

ScholarWorks@GSU

illuminating Actionable Biology in Breast Cancer: Novel Predictive and Prognostic Biomarkers

Authors	Bellos, Angela Ogden
Citation	Bellos, Angela Ogden. 2017. "Illuminating Actionable Biology in Breast Cancer: Novel Predictive and Prognostic Biomarkers." Georgia State University. https://doi.org/10.57709/9997490
DOI	https://doi.org/10.57709/9997490
Download date	2026-03-13 17:09:41
Link to Item	https://hdl.handle.net/20.500.14694/1986

ILLUMINATING ACTIONABLE BIOLOGY IN BREAST CANCER: NOVEL PREDICTIVE
AND PROGNOSTIC BIOMARKERS

by

ANGELA O. BELLOS

Under the Direction of Ritu Aneja, PhD

ABSTRACT

Assessing hormone receptors (the estrogen and progesterone receptors) and the human epidermal growth factor receptor 2 (HER2) to guide clinical decision making revolutionized treatment for breast cancer patients. However, in the years since these biomarkers were first incorporated into routine clinical care, only a few others have been validated as clinically useful in guiding adjuvant chemotherapy decisions and are recommended by the American Society of Clinical Oncology (ASCO) for patients with hormone-positive breast cancer. For patients with triple-negative breast cancer (TNBC), which lacks hormone and HER2 receptors, not any of these biomarkers are recommended by ASCO due to insufficient evidence that they meaningfully improve clinical outcomes. Breast cancer is the second-leading cause of cancer-related death

among women in the US, indicating an unmet need to improve treatments, which can be accomplished in part by identifying and validating novel predictive and prognostic biomarkers that yield actionable information about the clinical course of breast cancers, especially TNBCs. A major obstacle to improving outcomes for breast cancer patients is intratumor heterogeneity (ITH), which can be extensive in breast cancer and drives treatment resistance and relapse. I envision that assaying drivers of ITH can inform clinicians about which breast tumors may be intrinsically more aggressive and carry a greater risk of breast cancer-related morbidity and mortality. My research, presented here, primarily focuses on testing the impact of drivers of ITH (namely, centrosome amplification [CA], the clustering protein KIFC1, and mitotic propensity and its drivers) on clinical outcomes in breast cancer in multivariable models as well as the correlates of in vitro efficacy of centrosome declustering drugs (which can selectively eliminate cancer cells with CA). Collectively, these studies reveal gene signatures and immunohistochemical biomarkers that are independent predictors of aggressive breast cancer course and rational strategies to optimize targeted therapy to combat cancer cells exhibiting CA, thereby contributing to the literature on the development of precision medicine for breast cancer patients, including TNBC patients.

INDEX WORDS: Chromosomal instability, proliferation, Ki67, HSET, HER3, EGFR, RARA, racial disparity

ILLUMINATING ACTIONABLE BIOLOGY IN BREAST CANCER: NOVEL PREDICTIVE
AND PROGNOSTIC BIOMARKERS

by

ANGELA O. BELLOS

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

2017

Copyright by
Angela Ogden Bellos
2017

ILLUMINATING ACTIONABLE BIOLOGY IN BREAST CANCER: NOVEL PREDICTIVE
AND PROGNOSTIC BIOMARKERS

by

ANGELA O. BELLOS

Committee Chair: Ritu Aneja

Committee: Jaideep Chaudhary

Zhi-Ren Liu

Adam Marcus

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2017

DEDICATION

This dissertation is fondly dedicated to my family, without whose laughter, affection, and support this work would not have been possible, and in remembrance of my loving father, whose courage and faith in his hard-fought battle with cancer taught me what it means to be a person of character and to cherish the moments spent with those I love.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my advisor, Dr. Ritu Aneja, for the invaluable mentorship and opportunities she has provided to me in the course of my studies. I also want to thank her and senior research scientist Dr. Padmashree C. G. Rida for their enduring professional and personal guidance and support; my dissertation committee members, Drs. Jaideep Chaudhary, Adam Marcus, and Zhi-Ren Liu, and my qualifying committee members, Drs. Julia Hilliard and Andrew Gewirtz, for their constructive feedback and encouragement; and my current and former colleagues in the lab, especially Karuna, Nikita, Shristi, Sergey, and Drs. Sushma Gundala, Vaishali Pannu, Roopali Saxena, and Chunhua Yang, for their assistance and friendship.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF TABLES	xii
LIST OF FIGURES	xv
1 INTRODUCTION¹	1
1.1 Biomarkers in breast cancer: an overview	1
1.2 Triple-negative breast cancer	1
1.3 Centrosome amplification	3
1.4 Centrosome clustering and chromosomal instability	5
<i>1.4.1 KIFC1</i>	<i>8</i>
<i>1.4.2 Racial disparities in breast cancer</i>	<i>10</i>
1.5 Centrosome declustering	13
<i>1.5.1 Griseofulvin</i>	<i>14</i>
<i>1.5.2 Noscapinoids</i>	<i>15</i>
<i>1.5.3 Phenanthrene-derived PARP inhibitors</i>	<i>15</i>
<i>1.5.4 High-grade spindle multipolarity: its power and payoff</i>	<i>16</i>
1.6 Mitotic Propensity	18
1.7 Drivers of Proliferation: EGFR and HER3	18
1.8 Inhibitors of Proliferation: RARA	20
1.9 References	21

2	Prognostic value of CA20, a score based on centrosome amplification-associated genes, in breast tumors²	1
2.1	Abstract	1
2.2	Introduction	1
2.3	Results	3
2.3.1	<i>METABRIC dataset</i>	3
2.3.2	<i>TCGA dataset</i>	5
2.4	Discussion.....	7
2.5	Methods	10
2.5.1	<i>Dataset details and power analyses</i>	10
2.5.2	<i>Gene signatures and microarray probe selection</i>	12
2.5.3	<i>Survival analyses</i>	13
2.5.4	<i>Grade-wise comparison of average CA20 and CIN25 between TNBCs and non-TNBCs</i> 14	
2.5.5	<i>Gene Set Enrichment Analyses</i>	15
2.6	References	15
3	Multi-institutional study of nuclear KIFC1 as a biomarker of poor prognosis in African American women with triple-negative breast cancer³	1
3.1	Abstract	1
3.2	Introduction	1

3.3	Results	2
3.4	Discussion.....	4
3.5	Materials and Methods	8
3.5.1	<i>Specimen procurement and datasets</i>	<i>8</i>
3.5.2	<i>Immunohistochemistry and specimen scoring</i>	<i>8</i>
3.5.3	<i>Statistical analysis of clinical data</i>	<i>9</i>
3.5.4	<i>Statistical analysis of experimental data</i>	<i>10</i>
3.5.5	<i>Cell culture</i>	<i>10</i>
3.5.6	<i>siRNA.....</i>	<i>11</i>
3.5.7	<i>Western blotting and antibodies</i>	<i>11</i>
3.5.8	<i>Cell proliferation assay</i>	<i>12</i>
3.5.9	<i>Wound healing assay</i>	<i>12</i>
3.6	References	12
4	Quantitative multi-parametric evaluation of centrosome declustering drugs: centrosome amplification, mitotic phenotype, cell cycle and death⁴.....	1
4.1	Abstract	1
4.2	Results	3
4.2.1	<i>Characterization of MA induced by centrosome declustering drugs.....</i>	<i>3</i>
4.2.2	<i>Declustering drugs induce CA in cancer cell lines.....</i>	<i>4</i>

4.2.3	<i>Effect of putative declustering agents on centrosome declustering and spindle MP</i>	5
4.2.4	<i>Cross talk between drug-induced spindle MP, declustering, and drug efficacy</i>	6
4.3	Discussion.....	9
4.4	Materials and Methods	12
4.4.1	<i>Cell lines, cell culture, and drugs</i>	12
4.4.2	<i>Immunofluorescence microscopy and western blotting</i>	13
4.4.3	<i>Cell cycle analysis</i>	14
4.4.4	<i>Statistical analysis</i>	14
4.5	References	14
5	Multi-institutional study of triple negative breast cancer stratification by a novel biomarker of cell cycling kinetics.....	1
5.1	Abstract	1
5.2	Introduction	2
5.3	Materials and Methods	3
5.3.1	<i>Description of datasets</i>	3
5.3.2	<i>Formulation of KAMS</i>	4
5.3.3	<i>Statistical analyses</i>	5
5.3.4	<i>RNA sequencing and analysis</i>	6

5.4	Results	8
5.4.1	<i>KI and MI bear a highly variable relationship in breast tumors</i>	8
5.4.2	<i>KAMS is a superior predictor of treatment response than both Ki67 and mitotic indices in three unique cohorts.....</i>	9
5.4.3	<i>KAMS groups are characterized by different levels of immunohistochemistry-based biomarkers</i>	10
5.4.4	<i>KAMS groups are characterized by different molecular pathways</i>	11
5.5	Discussion.....	11
5.6	References	15
6	High HER3-EGFR Score Predicts Aggressive Disease Course in Triple-Negative Breast Cancer	1
6.1	Abstract	1
6.2	Introduction	2
6.3	Materials and methods.....	3
6.3.1	<i>Datasets.....</i>	3
6.3.2	<i>Sample scoring</i>	4
6.3.3	<i>Survival analyses.....</i>	4
6.3.4	<i>Biomarker comparisons.....</i>	5
6.3.5	<i>RNA sequencing.....</i>	5
6.3.6	<i>RNA-seq data analyses.....</i>	7

6.3.7	<i>Oncomine analyses</i>	7
6.4	Results	8
6.5	Discussion	11
6.6	References	13
7	Retinoic Acid Receptor Alpha Predicts Good Prognosis in Triple-Negative Breast Cancer	1
7.1	Abstract	1
7.2	Introduction	2
7.3	Materials and Methods	3
7.4	Results and Discussion	4
7.5	References	5
8	CONCLUSIONS	1
	Why are these findings clinically important?	1
	How do these findings align with the era of precision medicine?	4
	How do these findings relate to the next-generation of cancer hallmarks?	11

LIST OF TABLES

Table 2.1 CA20 genes	1
<i>Table 2.2 Descriptive statistics for the METABRIC breast discovery and validation sets.....</i>	<i>1</i>
<i>Table 2.3 Descriptive statistics for the TCGA breast dataset.....</i>	<i>3</i>
Table 2.4 Final multivariable Cox proportional-hazards models of breast cancer-specific survival including CA20 and CIN25 (based on average scores) in full models using METABRIC data	4
Table 2.5 Final multivariable Cox proportional-hazards models of breast cancer-specific survival including CA20 and CIN25 (based on optimal thresholds) in full models using METABRIC data	5
Table 2.6 Cox models of overall survival including CA20 and CIN25 (based on average or optimal thresholds) and AJCC stage in full models using TCGA data	6
<i>Table 2.7 Final multivariable Cox proportional-hazards models of overall survival including CA20 and CIN25 (based on average or optimal thresholds), AJCC stage, and age at diagnosis in full models using TCGA data</i>	<i>7</i>
Table 3.1 Relationships between nuclear KIFC1 weighted index and patient and clinicopathological factors	15
<i>Table 3.2 Impact of Nuclear KIFC1 weighted index on survival outcomes</i>	<i>16</i>
Table 3.3 Descriptive statistics for continuous variables by race.	17
Table 3.4 Descriptive statistics for categorical variables by race.	18
<i>Table 5.1 Patient and clinicopathologic characteristics of breast carcinoma patients in the Northside and Emory cohorts.</i>	<i>18</i>
Table 5.2 Patient and clinicopathologic characteristics of triple-negative breast cancer patients in the Nottingham, Stavanger, and Nigeria cohorts	20

Table 5.3 Patient and clinicopathologic characteristics of the 99 breast carcinoma samples from the Northside cohort used to determine the average number of mitoses for mitotic index categories as used in the Nottingham Grading System.....	22
Table 5.4 Predictive value of KAMS and Ki67 index in simple Cox models of breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort)	24
Table 5.5 Multivariable Cox models of breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort) that include KAMS or KI in full models.	25
Table 5.6 Genes overexpressed in the high KAMS group at $q < 0.05$	26
Table 5.7 Genes underexpressed in the high KAMS group at $q < 0.05$	29
Table 5.8 Differentially regulated pathways according to KAMS	31
Table 6.1 Clinicopathologic data for Nottingham, Norway, and Emory datasets	16
Table 6.2 Risk of worse breast cancer-specific or distant metastasis-free survival conferred by high levels of membranous HER3, EGFR, or HER3-EGFR H-scores in simple Cox models stratified by hospital.....	17
Table 6.3 Risk of worse breast cancer-specific or distant metastasis-free survival conferred by high levels of membranous HER3, EGFR, or HER3-EGFR H-scores, as defined by respective medians, in multivariable Cox models including each respective biomarker, adjusted for AJCC stage and age at diagnosis, and stratified by hospital	18
Table 6.4 Immunohistochemical biomarkers differentially expressed between HER3-EGFR-low and high groups at $q < 0.25$	19
<i>Table 7.1 Descriptive statistics for continuous and categorical patient and clinicopathologic variables and nRARa scores</i>	<i>6</i>

Table 7.2 Final multivariable Cox regression model of the impact of nRAR α on breast cancer-specific survival	7
Table 7.3 Final multivariable Cox regression model of the impact of nRAR α on distant metastasis-free survival.....	8

