that of patients of eight other breast cancer types; however, patients with Residual Disease burden have significantly worse progression-free survival (86).

The mainstay of TNBC treatment is chemotherapy with nonspecific cytotoxins, which does not eliminate the tumor in almost 80% of patients (86). These treatment-resistant patients have a dismal prognosis—almost 50% die within 5 years. Furthermore, TNBC patients with metastasized tumors only survive for about 1 year after treatment. Therefore, a diagnosis of metastatic TNBC is essentially a death sentence for the patient (87). Given the unreliability of nonspecific cytotoxins in suppressing TNBC metastasis, there is a dire need to identify specific molecules that drive TNBC metastasis so that targeted metastasis-suppressing drugs can be rationally designed. This strategy has been very effective for other breast cancer subtypes, which are treated with targeted drugs like Tamoxifen, Arimidex, or Herceptin alone or in combination with cytotoxins. Approximately 0.5 million women in the US have TNBC, which is ~15–20% of the nearly 3 million women living with breast cancer (88, 89). African American women are overrepresented among TNBC patients, as they are ~3 times more likely to develop TNBC as white women (90). Among TNBC patients, mortality risk is ~2 times more in African Americans compared with whites after adjustment for age, grade, stage, and poverty index (91). Because TNBC is defined by the biomarkers it lacks, it is difficult for clinicians to provide a clear prognosis to patients and
to predict which patients require more aggressive chemotherapeutic regimens. The wide survival gap between non-TNBC and TNBC patients implies that TNBC patients may not be receiving sufficiently aggressive treatment. A lack of sufficient knowledge on biomarkers in TNBC that can meaningfully guide therapeutic interventions remains a critical barrier to eliminating treatment disparity for TNBC patients. Risk-stratifying biomarkers would constitute revolutionary progress in TNBC treatment and might also mitigate racial health disparity in breast cancer.

There are currently no targeted therapies for TNBC, and chemotherapy remains the best therapeutic option. However, upon recurrence of chemoresistant disease, effective therapeutic options are limited. Although TNBC lacks hormone receptors traditionally associated with breast cancer, both molecular and immunohistochemical analyses have demonstrated that a subset of TNBCs express the androgen receptor (AR). Recent data suggest that AR is known to significantly influences breast cancer gene expression profiles and also affects tumorigenic properties of TNBC (92). Development of new-generation anti-androgens for the treatment of prostate cancer has led to renewed interest in hormonal therapy targeting AR in the AR+ TNBC subset and constitutes a novel therapeutic option that could improve prognosis with relatively few side effects. The following section reviews the role of AR in the biology of TNBC based on preclinical models and clinical data on the efficacy of targeting AR.

1.6 Androgen receptor signaling

1.6.1 AR: nuclear hormone receptors

AR is part of the nuclear receptor superfamily of proteins whose members are ligand-inducible transcription factors (TF) that control the transcription of target genes in response to receptor-specific ligands (93). AR consists of a N-terminus transactivation domain (A/B domain), a central DNA-binding domain (DBD), and a C-terminus ligand-binding domain (LBD), with a
hinge region between DBD and LBDs as a homodimer, AR binds to specific inverted androgen response elements (AREs); especially at the consensus ARE sequence GGTACAnnnTGTTCT which is most prevalent (94-96). Moreover, AR has a ligand-dependent bipartite nuclear localization sequence located in the hinge region (amino acids 617−633) (97). A point mutation in this region (R617P) found in three separate clinical samples did not affect the ability of AR to bind to its ligand or DNA but did appear to repress transcriptional activation; therefore, this mutation most likely suppresses the transcriptional activity of AR by preventing its translocation to the nucleus (98-100). The polyproline domain, the polymorphic polyglutamine region, and the surrounding sequence (amino acids 141−338) of the AR N-terminus are required for complete transcriptional activity of AR (101-103).

1.6.2 Mechanism of action

Similar to ER and PR, AR is a member of the nuclear steroid hormone receptor family and acts as a transcription factor to regulate target genes. Testosterone and dihydrotestosterone are androgens that directly or indirectly act on AR. Once these lipophilic hormones pass through the cell membrane, they bind to AR. This binding causes a conformational change in AR, which displaces it from the heat shock protein complex. AR then translocates to the nucleus, dimerizes, recruits specific coregulators, and influences transcription of its specific target genes. Prior to binding of the ligand, the LBD of AR stabilizes the interaction between unliganded-AR and the HSP63. After binding of the ligand, the LBD interacts with the N-terminus of AR to stabilize the bound androgen (104).

1.6.3 AR expression in TNBC

AR is expressed in approximately 7−75% of TNBCs depending on the cut-off value used to determine positivity and thus, has emerged as a promising therapeutic target for TNBC patients
Importantly, nuclear AR staining is indicative of active receptors, as AR translocates to the nucleus upon ligand binding as previously described. Discrepancies in AR expression in TNBC are largely attributed to variability in sample procuring methods, antibodies, staining and scoring methods, and AR positivity cut-off values. In two prospectively conducted clinical trials, rates of AR\(^+\) disease ranged from 12\% (AR \(\geq\) 10\%, DAKO) to 55\% (AR \(\geq\) 10\%, DAKO and Ventana (107, 113). Studies comparing AR expression in primary versus metastatic disease found that AR is frequently retained in metastatic samples from patients with AR\(^+\) primary tumors (114, 115). One subtype of TNBC that is dependent on AR signaling is LAR (76). Studies have shown that this subtype expresses high AR mRNA levels along with increased expression of AR target genes and exhibits sensitivity to AR-targeted therapies. However, with regard to TNBC, the prognostic significance of AR is controversial, as multiple studies have shown that AR expression can be associated with both good and bad prognosis while several studies have reported that AR loss is associated with worse prognosis in TNBCs (116-119).

Thus, to reconcile conflicting reports on the impact of AR loss on long-term prognosis of TNBC, in Chapter 4, we determined the prognostic value of AR in diverse TNBC cohorts from different countries.

1.6.4 AR expression in different TNBC subtypes

AR protein expression is ten-fold higher in LAR tumors than in non-LAR subtypes (76). LAR cell lines show a high frequency of PIK3CA mutations (120). Xenograft experiments showed that LAR TNBC cell lines demonstrated greater sensitivity to the AR antagonist bicalutamide than non-LAR xenografts. The combination of PI3K/mTOR and AR inhibitors showed additive growth inhibitory effects in vitro. Additive effects of bicalutamide with the PI3K inhibitor GDC-0941 or the PI3K/mTOR inhibitor GDC0980 were also observed in MDA453 and CAL-148 LAR
xenografts (120). In a comparison of global DNA-binding events, Robinson et al. found that AR binding in the LAR TNBC cell line MDA453 was more similar to ER binding in an ER⁺ breast cancer cell line (MCF7) than to AR binding in a prostate cancer cell line (LNCaP) (117). Thus, in the absence of ER, AR may function similarly to ER, accounting for the luminal transcriptome of the LAR subtype.

Although AR is most highly expressed in LAR tumors, it is also expressed in non-LAR subtypes (92, 121). Cell lines representing non-LAR and AR⁺ TNBC subtypes, including mesenchymal stem-like, mesenchymal-like, and basal-like, exhibit decreased anchorage-independent growth when treated with the AR antagonist Enza (92). Basal-like HCC1806 and mesenchymal stem-like SUM159PT xenografts with relatively low AR expression, which were treated with Enza, exhibited decreased viability and increased necrosis. The response of non-LAR xenografts with relatively low AR expression to Enza provides promising preclinical data that suggest TNBC patients with relatively low AR expression may also benefit from anti-androgen therapy. Indeed, results of the TBCRC011 trial demonstrate that some non-LAR TNBC patients benefit from treatment with bicalutamide.

1.7 Canonical Wnt signaling pathway

The Wnt/β-catenin signaling pathway, also named as the canonical Wnt pathway, refers to a molecular cascade, which is initiated in normal cells by the binding of a Wnt ligand to its cognate receptor complex consisting of a Frizzled receptor and an low-density lipoprotein-related protein (LRP) co-receptor. This is followed by the phosphorylation of the integrator molecule Disheveled. By associating with Axin, Disheveled blocks the formation of the scaffolding complex that is responsible for presenting β-catenin for phosphorylation by Glycogen synthase kinase 3 beta GSK-3β. Unphosphorylated β-catenin escapes ubiquitination and proteasomal degradation, accumulates
in the cytoplasm and, in certain circumstances, may translocate to the nucleus. In the nucleus, β-catenin cooperates with the T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) family of transcription factors to activate the transcription of target genes (122, 123). In general, the canonical Wnt signaling pathway regulates cell proliferation, differentiation, and cell fate determination during animal development. In adult organisms, it regulates tissue homeostasis and stem cell maintenance, and perturbation of the pathway leads to cancer. Extensive research in different animal systems has demonstrated the critical role of the canonical Wnt pathway in different aspects of embryonic development. Indeed, mutation studies of different genes in the canonical Wnt pathway have revealed its importance in the gastrulation process (124). Aberrations in that pathway are known to cause a wide range of pathologies in humans. In many different types of cancers, mutations in at least one of the canonical Wnt signaling components have been found. In colorectal cancer, around 85% of loss-of-function mutations in the adenomatous polyposis coli gene (APC) have been detected, resulting in elevated β-catenin levels. Furthermore, mutations in β-catenin have been detected in many different cancer types, in which the N-terminal phosphorylation sites of β-catenin are affected. These mutations affect the targeted degradation of β-catenin by the destruction complex, resulting in its stabilization (124, 125).

Although mutations in Wnt signaling pathway components are observed in many cancer types, mutations in β-catenin or other canonical Wnt signaling components like AXIN or APC are rarely found in breast cancer. However, there is evidence for aberrant Wnt/β-catenin signaling activation in a majority of breast cancers, which results in the translocation of β-catenin to the nucleus and the subsequent transcription of its 26 target genes (126). Therefore, immunohistochemical staining can be used to detect elevated levels of cytoplasmic or nuclear β-catenin, which is linked to reduced overall survival. Consistent with these findings, canonical Wnt
target genes (e.g., Cyclin-D1) have been reported to be upregulated in a majority of breast cancer patients (127). Therefore, in those patients, it is possible that upstream signaling pathways and/or other mechanisms that lead to the stabilization of β-catenin are deregulated.

Experimental evidence suggests that the Wnt signaling pathway in breast cancer may be altered due to a loss of expression of negative pathway regulators or overexpression of individual Wnt ligands. For example, upregulation of Wnt2, Wnt4, Wnt7b, and Wnt10b has been confirmed in a subset of breast cancers (128, 129). Epigenetic silencing of negative pathway regulators like Wnt inhibitory factor 1 (WIF1), secreted frizzled-related protein (SFRP), Dickkopf 1 and Dickkopf 3 (DKK1 and DKK3), as well as APC and E-cadherin (CDH1) are found to be hypermethylated and their expression to be significantly reduced in human breast cancers (130-134). Consistently, increased nuclear β-catenin expression is detected in many breast tumors. Furthermore, Wnt/β-catenin signaling is upregulated in TNBC/basal-like breast cancer compared to hormone receptor-positive/luminal breast cancer; and pharmacologic inhibition of Wnt in TNBC inhibits cell migration and invasion (135). AR is known to interact with β-catenin in an androgen-dependent manner (136, 137). Androgen-bound AR can inhibit β-catenin’s target gene expression as a result of competitive binding with TCF/LEF1 transcription factors. Furthermore, AR inactivation in female mice is known to upregulate oncogenic Wnt/β-catenin signaling (138). Thus, in Chapter 5, we validated the prognostic role of AR expression in a US cohort and identified the canonical Wnt signaling pathway components that can be therapeutically targeted in QNBCs.

1.8 Hypothesis and Aims

Recently, very few clinically useful breast cancer biomarkers – that is, those that offer actionable information that may change a breast cancer patient’s disease course. The overall hypothesis of this dissertation is that a biomarker-guided deeper stratification of patients will
improve prognostication; aid in tailored therapy and medicolegal decision making. These biomarkers will provide additional prognostic insight and rational strategies to optimize targeted therapy, thereby contributing to the literature on the development of precision medicine.

The aims of this dissertation are:

I. To identify predictors of BC in-vivo growth rate, evaluate the impact of BC growth rate on disease outcome and develop a surrogate model that robustly predicts pre-diagnosis in-vivo growth rate.

II. To evaluate biomarkers to better predict pCR in neoadjuvant chemotherapy, determine pCR rates associated with different major subtypes of breast cancer, and explore the association of TLIs in pCR in BC subtypes.

III. To determine the prognostic role of AR in multi-institutional TNBC cohort.

IV. To validate the impact of AR loss on prognosis of TNBC and uncover the molecular pathways/drivers that can be therapeutically targeted in QNBCs.
1.9 References


2 MACHINE LEARNING-BASED PREDICTION OF BREAST CANCER GROWTH RATE IN-VIVO

2.1 Abstract

Determining the rate of breast cancer (BC) growth in vivo, which can predict prognosis, has remained elusive despite its relevance for treatment, screening recommendations and medicolegal practice. We developed a model that predicts the rate of in vivo tumor growth using a unique study cohort of BC patients who had two serial mammograms wherein the tumor, visible in the diagnostic mammogram, was missed in the first screen. A serial mammography-derived in vivo growth rate (SM-INVIGOR) index was developed using tumor volumes from two serial mammograms and time interval between measurements. We then developed a machine learning-based surrogate model called Surr-INVIGOR using routinely assessed biomarkers to predict in vivo rate of tumor growth and extend the utility of this approach to a larger patient population. Surr-INVIGOR was validated using an independent cohort. SM-INVIGOR stratified discovery cohort patients into fast-growing versus slow-growing tumor subgroups, wherein patients with fast-growing tumors experienced poorer BC-specific survival. Our clinically relevant Surr-INVIGOR stratified tumors in the discovery cohort and was concordant with SM-INVIGOR. In the validation cohort, Surr-INVIGOR uncovered significant survival differences between patients with fast-growing and slow-growing tumors. Our Surr-INVIGOR model predicts in vivo BC growth rate during the pre-diagnostic stage and offers several useful applications.

2.2 Introduction

Breast cancer (BC) is a heterogeneous disease with tumors exhibiting variable morphology, molecular profiles, behavior, and response to therapy. Mounting evidence demonstrates that BC
shows variable rates of growth, which has important clinical and medicolegal implications (1-4). *In-vivo* growth rate is not only a quantifiable trait of the tumor but can also serve as a tool to plan and evaluate screening programs, clinical trials or epidemiologic studies. In addition, BC growth rate evaluated using tumor size from mammograms may predict tumor response to chemotherapy and may help in determining the likely time of tumor initiation and previous tumor size in medicolegal cases (5-7). BC growth rate is also associated with prognostic variables such as lymph node status, stage and vascular invasion (3, 4, 8); however, the prognostic and predictive value of BC growth rate has not been harnessed in routine practice due to the inherent difficulty in its assessment in the short intervals between diagnosis and treatment.

Although the growth rate of BC *in-vivo* is strictly regulated, it appears to be dependent on the balance between several variables including growth fraction (the tumor cells that are proliferating and leading directly to the addition of new tumor cells), the rate of tumor cell loss by apoptosis and/or necrosis, tumor cells’ doubling-time/kinetics, and the surrounding microenvironment including angiogenesis, blood supply, and host immune response to the proliferating tumor cells (9-12). The complexity of the processes controlling BC growth and the interaction with the tumor microenvironment make assessment and prediction of BC growth rate a challenging task. Therefore, serial imaging of BC at different time points is considered as the best model available for assessing the *in-vivo* growth rate and for determining associations between potential intrinsic growth rate determinants and BC behavior, including response to therapy.

This study utilizes a discovery cohort comprising clinically and molecularly well-characterized data from BC patients who underwent serial mammography. It is a unique and rare cohort because the second mammogram illuminted that the tumor was indeed “missed” during the first mammogram. We find that this one-of-a-kind cohort can be interrogated to (a) identify predictors
of BC *in-vivo* growth rate, (b) evaluate the impact of BC growth rate on disease outcome, and (c) develop a surrogate model that robustly predicts pre-diagnosis *in-vivo* growth rate for patients who would normally not have tumor volume data from two serial mammograms. In contrast to a matched first-presentation-only BC patients’ cohort, BC growth rate in this study is determined by the changes in tumor volume between sequential mammograms, wherein the first mammogram “mistakenly” reported the case as normal/benign and the cancer was identified in the screening mammogram on a retrospective review subsequent to the second (diagnostic) mammogram (Figure 2.1).

2.3 Methods

2.3.1 Study cohort

The study cohort comprised of 114 BC patients aged between 50-70 years who were presented at the Nottingham City Hospital from 1988 to 2008 with BC, and for whom review of the previous screening mammogram showed a previously undetected tumor at the same affected site. This may have been due to either a false-negative screening outcome, or due to minimal visible signs of malignancy on the previous mammogram. Mammographic abnormalities included measurable soft tissue abnormality (mass, distortion or asymmetry) on screening and diagnostic films. On retrospective review of the previous mammogram after the disease diagnosis, two radiologists (blinded to each other’s observations) confirmed the “missed” cancer. We selected patients in whom a soft tissue abnormality was detected (upon retrospective review of prior screening mammograms) at the site of the subsequent cancer. Due to a misdiagnosed mammogram, this cohort uniquely comes with an earlier screening measurement with a visible tumor. Clinicopathological data including age, histological tumor type, primary tumor size, lymph node status, histological grade, Nottingham Prognostic Index (NPI), vascular invasion and patients’
outcome data were obtained. BC-specific survival (BCSS) was defined as the time interval (in months) between the primary surgeries and death from BC. The mean survival time of this cohort of patients was 120 months. Clinicopathological variables were available for 92 cases and the BCSS was available in 90 cases; thus, we restricted our study to these cases (Figure 2.1A).