LIST OF FIGURES

Figure 1.1 Numerical and structural centrosome amplification in breast cancer.	27
Figure 1.2 Global centrosome clustering mechanisms	28
Figure 2.1 Plots of Kaplan-Meier product limit estimates of breast cancer-specific survival of patients in METABRIC discovery and validation sets	8
<i>Figure 2.2 Grade-wise comparisons of (A) average CA20 score and (B) average CIN25 score in non-TNBCs vs. TNBCs in the METABRIC dataset</i>	<i>9</i>
Figure 2.3 Plots of Kaplan-Meier product limit estimates of overall survival of patients in TCGA dataset	10
Figure 3.1 Representative images of nKIFC1 staining	19
Figure 3.2 Impact of KIFC1 knockdown on proliferation and migration of White (W) (HCC1143 and MDA-MB-231) and African American (AA) (HCC1906 and MDA-MB-468) triple-negative breast cancer (TNBC) cell lines.....	20
Figure 3.3 Kaplan-Meier curves of overall survival based on nKIFC1 weighted index (stratified by 1 standard deviation above the mean).....	21
Figure 3.4 Kaplan-Meier curves of progression-free survival based on nKIFC1 weighted index (stratified by 1 standard deviation above the mean)	22
<i>Figure 3.5 Kaplan-Meier curves of distant metastasis-free survival based on nKIFC1 weighted index (stratified by 1 standard deviation above the mean).....</i>	<i>23</i>
Figure 4.1 Mitotic arrest (MA) phenotypes observed upon treatment with putative centrosome declustering drugs.	18
Figure 4.2 Centrosome declustering drug-induced changes in expression levels of markers of centrosome amplification.....	21

Figure 4.3 Average CA observed over 24 h and its relationship with peak subG1 for each drug treatment regimen	22
Figure 4.4 Representative confocal micrographs depicting centrosome amplification in controls and drug-treated cancer cells	23
Figure 4.5 Phenotypes induced by declustering drugs at different concentrations.....	24
Figure 4.6 Correlates of peak subG1 percent within cell lines	25
Figure 4.7 Diversity of phenotypes produced by putative centrosome declustering drugs	26
Figure 5.1 Analysis of grade-wise relationship between KI and MI	32
<i>Figure 5.2 Stratification by KAMS, Ki67 index, and mitotic index according to breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort)</i>	<i>33</i>
Figure 5.3 Comparison of significantly different immunohistochemistry-based biomarkers between high- and low-KAMS TNBCs.....	34
Figure 7.1 Normalized expression levels of nRAR α in TNBCs vs. non-TNBCs in the METABRIC dataset.....	9
Figure 7.2 Kaplan-Meier survival plots of breast cancer specific-survival stratified by nRAR α score.	10
Figure 7.3 Kaplan-Meier survival plots of distant metastasis-free survival stratified by nRAR α score.	11

1 INTRODUCTION¹

1.1 Biomarkers in breast cancer: an overview

Very few clinically useful breast cancer biomarkers – that is, those that offer actionable information that may change a breast cancer patient’s disease course – are currently recommended by ASCO. The society recommends routine testing for the estrogen and progesterone hormone receptors (ER and PR, respectively) along with the ErbB family member human epidermal growth factor receptor 2 (HER2), which provide both prognostic and predictive information that revolutionized breast cancer treatment. Patients with hormone receptor-positive breast cancers (ER/PR+, HER2-) generally have a favorable prognosis and respond to treatment with endocrine therapy, such as hormone receptor antagonists and aromatase inhibitors. HER2+ tumors confer a less favorable prognosis, yet patients may respond to treatment with HER2 antagonists as well as cytotoxic chemotherapy. Patients whose tumors are negative for all three of these biomarkers (TNBCs) also have a less favorable prognosis than hormone receptor-positive breast cancers; however, 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. For patients with hormone receptor-positive early-stage invasive breast cancer, several biomarkers and genomic tests are available that may predict the risk of recurrence and guide the selection of adjuvant therapy. These biomarkers include urokinase-type plasminogen activator, plasminogen activator inhibitor type 1, Oncotype DX, PAM50, EndoPredict, and Breast Cancer Index [1]. For HER2+ and TNBC patients, these tests are contraindicated because there is insufficient evidence to suggest that they are clinically

¹Parts of this chapter have been published verbatim in *Cell Death and Differentiation* 2012 Aug;19(8):1255-67 as “Let’s huddle to prevent a muddle: centrosome declustering as an attractive anticancer strategy” and *Endocrine Related Cancer* 2017 (in press) as “Centrosome Amplification: Suspect in Breast Cancer and Racial Disparities.”

useful. Thus, there is an unmet need for biomarkers of clinical utility for these high-risk patients, preferably those that can offer both prognostic and predictive information for TNBC patients, whose tumors remain defined by what biomarkers and drug targets they lack.

1.2 Triple-negative breast cancer

TNBC is a significant public health concern owing to the high prevalence and grave nature of the illness. It afflicts nearly a half million women in the US – that is, about one fifth of the ~2.8 million women in the country with breast cancer [2, 3]. A disproportionate fraction of pre-menopausal African American women develop the disease, causing TNBC to be an important component of racial health disparity in this country as well [4]. At the time of diagnosis, which often occurs between mammograms, TNBCs tend to be of a more advanced histologic grade and larger size than non-TNBCs [5]. TNBC is also characterized by a rapid proliferation rate and high propensity to metastasize, typically to visceral organs [6]. In addition, the tripe-negative phenotype carries the greatest risk of brain metastasis, and brain metastases that are triple-negative carry the worst prognosis [7]. Women with TNBC exhibit significantly worse 5-year survival rates than those with non-TNBC regardless of the tumor stage at diagnosis [8]. This worse prognosis stems not only from the strikingly aggressive behavior of TNBCs but also from the poor efficacy of currently approved TNBC treatment regimens as compared with treatment regimens for other breast cancer subtypes. The mainstay of TNBC treatment is chemotherapy with non-specific cytotoxins, which do not eliminate the tumor in almost 80% of patients [6]. These treatment-resistant patients have a dismal prognosis: almost half of them die within 5 years. Furthermore, TNBC patients whose tumors metastasize only survive about 1 year. As a result, a diagnosis of metastatic TNBC is essentially a death sentence. Given the

unreliability of non-specific 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 was immensely effective for other breast cancer subtypes, which are treated with targeted drugs like Tamoxifen, Arimidex, or Herceptin alone or in combination with cytotoxins.

Approximately half a million women in the US have TNBC, or ~15-20% of the nearly 3 million women living with breast cancer [9-11]. AA women are overrepresented among TNBC patients, as they are ~3 times as likely to develop TNBC as white women [4]. Among TNBC patients, AAs have ~2 times the risk of mortality compared with whites after adjustment for age, grade, stage, and poverty index [12]. These data suggest that AAs have inherently more aggressive TNBCs, which contributes to racial health disparity in breast cancer [12]. Because TNBC is a disease defined by the targets it lacks, prognostication and treatment are particularly challenging. Assays that are indicated for other types of breast cancer perform poorly in risk-stratifying TNBCs. For instance, MammaPrint® tends to uniformly classify TNBCs as high risk [13]. While this characterization may be accurate relative to other breast cancer subtypes, it fails to capture the variability in outcomes among TNBC patients, a substantial proportion of whom experience favorable outcomes. Specifically, ~10-30% of patients achieve a pathologic complete response to chemotherapy depending on the regimen, nearly all of whom survive at least five years [6]. By contrast, the five-year survival probability for TNBC patients with residual disease is ~55% [6]. 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 survival gap between AAs and whites with TNBC implies that

AAs may not receive sufficiently aggressive treatment. Therefore, a critical barrier to eliminating racial health disparity in breast cancer is a lack of knowledge about biomarkers in AA TNBCs that can meaningfully guide therapeutic interventions. Biomarkers that could risk-stratify AA TNBCs would constitute revolutionary progress in TNBC and might mitigate racial health disparity in breast cancer. Breast tumors, especially TNBCs, are often genetically unstable and thus exhibit genotypic and phenotypic ITH, which underlies treatment resistance and relapse and which is more severe in AA TNBCs [14]. Chromosomal instability (CIN) is a driver of ITH that describes a dynamic state wherein the cell experiences an increased rate of gain or loss of whole or parts of chromosomes, defined as numerical and structural CIN, respectively [15]. A major cause of CIN is the presence of supernumerary centrosomes during cell division. Therefore, it stands to reason that biomarkers of increased centrosome number (called numerical CA) and enhanced cell cycle kinetics could reveal which tumors are chromosomally unstable and carry a poor prognosis, in part due to increased risk of treatment relapse, which can inform clinical decision making.

1.3 Centrosome amplification

At least 75% of malignant breast tumors exhibit CA, although the proportion of intratumoral cells with CA varies widely, from 1-100% [16]. CA is the presence of excessively voluminous or numerous centrosomes, referred to as structural and numerical CA, respectively (pictured in **Figure 1**). Structural CA arises from excessive recruitment of PCM, a highly ordered yet dynamic matrix of hundreds of proteins and nucleic acids. PCM size is regulated by centrioles, free cytoplasmic $\alpha\beta$ -tubulin, centrobins, kinases like PLK1 and CHK1, and several coiled-coil proteins like pericentrin and CPAP [17-19]. The mechanisms undergirding structural

CA in cancer are still poorly defined, although several stimulators of PCM assembly are overexpressed in cancer (e.g., PLK1 [20] and CHK1 [21]), cancer cells often harbor supernumerary centrioles (which could then recruit excessive PCM), and the PCM expands following DNA damage [22]. Centrosomes may also appear to have an increased size but actually consist of multiple centrosomes clustered together, called “megacentrosomes” [23] or “speckling” (>5 centrosomes clustered together) [24], with individual centrosomes only readily discerned by imaging of centrioles. Nearly half of breast tumors exhibit centrosomes with abnormal morphology, which may be associated with worse recurrence-free survival (p=0.062 in Kaplan-Meier analysis) [24].

Centrosomes in breast tumor cells frequently exhibit augmented volume compared with cells from normal breast tissue [24-27]. Furthermore, centrosomes from unstable aneuploid breast tumors have ~75% increased volume compared with stable aneuploid breast tumors [26], and centrosomal volume correlates with CIN in invasive breast tumors [28]. Centrosomes from breast tumor cells of the highly aggressive triple-negative subtype have ~60% greater volume than centrosomes from grade-matched non-triple-negative tumor cells [29]. These data suggest that larger centrosomes are associated with more aggressive phenotypes in breast cancer. Structural CA is found in ductal carcinomas in situ, suggesting it may be an early event in breast tumorigenesis [28]. Numerical CA (>1 centrosome before S-phase and >2 centrosomes after S-phase) can arise from various abnormal processes, such as cytokinesis failure, templated overduplication, *de novo* formation, or cell-cell fusion, such as can be induced by all known human oncogenic viruses [30, 31]. Because centrosome number correlates with ploidy in breast cancer, cell doubling events are a likely cause of numerical CA in breast cancer, which has been estimated as explaining at least 15% of CA (i.e., 15% of cells with ploidy >3 have CA, although

because chromosomes may subsequently be lost, this statistic may underrepresent the true prevalence of CA caused by doubling events) [24]. The molecules and pathways responsible for numerical CA have been comprehensively reviewed elsewhere [30, 32-38]. Average centrosome number/cell correlates with tumor grade, Ki67 index, and CIN in breast cancer and is highest in the aggressive TNBC subtype (nearly two-thirds of which exhibit >2 centrosomes/cell on average), suggesting that numerical CA is associated with aggressive breast cancer features [24]. Induction of CA by PLK4 overexpression in MCF10A breast epithelial cells results in decreased CD24 and increased CD44 expression, suggesting that CA drives cellular dedifferentiation rather than merely co-occurring with it [24]. In a study of $n=362$ predominantly white breast cancer patients with at least 5 years follow up, it was found that numerical CA confers worse overall, breast cancer-specific, and recurrence-free survival, although not independent of stage and hormone receptor status [24]. Clustering of centrosomes, which circumvents spindle multipolarity that jeopardizes cell survival, occurs in more than half of breast tumors and is associated with significantly worse overall and recurrence-free survival; however, the impact of clustering on these survival outcomes after adjusting for potential confounders is unclear [24]. In sum, CA is associated with more aggressive breast cancer features and may adversely impact survival, although further study to substantiate this paradigm in multivariable models is required. Testing the prognostic value of a gene expression-based CA score in multivariable models is the primary goal of the study in the first manuscript in this dissertation (Chapter 2).