2.3.2 Calculating tumor volumes and growth rates

The two measurements in the screening and diagnostic mammograms were assumed as tumor diameter and tumor height, which were then used to calculate tumor volumes at the time of screening and diagnosis. The greater mammogram dimension was assumed as height corresponding to the diameter of the semi-major axis, and the other dimension was regarded as diameter of the semi-minor axis. For tumor volume calculation, we considered the aforementioned dimensions as volume inputs for a cylinder, sphere, and an oblate spheroid (13). For tumor growth rates, we tested exponential growth (14, 15), the Gompertz model (16), and power law growth with the exponent set to both the classic value of 2/3 (17, 18) and 1/2 (19) as shown in Table 2.6.2. For all models, the initial volume for the growth rate was determined using the screening mammogram and the final volume was determined from the diagnostic mammogram, with the time variable denoted by the days between the two mammograms.

2.3.3 Selecting optimal tumor volume, growth rate combination and development of SM-INVIGOR

Multiple tumor volume/three-dimensional shape assumptions and growth rate functions used in previous studies (19), were tested to find the optimal combination that was prognostic. Growth rate indices that combined tumor volume (calculated assuming the tumor to be a sphere, cylinder, or spheroid) and individual growth functions (calculated assuming exponential growth,
two sets of the Power Law function (α =1/2 or 2/3), or Gompertz growth), were compared on the basis of their prognostic ability. Growth rates were used either as a continuous variable or through a fast/slow growth cutoff determined through optimizing the log-rank statistic (20, 21). Both forms of all growth rates were analyzed univariately in a Cox proportional hazard regression model using 10-year breast cancer specific survival (BCSS), and corresponding model fits were ranked with the Akaike Information Criterion (AIC) (22). The best-fitting growth rate index was chosen via the lowest relative AIC and was used in subsequent analyses. Data related to changes in volume of the lesion between the time of screening and at diagnosis, as well as the time between screening and diagnosis, were used to estimate the Serial Mammography-derived In-vivo Growth Rate (SM-INVIGOR) (Figure 2.1B). To control for common clinicopathological confounders, the growth rate model was also analyzed with multivariate Cox regression alongside grade, age, and estrogen receptor (ER) status. In addition, the tumor volumes at the screening and diagnostic time-points were tested prognostically to evaluate the prognostic significance of the change in tumor volume versus that of the screen- or diagnostic mammogram-calculated volume individually (Figure 2.1C).

2.3.4 Assessing and scoring immunohistochemical staining

For each patient, a representative formalin-fixed paraffin wax-embedded (FFPE) tumor block of the resected tumor was obtained from the Nottingham breast tumor bank (Figure 2.1D). Full-face sections 4 μm thick from the representative FFPE tumor blocks were prepared onto Xtra® Surgipath glass slides and were used for immunohistochemical (IHC) assessment of the following markers: estrogen receptor (ER), progesterone receptor (PR), HER2 (human epidermal growth factor receptor 2), the proliferation markers Ki67 and MCM2 (Minichromosome Maintenance 2), the basal markers CK5/6 (cytokeratin 5/6) and EGFR epidermal growth factor

*Parts of this chapter have been published verbatim in British Journal of cancer 2019 Sep; 121(6):497-504. Doi:10.1038/s41416-019-0539-x as “Machine learning-based prediction of breast cancer growth rate in vivo.”*
Parts of this chapter have been published verbatim in British Journal of cancer 2019 Sep; 121(6):497-504. Doi:10.1038/s41416-019-0539-x as “Machine learning-based prediction of breast cancer growth rate in vivo.”

receptor), the apoptosis markers BCL2 and cleaved caspase-3. IHC was performed on tissue sections using Novolink™ Max Polymer Detection System. (Leica, Newcastle, UK). Briefly, heat-assisted retrieval of antigen epitopes was performed in citrate buffer (pH 6) using a microwave for 20 minutes, followed by immediate cooling. The slides were rinsed with Tris-Buffered Saline (TBS, pH 7.6). The primary antibodies as summarized in Table 3 were applied for 30 minutes at room temperature except for cleaved caspase-3 staining. For cleaved caspase-3 a pre-fabricated detection kit (SignalStain® Cleaved Caspase-3 (Asp175) IHC Detection Kit #8120, Cell Signaling Technology) was used following manufacturer’s instructions. Other markers were stained using our protocols as previously published (23, 24).

Appropriate positive and negative controls were used for each marker and included in each staining run. Only the invasive tumor cells were scored independently by two observers (SB and MA) blinded to each other’s scores and clinicopathological data. Cases with discordant results were further reviewed by both observers to achieve scoring consensus. For each marker, the percent and intensity of staining were assessed, and H-scores were generated. For ER, PR, and HER2, cut-offs according to published guidelines were used (25, 26). Ki67, and cleaved caspase-3 were assessed and scored as previously described (23, 24). BC molecular subtypes were defined based on their IHC expression profile into: a) luminal (ER+ and/or PR+ /HER2-), b) HER2+ (HER2-positive), c) Triple negative (TN; ER-, PR-, HER2-) and d) Basal-like Breast cancer (BLBC: TN+ CK5/6+) (24). A total of 92 cases were informative for IHC biomarkers and these comprised the study cohort in the subsequent analyses including molecular markers (Figure 2.1E).
2.3.5 Development of the machine learning-based surrogate model (Surr-INVIGOR)

The above mentioned clinical and molecular variables, and immunohistochemical biomarkers (Table 2.6.4) were evaluated using machine learning algorithms to identify an optimal feature set that could serve as a surrogate model for SM-INVIGOR to predict fast or slow in-vivo growth rate for cases where only a single (diagnostic) mammogram is available (Figure 2.1F/G/H). The significance of mean differences for all potential surrogate variables, between fast- and slow-growing tumors, was first calculated using a 2-tailed t-test; this was followed by a ranking of the variables based upon their discriminating capacity. Multiple classification algorithms (support vector machines,, naïve Bayes, decision trees, discriminant analysis, ensemble), with optimized hyperparameters (27, 28) were then tested. The machine learning algorithm and feature set that resulted in the maximum 5-fold cross-validated accuracy (mean of 100 iterations) was chosen. For each trained machine learning model (combination of biomarkers), hyperparameters were fit through Bayesian optimization (27, 28) over 180 iterations (Table 2.6.5). Furthermore, a combination of variables was used, in an optimized regression model, to identify if the continuous growth rate value for each patient could be determined. Finally, the outputs from the machine learning-based approach were compared to the regression-based models which did not yield good R² values owing to small sample size.

2.3.6 Validation of Surr-INVIGOR

The prognostic performance of this surrogate model (Surr-INVIGOR) was tested in an independent, well-characterized large validation cohort of 1241 BC patients using Kaplan-Meier survival analysis (Figure 2.1I/J). Multivariate Cox regression was used to control for confounding effects of common clinicopathological variables.
2.3.7 Statistical analysis

All statistical analyses were carried out with SAS 9.4® software and MATLAB Version 9.2. Clinicopathological proportion differences between growth groups were determined using the \( \chi^2 \) test. Continuous clinicopathological variable differences were evaluated via a 2-tailed t-test. Prognostic time to event analysis was performed using Kaplan-Meier and Cox Proportional Hazard regression, wherein a death due to BC was considered as an event and every other outcome was censored. For all analyses, \( p<0.05 \) was considered significant.

2.4 Results

2.4.1 Clinicopathological and molecular features of cases in the study cohort

Most patients in the study cohort showed features associated with good prognosis including lower grade and negative (65%) or early positive (pN1; 26%) lymph nodes. Age at the time of diagnosis ranged from 50 to 73 years (mean=60.3 years, median=61.0 years). There was a predominance of the luminal A subtype with 85% positive for ER while HER2 overexpression was identified in only 6% of the patients. Ki67 staining ranged from 0 to 96%, with a mean expression of 19% (Table 2.6.1). Moreover, there was a significant correlation between the histological tumor size and the mammogram tumor size at time of diagnosis (Pearson’s correlation=0.58870; \( p<0.0001 \)).

2.4.2 Development of SM-INVIGOR, a significant predictor of breast cancer-specific survival

Since fast \( \text{in-vivo} \) growth prior to diagnosis is a sign of aggressive disease and could lead to poor outcomes, we reasoned that the growth rate model of choice would be the one that is most prognostic. Thus, we evaluated various combinations of growth rate functions and assumptions regarding the tumor’s three-dimensional shape. The best fitting model of tumor volume and growth
rate was obtained using the assumption that the study cohort comprises spherical tumors growing at a power law ($\alpha=0.5$) rate; this growth rate function (SM-INVI\textsc{gor}) stratified the tumors into slow-growing and fast-growing subgroups and produced a minimum cross validated AIC of 152.621 (Table 2.6.6). Using these assumptions, tumor volumes at the time of screening ranged from 53-56,115 mm$^3$ (mean of 2,742 ± 7,619 mm$^3$). This contrasted with tumor volumes at diagnosis, which ranged from 61 to 61,562 mm$^3$ (mean=5,573 ± 8,768 mm$^3$). The mean time difference between date of first screening and that of second diagnostic screening was 18 months, (range 4-37 months, median=17.5 months). Tumor growth rate differed considerably from patient to patient, ranging from 0 to 0.53 mm$^3$/day (mean=0.08 ± 0.13 mm$^3$).

SM-INVI\textsc{gor} used a cutoff of 0.045 mm$^3$/day to stratify tumors into slow-growing (n=53) and fast-growing (n=37) subgroups. Faster SM-INVI\textsc{gor} significantly associated with clinicopathological factors normally associated with poorer prognoses, such as larger histological tumor size (p=0.0023), high grade (Grade 3) (p=0.0186), more mitotic divisions (p=0.0134), apparent vascular invasion (p=0.0139), and a poor Nottingham Prognostic Index (p=0.011) (Figure 2A). SM-INVI\textsc{gor} varied significantly between BC molecular subtypes with the highest rate observed in triple-negative BC (TNBC) compared to other subtypes (p<0.05). Among the proliferation/apoptosis-related biomarkers that were immunohistochemically assessed (Table 2.6.4), only Ki67 showed a significant mean difference (p=0.0003) between the fast- (24%) versus slow- growing (11%) tumor subgroups. Furthermore, patients with higher tumor growth rate showed significantly poorer survival (BCSS=71.7%) relative to the slow-growing tumors (BCSS=91.9%) as shown in Kaplan Meier’s survival graph (Figure 2.2B). SM-INVI\textsc{gor} retained prognostic significance (p=0.0299, high growth rate HR=4.605) upon controlling for common
clinicopathological variables including grade, age and ER status. In fact, SM-INVIGOR was the only variable significantly associated with BCSS in our multivariable analysis (Figure 2.2C).

2.4.3 Development of a clinically-relevant surrogate model (Surr-INVIGOR) for in-vivo growth rate prediction

Unlike the patients in our unique discovery cohort, most begin therapy at an initial cancer diagnosis, and are therefore unlikely to have two serial mammograms with two tumor volume measurements. Because of this difference, SM-INVIGOR is limited in its utility to derive in-vivo tumor growth rate for most BC patients in routine clinical practice. Therefore, to extend the benefits of having growth rate data (or estimates) to a much larger group of patients lacking a second mammogram, we developed a machine learning-based surrogate growth rate model for SM-INVIGOR and called it Surr-INVIGOR (described in Suppl. data). Surr-INVIGOR non-linearly combines multiple clinicopathological variables and immunohistochemical biomarkers to predict in-vivo growth rate. First, we evaluated the ability of individual clinicopathological variables to serve as potential surrogate features and discriminate between the fast- and slow- growing tumor subgroups of our study cohort (p-values for mean difference between the subgroups is shown in Table 2.6.5. Ki67 (p=0.000265), mitotic score (MI; p=0.002479), tumor size (p=0.003619), NPI (p=0.004163), and grade (p=0.021128) differed significantly between the fast- and slow- growing tumors. The seven variables (Ki67, Mitotic score, tumor size, NPI, Grade, Stage and Tumor size) with p value <0.2 were then tested in multiple machine learning-based classification algorithms via sequential selection (Figure 2.3). The maximized cross-validated accuracy, which indicates the optimal Surr-INVIGOR model, was obtained when three features (Ki67, MI, and histological tumor size) were used in a K-nearest neighbor algorithm or KNN (accuracy or concordance with the classification yielded by SM-INVIGOR=0.706). The Ensemble also yielded a 70% accurate
classifier but required 4 additional features; the more parsimonious KNN was thus selected for use in *Surr-INVIGOR*. Fitting an optimal regression model to predict the growth rate continuously resulted in a poor $R^2$, peaking at 0.22, as shown in Figure 2.4, perhaps owing to the small sample size. Thus, our machine learning-based *Surr-INVIGOR* model was a clinically-relevant, superior choice compared to regression-based models.

### 2.4.4 Validation of Surr-INVIGOR in an independent BC case series demonstrates its robust prognostic value

We then evaluated the prognostic ability of *Surr-INVIGOR* in an independent BC case series ($n=1241$) from Nottingham University Hospital, UK. Patient age at the time of diagnosis ranged from 21-71 years (mean=53.6 years, median=54 years). Most patients showed features associated with good prognosis including negative lymph vascular invasion (55.3%), and negative (61%) or showed 1-3 positive (30%) lymph nodes. Patient follow up time ranged from 1 to 120 months (mean=100.237, median=120 The clinicopathological features of patients are summarized in Table 2.6.1.

The clinicopathological variables that discriminated between slow- and fast-growing tumors are depicted in Figure 2.2D. Applying the previously-trained *Surr-INVIGOR* model, using the same input parameters on this naïve validation cohort resulted in significant BCSS stratification. Patients in the fast growth rate group ($n=922$, BCSS=72.9%) had a significantly lower survival than patients in the slow growth rate group ($n=269$, BCSS=92.3%) Figure 2.2E. After accounting for potential clinicopathological cofounders, *Surr-INVIGOR* retained prognostic significance (HR=1.758, $p=0.0361$) alongside grade as shown in Figure 2.2F.
2.4.5 Surr-INVIGOR can be used to determine tumor age at diagnosis in a subset of breast tumors

Using the different growth rate groups, we can estimate tumor age and the time of inception of a subset of tumors. Assuming the highest (bounded) power law ($\alpha=0.5$) growth rate (0.04593 mm$^3$/day) for the slow-growing subgroup, we can estimate the date after which the tumor was definitely present within the patients in the slow-growing tumor subgroup. Using these assumptions, we determined that the average tumor age at diagnosis of slow-growing tumors was 4.7 years (Figure 2.6.5). Using this methodology, it may be possible to determine whether a patient possessing a slow-growing tumor undetected at earlier screenings, had received a true-negative or false-negative (i.e., tumor was missed) screening result.

2.5 Discussion

Although several studies have investigated variables associated with pre-diagnosis in-vivo BC growth rate, only clinicopathological variables and a few molecular biomarkers have been studied in this context and the available tumor dimensions were limited due to the measurement of the tumor’s long-axis only (2, 5, 29, 30). This study utilized a unique cohort of cases with tumor volume measurements (derived using tumor diameter and height data) available from a pair of serial mammograms to derive their in-vivo growth rates (SM-INVIGOR). We explored the potential association of a larger number of molecular biomarkers with their in-vivo BC growth rate, reaffirmed that fast tumor growth rate has a profound impact on prognosis, developed and validated a surrogate model (Surr-INVIGOR) that can predict a gross scale (fast versus slow) in-vivo growth rate accurately in routine practice, and its medicolegal consequences.

The success of breast screening lies in the timely detection of cancer on mammography. False negative mammography is among the principal reasons for delayed diagnosis of BC(31-34). Even
though some authors quote high sensitivity (>90%) for diagnostic mammography, such results are not universal (35). Among many factors, age appears to be one of the important factors underlying false negative reporting because the high radiographic density of breast in young women makes detection difficult (6). Mammograms are generally capable of detecting tumors as small as 2 mm in diameter, which equates to a tumor of approximately $10^7$ cells and about 23 tumor doublings (36). In our study cohort, however, patients with tumors ranging from 4-55 mm received false-negative diagnoses in their screening mammograms, showing the imperfection associated with this technology and inherent human limitations associated with reading radiology films. Whether the spread of a tumor is due to delays in diagnosis and initiation of treatment, or due to the inherently more aggressive nature of the tumor cells themselves (i.e., higher \textit{in-vivo} tumor growth rate) is another highly controversial matter. Natural fears that the delay in diagnosis has reduced their chances of survival or of avoiding the life-sapping effects of chemotherapy, or the feeling that cosmetic outcomes which would have been better had the tumor been detected earlier, are frequent causes of patients seeking legal redress. The importance of breast imaging in BC diagnosis and the use of mammography in screening has thus pushed breast radiologists into the frontline for medicolegal actions (37). Cancers missed at screening but followed by a positive diagnostic mammogram are not common yet false negative mammography is among the principal reasons for delayed diagnosis of BC (31-34). Only few population-screening programs have reported data on this group of cancers, which makes our study cohort uniquely valuable. This cohort allowed us to develop a model to predict pre-diagnostic \textit{in-vivo} tumor growth rate and provide insights into the potential prognostic consequences of delays in BC diagnosis.

Our study has yielded several key insights into features and the prognostic significance of the rate of tumor growth in its early stages. In our study, we found that \textit{SM-INVIGOR} varies considerably
and is consistent with findings by Weedon-Fekjaer and colleagues (5) who reported that the time BC takes to grow from 10 mm to 20 mm in diameter varied from less than 1.2 months to more than 6.3 years. Our current study also reinforced previous findings that higher grade and larger tumors with high proliferative activity are likely to have faster \textit{SM-INVIGOR} and that faster pre-diagnosis growth rate predicted shorter survival (2, 5, 29, 30, 38, 39). We also found that the status of lymphovascular invasion (LVI) correlated with growth rate; with highly proliferative and fast-growing tumors more likely to develop when there is increased provision of nutrients to the tumor cells from the leaky invaded blood vessels. Our results indicated that increasing \textit{SM-INVIGOR} increases the risk of mortality of the disease. However, \textit{SM-INVIGOR} cannot be included as a prognostic variable in routine clinical practice because of difficulty in evaluating it in the short interval between diagnosis and treatment.