1.4 Centrosome clustering and chromosomal instability

Much in vitro evidence suggests that CA actively drives tumorigenesis rather than merely being a consequence of it by promoting phenotypes such as CIN which can promote aggressive disease features [39], through centrosome clustering. Whether CIN promotes or inhibits

tumorigenesis depends on the type of cell (some being inherently more tolerant of DNA damage and aneuploidy than others), its genetic background (e.g., pre-existing p53 mutations), the specific karyotype that is acquired (e.g., gain vs. loss of an oncogene), and the rate of CIN (with moderate levels tolerated better than extreme levels). Overexpression of Aurora Kinase A (AURKA), which causes CA, results in CIN that precedes tumor formation in mouse mammary epithelium, the incidence of which is increased by a p53^{+/-} background [40]. Single-cell genome sequencing has revealed that major aneuploid rearrangements (which can be caused by CA) occur early in breast tumor evolution, followed by incremental clonal diversification over time [41, 42]. Furthermore, TNBC cells, which exhibit more extensive CA than non-TNBC cells [25], have a ~9-13-fold higher mutation rate than ER⁺ tumor cells and normal breast cells [41]. CA fuels tumor evolution by constructing a multipolar spindle in prophase, since supernumerary centrosomes are not initially clustered together [43]. This abnormal spindle geometry predisposes kinetochores to attach to microtubules emanating from two spindle poles, which is termed merotelly. Merotelic attachments can arise from syntelic attachments (where sister kinetochores attach to microtubules from the same spindle pole), which are converted to merotelic attachments to satisfy the spindle assembly checkpoint, or they can be formed from the outset. The quantity of microtubules involved in merotelic attachments dictates the behavior of the merotelically attached chromosome [44]. If few microtubules are oriented to the “wrong” spindle pole (pauci-merotelly), chromosome segregation proceeds without apparent impairment. If a roughly equal number of microtubules are attached to the right and wrong spindle poles (equi-merotelly), the chromosome lags during anaphase due to strong, opposite poleward forces, but ultimately it tends to segregate to the right cell as a micronucleus. Nevertheless, lagging chromosomes can become trapped in and damaged by the cleavage furrow, resulting in breaks

and unbalanced translocations [45]. Trapped chromosomes are either removed from the cleavage site or the cleavage furrow regresses, resulting in polyploidization, which may itself be tumorigenic [46]. Furthermore, even if the lagging chromosome segregates to the right cell as a micronucleus, micronuclear DNA replicates aberrantly and asynchronously with primary nuclear DNA, resulting in rapid accrual of complex, clustered chromosome rearrangements [47-50]. Chromoanasythesis can occur because micronuclear replication forks are prone to stalling and collapse, resulting in template switching and microhomology-mediated break-induced replication, respectively, which promote intricate, local chromosome rearrangement. Chromothripsis can occur when the cell enters mitosis despite the fact that micronuclear chromosomes are still slowly replicating, causing the micronuclear chromosomes to prematurely condense and shatter. Rearrangement then occurs when the fragments are stitched back together. In addition, under-replication of micronuclear DNA results in copy-number asymmetry. Thus, equi-merotely can result in “all-at-once,” catastrophic mutagenesis, which permits rapid karyotype evolution and might be an important cause of rapidly-progressing, interval breast cancers such as TNBCs [51]. If many microtubules are attached to the wrong pole (multi-merotely) [44], the force pulling the chromosome to the wrong pole is strong, causing it to missegregate without lagging, which results in aneuploidy.

Although the cell can correct merotelic attachments by converting them to amphitelic ones via Aurora B and MPS1 kinases [52], supernumerary centrosomes seem to induce so many merotelic attachments that cellular correction mechanisms are overwhelmed, and persistence of these errors into anaphase results in missegregation [43]. When centrosomes are clustered into two polar groups, the cell is able to complete bipolar mitosis, but chromosomes missegregate

more frequently than in cells with only two centrosomes. Accordingly, centrosome number and size correlate with CIN and aneuploidy in breast tumors [26]. Massive genomic alterations occur in punctuated bursts early in breast tumorigenesis (e.g., due to chromothripsis or chromosome missegregation), followed by the gradual accumulation of point mutations over time (e.g., due to defective DNA repair and replication) [41]. Thus, it stands to reason that CA is an early event in tumorigenesis and may subsequently be suppressed to inhibit further large-scale karyotypic changes that may be tumor-inhibiting. Indeed, CA is found in pre-invasive breast carcinomas [53, 54], and cells spontaneously lose supernumerary centrosomes over time in culture and become chromosomally more stable [43, 55, 56]. However, other studies have found that CA increases from breast hyperplasia to tumor [57, 58], so perhaps, at least in some cases, CA is not eliminated as the tumor evolves but rather its potentially deleterious effects (e.g., severe CIN, slowed proliferation) are suppressed. It has been demonstrated that centrosomes can be excluded from the mitotic spindle [59], so they can potentially be borne as passengers without driving CIN. A specific tumor's evolutionary trajectory may allow it to evolve to tolerate CA without sacrificing its malignant karyotype, whereas due to the mostly stochastic nature of mutations, other tumors may not acquire such mechanisms and thus must suppress CA to avoid prolonged, severe CIN. Thus, CA can drive tumor evolution and thus promote intratumor heterogeneity, which drives chemoresistance and disease relapse.

1.4.1 KIFC1

In order to promote intratumor heterogeneity, cancer cells must cluster supernumerary centrosomes. The top centrosome clustering protein in a fly screen was the Kinesin-14 family member, KIFC1 (also known as HSET), a minus end-directed microtubule motor [60]. It has a

critical role in centrosome clustering in some cell lines. Human KIFC1 localizes between microtubules within the mitotic spindle [61, 62], as depicted in **Figure 2**, which exhibits various of the molecular players in centrosome clustering. KIFC1 inhibition has no significant impact on bipolar anaphase or cell viability in human BJ fibroblasts, which exhibit virtually no CA (~1%) [62], as well as mouse NIH-3T3 fibroblast and human MCF-7 breast cancer cells, which exhibit only “low-level” CA [60]. However, in cells with a higher incidence of CA, such as isogenic tetraploids of the aforementioned cell lines or MDA-MB-231 cells, bipolar anaphase is inhibited and viability compromised by KIFC1 knockdown [63]. These data suggest that the presence of a “normonumerary” centrosome complement partly masks the requirement for KIFC1 in spindle organization. Although KIFC1 appears to serve no vital function when centrosome number is normal, the minus-end-directed motor was originally identified in embryonic mouse brain; consequently, the possibility that the protein may serve a non-redundant function during embryogenesis cannot be excluded [64]. In certain transformed cell lines, KIFC1 and NuMA play overlapping and therefore redundant cellular functions, in which case KIFC1 may be non-essential for centrosome coalescence [60, 65]. Differential dependence on KIFC1 may indicate that various cell types have evolved distinct clustering mechanisms. Owing to the seemingly nonessential role of KIFC1 in nontransformed adult human cells, KIFC1 offers immense promise as a novel chemotherapeutic target for “centrosome-rich” cancers, including those of the breast, prostate, bladder, colon, and brain. It was recently found that KIFC1, in particular, nuclear KIFC1 expression, confers worse prognosis in breast cancer. High nuclear KIFC1 expression (defined as H-scores above the median) was associated with significantly worse overall and progression-free survival after adjusting for age at diagnosis, Nottingham grade, and hormone receptor and HER2 statuses [66]. Average nuclear KIFC1 expression was also higher in TNBCs

than non-TNBCs. Thus, nuclear KIFC1 is associated with aggressive disease features in breast cancer.

1.4.2 Racial disparities in breast cancer

Given the connection between KIFC1 and aggressive breast cancer, KIFC1 may serve as a useful prognostic biomarker in AA breast cancer patients, who suffer a more aggressive disease course than white patients. Unfortunately, the AA-white breast cancer outcome gap appears to be widening despite the multifarious advancements in breast cancer medicine in recent years. According to SEER data from 2012, the age- and delay-adjusted incidences of invasive breast cancer among white and AA women were 135.0 and 135.2 (per 100,000), respectively, and the age-adjusted mortality rates were 20.7% and 29.4%, respectively [67]. From 1975-2011, AA women never had a greater age- and delay-adjusted incidence of invasive breast cancer than white women, so the year 2012 represents a regrettable inflection point for the health of AA women. Perhaps more disturbingly, between 1975-2012, the age- and delay-adjusted increase in the incidence of breast cancer was greater among AA women than white women (44.5% vs. 25.8%, respectively) and the decrease in age-adjusted mortality was less among AA women than white women (0.002% and 34.8%, respectively). These data suggest that improvements in breast cancer screening, diagnosis, and treatment disproportionately benefit white women, a trend that has been attributed to the “perfect storm” of aggressive tumor biology among AA women colliding with healthcare inequality [68]. Rates of hormone receptor-positive HER2-negative breast cancer, the least aggressive subtype, are highest among white women, whereas rates of TNBC, distant-stage disease, and poorly/undifferentiated grade, which are all associated with worse survival, are highest among AA women [69]. Although national efforts have been made to

try to mitigate differences in access to healthcare, clinical strategies addressing inherently more aggressive AA tumor biology are presently inadequate. Prognostic tools and treatments tailored to AA women are urgently needed to attenuate racial breast cancer disparity. We believe that assessment of centrosomal profiles and targeting cells with CA represent promising prognostic and therapeutic strategies with exceptional potential for AA women with breast cancer.

Intriguingly, the AA/white survival gap did not manifest until the mid-1980's, when use of the estrogen receptor antagonist Tamoxifen became widespread [70]. White women are more likely than AA women tend to have hormone receptor-positive breast cancers, affording them disproportionate opportunity to take advantage of this revolutionary and generally life-saving chemotherapeutic. AA women are ~2-3 times more likely than white women to develop TNBC [71], which constitutes ~20-46% of AA cases and for which their lifetime risk is ~2% [71, 72]. This breast cancer subtype is characterized by mutations in *TP53*, which are present in more than half of all TNBCs [73]. No targeted therapy is FDA-approved to treat this aggressive breast cancer subtype. Instead, cytotoxic chemotherapy is the standard of care, although it achieves a pathologic complete response in only ~20% of patients [6]. This aggressive breast cancer subtype is usually diagnosed at a more advanced grade and stage, is more likely to distantly metastasize, and carries a higher five-year risk of mortality relative to non-TNBCs [5]. Although the molecular mechanisms underpinning the greater risk of developing TNBC among AA women, sub-Saharan African women have an even greater risk of developing TNBC than AA women, strongly implicating ancestry [70].

The association of CA with TNBC and TNBC with AA race suggest that, transitively, CA may be associated with AA race. Hence, differences in centrosome homeostasis may account for some proportion of racial disparity in breast cancer, a novel and potentially impactful idea that merits investigation. A study of breast tumor gene expression data from The Cancer Genome Atlas provides circumstantial evidence that CA may differ based on race. Specifically, it was discovered that PLK1 and Aurora C signaling pathways are upregulated in age- and stage-matched breast tumors from AAs compared with EAs [74]. Since PLK1 and Aurora C drive CA [75-77], AA breast tumors may exhibit increased CA relative to white breast tumors. Whether AA breast tumors are enriched in cells with CA is a question of immense clinical importance because centrosome clustering inhibitors are available, some of which are FDA-approved drugs (e.g., griseofulvin, an antifungal). If AA tumors do exhibit disproportionate CA relative to white breast tumors, then therapeutically targeting this organelle-level difference could help to attenuate the race-based survival gap. Undoubtedly, a range of other molecular and cellular aberrations are involved in racial health disparity; therefore, targeting cells with supernumerary centrosomes alone is unlikely to entirely close the gap. That being said, targeting centrosome clustering mechanisms is nonetheless a highly appealing therapeutic strategy because these mechanisms generally are expendable to normal cells, as they lack CA [38]. Another vital and interesting question is whether CA compels breast tumors to evolve into TNBCs, or conversely whether the TNBC phenotype engenders CA. Finally, TNBC is a highly heterogeneous disease that comprises different molecular subtypes [78, 79]. It is unknown whether differences in CA exist between subtypes and, consequently, whether centrosome clustering inhibitors could prove superiorly efficacious in certain TNBC subtypes. AA TNBCs tend to belong to the basal-like 1 subtype, unlike white TNBCs, which tend to belong to the luminal androgen receptor subtype

[80]. Thus, it is reasonable to conjecture that the basal-like 1 subtype evinces greater CA than the luminal androgen receptor subtype. If CA is more extensive in AA TNBCs, it stands to reason that their cells may rely more staunchly on clustering molecules, such as KIFC1, for survival; thus, clustering molecules may be particularly valuable prognostic biomarkers and predictors of response to declustering drugs in AA TNBC patients. Evaluation of KIFC1 as a prognostic biomarker in racially distinct TNBC patients is the goal of the study described in the second manuscript in this dissertation (Chapter 3).

1.5 Centrosome declustering

Too many centrosomes may prove bane or boon to cancer cells depending on whether the cell is able to cluster them neatly at opposite poles. Clustering may confer survival advantages and promote malignancy by predisposing the cell to CIN via merotelic microtubule–kinetochore attachment and genome missegregation [43, 81]. When “low-grade” (i.e., survivable) missegregation results in the loss of a gene that promotes faithful chromosome segregation and maintenance (or gain of another copy of a gene that disturbs these processes), then the cell acquires CIN – essentially, the ability to shuffle its genome until a stable, malignant phenotype is procured [82]. By contrast, in the absence of clustering, supernumerary centrosomes result in spindle multipolarity, which may cause aneuploidy of a mortally high grade. Alternatively, multipolar cells may arrest in mitosis and succumb to death via other mechanisms [43, 83]. Given the lethality of the multipolar state, induction of high-grade spindle multipolarity constitutes a novel chemotherapeutic strategy whose efficacy holds much promise. Moreover, declustering of supernumerary centrosomes to achieve multipolarity should specifically target cancer cells and pose no apparent threat to most healthy tissues, unlike the majority of current

anticancer remedies, drugs and radiation alike. Understanding the mechanisms by which the few known declustering agents operate can pave the way for rational design and synthesis of cancer cell-specific and thus “kinder and gentler” chemotherapy.

1.5.1 Griseofulvin

The nontoxic antifungal, griseofulvin, induces declustering in various human cancer cell lines in a concentration-dependent manner [84]. Griseofulvin has garnered much attention for its anticancer potential, as it suppresses proliferation of tumor cells at doses that are nontoxic to nontransformed cells (*viz.*, normal human fibroblasts and keratinocytes). The antiproliferative effect of griseofulvin is correlated with both its antimitotic action and its ability to induce declustering. Among several 2'-substituted derivatives of griseofulvin, the one with the highest potency also has the greatest ability to induce declustering, suggesting that declustering contributes to inhibition of both mitosis and cell proliferation, although non-correlative studies are needed to validate this possibility. The precise mechanism by which griseofulvin accomplishes declustering remains largely unexamined; however, a novel idea is that attenuation of microtubule dynamicity may vitiate clustering ability and thereby execute declustering. Griseofulvin suppresses dynamic instability independently of MAPs and does so at doses below those necessary to cause microtubule depolymerization [85]. Griseofulvin may therefore realize declustering by hampering dynamic instability, a mechanism we suspect is integral to centrosome clustering.