Therefore, we developed \textit{Surr-INVIGOR} to predict the pre-diagnosis in-vivo BC growth rate after testing multiple clinicopathological and molecular variables (individually and in combination) using diverse machine learning algorithms. The optimal algorithm, a KNN which used Ki67, MI, and size, stratified both the study and validation cohorts into two subgroups with very distinct outcomes. \textit{Surr-INVIGOR} further allowed routine clinical parameters to be used in patients with slow-growing tumors to determine tumor size at various time-points before the diagnosis of the tumor. For fast-growing tumors, immediate surgery is often recommended, as delays may result in upgrading of clinical T stage. \textit{Surr-INVIGOR} may thus have a potential use in medicolegal cases, and may be used to guide screening and perhaps even follow-up intervals in selected groups of BC patients.

Consistent with previous studies (40, 41), results from our validation cohort showed a significant correlation between BC molecular subtypes and pre-diagnosis tumor growth rate wherein a higher
growth rate was observed in triple negative/basal-like BC patients. Previous studies have indicated that faster growing tumors lead to poorer survival (42-45). Our results compellingly demonstrated that high pre-diagnosis in-vivo BC growth rate increases the risk of mortality from the disease regardless of potential clinicopathological cofounders. Some previous studies did not find such statistically significant associations (3, 4), which might be because in those studies, the tumor volume was calculated using only one dimension—a method that can introduce considerable inaccuracy into growth rate calculations. In the current study, we utilized a combination of power law growth rate and spherical volume—both of which were significant in a previous study using 2-dimensional breast mammogram data (19), and showed the most significant prognostic relevance in our data.

Review of previous mammography is carried out as a routine practice at Nottingham Hospital, and cases that show an abnormality at the same site as the diagnosed tumors are considered as cancers potentially missed in the prior screening. Some of these tumors are only detectable in retrospect with knowledge of the diagnostic mammograms, and if all such subtle areas were recalled for further assessment, this would likely increase the false positive rate beyond what is regarded as acceptable in the NHS breast screening program. The impact of such delay in the diagnosis on the presentation and outcome of these tumors compared to matched population of women who presented for the first time as symptomatic or with screen-detected BC remains to be defined. Most tumors included in our study (similar to other studies looking at screen-detected tumors) by their very nature, were small, slow-growing luminal tumors, and infrequently expressed basal markers or HER2 with similar nodal status (30). This can be explained by the unique nature of these slow growing early-stage tumors in this study. By contrast, aggressive tumors are likely to present without prior mammographic abnormality (46). In line with these results, Kalager et al.
(47) have reported that BCs presenting as interval cancers were slightly larger than symptomatic BC but there was no difference between the two groups regarding lymph node status or patient outcome. Moreover, our results indicated that the impact of SM-INVIGOR on disease stage and development of LVI is limited. However, the present study holds a few limitations: due to the unique nature of the study cohort and the lack of similar missed cancer cohorts, the SM-INVIGOR growth index could not be readily validated. Additionally, this is a retrospective, single center study and adjuvant treatment regimens were not factored in our analyses. Validation of the model in diverse cohorts is necessary before it can be applied for the prediction of in-vivo growth rate and determination of the likely tumor initiation date and previous tumor size in clinico-legal cases. If validated in further studies, the model developed herein could potentially guide treatment selection as it prognostically distinguishes fast-growing tumors from slow-growing ones. For example, for fast growing tumors, immediate treatment in the form of primary systemic therapy (rather than surgery) may be required. Moreover, HER2 is known to be related to rapid growth of tumors and might be a good marker to add to the Surr-INVIGOR, however our study cohort was overwhelmingly HER2 negative and thus it’s impact within a prognostic model could not be properly measured. Further analysis may be required in a diverse cohort.

2.6 References


Parts of this chapter have been published verbatim in British Journal of cancer 2019 Sep; 121(6):497-504. Doi:10.1038/s41416-019-0539-x as “Machine learning-based prediction of breast cancer growth rate in vivo.”

Parts of this chapter have been published verbatim in British Journal of cancer 2019 Sep; 121(6):497-504. Doi:10.1038/s41416-019-0539-x as “Machine learning-based prediction of breast cancer growth rate in vivo.”
Figure 2.1 Schematic depicting sequences of steps in our study leading to the calculation of SM-INVIGOR and the development of Surr-INVIGOR that predicts in-vivo tumor growth rate in BC.

Briefly, tumor volumes from two serial mammograms and the time interval between measurements in a unique dataset of 92 patients (A), were used to develop a growth rate index SM-INVIGOR (B). The growth index significantly predicts BCSS and classifies tumors as slow- or fast-growing (C). When the tumors were resected after final diagnosis (D), tumor sections were immunohistochemically stained for a panel of BC biomarkers (E). A machine learning algorithm was used to develop a surrogate model (termed Surr-INVIGOR) for SM-INVIGOR that uses routinely assessed BC clinical biomarkers like Ki67, Mitotic Index and Histological size. The multivariable model non-linearly combines multiple clinicopathological variables and immunohistochemical biomarkers to predict the tumor’s in-vivo growth rate prior to diagnosis (F, G). Using the same growth rate threshold as SM-INVIGOR, the Surr-INVIGOR model was able to prognostically stratify patients in study cohort (H). Finally, Surr-INVIGOR was validated using an independent BC validation cohort of 1241 patients and was found to be strongly prognostic in the validation cohort (I, J).
Figure 2.2 Prognostic significance of SM-INVIGOR.

(A) Univariate associations between clinicopathological parameters and SM-INVIGOR. (B) Kaplan-Meier survival curve for study cohort patients stratified into high and low growth rate groups by SM-INVIGOR. (C) Multivariable analysis of the association between clinicopathological variables and patient’s outcome in the study cohort. (D) Univariate association between clinicopathological parameters and Surr-INVIGOR in validation cohort. (E) Kaplan-Meier survival curve for patients stratified into high and low growth rate subgroups by Surr-INVIGOR in validation cohort. (F) Multivariable analysis of the association between clinicopathological variables and BCSS in validation cohort.
Figure 2.3 Tumor Growth Rate group (fast versus slow) classification accuracies obtained using different machine learning algorithms and features (included sequentially as indicated on the x-axis).

The y-axis represents the cross-validated model performance (accuracy) obtained with inclusion of the indicated discriminatory biomarker and all the ones on its left. Optimal accuracy (~70%) is found using three markers (Ki67, MI, and size) within the KNN algorithm or all 7 variables with the Ensemble algorithm. The KNN was chosen as it minimized the features needed.
Figure 2.4 Scatterplot representing the optimal predicted in-vivo growth rate (through regression using surrogate markers) versus the SM-INVIGOR from the study cohort. Each dot represents a patient and the calculated $R^2 = 0.22$. 

Parts of this chapter have been published verbatim in British Journal of Cancer 2019 Sep; 121(6):497-504. Doi:10.1038/s41416-019-0539-x as “Machine learning-based prediction of breast cancer growth rate in vivo.”
Figure 2.5 Plot representing the duration before diagnosis that patients in the slow growth subgroup have had a tumor volume >0 mm3 (black lines). The end of black line represents time of tumor diagnosis. The red points indicate when the original screening was performed. As the growth rate used was the highest found in the slow-growing subgroup, it represented the time that a tumor was conservatively present.
### Table 2.6.1 Clinicopathological characteristics

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Study Cohort</th>
<th>Validation Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of cases (N; %)</td>
<td>Number of cases (N; %)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤65</td>
<td>75 (81.5)</td>
<td>1057 (85.2)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>17 (18.5)</td>
<td>184 (14.8)</td>
</tr>
<tr>
<td><strong>Tumor Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 (17.4)</td>
<td>325 (26.2)</td>
</tr>
<tr>
<td>2</td>
<td>42 (45.7)</td>
<td>501 (40.4)</td>
</tr>
<tr>
<td>3</td>
<td>34 (36.9)</td>
<td>415 (33.4)</td>
</tr>
<tr>
<td><strong>Tumor Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤15</td>
<td>32 (35.0)</td>
<td>969 (72.31)</td>
</tr>
<tr>
<td>&gt;15</td>
<td>60 (65.0)</td>
<td>371 (27.69)</td>
</tr>
<tr>
<td><strong>Lymph Node</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>60 (65.2)</td>
<td>763 (61.5)</td>
</tr>
<tr>
<td>2</td>
<td>24 (26.1)</td>
<td>382 (30.8)</td>
</tr>
<tr>
<td>3</td>
<td>8 (8.7)</td>
<td>96 (7.7)</td>
</tr>
<tr>
<td><strong>Hormone Receptor Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER Positive</td>
<td>78 (84.8)</td>
<td>915 (73.7)</td>
</tr>
<tr>
<td>ER Negative</td>
<td>14 (15.2)</td>
<td>326 (26.3)</td>
</tr>
<tr>
<td>PR Positive</td>
<td>59 (64.1)</td>
<td>675 (54.4)</td>
</tr>
<tr>
<td>PR Negative</td>
<td>33 (35.9)</td>
<td>566 (45.6)</td>
</tr>
<tr>
<td><strong>HER2 Expression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>5 (5.4)</td>
<td>151 (12.2)</td>
</tr>
<tr>
<td>Negative</td>
<td>81 (88.0)</td>
<td>1058 (85.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (6.5)</td>
<td>32 (2.6)</td>
</tr>
<tr>
<td><strong>Intrinsic Molecular Subtypes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>38 (41.3)</td>
<td>408 (32.9)</td>
</tr>
<tr>
<td>Luminal B</td>
<td>28 (30.4)</td>
<td>429 (34.6)</td>
</tr>
<tr>
<td>HER2</td>
<td>5 (5.4)</td>
<td>151 (12.2)</td>
</tr>
<tr>
<td>BLBC</td>
<td>4 (4.3)</td>
<td>138 (11.1)</td>
</tr>
<tr>
<td>Triple Negative</td>
<td>11 (12.0)</td>
<td>68 (5.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (6.5)</td>
<td>47 (3.8)</td>
</tr>
<tr>
<td><strong>Ki67</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>44 (47.8)</td>
<td>667 (53.7)</td>
</tr>
<tr>
<td>Low</td>
<td>48 (52.2)</td>
<td>574 (46.3)</td>
</tr>
<tr>
<td><strong>Tumor Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invasive No Special Type</td>
<td>50 (54.3)</td>
<td>761 (61.3)</td>
</tr>
<tr>
<td>Invasive lobular</td>
<td>17 (18.5)</td>
<td>93 (7.5)</td>
</tr>
<tr>
<td>Tubular</td>
<td>11 (12.0)</td>
<td>299 (24.1)</td>
</tr>
<tr>
<td>Mucinous</td>
<td>2 (2.2)</td>
<td>11 (0.8)</td>
</tr>
<tr>
<td>Mixed type</td>
<td>12 (13.0)</td>
<td>77 (6.2)</td>
</tr>
<tr>
<td><strong>Coexisting DCIS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>21 (23.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Low grade</td>
<td>20 (22.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Intermediate grade</td>
<td>22 (24.0)</td>
<td>NA</td>
</tr>
<tr>
<td>High grade</td>
<td>29 (31.0)</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Lympho-vascular Invasion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>60 (66.2)</td>
<td>686 (55.3)</td>
</tr>
<tr>
<td>Definite</td>
<td>21 (22.8)</td>
<td>397 (32.0)</td>
</tr>
<tr>
<td>Probable</td>
<td>11 (11)</td>
<td>158 (12.7)</td>
</tr>
<tr>
<td><strong>Outcome Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>62 (67.4)</td>
<td>650 (52.3)</td>
</tr>
<tr>
<td>Dead</td>
<td>30 (32.6)</td>
<td>591 (47.6)</td>
</tr>
</tbody>
</table>
Table 2.6.2: Growth rate functions evaluated for development of SM-INVIGOR

<table>
<thead>
<tr>
<th>Growth Rate</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Rate = ( \frac{\log (Volume_{Diagnosis}) - \log (Volume_{Screening})}{Time \text{ Between Diagnosis and Screening}} )</td>
</tr>
<tr>
<td>Power Law (( \alpha = 2/3 ))</td>
<td>Rate = ( \frac{Volume_{Diagnosis}^{(1-2/3)} - Volume_{Screening}^{(1-2/3)}}{(1 - 2/3) \times Time \text{ Between Diagnosis and Screening}} )</td>
</tr>
<tr>
<td>Gompertz</td>
<td>Rate = ( \frac{\log (\log(10^6mm^3/Volume_{Diagnosis})) - \log((\log(10^6mm^3/Volume_{Screening}))}{Time \text{ Between Diagnosis and Screening}} )</td>
</tr>
<tr>
<td>Power Law (( \alpha = 1/2 ))</td>
<td>Rate = ( \frac{Volume_{Diagnosis}^{(1-1/2)} - Volume_{Screening}^{(1-1/2)}}{(1 - 1/2) \times Time \text{ Between Diagnosis and Screening}} )</td>
</tr>
</tbody>
</table>
Table 2.6.3 Details of antibodies used and dilutions.

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Clone</th>
<th>Raised in</th>
<th>Monoclonal or Polyclonal</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER α</td>
<td>1D5</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>PR</td>
<td>636</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>HER-2 (cerb-2)</td>
<td>A0485</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>Ki67</td>
<td>MIB-1</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>CK 5/6</td>
<td>D5/16 B4</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>MCM2</td>
<td>HPA031496</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>BCl2</td>
<td>124</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:100</td>
</tr>
<tr>
<td>EGFR</td>
<td>C31G7</td>
<td>Mouse</td>
<td>Monoclonal</td>
<td>1:50</td>
</tr>
</tbody>
</table>
Table 2.6.4 The list of surrogate markers to develop Surr-INVIGOR.
Ki67, Mitosis and Histological tumor size served as the significant surrogate variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ki-67</td>
<td>0.000265</td>
</tr>
<tr>
<td>Mitosis</td>
<td>0.002479</td>
</tr>
<tr>
<td>Tumor size</td>
<td>0.003619</td>
</tr>
<tr>
<td>Nottingham prognostic Index</td>
<td>0.004163</td>
</tr>
<tr>
<td>Histologic Grade</td>
<td>0.021126</td>
</tr>
<tr>
<td>Tumor Stage</td>
<td>0.143553</td>
</tr>
<tr>
<td>Tubule formation</td>
<td>0.195234</td>
</tr>
<tr>
<td>Cleaved-caspase3</td>
<td>0.205719</td>
</tr>
<tr>
<td>BCL2</td>
<td>0.221934</td>
</tr>
<tr>
<td>Patient Age</td>
<td>0.226801</td>
</tr>
<tr>
<td>Estrogen Receptor</td>
<td>0.270581</td>
</tr>
<tr>
<td>MCM2</td>
<td>0.380132</td>
</tr>
<tr>
<td>Progesterone Receptor</td>
<td>0.573077</td>
</tr>
<tr>
<td>EGFR</td>
<td>0.657271</td>
</tr>
</tbody>
</table>

Table 2.6.5 Table with the chosen hyperparameters for the optimal KNN model.

Table 2.6.5: K Nearest Neighbor (KNN) hyperparameters optimized for Surr-INVIGOR

<table>
<thead>
<tr>
<th>Number of Neighbors</th>
<th>Distance</th>
<th>Distance Weight</th>
<th>Exponent</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Standardized Euclidean distance</td>
<td>Inverse</td>
<td>None</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Table 2.6.6 Table showing model fit statistics (Akaike information criterion) for various combinations of growth rate functions and tumor volumes.
The growth rate was treated as a categorical variable. The model with the best fit was the one with the lowest.

<table>
<thead>
<tr>
<th>Growth Rate</th>
<th>Volumes</th>
<th>AIC cat. BCSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Sphere</td>
<td>153.275</td>
</tr>
<tr>
<td></td>
<td>Cylinder</td>
<td>155.37</td>
</tr>
<tr>
<td></td>
<td>Spheroid</td>
<td>155.592</td>
</tr>
<tr>
<td>Power Law (α =2/3)</td>
<td>Sphere</td>
<td>155.598</td>
</tr>
<tr>
<td></td>
<td>Cylinder</td>
<td>155.057</td>
</tr>
<tr>
<td></td>
<td>Spheroid</td>
<td>155.243</td>
</tr>
<tr>
<td>Gompertz</td>
<td>Sphere</td>
<td>155.583</td>
</tr>
<tr>
<td></td>
<td>Cylinder</td>
<td>155.263</td>
</tr>
<tr>
<td></td>
<td>Spheroid</td>
<td>155.627</td>
</tr>
<tr>
<td>Power Law (α =1/2)</td>
<td>Sphere</td>
<td>152.621</td>
</tr>
<tr>
<td></td>
<td>Cylinder</td>
<td>155.179</td>
</tr>
<tr>
<td></td>
<td>Spheroid</td>
<td>155.306</td>
</tr>
</tbody>
</table>
3 BIOMARKERS PREDICTING PATHOLOGIC COMPLETE RESPONSE TO NEOADJUVANT CHEMOTHERAPY IN BREAST CANCER

3.1 Abstract

Recent studies have shown strong correlation of pathologic complete response (pCR) to neoadjuvant chemotherapy with survival and prognosis in breast cancers. Clinical data from 237 breast cancer patients who received neoadjuvant chemotherapy between 2012 and 2014 were reviewed. Correlations were sought between pCR and estrogen receptor (ER), progesterone receptor (PR), and HER2 status; Nottingham and nuclear grades; tumor tubule formation; mitotic score; Ki67 index; and tumoral and stromal lymphocytic infiltration (TLI and SLI, respectively). Of the 237 cases, 104 (43.9%) achieved pCR. The HER2+ and triple negative breast cancer (TNBC) subtypes had higher pCR rates compared with the luminal subtype (ER+ or PR+ and HER2-). ER and PR negativity, HER2 positivity, Nottingham grade 3, increased TLI and SLI, high mitotic count and Ki67 score correlated significantly with pCR in the overall cohort. TLI and SLI correlated significantly with pCR in the HER2+ and TNBC subtypes in multivariate analysis, whereas no biomarkers correlated with pCR in the luminal subtype. In addition to the pathologic parameters and biomarkers already routinely assessed, evaluation of TLI and SLI may help to better select patients with HER2+ and TNBC for neoadjuvant chemotherapy.

3.2 Introduction

Breast cancer is the most prevalent malignancy and second deadliest cancer in women. Currently, the main treatments include surgery, chemotherapy, radiation and targeted therapies. Chemotherapy as a systemic control regimen has dramatically increased the disease-free survival.
and overall survival rate (1,2). Chemotherapy can be administrated before or after surgery. When chemotherapy is given before surgery, it is referred as neoadjuvant therapy. For decades, neoadjuvant therapy has been used to treat locally advanced tumors to be operable. Recent research and clinical trials have shown strong correlation of breast cancer responses to neoadjuvant therapies with survival and prognosis (3-5). Patients who achieve pathologic complete response (pCR) to neoadjuvant therapy tend to have improved disease-free and overall survival compared with patients with residual invasive disease (6,7). pCR is defined as no invasive carcinoma in the breast and lymph nodes at the time of surgery. Because of the strong correlation between pCR and survival, the US Food and Drug Administration (FDA) now considers pCR to neoadjuvant chemotherapy a surrogate endpoint for clinical trials and drug approval (8). The strongest correlation between pCR and outcomes is found within the triple negative breast cancer (TNBC) and HER2þ breast cancers (9). Predicting which patients will achieve pCR to neoadjuvant chemotherapy is important because neoadjuvant chemotherapy is not without risk. For instance, although neoadjuvant chemotherapy can prolong disease-free and overall survival, it may increase the rate of ipsilateral tumor recurrence compared with adjuvant therapy (2). In addition, delaying surgery may decrease survival (10). A variety of methodologies have been investigated, such as magnetic resonance imaging (11), positron emission tomography (12), and gene expression profiling (13). However, none has been universally accepted, and these modalities are expensive and not routinely performed. We conducted a comprehensive evaluation of tumor morphology and biomarker status, and correlated these parameters with pCR rate in a neoadjuvant setting.
3.3 Methods

3.3.1 Patient Selection and Clinicopathologic Characteristics

A total of 2,691 consecutive excisional specimen cases were retrieved from the archives of the Department of Pathology and Laboratory Medicine at Emory University from 2012 to 2014 after protocol approval from the Emory Institutional Review Board. All surgical procedures were performed at two major teaching hospitals of Emory University. Among the 2,691 cases, 237 patients had neoadjuvant therapies. Of the 237 cases, 229 cases had the status of estrogen receptor (ER), progesterone receptor (PR) and HER2 expression available, and 195 cases had information of biopsy diagnoses. Among the 195 cases, 129 had core biopsy slides available, and all of these 129 cases were reviewed by a pathologist (XL). For the cases without biopsy slides, information was retrieved from the patient’s medical record and pathology report. The majority of the patients received four cycles of neoadjuvant chemotherapies. HER2+ cancers received HER2 targeted therapies in addition to neoadjuvant chemotherapies. Clinicopathologic characteristics of all patients were summarized in Table 3.6.1.

3.3.2 Pathologic Evaluation

The following morphological features and biomarkers were evaluated in the biopsy specimen: tubule formation (score 1-3), nuclear grade (score 1-3), mitotic count (score 1-3) per the College of American Pathologists (CAP) recommendation, Nottingham histologic grade (score 1-3), stromal and tumoral lymphocytic infiltration (SLI and TLI, respectively), fibrosis (scored 1-3 as follows: 1, mild; 2, moderate; 3, severe), ER and PR expression, and HER2 amplification (positive or negative per American Society of Clinical Oncology recommendation) and Ki67 score (high: ≥15%; low: <15). TLI was evaluated as percentage of tumor cells infiltrated with lymphocytes. SLI was evaluated as percentage of stromal area covered by lymphocytes.
stroma included both intratumoral stroma as well as the stroma adjacent to the periphery of the tumor.

### 3.3.3 Breast Cancer Classification and Definition of pCR

Tumors were classified as luminal, HER2+, or TNBC. The luminal subtype was defined as ER+ and/or PR+ and HER2−. HER2+ cancer was defined by either a score of 3+ from immunohistochemical (IHC) study or positive HER2 amplification by fluorescence in situ hybridization (FISH) regardless of ER or PR status. TNBC was defined by the absence of ER and PR expression and HER2 overexpression by IHC or FISH. Standard 1% expression rate was used as the cutoff to define positivity in ER and PR expression. pCR was defined as no invasive carcinoma in both breast and lymph nodes at the time of surgery. In situ carcinoma was allowed in the pCR cases.

### 3.3.4 Statistical Analysis

Logistic regression was performed on the total patient cohort as well as the subtypes with a pCR case indicating an event. Odds ratios (ORs), which indicate a multiplicative effect in the odds of achieving a pCR, were compared between categorical groups (using the lowest risk group as the reference). Multivariate models were fit via a backwards selection method, in which the full model was reduced and refit stepwise by removing the least significant variable based on Wald test for the individual parameters, until either all variables had been removed or all had a p value <0.10. Due to the variability in multivariate data completeness, we also used the Firth penalized likelihood approach. Correlations were analyzed using the Pearson product-moment correlation coefficient. Statistical analysis software was used for all statistical analysis.
3.4 Results

3.4.1 Clinicopathologic Features of the Cases

Of the total 237 patients, 91 (38.4%) were younger than 50 years of age, 78 (32.9%) had advanced-stage disease (pT3 or pT4), and 104 (43.9%) had pCR. Among the 229 cases with available biomarker information, 72 were luminal, 79 were HER2+ and 78 were TNBC (Table 3.6.1). In the non-pCR patients, the size of residual invasive carcinoma ranged from 0.5 to 9 cm.

3.4.2 HER2+ and TNBCs have higher pCR Rate than Luminal Breast Cancer

Subtype analysis revealed that 20 of 72 (27.8%) luminal type, 46 of 79 (58.2%) HER2+ and 37 of 78 (47.4%) TNBC exhibited pCR. On univariate analysis, the HER2+ and TNBC had 3.6 and 2.4 times the odds of achieving pCR compared with the luminal subtype (95% confidence interval [CI]= 1.83-7.17, p<0.001 and 1.19- 4.64, p=0.014, respectively) Table 3.6.2. Univariate analysis failed to identify any parameter or biomarker as significantly correlated with pCR rate in the luminal subtype.

3.4.3 High TLI and SLI are significantly associated with pCR rate in HER2+ breast cancers and TNBCs

TLI (both with 3% as threshold and as a continuous increasing value) as well as SLI (both with 5% as threshold and as a continuous increasing value) were significantly associated with pCR in the overall cohort (Table 3.6.2). The categorical cutpoints (3% in TLI and 5% in SLI) were chosen by using the threshold that was at the intersection of specificity and sensitivity when identifying pCR patients vs non pCR patients. On univariate analysis, high TLI and SLI were both found to be significantly associated with a greater probability of pCR in HER2+ and TNBC cancers (Table 3.6.2). This significant correlation persisted in HER2+ and TNBCs in multivariate analysis.

Table 3.6.3. No correlation between TLI or SLI and pCR was seen in the luminal type.
3.4.4  **ER and PR expression are associated with decreased pCR in the Overall Cohort and HER2+ Breast Cancers**

Positivity of ER and PR expression was found to be associated with decreased pCR in the overall cohort and within the HER2+ subtype (Table 3.6.2). With increased ER or PR expression, there was a small decrease in the odds of a pCR.

3.4.5  **High mitotic count and Ki67 score and Nottingham histologic Grade 3 are significantly associated with pCR rate in the overall cohort**

Nottingham histologic grade 3 and high mitotic count had a positive correlation with the probability of pCR in the overall cohort (OR 3.47; p=0.022 and OR 3.44; P< 0.001, respectively; Table 3.6.2). Along with the positive correlation of mitotic count and pCR, Ki67 score (both 15% as threshold and as increasing absolute value) was significantly associated with pCR (Table 3.6.2). The mitotic count remained a significant predictor of pCR in the overall cohort in multivariate analysis (Table 3.6.3).

3.4.6  **High correlation between stromal and intratumoral lymphocytic infiltration**

TLI and SLI are very highly correlated in the overall cohort (r =0.68, p<0.001) and within each subtype, HER2+ (r =0.52, p<0.001), luminal (r =0.85, p<0.001), triple negative (r =0.69, p<0.001). Such strong correlation was seen in multivariate analysis model. When we swapped SLI with TLI in the final multivariate model, very similar ORs (0.96 vs 0.95) and OR p values (0.013 vs 0.017) with correlation of pCR were obtained.
3.4.7  **Luminal type breast cancers have low pCR rate and no parameter is strongly associated with pCR**

Among all 3 subtypes, the luminal type had the lowest pCR rate of 27.8%, compared with 58.2% in HER2+ and 47.4% in TNBC. No pathologic parameter or biomarker was identified to have significant association with pCR in the luminal type.

3.5  **Discussion**

In our study, we have performed a comprehensive evaluation of tumor morphology and biomarker statuses, and correlated them with pCR rate in the neoadjuvant setting. We found that pCR was significantly higher in the HER2+ and triple-negative subtypes (58.2% and 47.4%, respectively) compared with the luminal subtype (27.8%). The odds of achieving a pCR in HER2+ cancers were 3.6 times higher than in luminal cancers. This result is similar to that found in the metaanalysis by Houssami *et al* of 30 studies encompassing 11,695 patients, which estimated that pCR occurred in 8.3% of hormonal receptor positive (HR+)/HER2–, 18.7% of HR+/HER2+, 38.9% of HR–/HER2+, and 31.1% of TNBCs. The pCR rates in our study were slightly higher than those in the metaanalysis by Houssami *et al*, but high pCR rates of TNBC and HER2+ breast cancers have been reported (14). The different pCR rates in different studies might be due to variation of subtypes of breast cancers. For example, TNBC has been classified into six subtypes, and these subtypes might have different clinicopathologic characteristics and respond differently to chemotherapies (15-19). Furthermore, Houssami *et al* found that the odds of achieving a pCR was 7 times higher for patients with HR–/HER2+ breast cancer and 5 times higher for patients with TNBC in comparison with patients with the HR+ subtype (9). Several other studies that investigated the use of anti-HER2 therapies in the neoadjuvant setting have found that the pCR...
rate varies from 20% to 65% in HER2+ cancers (1,20-24). Other studies have also reported an increased pCR rate in TNBCs as compared with non-TNBCs (22% vs 11%) (1,25-27). While Tan et al reported that ER and PR negativity significantly correlated with pCR in multivariate analysis, we found that ER and PR negativity were only associated with pCR in univariate analysis, indicating that ER and PR negativity is associated with HER2+ or triple-negative status (28). In our study, HER2+ cases included HR+ as well as HR− cases. Within the HER2+ group, we found that pCR positively correlated with ER and PR negativity. Von Minckwitz et al suggest that pCR may not be a suitable endpoint for the luminal subtypes. Specifically, they found that in low-proliferative subgroups (which included lobular, grade 1 and HR+ tumors) pCR conferred no predictive power in disease-free or overall survival, in contrast with the high-proliferative subgroup (which included ductal, grade 2/3, and HR− tumors) in which pCR was associated with improved disease-free and overall survival (7). We too determined that pCR is significantly associated with high histologic grade and mitotic activity in both univariate and multivariate analyses across the entire cohort. We report a strong correlation between pCR and Ki67 score both as a categorical variable (specifically, when 15% is set as the threshold for defining high proliferation index) and as a continuously increasing variable, similar to the findings of other studies (1,29-32). Brown et al, who measured Ki67 expression using quantitative automated quantitative analysis by immunofluorescence (AQUA), found that both average and maximum AQUA scores were significant predictors of pCR to neoadjuvant therapy in multivariate analysis (32) Kim et al found Ki67 expression to be the only independent predictor of pCR and also discovered that a Ki67 value of greater than 25% was a significant predictive factor for pCR (29). Yoshioka et al found that high Ki67 was a predictive marker for pCR and that all patients achieving pCR were disease-free by the study’s end (31). Furthermore, high Ki67 expression in tumors of
posttreatment was strongly correlated with poor disease-free and overall survival regardless of subtype in their study. We found high mitotic score positively correlated with pCR. A mitotic count of >9/mm² was reported to be significantly correlated with pCR by Balmativola et al (33). Based on these compelling data, mitotic count and Ki67 should be considered for inclusion in routine clinical evaluation of patients with HER2+ or TNBC to help determine whether neoadjuvant chemotherapy is indicated. Herein, we additionally found TLI and SLI to be significantly correlated with pCR in univariate analysis in HER2+ breast cancers (OR =0.94, p =0.028). Our results are similar to those reported by Mao et al, who carried out a systematic review and metaanalyses to evaluate the predictive roles of tumor infiltrating lymphocytes (TILs) in response to neoadjuvant chemotherapy in breast cancer (34). They evaluated a total of 13 studies that included 3,251 patients. In pooled analysis, they found that the detection of higher TIL numbers in the pretreatment biopsy was correlated with better pCR to neoadjuvant chemotherapy (OR 3.93; 95% CI =3.26-4.73). Moreover, TILs predicted higher pCR rates in HER2+ and TNBCs (OR =2.49 [95% CI =1.61-3.83] and OR =5.05 [95% CI =2.86-8.92], respectively), but not in ER+ breast cancer (OR =6.21 [95%CI: 0.86-45.15]). In multivariate analysis, they found that TILs were still an independent predictor for high pCR rate (OR =1.41 [95% CI =1.19-1.66]) (34). Furthermore, they found that, in three studies that examined lymphocyte-predominant breast cancer (LPBC; defined as having >50% or 60% lymphocytic infiltration of the tumor bed or stroma), LPBC patients had higher pCR rates compared with nonLPBC patients (OR =3.64 [95% CI =2.70-4.90]) (35,36). Other studies have evaluated the density of TILs as per 10% increase in the number of lymphocytes infiltrating either intratumoral or stromal compartments, both of which predicted better pCR (OR =1.35 [95% CI =1.27-1.44] and OR =1.26 [95% CI, 1.20-1.32], respectively). Altogether, their analysis indicated that TIL infiltration was an independent

Parts of this chapter have been published verbatim in Am J Clin Pathol 2016;145:871-878; DOI: 10.1093/ajcp/aqw045 as “Biomarkers predicting pathologic complete response to neoadjuvant chemotherapy to breast cancer.”
predictive marker for higher pCR rate (OR =1.41 [95% CI =1.19-1.66]), whether TILs were detected in intratumoral (OR =1.23 [95% CI =1.12-1.34]) or stromal (OR =51.22 [95% CI =1.09-1.36]) compartments (34). While other studies have used higher thresholds in evaluating TLI and SLI, we report significantly higher pCR rates even with very low thresholds (3% and 5%, respectively). Our study reveals that TLI and SLI are highly correlated with each other, and both are significantly associated with pCR in univariate analysis. Furthermore, TLI and SLI are each individually correlated with pCR in multivariate analysis, and assessment of both variables together does not give any additive value in predicting pCR. During our evaluation of the H&E slides, we found that TLI was challenging to assess, usually requiring tedious evaluation at high power. Therefore, evaluating SLI alone might be sufficient to assist in predicting pCR.

3.6 References


Parts of this chapter have been published verbatim in Am J Clin Pathol 2016;145:871-878; DOI: 10.1093/ajcp/aqw045 as “Biomarkers predicting pathologic complete response to neoadjuvant chemotherapy to breast cancer.”


human epidermal growth factor receptor 2-positive operable breast cancer: an update of the initial randomized study population and data of additional patients treated with the same regimen. Clin Cancer Res. 2007;13:228-233.


Parts of this chapter have been published verbatim in Am J Clin Pathol 2016;145:871-878; DOI: 10.1093/ajcp/awq045 as “Biomarkers predicting pathologic complete response to neoadjuvant chemotherapy to breast cancer.”

Table 3.6.1 Clinicopathological characteristics of study cohort.
ER, estrogen receptor; pCR, pathologic complete response; PR, progesterone receptor; SLI, stromal lymphocytic infiltration; TLI, tumoral lymphocytic infiltration; a. Some information was missing in a few cases.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No. of Cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, y</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>2</td>
<td>0.84</td>
</tr>
<tr>
<td>30-39</td>
<td>27</td>
<td>11.4</td>
</tr>
<tr>
<td>40-49</td>
<td>62</td>
<td>26.16</td>
</tr>
<tr>
<td>50-59</td>
<td>66</td>
<td>27.85</td>
</tr>
<tr>
<td>60-69</td>
<td>58</td>
<td>24.47</td>
</tr>
<tr>
<td>70+</td>
<td>22</td>
<td>9.28</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>121</td>
<td>51.1</td>
</tr>
<tr>
<td>White</td>
<td>105</td>
<td>44.3</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>4.64</td>
</tr>
<tr>
<td><strong>Tumor grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>9.01</td>
</tr>
<tr>
<td>2</td>
<td>89</td>
<td>38.2</td>
</tr>
<tr>
<td>3</td>
<td>123</td>
<td>52.79</td>
</tr>
<tr>
<td><strong>Clinical stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/II</td>
<td>158</td>
<td>66.95</td>
</tr>
<tr>
<td>III/IV</td>
<td>78</td>
<td>33.05</td>
</tr>
<tr>
<td><strong>Nodal status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>130</td>
<td>54.85</td>
</tr>
<tr>
<td>Negative</td>
<td>107</td>
<td>45.15</td>
</tr>
<tr>
<td><strong>Lymphocytic infiltration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLI</td>
<td>88</td>
<td>68.22</td>
</tr>
<tr>
<td>TLI</td>
<td>72</td>
<td>55.81</td>
</tr>
<tr>
<td><strong>Hormone receptor</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER+</td>
<td>108</td>
<td>47.16</td>
</tr>
<tr>
<td>PR+</td>
<td>74</td>
<td>32.31</td>
</tr>
<tr>
<td><strong>HER2 expression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>79</td>
<td>33.33</td>
</tr>
<tr>
<td>Negative</td>
<td>158</td>
<td>66.67</td>
</tr>
<tr>
<td><strong>Subtypes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HER2+</td>
<td>79</td>
<td>34.5</td>
</tr>
<tr>
<td>Triple negative</td>
<td>78</td>
<td>34.06</td>
</tr>
<tr>
<td>Luminal</td>
<td>72</td>
<td>31.44</td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCR</td>
<td>104</td>
<td>43.88</td>
</tr>
<tr>
<td>Non-pCR</td>
<td>133</td>
<td>56.12</td>
</tr>
</tbody>
</table>
### Table 3.6.2 Univariate logistic analysis of pCR in the total cohort and within subtypes.