1.5.2 Noscapinoids

In line with this notion, bromonoscapine (also known as EM011), a derivative of the nontoxic, poppy-derived antitussive, noscapine, suppresses dynamic instability and centrosome clustering [83]. This drug simply dampens microtubule dynamic instability without causing depolymerization, overpolymerization, or otherwise noticeably impacting microtubule ultrastructure. Like griseofulvin, bromonoscapine causes G2/M arrest followed by apoptosis; however, bromonoscapine-treated cells also succumb to death by another pathway: multipolar mitosis [83]. Synchronized HeLa cells treated with bromonoscapine experience CA followed by declustering and then, if able to overcome mitotic block, proceed to divide in an aberrant fashion to produce multiple, highly aneuploid, inviable daughter cells. Although bromonoscapine induces CA in cancer cells, there is no evidence to date that the drug has any untoward effects on centrosome copy number, spindle bipolarity, or the viability of noncancerous cells [86]. As a result, it seems that bromonoscapine selectively targets cancer cells for CA and declustering. The mechanism by which bromonoscapine attenuates microtubule dynamicity is currently under investigation. Some clues to its *modus operandi* come from the finding that bromonoscapine impairs plus-end association of the plus-end-tracking proteins, EB1 and CLIP-170 [83]. Whether the failure of these proteins to plus-end track is a cause or consequence of microtubule stabilization warrants further investigation.

1.5.3 Phenanthrene-derived PARP inhibitors

Some phenanthrene-derived poly-ADP-ribose polymerase (PARP) inhibitors also exhibit cancer cell-specific declustering [87]. PARP-1 expression is upregulated in various human

cancers [88] but downregulated in others [89], suggesting a complex role for this protein in tumorigenesis. Treatment of tumors exhibiting CA with the phenanthrene-derived PARP-1 inhibitor, PJ-34, results in clustering inhibition, spindle multipolarity, and death by mitotic catastrophe [90]. By contrast, treatment of normal proliferating cells with high drug concentration (i.e., 2–3 times greater than necessary for complete PARP-1 inhibition) for several days has no discernible effect on spindle morphology, centrosome integrity, mitosis, or cell viability. PARP-1 is involved in detection and base-excision repair of DNA strand breaks, initiation of the DNA damage checkpoint, and apoptosis [91], activities not plainly related to clustering. However, PARP 1 has been implicated in CA [89]. Furthermore, in their screen for clustering proteins in S2 cells, Kwon *et al.* identified a PARP, tankyrase-1, and a putative PARP-16 homolog as molecules critical for clustering [60]. Considering that PARP-1 inhibitors induce declustering, these drugs evince great potential for cancer cell-specific chemotherapy.

1.5.4 High-grade spindle multipolarity: its power and payoff

Given the clinical promise presented by declustering agents such as griseofulvin, noscapinoids, and PARP-1 inhibitors, the next logical step would be to enhance declustering efficacy by designing and quantitatively evaluating analogs of these lead small molecules for their ability to trigger death-inducing spindle multipolarity. Small molecules that trigger the most robust multipolarity would be desirable to minimize the chances of unwittingly producing any progeny that may escape death because they harbor “low-grade” tumor-promoting aneuploidy. A fairly common obligate intracellular bacterium, *C. trachomatis*, which causes the sexually transmitted disease chlamydia, may also have the potential to shed some light on mechanisms involved in declustering. In host cells, *C. trachomatis* forms an inclusion that associates closely

with the host centrosome, induces CA, and somehow inhibits the cell's clustering machinery [92]. Given the bacterium's controversial relationship with cervical cancer [93], much may be gained by analyzing its impact on centrosome declustering.

Contrary to present-day cancer treatments, which are notoriously toxic, declustering agents offer the prospect of mitigating chemotherapy-related side effects in a groundbreaking way. Declustering itself should not have any impact on cells that do not rely on clustering (such as normal adult cells), thus providing cancer cell-specific action. As cancer cells rely on centrosome clustering for survival, they represent an Achilles' heel in cancer cells, and this vulnerability can be exploited for chemotherapeutic ends. Given how frequently centrosomes are amplified in cancer cells – and not in adult human cells – agents that induce spindle multipolarity hold tremendous promise as the next generation of cancer cell-selective, non-toxic chemotherapeutics. An interesting avenue for investigation would be to examine whether cancer cells' increased need for centrosome clustering proteins translates into an upregulation of clustering protein expression and/or activity, which might make them invaluable cancer biomarkers. Thus, centrosome clustering mechanisms are attractive theranostic targets that could hold the key to unraveling, as well as subduing, the surreptitious enemy named cancer. Better understanding of the phenotypes imparted by centrosome declustering drugs, such as the grade of multipolarity and its correlation with cell death, can reveal mechanistic insights into how their efficacy can be improved, the goal of the study described in the third manuscript in this dissertation (Chapter 4).

1.6 Mitotic Propensity

CA and clustering only induce CIN if the cell is actively dividing, when the mitotic spindle apparatus is formed and microtubule-kinetochore attachments are made. Thus, the prognostic and potentially predictive power of CA and clustering rely, at least in part, on mitotic propensity, which may consequently also provide information about ITH. The utility of assaying proliferation markers, such as Ki67, in breast cancer is unclear, with studies yielding conflicting findings. Indeed, ASCO recommends against assessing Ki67 in early-stage breast cancer patients to guide treatment decisions [1]. On the other hand, histological grade, which includes quantitation of mitotic figures, is a well-established prognostic and predictive biomarker. Determinations of Ki67 index and mitotic count are normally made on separate slides on different scales (absolute percentage vs. ranking), which neglects the fact that mitosis is a subset of the proliferative cell cycle. This disintegration discards potentially useful information of cell cycle kinetics, specifically, what fraction of the proliferative cell pool is actually undergoing mitosis. Tumors with a large fraction of actively mitotic cells amongst all proliferating cells (that is, a high mitosis:proliferation or M:P ratio) likely pose a more serious threat to survival than tumors with a smaller fraction of actively mitotic cells. Testing the clinical utility of a M:P score, an indicator of mitotic propensity, in breast cancer prognostication is the subject of the study detailed in the fourth manuscript in this dissertation (Chapter 5).

1.7 Drivers of Proliferation: EGFR and HER3

Proteins that drive proliferation, and thus which impact mitotic propensity, can also offer insights into the potential clinical course of breast cancers. Two such potential prognostic biomarkers are the ErbB family members epidermal growth factor receptor (EGFR) and human

epidermal growth factor receptor 3 (HER3), which promote proliferation and have recently been identified as potential drivers of aggressive disease course in TNBC [94, 95]. A majority of TNBCs stain positive for EGFR expression [96, 97]. However, the literature is mixed regarding the utility of EGFR as a prognostic biomarker in TNBC [96, 98-102], and ASCO currently recommends against its assessment to guide adjuvant treatment decisions in early-stage breast cancer patients, including TNBC patients [1]. Despite compelling pre-clinical findings, early-phase clinical trials of EGFR antagonists in unselected TNBC patients failed to show substantial activity, and any significant increases in survival time were modest [103]. One possible explanation for the disappointing results of EGFR inhibitors in TNBC patients is that EGFR antagonism induces compensatory upregulation of HER3 signaling, which causes EGFR resistance in pre-clinical models of TNBC [104]. HER3 expression can stratify TNBCs into groups with significantly different outcomes [105], but it is not established whether HER3 expression is an independent prognostic biomarker in TNBC after adjusting for confounding variables in multivariable analysis. HER3 exhibits impaired kinase activity compared with other ErbB family and must heterodimerize with EGFR or HER2 to effect signaling [106]. Therefore, it stands to reason that assessment of HER3 levels in the context of EGFR or HER2 could improve its prognostic value. As TNBCs do not overexpress HER2 by definition, insights into TNBC prognosis may be greater by considering HER3 and EGFR in combination. Previously, it was shown that coexpression of two or more ErbB family members (EGFR, HER2, and/or HER3) offers synergistic information about breast cancer prognosis after adjusting for tumor size and lymph node status [107]. However, analysis in the TNBC subgroup was not made, nor was adjustment for receptor status, so the independent prognostic value of a combined HER3-EGFR

score in the setting of TNBC unknown. This research question was addressed by the study described in the fifth manuscript in this dissertation (Chapter 5).

1.8 Inhibitors of Proliferation: RARA

Just as proteins that drive proliferation can provide insights into the course of breast cancers, so can proteins that inhibit proliferation, the expression of which should portend a favorable prognosis. Recently, the importance of retinoic acid receptor α (RARA) in TNBC due to its involvement in suppressing proliferation has come into light. When this nuclear hormone receptor heterodimerizes with retinoic X receptors (RXRs), a transcriptional program that suppresses cell growth and induces differentiation is launched, although the opposite effect occurs when RARA binds to ER α [108]. Thus, in ER-negative cancers, such as TNBCs, RARA can put the brakes on proliferation and inhibit tumorigenesis, so RARA express could be a biomarker of good prognosis in TNBC and treatment with RARA agonists (such as all-trans retinoic acid) could exhibit anti-cancer efficacy. However, RARA is often expressed at lower levels in TNBCs than breast cancers of other subtypes [109], in part due to hypermethylation not found in hormone receptor-positive breast cancers [110]. As a result, TNBCs tend to resist the effects of RARA stimulation [109]. Whether RARA can serve as a predictor of good prognosis in TNBCs, a subgroup of which may express RARA (and thus potentially be susceptible to RARA agonists and perhaps tamoxifen when coadministered with retinoic acid, as suggested by a recent abstract [111]), was the goal of the study in the sixth and final manuscript in this dissertation (Chapter 6).

1.9 References

1. Harris, L.N., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. Journal of Clinical Oncology, 2016. **34**(10): p. 1134-1150.
2. Criscitiello, C., et al., *Understanding the biology of triple-negative breast cancer*. Ann Oncol, 2012. **23 Suppl 6**: p. vi13-8.
3. Howlander N, N.A., Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). , *SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site, April 2014*. 2014.
4. Boyle, P., *Triple-negative breast cancer: epidemiological considerations and recommendations*. Ann Oncol, 2012. **23 Suppl 6**: p. vi7-12.
5. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence*. Clin Cancer Res, 2007. **13**(15 Pt 1): p. 4429-34.
6. Liedtke, C., et al., *Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer*. J Clin Oncol, 2008. **26**(8): p. 1275-81.
7. Niikura, N., et al., *Treatment outcomes and prognostic factors for patients with brain metastases from breast cancer of each subtype: a multicenter retrospective analysis*. Breast Cancer Res Treat, 2014. **147**(1): p. 103-12.
8. Bauer, K.R., et al., *Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry*. Cancer, 2007. **109**(9): p. 1721-8.
9. Carey, L., et al., *Triple-negative breast cancer: disease entity or title of convenience?* Nat Rev Clin Oncol, 2010. **7**(12): p. 683-92.
10. Stead, L.A., et al., *Triple-negative breast cancers are increased in black women regardless of age or body mass index*. Breast Cancer Res, 2009. **11**(2): p. R18.
11. Howlander N, N.A., Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2010, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, April 2013*.
12. Lund, M.J., et al., *Race and triple negative threats to breast cancer survival: a population-based study in Atlanta, GA*. Breast Cancer Res Treat, 2009. **113**(2): p. 357-70.
13. Straver, M.E., et al., *The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer*. Breast Cancer Res Treat, 2010. **119**(3): p. 551-8.
14. Keenan, T., et al., *Comparison of the Genomic Landscape Between Primary Breast Cancer in African American Versus White Women and the Association of Racial Differences With Tumor Recurrence*. Journal of Clinical Oncology, 2015. **33**(31): p. 3621-3627.
15. A'Hern, R.P., et al., *Taxane benefit in breast cancer--a role for grade and chromosomal stability*. Nat Rev Clin Oncol, 2013. **10**(6): p. 357-64.

16. Chan, J.Y., *A clinical overview of centrosome amplification in human cancers*. Int J Biol Sci, 2011. **7**(8): p. 1122-44.
17. Woodruff, J.B., O. Wueseke, and A.A. Hyman, *Pericentriolar material structure and dynamics*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2014. **369**(1650): p. 20130459.
18. Antonczak, A.K., et al., *Opposing effects of pericentrin and microcephalin on the pericentriolar material regulate CHK1 activation in the DNA damage response*. Oncogene, 2015.
19. Conduit, P.T., et al., *Centrioles regulate centrosome size by controlling the rate of Cnn incorporation into the PCM*. Curr Biol, 2010. **20**(24): p. 2178-86.
20. Degenhardt, Y. and T. Lampkin, *Targeting Polo-like Kinase in Cancer Therapy*. Clinical Cancer Research, 2010. **16**(2): p. 384-389.
21. Zhang, Y. and T. Hunter, *Roles of Chk1 in cell biology and cancer therapy*. International Journal of Cancer, 2014. **134**(5): p. 1013-1023.
22. Mullee, L.I. and C.G. Morrison, *Centrosomes in the DNA damage response-the hub outside the centre*. Chromosome Res, 2015.
23. Mittal, K., et al., *Abstract B05: Centrosomal profiles of pancreatic adenocarcinoma from African American and European American patients: A comparative analysis*. Cancer Epidemiology Biomarkers & Prevention, 2015. **24**(10 Supplement): p. B05.
24. Denu, R.A., et al., *Centrosome amplification induces high grade features and is prognostic of worse outcomes in breast cancer*. BMC Cancer, 2016. **16**: p. 47.
25. Pannu, V., et al., *Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers*. Oncotarget, 2015. **6**(12): p. 10487-10497.
26. Lingle, W.L., et al., *Centrosome amplification drives chromosomal instability in breast tumor development*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 1978-1983.
27. Lingle, W.L., et al., *Centrosome hypertrophy in human breast tumors: Implications for genomic stability and cell polarity*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(6): p. 2950-2955.
28. Lingle, W.L., et al., *Centrosome amplification drives chromosomal instability in breast tumor development*. Proc Natl Acad Sci U S A, 2002. **99**(4): p. 1978-83.
29. Vaishali Pannu, K.M., Guilherme Cantuaria, Michelle D. Reid, Xiaoxian Li, Shashikiran Donthamsetty, Michelle McBride, Sergey Klimov, Remus Osan, Meenakshi V. Gupta, Padmashree C. G. Rida, Ritu Aneja *Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers*. Oncotarget (In press), 2015.
30. Godinho, S.A. and D. Pellman, *Causes and consequences of centrosome abnormalities in cancer*. Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1650).
31. Gao, P. and J. Zheng, *Oncogenic virus-mediated cell fusion: new insights into initiation and progression of oncogenic viruses--related cancers*. Cancer Lett, 2011. **303**(1): p. 1-8.
32. Nigg, E.A., L. Cajanek, and C. Arquint, *The centrosome duplication cycle in health and disease*. FEBS Lett, 2014. **588**(15): p. 2366-72.
33. Marina, M. and H.I. Saavedra, *Nek2 and Plk4: prognostic markers, drivers of breast tumorigenesis and drug resistance*. Front Biosci (Landmark Ed), 2014. **19**: p. 352-65.
34. Ogden, A., P.C. Rida, and R. Aneja, *Heading off with the herd: how cancer cells might maneuver supernumerary centrosomes for directional migration*. Cancer Metastasis Rev, 2013. **32**(1-2): p. 269-87.