CI, confidence interval; ER, estrogen receptor; OR, odds ratio; PR, progesterone receptor; SLI, stromal lymphocytic infiltration; TLI, tumoral lymphocytic infiltration. A dash (-) indicates a missing value while a (*) indicates nonconversion for that variable. Bold values indicate significance.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Overall cohort</th>
<th>Hormone receptor status</th>
<th>PIK3CA (exon 20)</th>
<th>Triple negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
</tr>
<tr>
<td><strong>HER2</strong></td>
<td>n</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td><strong>ER</strong></td>
<td>n</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td><strong>KI67</strong></td>
<td>n</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td><strong>Triple negative</strong></td>
<td>n</td>
<td>Odds Ratio (95% CI)</td>
<td>OR P Value</td>
<td>Odds Ratio (95% CI)</td>
</tr>
</tbody>
</table>

Parts of this chapter have been published verbatim in *Am J Clin Pathol* 2016;145:871-878; DOI: 10.1093/ajcp/aqw045 as “Biomarkers predicting pathologic complete response to neoadjuvant chemotherapy to breast cancer.”
Table 3.6.3 Multivariate Logistic Regression Analysis for pCR in the Overall Cohort and HER2+ and TNBC Subtypes.

pCR, pathologic complete response; PR, progesterone receptor; SLI, stromal lymphocytic infiltration; TLI, tumoral lymphocytic infiltration; TNBC, triple negative breast cancer. Backward elimination (a=0.10) was used for covariate selection. Bold values indicate significance.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Covariate</th>
<th>Level</th>
<th>Odds Ratio (95% CI)</th>
<th>Odds Ratio p Value</th>
<th>Type 3 p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall cohort</td>
<td>Stage</td>
<td>I/II</td>
<td>reference</td>
<td>–</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III/IV</td>
<td>0.20 (0.063-0.625)</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>9.40 (2.897-30.295)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>18.60 (2.521-136.510)</td>
<td>0.009</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>Mitotic score</td>
<td>3</td>
<td>94.89 (7.316-999.999)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HER2 Positive</td>
<td>1</td>
<td>0.77 (0.632-0.943)</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.10 (1.025-1.122)</td>
<td>0.017</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>9.10 (1.352-60.684)</td>
<td>0.023</td>
<td>.023</td>
</tr>
<tr>
<td>TNBC</td>
<td>PR</td>
<td>Positive</td>
<td>reference</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>HER2</td>
<td>TLI/SLI</td>
<td>1</td>
<td>0.77 (0.632-0.943)</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1.10 (1.025-1.122)</td>
<td>0.017</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>TLI/SLI</td>
<td>3</td>
<td>9.10 (1.352-60.684)</td>
<td>0.023</td>
<td>.023</td>
</tr>
</tbody>
</table>
4 PROGNOSTIC ROLE OF ANDROGEN RECEPTOR IN TRIPLE NEGATIVE BREAST CANCER: A MULTI-INSTITUTIONAL

4.1 Abstract

The androgen receptor (AR) has emerged as a potential therapeutic target for AR-positive triple-negative breast cancer (TNBC). However, conflicting reports regarding AR’s prognostic role in TNBC are putting its usefulness in question. Some studies conclude that AR positivity indicates a good prognosis in TNBC, whereas others suggest the opposite, and some show that AR status has no significant bearing on the patients’ prognosis. Methods: We evaluated the prognostic value of AR in resected primary tumors from TNBC patients from six international cohorts {US (n = 420), UK (n = 239), Norway (n = 104), Ireland (n = 222), Nigeria (n = 180), and India (n = 242); total n = 1407}. All TNBC samples were stained with the same anti-AR antibody using the same immunohistochemistry protocol, and samples with ≥1% of AR-positive nuclei were deemed AR-positive TNBCs. Results: AR status shows population-specific patterns of association with patients’ overall survival after controlling for age, grade, population, and chemotherapy. We found AR-positive status to be a marker of good prognosis in US and Nigerian cohorts, a marker of poor prognosis in Norway, Ireland and Indian cohorts, and neutral in UK cohort. Conclusion: AR status, on its own, is not a reliable prognostic marker. More research to investigate molecular subtype composition among the different cohorts is warranted.

4.2 Introduction

Triple-negative breast cancer (TNBC) is an aggressive subtype characterized by the lack of estrogen (ER), progesterone (PR), and Her2 receptors. This designation masks the heterogeneity
of this patient population and the challenge of stratifying them for optimal treatment selection [1,2]. Due to the paucity of treatment targets, cytotoxic chemotherapy is still the standard of care for TNBC, and there is a need to develop new and more effective targeted treatments for these patients. Among several therapeutic targets currently under study for the management of TNBC is the androgen receptor (AR) [3–5]. The AR, a nuclear steroid hormone receptor, is expressed in 10–43% of TNBCs [6,7]. In the absence of ERα, AR drives “luminal-like” gene expression patterns. One of the TNBC molecular subtypes consistently identified via gene expression profiling is the Luminal Androgen Receptor (LAR) subtype [3,4]. LAR TNBCs express full-length AR mRNA and AR target genes at high levels; however, they tend to be less proliferative and generally respond poorly to chemotherapy in both the neoadjuvant [8,9] and adjuvant settings [10]. Thus, AR expression reclassifies TNBCs into AR-positive TNBCs and AR-negative TNBCs. Because LAR TNBCs are dependent on AR signaling for their growth, AR-driven TNBC is considered an actionable subtype and targeting the AR pathway is an area of active investigation [11,12].

Studies have explored the prognostic role of AR in TNBC to better understand androgen action in TNBC, identify actionable factors that drive outcomes, and determine if testing for AR status should become part of routine clinical practice for TNBCs. However, there are conflicting reports about AR’s prognostic value in TNBC. While some studies report that AR-positivity is associated with better prognosis [13,14], others either contend that an AR-positive phenotype portends worse long-term outcomes [15,16] or that AR status has no significant impact on TNBC prognosis [17,18]. To some extent, these discrepant results can be attributed to small sample sizes, differences in the ethnic composition of cohorts, the anti-AR antibodies used for staining,

*Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study*
staining/scoring method, and use of different thresholds to define AR-positivity. Together these factors make comparisons of results across studies and cohorts challenging. In this study, we evaluated the prognostic role of AR in 1,407 TNBC tumors from seven ethnically and racially diverse populations. The tumor samples were processed identically, and the differences in treatment protocols were accounted for.

4.3 Methods

4.3.1 Study Cohorts and Samples

All aspects of this study were approved by Institutional Review Boards of the institutions involved (IRB number: H19306). Materials and data were shared in accordance with the stipulations of Material Transfer Agreements and Data User Agreements between Georgia State University (GSU) and the other participating institutions as listed in Table 4.5.1). A total of 1407 cases consecutively diagnosed with TNBC were identified from the electronic health records of multiple institutions (Table 4.5.1). Samples were collected from institutions in the US (n = 420), UK (n = 239), Norway (n = 104), Ireland (n = 222), Nigeria (n = 180), and India (n = 242). Resection samples from primary tumors were formalin-fixed and paraffin-embedded (FFPE). Samples from Nigeria, US, and Norway were processed at GSU, and the samples from the UK, India and Ireland were processed at their respective institutions.

Clinicopathological data (Table 4.5.2; reviewed and provided by a pathologist from the respective hospitals), race information (for US cohort only), overall survival (OS), and deidentified tissue blocks were available for all cases. All patients in the Nigerian cohort received adjuvant chemotherapy; by contrast, the remaining cohorts included a mix of patients who did not receive
any systemic treatment in the adjuvant setting and patients who received adjuvant chemotherapy. Patient consent was not required, because all samples were archival.

### 4.3.2 Immunohistochemistry (IHC)

All samples in the six cohorts were stained using the following protocol: Briefly, the FFPE TNBC primary tumor resection samples were deparaffinized following by rehydration in a series of ethanol baths (100%, 90%, 75%, and 50%). Heat-induced retrieval of antigen epitopes was performed in citrate buffer (pH 6.0) using pressure cooker at 15 psi for 30 min. Next, the samples were quenched using hydrogen peroxide for 20 min, followed by blocking with the Ultra-Vision protein block (Life Sciences, Fremont, CA, USA) for 10 min. Samples were then incubated for 60 min at room temperature with anti-AR primary antibody (Monoclonal Mouse Anti-Human Androgen Receptor, clone AR 441, DAKO) at 1:50 dilution. Next, the samples were incubated with MACH2 HRP-conjugated secondary antibody (BioCare Medical, Pacheco, CA, USA) for 30 min. The AR antigen was visualized using the Betazoid 3, 3-diaminobenzidine Chromogen Kit (BioCare Medical). The tissue sections were counter stained with Mayer’s hematoxylin for 1 min. Slides were then dehydrated in alcohol, cleared in xylene, and mounted with mounting medium. Appropriate negative and positive controls were used during staining.

### 4.3.3 Assessment of IHC Staining

Nuclear AR staining is indicative of active receptors, because AR can translocate to the nucleus upon ligand binding. Therefore, the stained slides were scored for the percentage of tumor-cell nuclei that showed AR-positivity. TNBC samples were considered AR-negative if they had
<1% of the tumor nuclei positive for AR; samples that had AR in \( \geq 1\% \) of the nuclei were considered AR-positive. Samples from the Nigerian, US, and Norwegian cohorts were stained and scored centrally (at GSU) and the samples in the UK, Indian, and Irish, cohorts were stained and scored at the respective institutions by two independent pathologists without prior knowledge of the patients’ pathologic or outcome data. For cross validation of the IHC scores, stained slides of grade- and stage-matched cases (representing 20–30% of each cohort) were selected from each cohort (along with the positive and negative control slides from each round of staining). The slides were reviewed and re-scored at Nottingham University Hospital by two pathologists who were blinded to the clinical annotation and previous IHC scores. The scores from the second round of scoring showed >99% concordance with the original IHC scores.

4.3.4 Statistical Analyses

Bar graphs were plotted using Excel and SPSS. Experimental groups were compared using Student’s t-test, and p-values were calculated with \( p < 0.05 \) considered statistically significant. Statistical analyses were carried out with SAS 9.4® software (Cary, NC, USA). Differences between clinicopathological proportions were determined using the \( \chi^2 \) test. Differences between baseline IHC biomarker expressions or continuous clinicopathological variables were evaluated via a 2-tailed t-test. Univariate Kaplan-Meier curves were used to investigate the effect of AR status (positive vs. negative) on OS, and the log-rank test was used to assess the statistical significance of between-group survival differences. Multivariable Cox Proportional Hazard models were used to adjust for grade, chemotherapy, population, and age with significance being
determined with the Wald chi-square test. For all clinical survival analysis, we used the OS, which was calculated as the time interval from surgery until death, and death was used as an event.

4.4 Results and Discussion

AR expression varied widely in our cohorts (from 8.3% in the Nigerian TNBCs, to 55% in the UK cohort). In the US cohort, 25% of all cases were AR-positive: Twenty percent among African Americans (AA) and thirty percent among those of European decent (EA; \( p = 0.02 \); Figure 4.1). This finding is consistent with previous reports [19]. The diversity in AR-positivity suggests our global cohorts could differ substantially in their TNBC molecular subtypes. Population-specific differences between cohorts became more apparent upon evaluating the prognostic value of AR among the cohorts in our study. We found that AR-positive TNBCs showed better OS than AR-negative TNBCs in the US (\( p = 0.03 \)) and Nigerian (\( p = 0.01 \)) cohorts among all patients (Figure 4.2A), as well as among adjuvant chemotherapy-treated patients (Figure 4.2B). By contrast, AR-positive TNBCs showed poorer prognosis (OS) than AR-negative TNBCs in the cohorts from Ireland (\( p = 0.08 \)), Norway (\( p = 0.08 \)), and India (\( p = 0.02 \)) There was no statistically significant difference in OS between AR-positive TNBCs and AR-negative TNBCs in the UK (\( p = 0.79 \)) cohort among all patients, or among adjuvant chemotherapy-treated patients (Figure 4.2A, B).

These survival trends did not change appreciably when the cut-point for AR-positivity was changed from 1% to 10%, or when an optimal cut-point was used to see if changing the cut-point affected the prognostic trend (Table 4.5.3). Multivariable Cox regression analyses that adjusted for potentially confounding variables, such as age, grade, chemotherapy, and population revealed two main trends. (1) In the US and Nigerian cohorts (wherein AR-positive TNBCs have a better prognosis compared with AR-negative TNBCs), the only variables significantly and independently
associated with OS were being AA, being native African, and AR positive/negative status. (2) In the cohorts from Norway, Ireland, and India (wherein AR-positive TNBCs have a poorer prognosis compared with AR-negative TNBCs), age, being from India, and AR positive/negative status were all significantly and independently associated with OS. Thus, in five TNBC cohorts from four continents, AR positive/negative status shows population-specific patterns of association with OS (Table 4.5.4). One cohort showed no pattern of association with OS.

Some noteworthy limitations of our study include potential cohort-to-cohort differences in tissue fixation and the lack of centralized AR staining for some of the cohorts. Notwithstanding these caveats, we believe that by rigorously addressing previously reported inconsistent results regarding AR’s prognostic role in TNBC, our multi-institutional study has allowed the TNBC field to move forward.

We have demonstrated that AR’s prognostic value likely hinges on the proportions of the TNBC molecular subtypes (especially non-LAR subtypes) present in the cohorts, and perhaps other covert modifiers of AR biology. Candidate modifiers in different populations might include patients’ biogeographic ancestry, AR splice variants, epigenetic factors, tumor microenvironment, and repeat length polymorphisms (RLPs) in the AR gene. For example, among the alternatively spliced transcripts known to be generated from the AR gene, the constitutively active variant AR-V7 is often expressed in TNBC cells along with the full-length mRNA, but it regulates a transcriptional program distinct from full-length AR, and may affect outcomes. At present, little is known about the expression of AR-V7 and other splice variants in TNBC tumors in different populations. Few studies have examined the RLPs or tumor microenvironment in AR
positive/negative TNBC tumors and correlated these features with the tumors’ molecular subtypes. Moreover, amplification of and mutations in the AR gene have been shown to drive castration-resistant prostate cancer; thus, further analysis of the mutational status of the AR gene and its downstream signaling components would shed some light on the biology of AR in different TNBCs arising in different populations. The aforementioned under-studied potential modifiers of AR biology could affect not only patients’ prognosis following chemotherapy, but also responses to AR-targeting agents, and are, therefore, very clinically relevant.

Another potential confounder of our results may be the molecular apocrine (MA) nature of the TNBCs in our cohorts. Studies have shown that AR binds ER-binding cis-regulatory elements in MA tumors and drives expression of genes normally regulated by ER. More importantly, a study of 58 transcriptionally defined MA tumors found that a significant proportion of these tumors showed expression of AR mRNA via qRT-PCR even though they were AR-negative via IHC [20]. These intriguing findings raise the possibility that some of the TNBC tumors in our cohorts (as well as in cohorts used in previous studies) that were classified as AR-negative based on IHC, may, in fact, be MA tumors that express detectable levels of AR mRNA and show activation of AR-target genes. In other words, IHC-based detection of AR expression may not be sensitive enough to identify all TNBC tumors in which AR-mediated signaling is active, nor to identify patients likely to respond to the AR-targeted treatments. Moreover, studies have shown AR as an independent predictor for the complete pathological response in breast cancer [21]. Since all our study cohorts are comprised of the adjuvant chemotherapy-treated patients and patients who did not receive any systemic therapy, it will be worth exploring the prognostic and predictive role of AR in similarly diverse cohorts in the context of neoadjuvant treatment.

*Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study*
4.5 References


*Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study*


*Parts of this chapter have been published verbatim in* Cancers (Basel) 2019 Jul 17; 11(7):995. *Doi:10.3390/cancers11070995* as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study”
apocrine breast cancers are aggressive estrogen receptor negative tumors overexpressing either HER2 or GCDFP15. Breast Cancer Res. 2013, 15, R37. doi:10.1186/bcr3421.

Figure 4.1 Androgen receptor (AR) expression in the seven cohorts of our multi-institutional study.

(A.) Representative micrographs showing AR expression in triple-negative breast cancer (TNBC) and their adjacent normal tissues, AR (brown) and nuclei (blue). Insets: 20X magnification. (B.) Expression of AR in the different study cohorts. (C.) Expression of AR among African Americans (AAs) and European Americans (EAs) in the US cohort. The p-value shows the significant difference in AR expression among AAs and EAs.
**Figure 4.2 Prognostic significance of the androgen receptor (AR) in different study cohorts.**

(A) Kaplan Meier survival curves for AR-negative TNBC (blue) and AR-positive TNBC (red) patients from (i) UK overall cohort, (ii) US overall cohort, (iii) Nigeria overall cohort, (iv) Norway overall cohort, (v) Ireland overall cohort, and (vi) India overall cohort. (B) Kaplan Meier survival curves for adjuvant-chemotherapy-treated AR-negative TNBC (blue) and AR-positive TNBC (red) patients from (i) UK cohort, (ii) US cohort, (iii) Nigeria cohort, (iv) Norway cohort, (v) Ireland cohort, and (vi) India cohort.

*Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study.”*
Table 4.5.1 Table representing number of cases and sample type evaluated for AR expression in global TNBC cohort and diverse institutions from US.