35. Brownlee, C.W. and G.C. Rogers, *Show me your license, please: deregulation of centriole duplication mechanisms that promote amplification*. Cell Mol Life Sci, 2013. **70**(6): p. 1021-34.
36. Korzeniewski, N., M. Hohenfellner, and S. Duensing, *The centrosome as potential target for cancer therapy and prevention*. Expert Opin Ther Targets, 2013. **17**(1): p. 43-52.
37. Anderhub, S.J., A. Kramer, and B. Maier, *Centrosome amplification in tumorigenesis*. Cancer Lett, 2012. **322**(1): p. 8-17.
38. Ogden, A., P.C. Rida, and R. Aneja, *Let's huddle to prevent a muddle: centrosome declustering as an attractive anticancer strategy*. Cell Death Differ, 2012. **19**(8): p. 1255-67.
39. Funk, L.C., L.M. Zasadil, and B.A. Weaver, *Living in CIN: Mitotic Infidelity and Its Consequences for Tumor Promotion and Suppression*. Developmental Cell, 2016. **39**(6): p. 638-652.
40. Wang, X., et al., *Overexpression of aurora kinase A in mouse mammary epithelium induces genetic instability preceding mammary tumor formation*. Oncogene, 2006. **25**(54): p. 7148-7158.
41. Wang, Y., et al., *Clonal evolution in breast cancer revealed by single nucleus genome sequencing*. Nature, 2014. **512**(7513): p. 155-160.
42. Navin, N., et al., *Tumor Evolution Inferred by Single Cell Sequencing*. Nature, 2011. **472**(7341): p. 90-94.
43. Ganem, N.J., S.A. Godinho, and D. Pellman, *A mechanism linking extra centrosomes to chromosomal instability*. Nature, 2009. **460**(7252): p. 278-282.
44. Thompson, S.L. and D.A. Compton, *Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(44): p. 17974-17978.
45. Janssen, A., et al., *Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations*. Science, 2011. **333**(6051): p. 1895-1898.
46. Steigemann, P., et al., *Aurora B-Mediated Abscission Checkpoint Protects against Tetraploidization*. Cell. **136**(3): p. 473-484.
47. Zhang, C.Z., M.L. Leibowitz, and D. Pellman, *Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements*. Genes Dev, 2013. **27**(23): p. 2513-30.
48. Holland, A.J. and D.W. Cleveland, *Mechanisms and consequences of localized, complex chromosomal rearrangements in cancer and developmental diseases*. Nature medicine, 2012. **18**(11): p. 1630-1638.
49. Zhang, C.Z., et al., *Chromothripsis from DNA damage in micronuclei*. Nature, 2015. **522**(7555): p. 179-84.
50. Leibowitz, M.L., C.Z. Zhang, and D. Pellman, *Chromothripsis: A New Mechanism for Rapid Karyotype Evolution*. Annu Rev Genet, 2015. **49**: p. 183-211.
51. Goncalves, R., et al., *New concepts in breast cancer genomics and genetics*. Breast Cancer Res, 2014. **16**(5): p. 460.
52. Santaguida, S. and A. Amon, *Short- and long-term effects of chromosome mis-segregation and aneuploidy*. Nat Rev Mol Cell Biol, 2015. **16**(8): p. 473-85.

53. Lingle, W.L., et al., *Centrosome amplification drives chromosomal instability in breast tumor development*. Proceedings of the National Academy of Sciences, 2002. **99**(4): p. 1978-1983.
54. Pihan, G.A., et al., *Centrosome abnormalities and chromosome instability occur together in pre-invasive carcinomas*. Cancer Res, 2003. **63**(6): p. 1398-404.
55. Chiba, S., et al., *Genomic convergence and suppression of centrosome hyperamplification in primary p53^{-/-} cells in prolonged culture*. Exp Cell Res, 2000. **258**(2): p. 310-21.
56. Oikawa, T., et al., *Transcriptional control of BubR1 by p53 and suppression of centrosome amplification by BubR1*. Mol Cell Biol, 2005. **25**(10): p. 4046-61.
57. Bergmann, S., et al., *YB-1 Provokes Breast Cancer through the Induction of Chromosomal Instability That Emerges from Mitotic Failure and Centrosome Amplification*. Cancer Research, 2005. **65**(10): p. 4078-4087.
58. Li, J.J., et al., *Estrogen mediates Aurora-A overexpression, centrosome amplification, chromosomal instability, and breast cancer in female ACI rats*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(52): p. 18123-18128.
59. Kleylein-Sohn, J., et al., *Acentrosomal spindle organization renders cancer cells dependent on the kinesin HSET*. J Cell Sci, 2012. **125**(Pt 22): p. 5391-402.
60. Kwon, M., et al., *Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes*. Genes Dev, 2008. **22**.
61. Mountain, V., et al., *The kinesin-related protein, HSET, opposes the activity of Eg5 and cross-links microtubules in the mammalian mitotic spindle*. J Cell Biol, 1999. **147**(2): p. 351-66.
62. Cai, S., et al., *Kinesin-14 family proteins HSET/XCTK2 control spindle length by cross-linking and sliding microtubules*. Mol Biol Cell, 2009. **20**(5): p. 1348-59.
63. Godinho, S.A., M. Kwon, and D. Pellman, *Centrosomes and cancer: how cancer cells divide with too many centrosomes*. Canc Met Rev, 2009. **28**.
64. Yang, W.X. and A.O. Sperry, *C-terminal kinesin motor KIFC1 participates in acrosome biogenesis and vesicle transport*. Biol Reprod, 2003. **69**(5): p. 1719-29.
65. Gordon, M.B., L. Howard, and D.A. Compton, *Chromosome movement in mitosis requires microtubule anchorage at spindle poles*. J Cell Biol, 2001. **152**(3): p. 425-34.
66. Pannu, V., et al., *HSET overexpression fuels tumor progression via centrosome clustering-independent mechanisms in breast cancer patients*. Oncotarget, 2015. **6**(8): p. 6076-6091.
67. Howlader N, N.A., Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2012, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site, April 2015.
68. Daly, B. and O.I. Olopade, *A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change*. CA Cancer J Clin, 2015. **65**(3): p. 221-38.
69. Kohler, B.A., et al., *Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State*. J Natl Cancer Inst, 2015. **107**(6): p. djv048.

70. Newman, L.A., *Breast cancer disparities: high-risk breast cancer and African ancestry*. Surg Oncol Clin N Am, 2014. **23**(3): p. 579-92.
71. Dietze, E.C., et al., *Triple-negative breast cancer in African-American women: disparities versus biology*. Nat Rev Cancer, 2015.
72. Kurian, A.W., et al., *Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups*. Breast Cancer Res, 2010. **12**(6): p. R99.
73. Turner, N., et al., *Targeting triple negative breast cancer: Is p53 the answer?* Cancer Treatment Reviews, 2013. **39**(5): p. 541-550.
74. Stewart, P.A., et al., *Differentially Expressed Transcripts and Dysregulated Signaling Pathways and Networks in African American Breast Cancer*. PLoS ONE, 2013. **8**(12): p. e82460.
75. Zou, J., et al., *BRCA1 and FancJ cooperatively promote interstrand crosslinker induced centrosome amplification through the activation of polo-like kinase 1*. Cell Cycle, 2014. **13**(23): p. 3685-97.
76. Izumi, H., et al., *BubR1 localizes to centrosomes and suppresses centrosome amplification via regulating Plk1 activity in interphase cells*. Oncogene, 2009. **28**(31): p. 2806-20.
77. Khan, J., et al., *Overexpression of Active Aurora-C Kinase Results in Cell Transformation and Tumour Formation*. PLoS ONE, 2011. **6**(10): p. e26512.
78. Burstein, M.D., et al., *Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer*. Clin Cancer Res, 2015. **21**(7): p. 1688-98.
79. Lehmann, B.D., et al., *Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection*. PLoS ONE, 2016. **11**(6): p. e0157368.
80. Lindner, R., et al., *Molecular phenotypes in triple negative breast cancer from African American patients suggest targets for therapy*. PLoS One, 2013. **8**(11): p. e71915.
81. Silkworth, W.T., et al., *Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells*. PLoS One, 2009. **4**(8): p. e6564.
82. Ricke, R.M., J.H. van Ree, and J.M. van Deursen, *Whole chromosome instability and cancer: a complex relationship*. Trends Genet, 2008. **24**(9): p. 457-66.
83. Karna, P., et al., *A novel microtubule-modulating noscapinoid triggers apoptosis by inducing spindle multipolarity via centrosome amplification and declustering*. Cell Death Differ, 2011. **18**(4): p. 632-44.
84. Rebacz, B., et al., *Identification of griseofulvin as an inhibitor of centrosomal clustering in a phenotype-based screen*. Cancer Res, 2007. **67**(13): p. 6342-50.
85. Panda, D., et al., *Kinetic suppression of microtubule dynamic instability by griseofulvin: implications for its possible use in the treatment of cancer*. Proc Natl Acad Sci U S A, 2005. **102**(28): p. 9878-83.
86. Rathinasamy, K., et al., *Griseofulvin stabilizes microtubule dynamics, activates p53 and inhibits the proliferation of MCF-7 cells synergistically with vinblastine*. BMC Cancer, 2010. **10**: p. 213-213.
87. Castiel, A., et al., *A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells*. BMC Cancer, 2011. **11**: p. 412-412.

88. Miwa, M. and M. Masutani, *PolyADP-ribosylation and cancer*. *Cancer Sci*, 2007. **98**(10): p. 1528-35.
89. Tong, W.M., et al., *Poly(ADP-ribose) polymerase-1 plays a role in suppressing mammary tumourigenesis in mice*. *Oncogene*, 2007. **26**(26): p. 3857-67.
90. Castiel, A., et al., *A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells*. *BMC Cancer*, 2011. **11**(1): p. 412.
91. Kanai, M., et al., *Involvement of Poly(ADP-Ribose) Polymerase 1 and Poly(ADP-Ribosylation) in Regulation of Centrosome function*. *Mol Cell Biol*, 2003. **23**.
92. Knowlton, A.E., et al., *Chlamydia trachomatis infection causes mitotic spindle pole defects independently from its effects on centrosome amplification*. *Traffic (Copenhagen, Denmark)*, 2011. **12**(7): p. 854-866.
93. Lehtinen, M., et al., *Chlamydia trachomatis infection and risk of cervical intraepithelial neoplasia*. *Sex Transm Infect*, 2011. **87**(5): p. 372-6.
94. Baselga, J., *Why the Epidermal Growth Factor Receptor? The Rationale for Cancer Therapy*. *The Oncologist*, 2002. **7**(suppl 4): p. 2-8.
95. Mujoo, K., et al., *Regulation of ERBB3/HER3 signaling in cancer*. *Oncotarget*, 2014. **5**(21): p. 10222-10236.
96. Viale, G., et al., *Invasive ductal carcinoma of the breast with the "triple-negative" phenotype: prognostic implications of EGFR immunoreactivity*. *Breast Cancer Res Treat*, 2009. **116**(2): p. 317-28.
97. Changavi, A.A., A. Shashikala, and A.S. Ramji, *Epidermal Growth Factor Receptor Expression in Triple Negative and Nontriple Negative Breast Carcinomas*. *J Lab Physicians*, 2015. **7**(2): p. 79-83.
98. Lakis, S., et al., *Interaction Between Beta-Catenin and EGFR Expression by Immunohistochemistry Identifies Prognostic Subgroups in Early High-risk Triple-negative Breast Cancer*. *Anticancer Res*, 2016. **36**(5): p. 2365-78.
99. Yue, Y., et al., *Stratification of Prognosis of Triple-Negative Breast Cancer Patients Using Combinatorial Biomarkers*. *PLoS One*, 2016. **11**(3): p. e0149661.
100. Jacot, W., et al., *High EGFR protein expression and exon 9 PIK3CA mutations are independent prognostic factors in triple negative breast cancers*. *BMC Cancer*, 2015. **15**: p. 986.
101. Nogi, H., et al., *EGFR as paradoxical predictor of chemosensitivity and outcome among triple-negative breast cancer*. *Oncol Rep*, 2009. **21**(2): p. 413-7.
102. Rakha, E.A., et al., *Prognostic markers in triple-negative breast cancer*. *Cancer*, 2007. **109**(1): p. 25-32.
103. Costa, R., et al., *Targeting Epidermal Growth Factor Receptor in triple negative breast cancer: New discoveries and practical insights for drug development*. *Cancer Treatment Reviews*, 2017. **53**: p. 111-119.
104. Verma, N., et al., *Targeting of PYK2 Synergizes with EGFR Antagonists in Basal-like TNBC and Circumvents HER3-Associated Resistance via the NEDD4-NDRG1 Axis*. *Cancer Res*, 2017. **77**(1): p. 86-99.
105. Bae, S.Y., et al., *HER3 status by immunohistochemistry is correlated with poor prognosis in hormone receptor-negative breast cancer patients*. *Breast Cancer Res Treat*, 2013. **139**(3): p. 741-50.