<table>
<thead>
<tr>
<th>Study Cohorts</th>
<th>No. of Cases</th>
<th>Sample Type</th>
<th>Years of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>239</td>
<td>TMA</td>
<td>1987–1998</td>
</tr>
<tr>
<td>Nottingham Breast Cancer Research Center</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>104</td>
<td>Full-Face</td>
<td>1978–2004</td>
</tr>
<tr>
<td>Stavanger University Hospital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>222</td>
<td>TMA</td>
<td>2000–2015</td>
</tr>
<tr>
<td>NUI Galway, Lambe Institute for Translational Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>420</td>
<td>Full-Face</td>
<td>2000–2014</td>
</tr>
<tr>
<td>Emory University Hospital</td>
<td>142</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>Grady Memorial Hospital</td>
<td>114</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>Northside Hospital Cancer Institute</td>
<td>95</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>West Georgia Medical Center</td>
<td>32</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>University of Alabama at Birmingham</td>
<td>23</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>University of South Alabama College of Medicine</td>
<td>14</td>
<td>Full-Face</td>
<td>2000–2008</td>
</tr>
<tr>
<td>Nigeria</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olabisi Onabanjo University</td>
<td>77</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>University of Maiduguri</td>
<td>16</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>Ekiti State University</td>
<td>45</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>Ahmadu Bello University</td>
<td>42</td>
<td>Full-Face</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>242</td>
<td></td>
<td>2005–2014</td>
</tr>
<tr>
<td>Tata Memorial Hospital</td>
<td>139</td>
<td>TMA</td>
<td></td>
</tr>
<tr>
<td>Rajiv Gandhi Cancer Institute and Research Center</td>
<td>103</td>
<td>Full-Face</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.5.2 Patient demographic and tumor data stratified by cohort; n (%).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Clinicopathological Variables of Study Cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK</td>
</tr>
<tr>
<td></td>
<td>n = 239</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>128 (53.55)</td>
</tr>
<tr>
<td>≥50</td>
<td>111 (46.45)</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
</tr>
<tr>
<td>I/II</td>
<td>212 (88.7)</td>
</tr>
<tr>
<td>III/IV</td>
<td>26 (10.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 (2.1)</td>
</tr>
<tr>
<td>2</td>
<td>12 (5.0)</td>
</tr>
<tr>
<td>3</td>
<td>221 (92.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
</tr>
<tr>
<td>Treated</td>
<td>106 (44.4)</td>
</tr>
<tr>
<td>Untreated</td>
<td>110 (46)</td>
</tr>
<tr>
<td>Missing</td>
<td>23 (9.6)</td>
</tr>
<tr>
<td>Vital status</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>112 (46.9)</td>
</tr>
<tr>
<td>Alive</td>
<td>127 (53.1)</td>
</tr>
<tr>
<td>Androgen Receptor</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>130 (54.9)</td>
</tr>
<tr>
<td>Negative</td>
<td>109 (45.1)</td>
</tr>
</tbody>
</table>
Table 4.5.3 Table showing overall survival of AR positive TNBC at different thresholds of AR.

<table>
<thead>
<tr>
<th>Study</th>
<th>AR+ TNBC</th>
<th>AR- TNBC</th>
<th>5% Cutoff</th>
<th>AR+ TNBC</th>
<th>AR- TNBC</th>
<th>10% Cutoff</th>
<th>AR+ TNBC</th>
<th>AR- TNBC</th>
<th>Optimal Cutoff</th>
<th>AR+ TNBC</th>
<th>AR- TNBC</th>
<th>Optimal Cutoff</th>
<th>Optimal Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>82.42</td>
<td>70.21</td>
<td>0.0402</td>
<td>82.5</td>
<td>70.59</td>
<td>0.0471</td>
<td>82.98</td>
<td>69.94</td>
<td>0.0248</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>UK</td>
<td>70.27</td>
<td>67.19</td>
<td>0.3711</td>
<td>70.53</td>
<td>67.36</td>
<td>0.3677</td>
<td>73.33</td>
<td>66.46</td>
<td>0.1613</td>
<td>30</td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Nigeria</td>
<td>85.71</td>
<td>22.54</td>
<td>0.0846</td>
<td>80</td>
<td>23.43</td>
<td>0.2682</td>
<td>68.75</td>
<td>20.73</td>
<td>0.0117</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Ireland</td>
<td>60.34</td>
<td>75.43</td>
<td>0.0738</td>
<td>60.78</td>
<td>74.9</td>
<td>0.1127</td>
<td>58.93</td>
<td>75.64</td>
<td>0.0472</td>
<td>6</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Norway</td>
<td>63.33</td>
<td>77.03</td>
<td>0.0533</td>
<td>60.71</td>
<td>77.63</td>
<td>0.0211</td>
<td>57.69</td>
<td>78.21</td>
<td>0.0064</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>India</td>
<td>75</td>
<td>91.55</td>
<td>0.0157</td>
<td>75</td>
<td>91.24</td>
<td>0.0304</td>
<td>80.77</td>
<td>92.06</td>
<td>0.0212</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study.”
Table 4.5.4 Multivariate analysis to reveal population-specific differences between cohorts.

*AA = African American; AR = androgen receptor.*

<table>
<thead>
<tr>
<th>Variables</th>
<th>US and Nigeria Study Cohorts</th>
<th>Norway, Ireland and India Study Cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard Ratio (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td>Hazard Ratio (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>Overall Survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt;50 vs. ≥50</td>
<td>1.00 (0.99–1.01)</td>
<td>0.3934</td>
</tr>
<tr>
<td>Grade 2</td>
<td>0.77 (0.86–8.84)</td>
<td>0.085</td>
</tr>
<tr>
<td>Grade 3</td>
<td>3.63 (1.15–11.44)</td>
<td>0.0275</td>
</tr>
<tr>
<td>Adjuvant Chemotherapy</td>
<td>0.72 (0.39–1.32)</td>
<td>0.2877</td>
</tr>
<tr>
<td>Population AA</td>
<td>1.82 (1.30–2.53)</td>
<td><strong>0.0004</strong></td>
</tr>
<tr>
<td>Population African</td>
<td>11.22 (8.11–15.51)</td>
<td><strong>&lt;0.0001</strong></td>
</tr>
<tr>
<td>Population Indian</td>
<td>-</td>
<td>0.49 (0.27–0.89)</td>
</tr>
<tr>
<td>Population Irish</td>
<td>-</td>
<td>0.79 (0.48–1.28)</td>
</tr>
<tr>
<td>AR AR-positive vs AR-negative</td>
<td>1.70 (1.10–2.61)</td>
<td><strong>0.0157</strong></td>
</tr>
</tbody>
</table>

*Parts of this chapter have been published verbatim in Cancers (Basel) 2019 Jul 17; 11(7):995. Doi:10.3390/cancers11070995 as “Prognostic Role of Androgen Receptor in Triple Negative Breast Cancer: A Multi-Institutional Study*
5 UPREGULATION OF WNT/β-CATENIN SIGNALING UNDERLIES AGGRESSIVE DISEASE COURSE IN QUADRUPLE NEGATIVE BREAST CANCER

5.1 Abstract

Based on AR expression, triple negative breast cancer (TNBC) is further subdivided into androgen receptor-positive (AR+) and androgen receptor-negative (AR- or Quadruple negative breast cancer, QNBC) subtypes. While, AR has emerged as a potential therapeutic target for AR+ TNBC, its prognostic role in TNBC remains controversial. Moreover, AR+ TNBC responds well to AR antagonists but there are no targeted drugs available for QNBC patients beyond chemotherapy. Thus, this study validates the impact of AR loss on prognosis of TNBC and uncover the molecular pathways/drivers that can be therapeutically targeted in QNBCs. AR expression was immunohistochemically evaluated in well-annotated formalin-fixed paraffin-embedded samples from 136 TNBC patients procured from two different institutes in the US. Samples with ≥1% of AR+ nuclei were deemed AR+ TNBCs. Finally, using a combination of in silico analysis, in vitro assays and RNA-sequencing, the study uncovered the molecular drivers/pathways upregulated in QNBCs. AR expression was observed in only around 29.4% of the total cases and that loss of AR was significantly associated with poor overall survival (OS) in TNBCs patients (p<0.001; n=96 for QNBCs, n=40 for AR+ TNBCs) among all patients. An RNA-sequencing on TNBC cases and Gene Set Enrichment Analysis revealed an upregulation of Wnt/β-catenin axis among QNBC cases. Moreover, proteomic data from TCGA exhibited significant upregulation of β-catenin and Dvl3 expression—evidence of Wnt signaling overdrive in QNBCs. Furthermore, β-catenin, HMGA2, VEGF-A and FOSL1 mRNA levels were higher in QNBCs when compared with AR+ TNBCs (p<0.05), suggesting aberrant activation of Wnt/β-catenin signaling in QNBCs. Using
reporter gene assays, the study confirms the activation of Wnt signaling in QNBC cells compared to AR+ TNBCs along with elevated levels of Wnt target genes (Axin2, CD44, and Cyclin D1). Interestingly, an enhanced nuclear accumulation of β-catenin in QNBCs was observed, indicative of attenuated β-catenin phosphorylation. Upon AR knockdown in TNBC cells, we could induce QNBC phenotype in AR+TNBCs wherein we observed a reduced expression of E-cadherin (suggesting enhanced epithelial-mesenchymal transition or EMT) and increased β-catenin expression. Concomitantly, β-catenin target genes were also up-regulated along with enhanced proliferation, migration and invasion capacities (p<0.05). This study validates the striking association of AR loss with worse clinical outcome. The collective data offers compelling evidence to support mis regulation of oncogenic Wnt/β-catenin in AR negative scenario. Finally, the study elucidates a previously unknown link of AR loss to more aggressive disease course while uncovering actionable targets in QNBCs.

5.2 Introduction

Triple Negative Breast Cancer (TNBC) is the most aggressive subtype of breast cancer lacking the expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). It comprises of 10-20% of invasive breast carcinomas. TNBC patients exhibit poor 5-year survival due to relatively higher rates of metastasis and recurrence when compared with other breast cancer subtypes (1-3). TNBC diagnosis occurs at an earlier age and more advanced stage of the disease (4). These patients are currently treated with Anthracycline-taxane-based chemotherapeutic regimens due to the lack of recognized molecular targets. Though the patients in early stages respond well to chemotherapy, the varied nature of tumor microenvironment in TNBC considerably influences the risk of relapse (1). Based on the gene expression profiling, TNBC is further divided into AR-positive (AR+) TNBCs and AR-
negative TNBCs (or Quadruple Negative Breast Cancer, QNBC) (5,6). Though AR is an emerging
therapeutic target for TNBCs, the prognostic significance of AR remains controversial. Multiple
groups have reported AR expression to be associated with significantly longer disease-free,
overall, and recurrence-free survival compared with a lack of AR expression among TNBC
patients (7-11). This underscores that loss of AR in TNBC tumors likely confers an aggressive
disease course. However, there are some conflicting studies showing that AR+ phenotype portends
worse long-term outcomes or that AR status has no significant impact on TNBC prognosis (12-
17). A recent study demonstrated that the prognostic significance of AR is population specific
(18). Note worthily, where AR+ TNBC responds well to AR antagonists, there are no targeted
drugs available for QNBC patients beyond chemotherapy, attributed primarily to the lack of
actionable pathways and targets.

Canonical Wnt/β-catenin signaling is a fundamental growth pathway essential for cell
proliferation, embryonic development and tissue homeostasis. Upon pathway activation, i.e., the
binding of Wnt ligands to the bonafide receptors, β-catenin is stabilized and translocated to the
nucleus to activate the expression of Wnt target genes. Hyper-activated Wnt pathway is associated
with several tumor types, including breast cancer (19-21). This induces unchecked expression of
Wnt target genes responsible for tumor initiation, progression and cell cycle regulation:
cyclin D1 and c-Myc, stem cell gene Bmi-1, matrix metalloproteinase Mmp-7, and Axin2 (22-27). Apart from
serving as a key molecular player in Wnt pathway, β-catenin via binding to the cytoplasmic domain
of E-cadherin—a cell surface protein—also maintains intercellular adhesiveness and plays a
critical role in cancer invasion and metastasis (28-29). Accordingly, imbalance in the structural
and signaling properties of β-catenin has been shown to result in deregulated growth connected to
cancer and metastasis (30). Furthermore, Wnt/β-catenin signaling is upregulated in TNBC/basal-
like breast cancer compared to hormone receptor-positive/luminal cases; and pharmacologic inhibition of Wnt in TNBC has been shown to inhibit cell migration and invasion (31). Thus, Wnt signaling is an attractive therapeutic target in breast cancer and drives the quest for discovery of new pathway members, signaling network cross-talks and small-molecules modulators.

In this lieu, AR is known to interact with β-catenin in an androgen dependent manner (32-33). Androgen bound AR can inhibit β-catenin target gene expression as a result of competitive binding with TCF/LEF1 transcription factors (43-37). Furthermore, AR inactivation in female mice is known to upregulate oncogenic Wnt/β-catenin signaling (38). Collectively these and other findings point towards a previously uncharacterized role of AR in Wnt pathway. Thus, in this study we validate the prognostic role of AR expression in US cancer cohort and using combination of molecular and cellular biology uncover the regulatory role of AR in aberrant Wnt pathway activity in QNBCs as a potential therapeutic avenue.

5.3 Methods

5.3.1 Study Cohorts and Samples

We identified a total of 136 TNBC cases from the database from multiple institutes [Emory Decatur Hospital (n=81), Georgetown Hospital (n=55)] in United States. Formalin-fixed paraffin-embedded samples for this study were obtained with information on clinical outcomes from the above-mentioned institutions. All the aspects of the study including protocols, sample procurement, and whole study design were approved by respective Institutional Review Boards of the hospitals. Patient consent was not required since all the samples were archival and were de-identified to maintain the patient privacy and anonymity. The clinicopathological characteristics were reviewed and provided by the pathologist from the respective hospitals. All the information of each patient was recorded to generate a database of all the patients. Descriptive statistics for demographical and clinicopathologic characteristics of patients are provided in Table 5.6.1.
5.3.2 **Immunohistochemistry (IHC)**

Tissue staining was performed as previously described (18). Briefly, Formalin fixed Paraffin embedded (FFPE) full-face tissue sections were deparaffinized following by rehydration in a series of ethanol baths (100%, 90%, 75% and 50%). Heat-induced retrieval of antigen epitopes was performed in citrate buffer (pH 6.0) using pressure cooker at 15 psi for 30 mins. The samples were quenched using Hydrogen Peroxide for 15 mins followed by incubation for 40 min at room temperature with anti-AR primary antibody (Monoclonal Mouse Anti-Human Androgen Receptor, clone AR 441) at 1:50 dilution. The samples were then incubated with MACH2 HRP-conjugated secondary antibody (Biocare Medical) for 30 min. The enzymatic antibody detection was done using Betazoid DAB Chromogen Kit (Biocare Medical). The tissue sections were counter stained with Mayer’s hematoxylin for 1 min. Slides were then dehydrated in alcohol, cleared in Xylene and mounted with mounting media.

5.3.3 **Assessment of IHC staining**

The semi-quantitative immunohistochemical scoring method was used, which takes the intensity and percentage of stained invasive tumor tissue stained into account (39). Two independent pathologists without prior knowledge of the patients’ pathologic or outcome data scored the biomarker. TNBC samples were considered QNBC if they had less than 1 percent of the cells positive for AR; samples that had AR in more than 1 percent of the cells were considered positive for AR.

5.3.4 **Cell culture**

AR positive (MFM 223 and HCC 70) and AR negative (HCC1806 and BT-20) cell lines were purchased from American Type Cell Culture (ATCC) and were grown in standard conditions. Briefly, BT-20 was cultured in Modified Eagle's medium with 10% fetal bovine serum (FBS)
HCC1806 and HCC 70 were cultured in Roswell Park Memorial Institute medium with 10% (FBS) (ATCC). MFM 223 was cultured in Minimal Essential medium 10% (FBS). Cells were maintained in humidified 5% CO$_2$ at 37°C.

### 5.3.5 AR knockdown and overexpression

RNAi knockdown of AR was performed using AR siRNA (SC, 29204). Non-targeting control siRNAs (SC, 37007) were used as controls. siRNA pools were transfected using 25 picomolar siRNA and 7.5 µl Lipofectamine RNAimax were used per well of a 6-well plate (resulting in a final siRNA concentration of 25 picomolar/well) according to the manufacturer’s protocol. All reagents were purchased from Thermofisher scientific (Waltham, MA, USA). Similarly, AR was genetically overexpressed by transfecting cells with AR (Myc-DDK-tagged)-Human androgen receptor, transcript variant 1 (Origene, RC215316). Cells at a confluency of around 65% were transfected using Lipofectamine 2000 according to manufacturer’s instructions.

Efficiency of RNAi knockdown and overexpression was analyzed after 36 h of transfection by Western blot analysis using antibodies specific to AR. The knockdown efficiency was >90%.

### 5.3.6 Western blotting and antibodies

After 36 h of transfection, whole cell extracts were prepared with Lysis buffer (5% SDS, 10% glycerol, 0.5 M Tris-Cl) supplemented with protease and phosphatase inhibitors and equivalent amounts of proteins were separated by SDS-PAGE. Proteins were then transferred to polyvinylidene difluoride (PVDF) membrane (Bio-Rad, Hercules, CA, USA) followed by blocking with 5% BSA in tris-buffered saline (TBS) supplemented with 0.1% Tween-20 for 1 h. Membranes were then incubated with primary antibodies {AR 441, Agilent (1:800), overnight on a shaker at 4°C. β-actin was used as an internal loading control and was detected using mouse monoclonal anti-B-actin antibody (from SC). The Pierce ECL detection kit (Thermo Scientific;
Waltham, MA, USA) was used to visualize bands. Cells were lysed using the CERI and CERII reagents of the cytoplasmic-nuclear extraction kit (ThermoFisher), to obtain cytoplasmic and nuclear fractions. Membrane was incubated overnight with active β-catenin (MilliporeSigma, Burlington, MA, USA) and Lamin A/C antibodies.