106. Gala, K. and S. Chandarlapaty, *Molecular Pathways: HER3 targeted therapy*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2014. **20**(6): p. 1410-1416.
107. Wiseman, S.M., et al., *Coexpression of the type I growth factor receptor family members HER-1, HER-2, and HER-3 has a synergistic negative prognostic effect on breast carcinoma survival*. Cancer, 2005. **103**(9): p. 1770-1777.
108. Connolly, R.M., N.K. Nguyen, and S. Sukumar, *Molecular Pathways: Current Role and Future Directions of the Retinoic Acid Pathway in Cancer Prevention and Treatment*. Clinical Cancer Research, 2013. **19**(7): p. 1651-1659.
109. Centritto, F., et al., *Cellular and molecular determinants of all-trans retinoic acid sensitivity in breast cancer: Luminal phenotype and RARalpha expression*. EMBO Mol Med, 2015. **7**(7): p. 950-72.
110. Conway, K., et al., *DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival*. Breast Cancer Res, 2014. **16**(5): p. 450.
111. Coyle, K.M., et al., *Abstract P1-12-14: Retinoic acid sensitizes triple-negative breast cancer cells to tamoxifen treatment*. Cancer Research, 2015. **75**(9 Supplement): p. P1-12-14-P1-12-14.

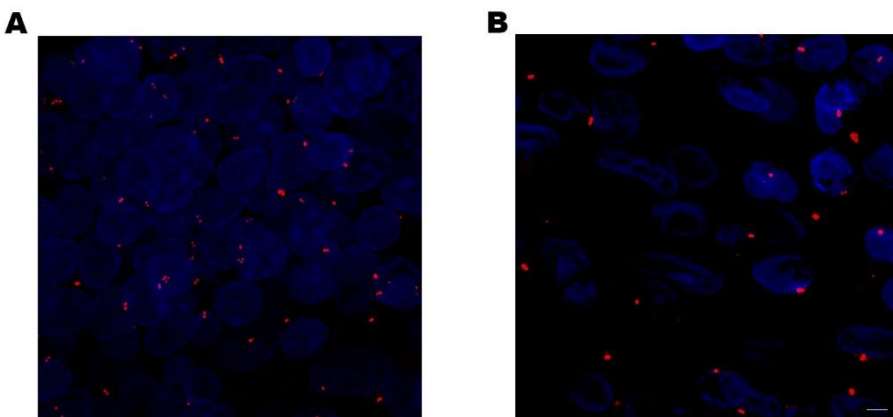


Figure 1.1 Numerical and structural centrosome amplification in breast cancer.

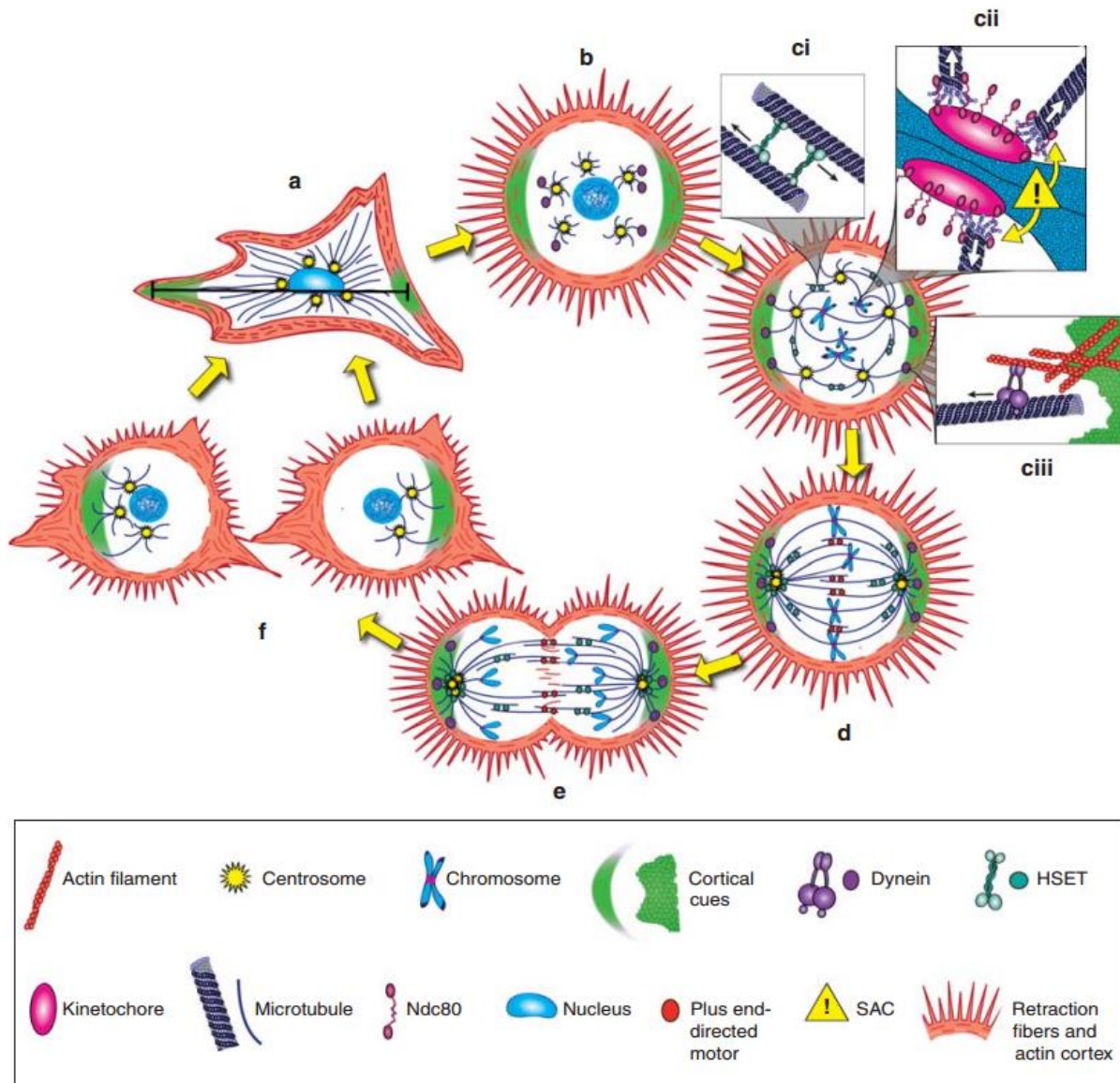


Figure 1.2 Global centrosome clustering mechanisms

(a) During interphase, cell adhesion geometry results in establishment of a pattern of cortical cues, which are preferentially distributed at the termini of the cell's long axis (black line). (b) When the cell rounds up for mitosis, actin-rich retraction fibers retain a memory of the cell's interphase adhesion pattern and maintain contacts with the substratum. Microtubule dynamic instability increases. (c) During prophase, astral microtubules probe the cytosol, making contacts with other microtubules, chromosomes, and the actin-rich cortex. (i) HSET may localize between antiparallel astral microtubules. Minus-end directed movement of HSET would thus promote centrosome coalescence. (ii) Astral microtubules are captured at kinetochores by the Ndc80 complex, thus forming kinetochore microtubules. Low-tension attachments activate the SAC, which provides time to correct the errors. By contrast, merotelic attachments are poorly sensed by the SAC and may persist. (iii) Dynein is delivered to the actin-enriched cortex preferentially in the vicinity of cortical cues, which are concentrated at opposite cellular poles. Cortical dynein then captures astral microtubules, and minus-end directed movement of dynein exerts tension

that pulls centrosomes to the poles. **(d)** Tug-of-war between kinetochore microtubules and chromosomes results in alignment of chromosomes along the metaphase plate, although some chromosome lagging occurs, signifying that the cellular milieu is in some way abnormal for attachment. Tug-of-war between dynein, HSET, and other microtubule motors results in a net force that tends to pull some of the centrosomes to one pole and some to the other. Owing to their common final destination, each group of centrosomes forms a cluster. **(e)** Attainment of the requisite tension or stretch across kinetochores satisfies the SAC and initiates anaphase onset. Sister chromatids separate, although some missegregation occurs. **(f)** Daughter cells experience low-grade aneuploidy, permitting survival and possibly promoting tumorigenic phenotype

2 PROGNOSTIC VALUE OF CA20, A SCORE BASED ON CENTROSOME AMPLIFICATION-ASSOCIATED GENES, IN BREAST TUMORS²

2.1 Abstract

Centrosome amplification (CA) is a hallmark of cancer, observable in $\geq 75\%$ of breast tumors. CA drives aggressive cellular phenotypes such as chromosomal instability (CIN) and invasiveness. Thus, assessment of CA may offer insights into the prognosis of breast cancer and identify patients who might benefit from centrosome declustering agents. However, it remains unclear whether CA is correlated with clinical outcomes after adjusting for confounding factors. To gain insights, we developed a signature, “CA20,” comprising centrosome structural genes and genes whose dysregulation is implicated in inducing CA. We found that CA20 was a significant independent predictor of worse survival in two large independent datasets after adjusting for potentially confounding factors. In multivariable analyses including both CA20 and CIN25 (a gene expression-based score that correlates with aneuploidy and has prognostic potential in many types of cancer), only CA20 was significant, suggesting CA20 captures the risk-predictive information of CIN25 and offers information beyond it. CA20 correlated strongly with CIN25, so a high CA20 score may reflect tumors with high CIN and potentially other aggressive features that may require more aggressive treatment. Finally, we identified processes and pathways differing between CA20-low and high groups that may be valuable therapeutic targets.

²Parts of this chapter have been published verbatim in *Scientific Reports* 2017 Mar 21;7(1):262. doi: 10.1038/s41598-017-00363-w as “Prognostic value of CA20, a score based on centrosome amplification-associated genes, in breast tumors.”

2.2 Introduction

CA is a hallmark of cancer observable in $\geq 75\%$ of breast tumors [1] that promotes invasive behavior [2] and enhanced migratory ability [3] in cancer cells. In addition, the presence of supernumerary centrosomes results in a transient multipolar intermediate in mitosis that promotes merotelic microtubule-kinetochore attachments [4]. To resolve spindle multipolarity and thereby avoid mitotic catastrophe or multipolar mitosis, which could lead to cell death, the cell clusters centrosomes into two polar groups, allowing bipolar division to occur; however, attachment errors persist in the spindle and chromosome missegregation occurs. CIN allows the cell to sample the fitness landscape and acquire a more aggressive karyotype and also promotes intratumor heterogeneity, which fosters chemoresistance [5]. It was recently demonstrated that transient induction of CA in p53-deficient epidermis causes aneuploidy and spontaneous skin cancer development in mice [6]. Given that CA promotes tumorigenesis and aggressive phenotypes and is common among breast tumors, it may have value as a prognostic biomarker in breast cancer and could guide treatment decisions.

Although several groups have performed semi-quantitative assessments of CA in patient tumors using microscopy, few have correlated CA with clinical outcomes, and none of these data are in the public domain. It would thus be valuable to be able to assess CA in publicly available datasets, such as microarray datasets, many of which have clinicopathologic and outcome annotation for breast cancer patients. Our lab previously developed a four-gene signature, which includes two genes for centrosome structural proteins and two genes whose overexpression induces CA, called the Centrosome Amplification Index (CAI), which we found stratifies breast cancer patients into two groups with significantly different overall survival (OS) in Kaplan-

Meier analysis [3]. Another group developed a Centrosome Index (CI), comprising four centrosome structural genes, that correlates with CA and is an independent predictor of poor OS in multiple myeloma patients in multivariable analysis [7, 8]. Given that CA can be caused by dysregulation of the expression of many different genes, there is a need to define a more comprehensive gene signature that may be able to identify a greater proportion of tumors with CA, which may arise through a variety of molecular pathways. Thus, in the present study, we define a gene expression signature, “CA20,” that includes 19 genes that have been experimentally demonstrated to induce CA when dysregulated, many of which also have known structural roles in the centrosome (such as SASS6, the primary component of the centriolar cartwheel structure [9] and CEP152, a key pericentriolar material component [10], both of which are among the most abundant proteins in the centrosome in several cell lines [11]), along with *TUBG1*, which encodes the most abundant centrosomal protein and is primarily responsible for microtubule nucleation, key to centrosomal function [11] (see **Table 2.1** for details regarding CA20 component genes). Our objective was to test the prognostic value of CA20 after adjusting for potentially confounding factors in multiple breast cancer cohorts and to explore processes, pathways, and oncogenic signatures that are associated with a high CA20 score. Because CA causes CIN, we were also interested in comparing the prognostic value of CA20 with that of the CIN score “CIN25,” which correlates with total functional aneuploidy and predicts worse outcomes in a variety of cancers [12], determining which of these two scores has the most significant impact on outcomes when included together in multivariable models of survival, and comparing processes, pathways, and oncogenic signatures that are enriched in tumors with high CA20 and CIN25 scores.

2.3 Results

We tested the ability of CA20 and CIN25 to risk-stratify breast cancer patients in two datasets, the METABRIC and TCGA datasets, comprising n=1,969 primary breast cancers with breast cancer-specific survival (BCSS) annotation and n=524 primary invasive breast cancers with OS annotation, respectively. The METABRIC dataset was split into discovery and validation sets. The TCGA dataset was not split because power analysis suggested the subsets would potentially be too small, so bootstrapping was instead used to obtain more reliable estimates of population parameters.

2.3.1 *METABRIC dataset*

Stratification was conducted according to average CA20 and CIN25 scores found in the discovery set as well as optimal cutpoints in CA20 and CIN25 scores found in the discovery set based on the log-rank test. In Kaplan-Meier plots, stratification into high- and low-BCSS groups based on the average and optimal cutpoints in CA20 and CIN25 scores was significant in both the discovery and validation sets ($p < 10^{-6}$ for all, **Figure 2.1**; see **Tables 2.2** and **2.3** for descriptive statistics of study datasets). When both CA20 and CIN25 (both stratified by the average score) were entered as covariates in full multivariable models using discovery set data, only CA20 (stratified by the average score) appeared in the final model, and it was a significant predictor of BCSS (**Table 2.4**). In the validation set too, CA20 (stratified by the average score) remained a significant predictor in the final model. Common significant covariates between discovery and validation set final models included tumor stage and chemotherapy. When both CA20 and CIN25 (both stratified by the optimal cutpoint) were entered as covariates in full multivariable models using discovery set data, both covariates appeared in final models but only

CA20 (stratified by the optimal cutpoint) significantly affected BCSS (**Table 2.5**). In the validation set, CA20 (stratified by the optimal cutpoint) remained a significant predictor, whereas CIN25 (stratified by the optimal cutpoint) did not significantly impact BCSS. Common significant covariates between discovery and validation final multivariable models of BCSS included tumor stage and chemotherapy, as was found when stratifying by average signature scores. Thus, CA20 (whether stratified by the average score or optimal cutpoint) is a significant predictor of BCSS after adjusting for stage and chemotherapy, whereas CIN25 (whether stratified by the average score or optimal cutpoint) is not an independent predictor in these models.

CA20 score was highly correlated with CIN25 score ($\rho=0.93$, $p<10^{-6}$), which may reveal that breast tumors with high CA20 scores have high levels of CIN. Although breast cancer subtype was not a common independent predictor of outcomes, we were interested to test whether CA20 and CIN25 scores differed grade-wise between TNBCs and non-TNBCs, which differ in aggressiveness. No grade 1 TNBCs were present in the dataset for comparison, but we found that average CA20 and CIN25 scores were higher in TNBCs than non-TNBCs in both grade 2 and 3 tumors per two-tailed independent samples t-tests, equal variances assumed ($p<0.001$ for all) (**Figure 2.2**), consistent with the more aggressive behavior of TNBCs compared with non-TNBCs and mirroring what we previously found for CA as assessed by microscopy [13].

2.3.2 TCGA dataset

To confirm the prognostic value of CA20 in a separate cohort, we analyzed the TCGA breast dataset. Stratification was conducted according to the average CA20 and CIN25 scores found in the entire dataset as well as optimal cutpoints in CA20 and CIN25 found in the entire dataset based on the log-rank test. In Kaplan-Meier plots, stratification into high- and low-OS groups based on CA20 (average and optimal cutpoints) was significant ($p=0.025$ and $p=0.024$, respectively), with high CA20 conferring a worse prognosis. For comparison, stratification by CIN25 (optimal cutpoint) was significant ($p=0.029$), whereas stratification by CIN25 (average cutpoint) was not (**Figure 2.3**). In stage-adjusted models, high CA20 scores (based on both average and optimal cutpoints) were associated with 2.72- and 2.79-fold worse OS (bootstrap- $p=0.016$ and 0.008 , respectively). For comparison, in stage-adjusted models, high CIN25 scores (based on both average and optimal cutpoints) were also associated with worse OS, HR=2.31 and 4.65 (bootstrap- $p=0.026$ and 0.035 , respectively). However, as in the METABRIC dataset, when both CA20 and CIN25 (stratified by average or optimal cutpoints) were entered along with stage in full models, following backward variable selection (based on an $\alpha=0.10$ removal criterion), only CA20 and stage remained as predictors in full models (CA20 [stratified by average cutpoint]: HR=2.72, $p=0.010$; CA20 [stratified by optimal cutpoint]: HR=2.79, $p=0.010$), and they remained significant following bootstrapping (**Table 2.6**).

Although age at diagnosis was not a significant predictor of BCSS in the METABRIC dataset in final models, we recognized the possibility that it could confound analyses of OS in this independent dataset. We thus refit multivariable models entering CA20 and CIN25 (stratified by average or optimal cutpoints), AJCC stage, and age at diagnosis. In final models,

CA20 remained a significant predictor, along with stage and age but not CIN25, and the hazard associated with high CA20 was even greater than in models not adjusted for age (CA20 [stratified by average cutpoint]: HR=3.82, p=0.001; CA20 [stratified by optimal cutpoint]: HR=3.67, P=0.002); furthermore, significance was retained after bootstrapping (**Table 2.7**). Because all the cases in the multivariable models were annotated with age at diagnosis, our sample size and, thus, statistical power were not diminished. Therefore, the prognostic value of CA20 adjusting for stage was upheld in this separate dataset adjusted for confounding variables, suggesting broad clinical utility for this score to predict outcomes in female breast cancer patients. Similar to our findings in the METABRIC dataset, in the TCGA dataset CA20 was very strongly correlated with CIN25 ($\rho=0.95$, $p<10^{-6}$), suggesting that breast tumors with high CA20 scores also have high levels of CIN.

Finally, we were interested in exploring differences in biological processes, molecular pathways, and oncogenic signatures between CA20-high and low groups (defined by the average CA20 value), which may reveal potentially actionable biology. To this end, we performed Gene Set Enrichment Analysis (GSEA) [14] using the TCGA dataset and explored differentially enriched biological processes, Reactome pathways, and oncogenic signatures. For the CA20-high group, 262 biological process gene sets were enriched at false discovery rate (FDR) $q<0.05$. Among the most significant results, the CA20-high group was enriched in DNA repair processes, the DNA integrity checkpoint, many cell cycle processes (e.g., mitotic nuclear division, cell cycle phase transition, cell division, spindle assembly, regulation of sister chromatid segregation, mitotic spindle organization), and regulation of microtubule polymerization/depolymerization. Regarding Reactome pathways, the CA20-high group was enriched in 96 gene sets at FDR

$q < 0.05$. Top enriched Reactome pathways in the CA20-high group exhibited much overlap with biological processes, including DNA repair and cell cycle pathways. For the purposes of comparison, we also compared biological processes and Reactome pathways between CIN25-high and low groups (stratified by the average CIN25 score), and it was found that results overlapped greatly with those from the CA20 analyses. Regarding enriched biological processes, only one gene set found in the CA20-high group (<1% of gene sets) was not found in the CIN25-high group, and only 14 gene sets found in the CIN25-high group (~5% of gene sets) were not found in the CA20-high group. Similarly, only four Reactome pathways enriched in the CA20-high group and four in the CIN25-high group differed (~4% of gene sets). We also explored differences in oncogenic signature gene sets. We found that the CA20-high group was enriched in 13 such gene sets, including genes upregulated upon overexpression of E2F1, stimulation with sonic hedgehog (SHH) protein, and loss of retinoblastoma protein (pRb). The CIN25-high group was enriched in 12 oncogenic signature gene sets, all of which were found in the CA20-high group. These data suggest the CA20- and CIN25-high groupings may capture rather similar molecular tumor profiles. Finally, we tested whether the CIN25-high group was enriched in the centrosome gene ontology cellular component, and we found that it was at FDR $q < 0.001$ (Normalized Enrichment Score = 2.29), suggesting this group is enriched in centrosomal genes, consistent with the strong correlation we found between CA20 and CIN25 scores.

2.4 Discussion

CA is a well-characterized hallmark of cancer [15] especially breast cancer. Indeed, $\geq 75\%$ of breast tumors (ductal carcinomas in situ, adenocarcinomas, invasive ductal carcinomas, or breast tumors not otherwise specified) exhibit CA [1]. Because CA promotes CIN and other

aggressive phenotypes, it may be a driving force in tumorigenesis and tumor evolution that can offer insights into the clinical course of breast tumors, but only a few studies have investigated the potential prognostic value of CA. Our lab previously developed a four-gene signature, the CAI, which we demonstrated could stratify n=162 breast cancer patients into two groups with significantly different clinical outcomes in Kaplan-Meier analysis, with high CAI based on an optimal cutpoint correlating with worse OS [3]. In the same study, we also found a non-significant trend among n=120 breast cancer patients towards worse progression-free survival (PFS) in Kaplan-Meier analysis for tumors with high levels of CA (defined as the sum of the percentage of cells with >2 centrosomes and the percentage of cells with abnormally voluminous centrosomes based on microscopy, using an optimal cutpoint). Another study of n=362 breast tumors found that large centrosomal size was not associated with OS or recurrence-free survival (RFS) after adjusting for tumor stage and subtype in multivariable Cox models; however, it is not known whether 2D (i.e., cross-sectional) measurements reliably estimate centrosome size, given that centrosomes are 3D structures, and numerical CA was not considered in multivariable analyses.[16] In the same study, however, Kaplan-Meier analyses revealed that high numerical CA (defined as >2 centrosomes per cell on average) was associated with worse BCSS, OS, and RFS. Thus, there is limited evidence that CA may be associated with worse outcomes in breast cancer, but it is unclear what impact CA has on survival after adjusting for potential confounders and what biological processes and pathways could be targeted therapeutically in tumors with high levels of CA.

To shed light on these questions, we developed the CA20 score based on genes encoding centrosome structural proteins and genes that have been demonstrated to induce CA following

experimental perturbations in their expression. As we found for CA previously [13], CA20 (and CIN25) were higher in the aggressive TNBC subtype than non-TNBCs in grade-matched comparisons. In analyses of two large and well-annotated breast cancer datasets (the METABRIC and TCGA breast datasets), we found that high CA20 score was associated with worse BCSS and OS after adjusting for potentially confounding factors, suggesting that CA20 could be a useful clinical tool to identify breast cancer patients at greater risk of poor outcomes. When both CA20 and CIN25 were factored into multivariable models, only CA20 was significantly associated with outcomes. This finding suggests that when CA20 is accounted for CIN25 no longer holds prognostic value. Given that we found a very strong correlation between CA20 and CIN25 in breast tumors and it has been shown by others that CA and CIN are correlated in breast tumors [16], it is tempting to speculate that CA20 captures CIN, thus rendering CIN25 redundant, and perhaps also captures other aggressive phenotypes not encompassed by CIN25 that are consequences of CA. Given that CIN engenders karyotypic diversity within tumors, we assert that CA20 may perhaps even serve as an indirect measure of intratumor heterogeneity in breast tumors. The overlap in biological processes, Reactome pathways, and oncogenic signatures that are enriched in CA20- and CIN25-high groups is striking given that the two signatures only share one gene in common (CDK1) and suggests that they reflect relatively similar molecular tumor biology (namely, potential activation of DNA repair pathways, perhaps to cope with DNA damage occurring due to chromosome missegregation, enhanced cell cycle kinetics and microtubule dynamics, and activated E2F1 signaling), although perhaps with subtle but prognostically important qualitative and quantitative distinctions.

An exciting avenue for future research would be to test whether breast tumors with high CA20 are more susceptible to E2F1 or SHH inhibitors, drugs targeting DNA repair mechanisms (e.g., PARP inhibitors), chemotherapeutics that target the cell cycle (e.g., taxanes), or centrosome declustering drugs (such as griseofulvin, noscapinoids, PJ34, and KifC1/HSET inhibitors), which preferentially eliminate cells with CA by forcing them to construct a multipolar spindle during mitosis [17-21]. Because most normal cells do not have amplified centrosomes, declustering drugs exhibit low to no apparent toxicity to them. It will also be important to validate (through careful microscopy and rigorous quantitation) that CA20 scores indeed correlate with CA in breast tumors in future studies.