5.3.7 Cell proliferation assay

Equal quantities (5000 cells/ml) of cells from all four cell lines were seeded into 6 well plates. The plates were incubated at 37 °C in 5% CO₂ for a week. Following incubation, the colonies were fixed in ice-chilled 70% methanol followed by crystal violet staining. The colonies were then counted under the microscope. The experiment was performed in duplicate.

5.3.8 Invasion assay

The cells were suspended in media in $5 \times 10^4$ cells/ml density. We carried out transmigration assay on Invasion chamber with pore size of 8 micrometers. This assay detects cells that have migrated to the bottom of a chamber-based apparatus in response to a chemoattractant or at random, depending on the setup. The cells were serum starved for 24 h and were suspended at a density of $5 \times 10^4$ in media (serum free). Cells were then loaded into the upper chamber (which contains serum free medium) of the invasion chambers (8.0 um of the pore size) and incubated for 24 h. The lower part of the well was supplemented with respective mediums of the cell lines containing 10% FBS (to assess random migration). After 24 h, the cells and medium in the upper compartment of the chamber was removed with the help of cotton swab, and the cells on the lower side of the chamber were fixed with paraformaldehyde. Following fixation, the chambers were transferred to wells with crystal violet to stain the cells (migrated) on the lower part of the chamber membrane. The migrated cells were quantified in 10 random fields imaged using a light microscope. The experiment was repeated two times maintaining same conditions.
5.3.9  Wound healing assay

Cells were seeded in 12-well plates and cultured until they reached 90% confluence, then they were serum starved for 12 h prior to the assay. Using a 100 ul pipette tip, a scratch was made keeping the pipette tip at an angle of around 30 degrees to create a wound. After scratching, the cells were washed in 1X Dulbecco's phosphate-buffered saline followed by addition of serum-containing medium and incubating the cells for 24 h in 5% CO₂ at 37 °C. Wound edges were imaged using a 10X objective. The experiments were performed in duplicate and we measured three values from each scratched region.

5.3.10  Wound healing assay

Cells were plated in a 24-well plate at a density of 1x10⁴ cells/well and transfected using Lipofectamine (Invitrogen) with 5 μg of TOPflash or FOPflash (obtained from S. Kapoor IIT, India) along with RL-TK (S. Kapoor, IIT, India) for determination of transfection efficiency. Lysates were harvested after 24 h and analyzed using the Dual Luciferase Assay System kit (Promega). Luciferase activity was normalized for transfection efficiency and graphed as ratio of TOPflash/FOPflash activity.

5.3.11  RNA isolation and quantitative RT PCR

Total cellular RNA was isolated from cultured cells using the RNAeasy mini kit (Qiagen, Montgomery, MD). Reverse transcriptase reactions were done using the SuperScript first strand synthesis system (Biorad) with 1 μg of RNA and 1 μl of oligo (dT) 12–18 as the reverse transcription primer. Real-time PCR was performed in an apparatus using the Light-Cycler-FastStart DNA Master plus SYBR Green I Kit (Biorad). cDNA primers for MYC, AXIN2, MMP7, BMI were used. The β-actin gene was used as a normalization control. At the end of the PCR cycles, melting curve analyses was performed to validate the generation of the specific PCR
product expected. The fold change in target gene relative to the \( \beta\)-actin control was determined by \( 2^{-\Delta\Delta C_t} \) method.

5.3.12 Statistical analysis of experimental data

We plotted statistical bar graphs using excel and SPSS. Experimental groups were compared using Student’s t-test, and p-values were calculated. \( p<0.05 \) was considered statistically significant.

5.3.13 Statistical analysis of clinical data

Statistical analysis was carried out with SAS 9.4 ® software. Differences between clinicopathological proportions were determined using the \( \chi^2 \) test. Differences between baseline IHC biomarker expressions, or continuous clinicopathological variables, were evaluated via a 2-tailed t-test. Univariate Kaplan-Meier curves were used to investigate the effect of AR status (positive vs. negative) on overall survival and using the log-rank test for statistical significance. Multivariable Cox Proportional Hazard models were used to correct for stage, grade and age with significance being determined with the Wald chi-square test. For all clinical survival analysis, we used the overall survival rate wherein time to death from surgery was used as the time interval and death was used as an event.

5.3.14 Insilico data analysis

The results shown here are in part based upon data generated by the TCGA Research Network. We obtained TCGA mRNA-seq data, alongside the annotated clinicopathological data, through the TCGA program (http://cancergenome.nih.gov) whose sequencing was done as previously detailed (41), 127 TNBC patients were used for mRNA analysis of 20,437 genes. The threshold for what determines AR probe “negativity” was the cutoff, which optimally (using Contal’s and O’Quigley’s approach: http://www2.sas.com/proceedings/sugi28/261-28.pdf)
stratified the overall survival of the patient cohort. Protein data was obtained via TCPA BRCA dataset and contained expression levels of 901 proteins (42). Survival data for both mRNA and protein TCGA datasets was analyzed as previously discussed.

5.4 Results

5.4.1 Lack of AR in TNBC patients is associated with poor overall survival

To validate the role of AR in TNBCs, we stained 136 full face TNBC tissue sections for AR from multiple institutes (for details see materials and methods). Descriptive statistics regarding patient and clinicopathologic characteristics for tumor samples utilized are given in Table 5.6.1. Consistent with our previous studies, we observed that expression of AR in US cohort was 29.4% (Figure 5.1A, B). We observed that AR loss was associated with high Ki67 group (Ki67 expression>14%) and higher age. As for grade, significantly higher proportion of QNBC patients were observed in Grade 3. In addition, Kaplan Meier’s survival analysis showed that AR loss was significantly associated with the poor overall survival (Figure 5.1C). Next, we performed the multivariate regression analysis to reveal that AR status retained its relative significant prognostic value (HR= 4.60; p = 0.0005) when controlling for common clinicopathological variables such as race, grade, and age (Table 5.6.2). Collectively our data implies that loss of expression of AR is associated with worse clinical outcome in QNBCs.

Having established the association of QNBCs with worse outcome and in clinical samples, we then investigated the effects of AR status in well-established TNBC cell lines. To this end, we first analyzed AR protein expression in a panel of 13 TNBC cell lines and subsequently four cell lines (two AR negative (HCC 1806 and BT-20) and two AR positive) were selected for further experiments, Figure 5.2A (i, ii). Next, we evaluated the role of AR in modulating cell proliferation using clonogenic assay for AR+ and QNBC cells. We observed significantly higher number of
colonies in QNBC than TNBC (p=0.03), Figure 5.2B (i, ii). Using invasion assay we demonstrated that loss of AR impaired the motility of TNBC cells, alternatively, an increase in invasion capacity of QNBC cells. Our results show that significantly higher number of QNBC cells migrate to lower compartment of the invasion chamber after 12 h (p=0.002), Figure 5.2C (i, ii). Additionally, using the classical wound healing assay to assess the cell migratory capacity, we revealed that cells lacking AR filled (~80% wound in 24 h) the scratch wound in less time compared to AR expressing cells. The p value was significant for TNBC vs QNBC (p<0.051), Figure 5.2D (i, ii). Based on the above findings, we proceeded to investigate whether AR loss is associated with cellular phenotype in the chosen cell line models. We transiently knocked down AR in AR+ TNBC cell lines (MFM 223 and HCC 70) and performed proliferation, migration and invasion assay. We found that AR KD increased the proliferation and migration capacity of the TNBC cells and OE if AR rescued these effects in QNBC cell lines (Figure 5.3 A,B). Moreover, the percentage of wound closure increased drastically when AR was KD in these cell lines (Figure 5.3C). Taken together, these data collectively suggest that loss of AR facilitates direct cell migration, invasion and proliferation. This observation lays a strong foundation that links the loss of AR to a more aggressive cancer phenotype.

5.4.2 In silico analysis reveals upregulation of Wnt signaling in QNBCs

AR loss linked with poorer outcomes and aggressive disease phenotype in QNBCs hints toward de-regulated cellular pathways leading to disease progression. Thus, using publically available proteomics data from TCGA, we compared the levels of 187 proteins involved in breast tumorigenesis in AR-high vs AR-low TNBCs (using median AR expression as a cutpoint, as the IHC-derived AR status of these samples is unknown). We observed that AR low group expressed significantly higher levels of β-catenin and Dvl3 when compared to the AR high group. To
confirm this observation, we performed gene set enrichment analysis (GSEA) to identify gene ontologies associated with the low AR group. These were found to be enriched in Wnt/β-catenin signaling pathway (Figure 5.4 A) and constitute the top-ranked gene ontology among biological processes highly active in aggressive TNBC (False discovery rate q-values <0.25 were considered as significant). We further observed that β-catenin m-RNA level was significantly higher in QNBCs compared to TNBCs by probing the publicly available data generated by the TCGA Research Network (Figure 5.4 B). Furthermore, TCGA data analysis reveals upregulation of HMGA2, VEGF-A and FOSL1 genes in QNBCs. Higher m-RNA level of β-catenin, HMGA2, VEGF-A and FOSL1 suggests upregulated Wnt signaling in QNBCs (Figure 5.4 Ci, ii and iii). Thus, oncogenic Wnt/β-signaling appears to be a silent feature of the AR-negative TNBC subgroup and potentially presents clinically actionable target.

### 5.4.3 Upregulation of Wnt/β-catenin is associated with loss of AR expression

To validate hyperactive canonical Wnt signaling in QNBCs, we measured the basal Wnt activity in QNBC cell lines using STF Wnt reporter assay. Upon Wnt activation, β-catenin is stabilized and joins TCF inside the nucleus. Together, they bind to TCF sites and induce the expression of Wnt target genes. The STF plasmid contains 8 of these TCF binding sites which control the expression of a firefly luciferase, whereas Fop-Luc contains mutated TCF/LEF binding sites upstream of the luciferase reporter. In addition, pRL-Renilla was used as normalizer that codes for Renilla luciferase. QNBC cells demonstrated a significantly high level of basal Wnt activity compared to TNBC cells (Figure 5.5 Ai). Next, we measured the expression of downstream direct targets of Wnt/β-catenin signaling pathway via qPCR. In concordance with the literature, we could show a higher expression of Cyclin D1, AXIN2, Myc and MMP7 in QNBC cell lines when compared to TNBC cell lines (Figure 5.5 Aii). Moreover, to gain insight into the
cross talk between AR and Wnt signaling pathway, we KD AR and measured the basal Wnt activity. We observed that AR KD significantly increased the basal Wnt activity in MFM 223 and HCC 70 (Figure 5.5 Bi). Concomitantly, we were also able to demonstrate higher expression of Cyclin D1, AXIN2, Myc, and MMP7 after KD of endogenous AR in TNBC cell lines (Figure 5.5 Bii). Because Wnt signaling and β-catenin expression are upregulated in QNBC patients, we investigated the relationship between AR loss and Wnt protein levels. AR KD in TNBC cell lines reduced E-cadherin expression, suggesting enhanced epithelial-mesenchymal transition, and increased the expression of β-catenin and its target genes CD44 and cyclin D1 (Figure 5.5 Ci A). Conversely, overexpression (OE) of AR in QNBC cell lines resulted in a striking upregulation of E-cadherin and downregulation of β-catenin, CD44 and cyclin D1 (Figure 5.5 Cii B). AR interacts with β-catenin in an androgen-dependent manner and androgen-bound AR competes with the transcription factors TCF and LEF1 for binding to β-catenin and suppressing the expression of β-catenin target genes. Thus, we also observed β-catenin upregulation in QNBC cells, suggesting Wnt/β-catenin signaling activation (Figure 5.5 Cii). In addition, QNBC cell lines showed an increase nuclear accumulation of β-catenin and also upon KD of AR in TNBC suggesting an upregulated Wnt signaling cascade in AR low or QNBCs (Figure 5.5 Ciii). Finally, we analyzed total and “activated”, or non-phosphorylated, β-catenin by immunoblotting. Total and “activated” β-catenin were expressed at high levels in QNBC cell lines (Figure 5.5 Civ). Taken together, our results substantiate that hyper-activated Wnt/β-catenin signaling likely represents a promising therapeutic target in QNBCs.
5.4.4 Aberrant localization of membrane $\beta$-catenin staining is associated with poor overall survival

In order to analyze the association between aberrant expression of $\beta$-catenin with aggressive disease course, we stained all the 136 cases for $\beta$-catenin. Our results showed a strong membranous expression of $\beta$-catenin in the TNBCs. Faint cytoplasmic staining was infrequently detected, but no nuclear staining was demonstrated in the TNBC tissue. In a total of 40 TNBC cases, 31 (77.5%) had intact membrane $\beta$-catenin staining while 9 (22.5%) cases depicted loss of membrane (LOM) expression of $\beta$-catenin. Normal membranous $\beta$-catenin expression was detected in 29/96 (28.2%) cases of QNBCs, and loss of membrane (LOM) expression of $\beta$-catenin in 67/96 (71.8%) cases of QNBC. Out of sixty-seven cases with LOM $\beta$-catenin expression, 38 cases (56.7%) showed LOM expression without cytoplasmic and/or nuclear staining, 27 cases (40.3%) demonstrated LOM with cytoplasmic staining, and 2 case (3%) showed LOM with nuclear and cytoplasmic staining (Figure 5.6 A, B). Cytoplasmic staining of $\beta$-catenin was diffuse and coarsely granular or dot-like. Cytoplasmic $\beta$-catenin was frequently accumulated in the perinuclear areas. In addition, Kaplan Meier’s survival analysis revealed that abnormal $\beta$-catenin expression was significantly correlated with worse overall survival only in the QNBC subgroup of patients and not among TNBCs indicating that the absent or decreased membrane expression of $\beta$-catenin in QNBCs was associated with tumor invasiveness and a poor prognosis (Figure 5.7).

5.5 Discussion

Breast cancer tumors are associated with various clinical variables including tumor size, histological subtype and grade, lymph node status and the expression of estrogen receptor alpha (ERa), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), all of which currently assist the routine clinical management (49). In lieu of the vast heterogeneity in
breast cancer, there is an unmet need to discover novel biomarkers aimed towards accurate and personalized prognosis (50). TNBC is arguably the most aggressive molecular subtype of breast cancer due to its overall aggressiveness and lack of targets for hormonal therapy. This leaves the clinicians to treat the disease with conventional chemotherapy only. In recent years, TNBC has been classified to various subtypes based on their gene expression profiles. Luminal Androgen Receptor (LAR) is one of the well-known subtypes with less response to chemotherapy and delayed recurrences when compared with other subtypes. The underlying reason of this difference is presence of the AR in this group. The expression of AR in TNBC patients is highly variable, ranging from 6-75% (16, 51-55) which is consistent with our findings. Several studies in TNBC cohorts have shown that AR positive TNBCs are good prognostic markers associated with lower stage, grade and mitotic scores (7,56). On the other hand, some studies mention that loss of AR associates with increased risk for recurrence, distant metastasis and high chances of mortality consistent with what we observed in our cohort (15,16). Studies have highlighted that multiple factors like samples sizes, source and sensitivity of antibody and scoring methods (cut off values) might be some critical reasons underlying these discrepancies (57,58). Nonetheless, the lack of molecular targets in AR negative TNBCs (QNBCs) bring to the forefront the dire need to discover novel bio-molecular players or pathways amenable for therapeutic targeting.

Interestingly, our in-silico analysis shows that loss of AR correlates with upregulated Wnt/\(\beta\)-catenin pathway and poor survival. In this regard, studies have revealed an increase in Wnt signaling being associated with poor clinical outcomes (59-61). Moreover, in TNBCs there is an upregulation of Wnt/\(\beta\)-catenin signaling compared to ER/PR-positive/luminal BC, and studies show that pharmacologic inhibition of Wnt signaling inhibits TNBC cell migration and invasion (59). Our study demonstrates a higher \(\beta\)-catenin expression in the AR negative TNBCs implying
a de-regulated Wnt signaling in QNBCs. Further, we could demonstrate that QNBC cell lines proliferate faster when compared with AR+ TNBCs and have higher migratory and invasive capabilities which further suggests an upregulated Wnt/ β-catenin signaling in QNBCs.

β-catenin is a co-activator of AR and regulates the AR gene transcriptionally through multiple TCF binding sites within the AR promoter region (62-65). However, AR pathway inhibition via β-catenin is still to be elucidated. Our study suggests that QNBCs have higher expression of basal Wnt activity and this expression increases in TNBCs after AR KD. Canonical Wnt signaling activity transcriptionally induces Wnt target genes that contain consensus TCF/LEF binding element. Cyclin D1 gene is a direct target for transactivation by the β-catenin pathway through a LEF-1 binding site in the promoter of Cyclin D1 (66). Aberrant activation of Wnt/ β-catenin signaling pathway is known to upregulate c-Myc and induces cell proliferation and differentiation (67); furthermore, nuclear accumulation of β-catenin protein together with the DNA binding protein TCF-4 functions as transcriptional activator of MMP-7 (48). In line with the upregulated Wnt pathway activity revealed using reporter gene assay, additional data show a higher expression of Cyclin D1, AXIN2, c-Myc, MMP7 and BMI-1 genes in QNBC cell lines when compared to TNBC cell lines. The same was also observed upon AR KD in TNBCs, further supporting the functional role of AR loss in deregulating Wnt pathway.

In the present study, the abnormal β-catenin expression was noted in more than 50% of cases, but it was higher among the QNBCs. This incidence is in accord with those of the previous reports (26,68). This fact implies that the Wnt/ β-catenin pathway is activated in breast cancers (69-70). The subcellular localization of β-catenin serves as a ruler of its activity. The cytosolic pool of β-catenin which is free, undergoes rapid proteolytic degradation while the accumulated cytoplasm and/or nucleus fraction is accounted by the stabilized β-catenin pool propelling the β-
catenin/TCF activity. β-Catenin localizes in the membrane when the transactivation activity is very low (71-73). Expression of nuclear β-catenin has been rarely demonstrated in breast cancers (26), when compared to other cancers such as colon, stomach, gallbladder and ovarian cancer (73-78). In our study, focal/weak nuclear expression of β-catenin was observed only in two case. Frequent cytoplasmic expression and rare nuclear expression in breast cancers reflect that the different signaling pathways may affect β-catenin in QNBCs. Moreover, our study reveals that loss of membrane expression of β-catenin was associated with worse clinical outcome in concordance with the plethora of literatures indicating that the absence or decreased membrane expression of β-catenin is associated with tumor invasiveness and a poor prognosis (74-76, 78). These results suggest that loss of membrane β-catenin staining with the cytoplasmic and/or nuclear expression of β-catenin is a stronger biologic indicator for poor prognosis.

In conclusion, our study demonstrates the loss of AR in TNBC being associated with worse clinical outcome. The collective data supports the misregulation of oncogenic Wnt/β-catenin in QNBC group of patients. Finally, we elucidate a previously unknown link of AR loss to a more aggressive disease course governed by Wnt pathway. These findings are expected to inspire further investigations that may lead to uncovering novel and hitherto unknown actionable targets in QNBCs by exploiting the now validated AR-Wnt axis.

5.6 References


Table 5.6.1 Clinicopathological variables of the study cohort.

<table>
<thead>
<tr>
<th>Clinicopathological variables of study cohort</th>
<th>N=136</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>128 (53.55)</td>
</tr>
<tr>
<td>≥50</td>
<td>111 (46.45)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>2</td>
<td>14 (10.3)</td>
</tr>
<tr>
<td>3</td>
<td>103 (75.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>17 (12.5)</td>
</tr>
<tr>
<td>Ki67</td>
<td></td>
</tr>
<tr>
<td>≤14 (low)</td>
<td>22 (16.2)</td>
</tr>
<tr>
<td>&gt;14 (high)</td>
<td>114 (83.8)</td>
</tr>
<tr>
<td>Vital status</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>58 (42.6)</td>
</tr>
<tr>
<td>Alive</td>
<td>71 (52.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>7 (5.2)</td>
</tr>
<tr>
<td>Androgen Receptor</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>40 (29.4)</td>
</tr>
<tr>
<td>Negative</td>
<td>96 (70.6)</td>
</tr>
</tbody>
</table>
Table 5.6.2 Multivariate analysis after adjusting the confounding variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hazard Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt;50 vs. ≥50</td>
<td>1.013</td>
<td>0.2064</td>
</tr>
<tr>
<td>Grade 2</td>
<td>1.034</td>
<td>0.976</td>
</tr>
<tr>
<td>Grade 3</td>
<td>1.087</td>
<td>0.9361</td>
</tr>
<tr>
<td>Tumor size</td>
<td>1.084</td>
<td>0.0422</td>
</tr>
<tr>
<td>AR AR-positive vs AR-negative</td>
<td>4.6</td>
<td><strong>0.0005</strong></td>
</tr>
</tbody>
</table>
Figure 5.1 AR expression and its association with overall survival.

(A) Representative micrographs showing AR expression in TNBC and their adjacent normal tissues, AR (brown) and nuclei (blue); (B) Pie-chart showing expression of AR in study cohort, (C) Kaplan Meier’ survival curves for QNBC (blue) and TNBC (red) patients.
Figure 5.2 QNBCs are associated with higher proliferation, invasion and migration.

A(i) Immunoblots showing the levels of AR and β-actin in different TNBC cell lines. A(ii) Immunoblots showing the levels of AR and β-actin in two TNBC cell lines (MFM 223 and HCC 70) and two QNBC cell lines (HCC 1806 and BT-20); B(i) Bar graph representing the number of colonies formed when proliferation assay was performed for 48 h; B(ii) Micrographs showing Proliferation assay TNBC and QNBC cell lines; C(i) Bar graph representing percentage of cell invaded when invasion assay was performed; C(ii) Micrographs showing Invasion assay TNBC and QNBC cell lines; D(i) Bar Graph showing percentage of wound healing capacity in 24 h; D(ii) Micrographs showing Invasion assay, Wound healing assay in TNBC and QNBC cell lines.
Figure 5.3 Loss of AR is associated with higher proliferation, invasion and migration.

A. Bar graph representing percentage of cell invaded when invasion assay was performed after transiently knocking down and overexpressing AR; B. Bar graph representing the number of colonies formed when proliferation assay was performed for 48 h after transiently knocking down and overexpressing AR; C. Bar Graph showing percentage of wound healing capacity in 24 h after genetic manipulation of AR.
Figure 5.4 Upregulation of Wnt/β-catenin signaling pathway in QNBCs.
(A) Enrichment plot showing QNBC group is enriched in gene sets associated with Wnt/β-catenin signaling; (B) β-catenin expression was significantly higher in QNBCs compared to TNBCs in TCGA dataset; (Ci, ii and iii) HMG2A, VEGFA and FOSL1 expression are significantly higher in QNBCs.
Figure 5.5 Upregulated of Wnt/β-catenin is associated with loss of AR.

(Ai) Dual Luciferase gene assay shows high basal TCF reporter activity among QNBCs; (Aii) Upregulation of downstream targets of Wnt/β-catenin signaling in QNBC cells; (Bi) Dual Luciferase gene assay shows higher TCF reporter activity after AR KD; (Bii) Upregulation of downstream target genes of Wnt pathway after AR KD; (Ci) Expression of E-cadherin, b-catenin, and target genes Cyclin D1 and CD44, (A) siRNA mediated KD of AR in TNBC cell lines, (B) OE of AR in QNBC cell lines; (Ciii) Immunoblots showing expression of β-catenin in TNBC and QNBC cell lines; (Civ) Immunoblots showing nuclear expression of β-catenin after AR KD in TNBC cell lines; (Cv) Immunoblots showing active β-catenin expression in TNBC and QNBC cells lines.
Figure 5.6 Aberrant localization of β-catenin is associated with QNBC patients.

(A). Representative micrographs showing (i) membrane, (ii) loss of membrane with cytoplasmic but no nuclear and (iii) nuclear expression of β-catenin in QNBC tumors; (B) Bar graph showing membrane and loss of membrane expression of β-catenin in TNBC and QNBC tumors.
Figure 5.7 Loss of membrane β-catenin expression is associated with poor survival among QNBCs. Kaplan Meier’ survival curves for membrane expression (red) and loss of membrane (blue) expression of β-catenin in (A) TNBCs and (B) QNBCs.
6 CONCLUSIONS

Significant advances in breast cancer survival have been achieved in the past decade, which can be attributed to a better understanding of disease development and progression. However, to date, it still cannot be explained why one patient may have a disease relapse and die, whereas another patient, with a seemingly similar tumor, survives. Therefore, the identification of novel breast cancer biomarkers and further evaluation of presently known biomarkers are crucial for accurately predicting and improving survival in breast cancer patients. In this dissertation, I have presented findings that describe an in-depth characterization of biological and clinico-pathological biomarkers in breast cancer subgroups, which could help refine diagnostics and define potential treatment targets for breast cancer.

The study of tumor growth rates is necessary to understand the biology and natural history of malignant diseases (1), and it is widely assumed that the gross tumor growth rate decelerates with increasing tumor mass. Breast cancer tends to show high variability in tumor growth rates. In developed countries, mammography is fast becoming an established part of routine health screening. However, questions related to the optimization of mammography screening, such as how to determine the optimal time intervals between screening and at what age should screening be initiated, remain unresolved. Knowledge of tumor growth rates is, therefore, important for planning and evaluating screening programs (2). Although breast cancer screening, with mammography at the forefront, has been proven to save lives and result in better treatment regimens in the US, breast cancer-related mortality following screening has not been reduced; on the contrary, it is on the rise. Despite its many advantages, a major limitation of mammography is that it yields false-positive results, thus leading to an increased number of breast biopsies, especially in younger patients (Baker et al, 1982). On the other hand, occasionally, cancers fail to
get detected at the initial mammogram screening but are detected during diagnostic mammography; therefore, false-negative mammogram results are amongst the principal reasons for the delayed diagnosis of breast cancer (3-6). Such breast cancers that are detected in the interval after a negative mammographic result are termed interval cancers (7). Although the tumor growth rate has not been implicated as a prognostic variable in clinical practice because of the difficulty in evaluating it in the short time frame of diagnosis and treatment, it has been shown that tumors with a faster growth rate in vivo are more aggressive and, therefore, associated with a poor prognosis (8). This may explain why the impact of missing subtle cancers in screening mammography appears largely to depend on whether the tumor was slow- or fast-growing prior to diagnosis. Therefore, the tumor growth rate may indeed be an important predictor of prognosis.

My dissertation work on developing a prognostic model that can predict in vivo breast tumor growth is uniquely valuable, considering that only a few population-screening programs, so far, have reported data on interval cancers. Because multiple factors control breast cancer growth, our model, Surr-INVIGOR, integrates Ki67, mitotic index, and tumor size, which makes it a robust prediction model of prediagnostic growth rate that can classify breast cancers as slow-growing or fast-growing. Breast radiologists are at the frontline of medicolegal actions due to the importance of breast imaging in the diagnosis of breast cancer and the increased use of mammography as a screening tool (9). Our prognostic model also has the potential to mitigate such medicolegal issues, as it could be used to guide screening procedures and perhaps even establish follow-up intervals in select groups of breast cancer patients. As part of my postdoctoral work, I plan to use multiple and more diverse cohorts to further validate the potential use of this model as a robust indicator of tumor prognosis to help solve medicolegal cases.
Unlike adjuvant chemotherapy, neoadjuvant therapies provide the opportunity to evaluate tumor response to therapies and treatment outcomes. It has been consistently shown that breast cancer patients who achieve pathologic complete response (pCR) have a better disease-free survival rate and overall survival rate (10) than those who do not. Because of its strong correlation with prognosis, pCR has been approved by the Food and Drug Administration as an endpoint for drug approval. Although treatment options and outcomes for breast cancer have rapidly developed over the last years, it is still a challenge for modern medicine to timely diagnose and treat invasive breast cancer. As consistent therapy options are still being developed, it remains difficult for physicians to prognosticate survival rates. Prognosis after a neoadjuvant systemic treatment is of great importance to the patient and the physician in charge to better determine the subsequent course of treatment. The residual cancer burden (RCB) score is a relatively new tool developed in 2007 by Symmans et al. to help physicians arrive at an accurate prognosis based on clinical and histopathological findings (11). In addition to prognosis, evaluation of tumor response to neoadjuvant therapy also guides physicians to choose optimal treatment options. However, identifying breast cancers that are most likely to respond to neoadjuvant therapies is important.

The National Cancer Institute defines precision medicine as “a form of medicine that uses information about a person’s genes, proteins, and environment to prevent, diagnose, and treat disease.” It further states that “In cancer, personalized medicine uses specific information about a person’s tumor to help diagnose, plan treatment, find out how cell treatment is working, or make a prognosis.” Currently, immunohistochemistry-based biomarkers are at the forefront of precision medicine in breast cancer patients because they assist oncologists in prognostication and planning treatment. Notably, the American Association of Clinical Oncology (ASCO) recommends the use of Estrogen receptor (ER), Progesterone receptor (PR), and Human Epidermal receptor2 (HER2)
biomarkers to guide the choice of specific treatment regimens in early-stage breast cancer patients (12). Compared with next-generation tests, these biomarkers are cheaper and easier to use, and their results can be easily interpreted without the need for complex bioinformatics analyses (13). Moreover, for triple-negative breast cancer (TNBC) patients, the development of next-generation testing to guide treatment decisions is at a very naive stage, and no next-generation test currently exists for TNBC, whether early-stage or metastatic.

The College of American Pathologists mandates the evaluation of ER, PR, and HER2 statuses in breast cancer patients; however, routine testing of other breast cancer biomarkers is currently not required by either the ASCO or the National Comprehensive Cancer Network. My dissertation work has investigated other biomarkers that, collectively, might aid better prediction of tumor response to neoadjuvant chemotherapy and better selection of patients for this treatment. This will allow patients who are unlikely to respond to neoadjuvant chemotherapy to be prescribed alternate therapy regimens and will also help patients avoid the side effects of a potentially futile treatment program that would delay a more successful strategy.

For TNBCs, which are characterized by the absence of the three markers ER, PR, and HER2, there is a lack of risk-predictive biomarkers and approved targeted treatment options. TNBC is often characterized by a more aggressive clinical disease course and poorer patient prognosis, which is evident from its higher distant metastasis and recurrence rates within 5 years of diagnosis compared to other breast cancer subtypes (14). Although TNBCs exhibit greater chemosensitivity than non-TNBCs, the vast majority of TNBC patients succumb to relapse following a traditional anthracycline/taxane-based chemotherapy regimen (15-17). Hence, novel alternative prognostic biomarkers and therapeutic targets are urgently needed to manage this subgroup of breast cancer patients. Recently, androgen receptor (AR), expressed in about 10–43%
of TNBCs, has emerged as a therapeutic target. However, the remaining 67–90% of TNBCs lack AR expression, deeming the disease a “quadruple threat”, often referred to as quadruple-negative breast cancer (QNBC). Thus, QNBCs are more immune to novel AR-targeted therapeutic agents compared with AR-expressing TNBCs, and some studies have reported a worse prognosis for QNBC patients than that for TNBC patients (18-22). Therefore, there is a dire need for alternative therapeutic options and risk-predictive biomarkers for QNBCs. Cutting-edge research in TNBC over the past decade, consisting of a plethora of ongoing preclinical and clinical trials, have yielded promising biomarkers and therapeutic targets. However, such research on alternative biomarkers and treatment strategies for QNBC patients remains scarce. Recent evidence suggests the existence of genetic differences between TNBC and QNBC tumors, and QNBC has emerged as a distinct molecular subtype of TNBC that may be considered as a “separate disease entity”. However, crippled by a lack of prognostic markers and therapeutic targets, this novel TNBC subtype has emerged as perhaps the most aggressive but poorly understood form of breast cancer and is in urgent need of alternative predictive and pharmacologically targetable biomarkers. The biomarkers characterized in my dissertation work may serve as novel therapeutic targets for QNBC patients and allow for the precise management of AR-positive and AR-negative TNBCs. My findings add further evidence to support the existing suggestion that AR should be added to the current set of ER, PR, and HER2 markers for breast cancer subtyping.

Recent evidence suggests that QNBCs may be highly proliferative and immunogenic, rendering them ideal candidates for cytotoxic chemotherapy, anthracyclines, and immunotherapeutic intervention (23). However, there remains a stark gap in the knowledge on QNBC tumor biology and etiology. Studies analyzing differential expression of biomarkers between QNBC and TNBC tumors may yield novel biomarkers unique to QNBC. These
discoveries may also benefit TNBC patients of African ancestry, who are at a higher risk of developing QNBC. Robust validation of existing and newly discovered biomarkers for QNBC patients will be pertinent for the successful management of this disease in the clinical setting and may uncover novel therapeutic strategies for QNBC patients. To promote the advancement of QNBC disease research and integration of the findings into routine clinical intervention, it may be useful to build an online platform dedicated to sharing groundbreaking discoveries in QNBC among research institutions and hospitals worldwide. Furthermore, preclinical and clinical testing of the efficacy of novel biomarkers in QNBCs will be essential to validate their clinical utility. Testing the prognostic role of these biomarkers during the early stages of QNBC will enhance risk prediction and therapeutic intervention. Additionally, investigating the upstream and downstream molecules of newly discovered QNBC biomarkers may improve rational design of targeted therapeutic interventions and their efficacy. As the basal-like enriched phenotype of QNBC tumors makes them ideal candidates for cytotoxic chemotherapy, administering targeted therapies along with chemotherapy may elicit a favorable response in QNBC patients. Generating a biomarker profile unique to each QNBC patient through advanced techniques, such as Next-Generation Sequencing, for personalized treatment may also improve QNBC clinical outcomes. The heterogeneous landscape of TNBC has been unraveled through genome-wide microarray expression studies and clustering analyses. Applying these techniques for QNBC subtyping will yield more precise targeted therapies. We also propose to examine the role of epidemiological and non-biological factors in QNBC to gain a holistic understanding of the disease etiology. These studies may uncover novel risk factors of QNBC, which can potentially be modified to prevent onset of the disease or dismal outcomes.
However, this dissertation work holds a few limitations: due to the unique nature of the study cohort in chapter 2, and the lack of similar missed cancer cohorts, the SM-INVIGOR growth index could not be readily validated. In addition, this is a retrospective, single center study and adjuvant treatment regimens were not factored in the analysis. Moreover, due to the unavailability of tissue samples, it was not possible to look at the various immune markers that may play a role in predicting pCR response to neoadjuvant chemotherapy. Potential cohort-to-cohort differences in tissue fixation and lack of centralized AR staining for some of the TNBC cohorts limited the ability to define prognostic significance. Furthermore, the IHC scoring method used in this dissertation is a semi-quantitative assessment that can be influenced by the subjective judgment of the scorers; however, double blind scoring was used in this study.

Collectively, my dissertation work opens up pertinent conversation on how to tackle a challenging new disease that has thus far been subject to very limited attention. My work encourages further dialogue and provides a platform for nurturing fresh ideas aimed at improving the management of this disease. I anticipate my work to generate increased interest among the scientific community toward this breast cancer subtype, encourage further investigation into the etiology and progression of QNBC, and increase its visibility as a mainstream disease to ensure that this unique subgroup of breast cancer patients is not left behind. Detailed analysis of the pathways upstream/downstream of AR, along with transcriptomic, genomic, and epigenetic profiling of samples from TNBC/QNBC patient populations, could move us towards finer signatures that may reclassify AR-positive/negative TNBCs into more clinically actionable subgroups. This is the next avenue I would like to address as part of my future work. Additionally, I am interested in exploring the QNBC landscape/profile that could inform rational development
of biomarkers to aid risk stratification for this unique subtype and guide the choice of preventative treatments that are most likely to prolong patient survival.
References