2.5 Methods

2.5.1 *Dataset details and power analyses*

Microarray datasets were chosen based on their availability in Oncomine [22] and the presence of annotation regarding survival time (measured in days) and statuses and signature gene expression levels. Three microarray datasets met these criteria, including the METABRIC [23], TCGA [24], and Esserman [25] breast datasets; however, power analysis suggested the Esserman dataset was too small, so it was excluded from analyses (see “Esserman dataset” below). The clinical data and \log_2 median-centered signature gene expression levels of the METABRIC and TCGA datasets were thus downloaded from Oncomine. **METABRIC dataset:** Normal breast, benign breast neoplasms, and cases without BCSS annotation were excluded from analyses, resulting in a sample size of $n=1,969$ primary breast cancers. A majority of the cases were annotated for AJCC stage and whether adjuvant chemotherapy was given. The dataset was then split randomly (via random number assignment) and approximately equally into

discovery and validation sets (n=985 and n=984, respectively; **Table 2.2**, descriptive statistics). Neither significant differences nor non-significant trends (i.e., $0.05 < p < 0.10$) were found between these two sets for continuous variables (age, CA20 score, and CIN25 score; 2-tailed t-tests), ordinal variables (Nottingham grade, tumor stage, CA20 group [optimal], CA20 group [average], CIN25 group [optimal], and CIN25 group [average]; Mann-Whitney tests), or nominal variables (breast cancer subtype, chemotherapy, radiotherapy, and hormone therapy; Chi-square tests) (data not shown). **TCGA dataset:** Normal breast specimens, metastases, and male breast cancers were excluded from analyses, resulting in a sample size of n=524 primary invasive breast cancers (**Table 2.3**, descriptive statistics). OS annotation was incomplete, so we supplemented it with clinical data downloaded from the TCGA data portal, after which all cases had OS time and status. 395 cases had AJCC stage annotation, but information about adjuvant chemotherapy was not available. We analyzed the METABRIC data to estimate whether this sample size would potentially achieve statistical power ≥ 0.80 in a study of the effect of CA20 on OS in stage-adjusted models with an average follow-up time of approximately 3 years, as in the TCGA study. Among METABRIC patients with invasive breast cancers (n=1,030), the overall probability of an event (death) within 3 years was $p_E=0.12$, the probability of belonging to the CA20 (optimal)-high group was $p_H=0.70$, and the relative risk of death was $HR=2.34$. Based on these data, it was estimated that a sample size of n=339 would be needed to detect a $HR=2.34$ with a Type I error rate of $\alpha=0.05$ and Type II error rate of $\beta=0.80$, based on the formula to calculate the one-sided sample size in Cox proportional-hazards models [26]: $n = \frac{1}{p_A p_B p_E} \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\ln(\theta) - \ln(\theta_0)} \right)^2$. Thus, we elected not to split the data into discovery and validation sets to preserve statistical power ≥ 0.80 and rather implemented bootstrapping methods to more reliably estimate population parameters. **Esserman dataset:** This dataset includes n=120 primary breast carcinomas with OS annotation

and expression values for all the signature gene probes selected. Average follow up time was ~4 years, so we based power analysis on the METABRIC data 4-year OS probabilities for invasive breast cancer patients, where $p_D=0.18$ and $p_H=0.70$. Based on these criteria and using the formula as in the power analysis for the TCGA data, we estimated that $n=227$ patients would be needed to detect a $HR=2.34$ with a Type I error rate of $\alpha=0.05$ and a Type II error rate of $\beta=0.80$, suggesting that the Esserman dataset would be too small for our purposes. Thus, it was excluded from analyses.

2.5.2 Gene signatures and microarray probe selection

A CA gene signature was derived by searching Pubmed (September, 2015) using the search term “centrosome amplification” and filtering for experimental studies wherein manipulation of a specific gene’s expression was found to induce CA, resulting in a set of 19 genes: *AURKA* [27 28], *CCNA2* [29], *CCND1* [30], *CCNE2*[31, 32], *CDK1*[33], *CEP63* [33], *CEP152* [34], *E2F1*[35], *E2F2*[35], *LMO4*[36], *MDM2*[37, 38], *MYCN*[38], *NDRG1*[39], *NEK2*[40], *PIN1*[41], *PLK1*[42, 43], *PLK4*[34, 44], *SASS6*[45], and *STIL* [46]. In addition, the gene encoding the primary centrosome structural protein, *TUBG1*, was included, resulting in a set of 20 genes. The CIN25 gene signature is described in Carter et al. [12] For both datasets, many genes were represented by multiple probes. To select the probe most likely to represent the gene, probes were filtered by rational selection processes that differed by dataset since they are based on different platforms (the Illumina HumanHT-12 V3.0 R2 Array for the METABRIC dataset and the *Agilent custom 244K* for the TCGA dataset). The CA20 and CIN25 scores were calculated as the sum of the normalized (\log_2 median-centered) expression levels of the signature genes. **METABRIC dataset:** Probes were filtered by preferentially selecting those targeting all

isoforms (A designation). For some genes, only probes targeting some isoforms (S designation) were available. In addition, probes mapping only to the gene of interest according to a BLAST-like Alignment Tool (BLAT) search against the reference genome GRCh38 using Ensembl [47] were preferentially selected. When multiple probes mapped to the gene of interest, the average expression level was calculated to represent that gene. **TCGA dataset:** In the absence of A and S designations, probes were filtered by performing a BLAT search as for the METABRIC data, also with averaging of normalized expression levels when multiple probes mapped to the gene of interest. Scores exhibited negative values, so for ease of interpretation, scores were converted to non-negative values by adding the minimum score value to all scores, which did not alter the results of statistical analyses.

2.5.3 Survival analyses

Stratification of cases into high- and low-survival groups both by the average (as performed in the CIN25 analyses previously [12]) and by an optimal cutpoint (based on the most significant log-rank test statistic found using Cutoff Finder [48]) per the Kaplan-Meier method. Prior to fitting Cox proportional-hazards models, the proportional-hazards assumption was tested by defining each covariate as a function of time and entering this time-dependent term into a simple Cox model and determining whether there was a significant hazard in the discovery and validation sets. For no covariate was the assumption violated (data not shown). Pearson correlation (2-tailed) was performed to determine the correlation between CA20 and CIN25 scores. IBM SPSS Statistics version 21 was used for all analyses, and $p < 0.05$ was considered statistically significant. **METABRIC dataset:** Multivariable Cox models were fit using both discovery set data via backward-stepwise elimination of covariates (subject to an $\alpha = 0.10$ removal

criterion) and validation set data by entering final discovery model covariates. Full multivariable discovery model covariates included age at diagnosis (years), Nottingham grade (1, 2, or 3), AJCC stage (0, I, II, or III/IV, the latter two categories combined due to the relatively small number of stage IV cases), breast cancer subtype (luminal: ER and/or PR+, HER2-; HER2-enriched: ER/PR+/-, HER2+; triple-negative: ER/PR/HER2-), chemotherapy (yes/no), hormone therapy (yes/no), and radiotherapy (yes/no). Also, depending on the analysis, the full model either contained CA20 and CIN25 categorized based on the average score as found in the discovery set or CA20 and CIN25 categorized based on the optimal cutpoint as found in the discovery set. **TCGA dataset:** To confirm the prognostic ability of CA20 in an independent dataset, multivariable Cox models were fit using TCGA data by entering CA20 and CIN25 (average or optimal, depending on the model) and AJCC stage (categorized as I/II vs. III/IV due to relatively low numbers of stage I and IV patients) into full models (chemotherapy information was not available). Covariates were then subjected to backward-stepwise elimination ($\alpha=0.10$ removal criterion). Multivariable models were also fit including age at diagnosis (years). To more robustly estimate population parameters, the final model covariates were entered into Cox models with simple bootstrapping (1,000 iterations).

2.5.4 Grade-wise comparison of average CA20 and CIN25 between TNBCs and non-TNBCs

Using the METABRIC dataset, as grade information was not available for the TCGA dataset, we compared average CA20 and average CIN25 between TNBCs and non-TNBCs grade-wise using two-tailed independent samples t-tests, guided by F-test results, and $p<0.05$ was considered statistically significant.

2.5.5 Gene Set Enrichment Analyses

Normalized (level 3) TCGA Hi-Seq data downloaded from the TCGA Data Portal were used for GSEA, although CA20 and CIN25 groups were specified based on average scores obtained from normalized OncoPrint data. The Broad Institute GSEA software version 2.2.3 was used. All 20,530 genes in the dataset were used. With the exception of not collapsing the dataset to gene symbols, all other default settings were used. Gene set databases included biological process gene ontologies (c5.bp.v5.2.symbols), Reactome pathways (c2.cp.reactome.v5.2.symbols), and oncogenic signatures (c6.all.v5.2.symbols). For the CIN25 analysis, the centrosome gene set was also used (<http://amigo.geneontology.org/amigo/term/GO:0005813>). FDR $q < 0.05$ was considered statistically significant.

2.6 References

- 1 Chan, J. Y. A Clinical Overview of Centrosome Amplification in Human Cancers. *Int. J. Biol. Sci.* **7**, 1122-1144 (2011).
- 2 Godinho, S. A. *et al.* Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* **510**, 167-171, doi:10.1038/nature13277 (2014).
- 3 Pannu, V. *et al.* Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers. *Oncotarget* **6**, 10487-10497 (2015).
- 4 Ganem, N. J., Godinho, S. A. & Pellman, D. A Mechanism Linking Extra Centrosomes to Chromosomal Instability. *Nature* **460**, 278-282, doi:10.1038/nature08136 (2009).
- 5 Bakhoun, S. F. & Compton, D. A. Chromosomal instability and cancer: a complex relationship with therapeutic potential. *J. Clin. Invest.* **122**, 1138-1143, doi:10.1172/JCI59954 (2012).
- 6 Sercin, O. *et al.* Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nat. Cell Biol.* **18**, 100-110, doi:10.1038/ncb3270 (2016).
- 7 Chng, W. J. *et al.* The centrosome index is a powerful prognostic marker in myeloma and identifies a cohort of patients that might benefit from aurora kinase inhibition. *Blood* **111**, 1603-1609, doi:10.1182/blood-2007-06-097774 (2008).
- 8 Chng, W. J. *et al.* Clinical implication of centrosome amplification in plasma cell neoplasm. *Blood* **107**, 3669-3675, doi:10.1182/blood-2005-09-3810 (2006).
- 9 Bauer, M., Cubizolles, F., Schmidt, A. & Nigg, E. A. Quantitative analysis of human centrosome architecture by targeted proteomics and fluorescence imaging. *EMBO J.* **35**, 2152-2166, doi:10.15252/embj.201694462 (2016).

- 10 Pihan, G. Centrosome Dysfunction Contributes to Chromosome Instability, Chromoanagenesis, and Genome Reprograming in Cancer. *Front. Oncol.* **3**, doi:10.3389/fonc.2013.00277 (2013).
- 11 Lawo, S., Hasegan, M., Gupta, G. D. & Pelletier, L. Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material. *Nat. Cell. Biol.* **14**, 1148-1158, doi:10.1038/ncb2591 (2012).
- 12 Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043-1048 (2006).
- 13 Pannu, V. *et al.* Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers. *Oncotarget* **6**, 10487-10497, doi:10.18632/oncotarget.3402 (2015).
- 14 Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 15 Godinho, S. A. & Pellman, D. Causes and consequences of centrosome abnormalities in cancer. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369**, 20130467, doi:10.1098/rstb.2013.0467 (2014).
- 16 Denu, R. A. *et al.* Centrosome amplification induces high grade features and is prognostic of worse outcomes in breast cancer. *BMC Cancer* **16**, 47, doi:10.1186/s12885-016-2083-x (2016).
- 17 Ogden, A. *et al.* Quantitative multi-parametric evaluation of centrosome declustering drugs: centrosome amplification, mitotic phenotype, cell cycle and death. *Cell Death Dis.* **5**, e1204, doi:10.1038/cddis.2014.164 (2014).
- 18 Pannu, V. *et al.* Centrosome-declustering drugs mediate a two-pronged attack on interphase and mitosis in supercentrosomal cancer cells. *Cell Death Dis.* **5**, e1538, doi:10.1038/cddis.2014.505 (2014).
- 19 Rebacz, B. *et al.* Identification of griseofulvin as an inhibitor of centrosomal clustering in a phenotype-based screen. *Cancer Res.* **67**, 6342-6350, doi:10.1158/0008-5472.can-07-0663 (2007).
- 20 Castiel, A. *et al.* A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells. *BMC Cancer* **11**, 412, doi:10.1186/1471-2407-11-412 (2011).
- 21 Watts, Ciorsdaidh A. *et al.* Design, Synthesis, and Biological Evaluation of an Allosteric Inhibitor of HSET that Targets Cancer Cells with Supernumerary Centrosomes. *Chem. & Biol.* **20**, 1399-1410, doi:10.1016/j.chembiol.2013.09.012 (2013).
- 22 Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1-6 (2004).
- 23 Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352, doi:10.1038/nature10983 (2012).
- 24 Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
- 25 Esserman, L. J. *et al.* Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res. Treat.* **132**, 1049-1062, doi:10.1007/s10549-011-1895-2 (2012).

- 26 Chow, S.-C. Sample size calculations for clinical trials. *Wiley Interdisc. Rev. Comput. Stat.* **3**, 414-427, doi:10.1002/wics.155 (2011).
- 27 Zhou, H. *et al.* Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.* **20**, 189-193, doi:10.1038/2496 (1998).
- 28 Meraldi, P., Honda, R. & Nigg, E. A. Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53^{-/-} cells. *EMBO J.* **21**, 483-492 (2002).
- 29 Hanashiro, K., Kanai, M., Geng, Y., Sicinski, P. & Fukasawa, K. Roles of cyclins A and E in induction of centrosome amplification in p53-compromised cells. *Oncogene* **27**, 5288-5302, doi:10.1038/onc.2008.161 (2008).
- 30 Nelsen, C. J. *et al.* Short term cyclin D1 overexpression induces centrosome amplification, mitotic spindle abnormalities, and aneuploidy. *J. Biol. Chem.* **280**, 768-776, doi:10.1074/jbc.M407105200 (2005).
- 31 Kawamura, K. *et al.* Induction of Centrosome Amplification and Chromosome Instability in Human Bladder Cancer Cells by p53 Mutation and Cyclin E Overexpression. *Cancer Res.* **64**, 4800-4809, doi:10.1158/0008-5472.can-03-3908 (2004).
- 32 Fukasawa, K. Oncogenes and tumour suppressors take on centrosomes. *Nat. Rev. Cancer* **7**, 911-924, doi:10.1038/nrc2249 (2007).
- 33 Löffler, H. *et al.* Cep63 Recruits Cdk1 to the Centrosome: Implications for Regulation of Mitotic Entry, Centrosome Amplification, and Genome Maintenance. *Cancer Res.* **71**, 2129-2139, doi:10.1158/0008-5472.can-10-2684 (2011).
- 34 Dzhindzhev, N. S. *et al.* Asterless is a scaffold for the onset of centriole assembly. *Nature* **467**, 714-718, doi:10.1038/nature09445 (2010).
- 35 Lee, M.-Y., Moreno, C. S. & Saavedra, H. I. E2F Activators Signal and Maintain Centrosome Amplification in Breast Cancer Cells. *Mol. Cell. Biol.* **34**, 2581-2599, doi:10.1128/MCB.01688-13 (2014).
- 36 Montanez-Wiscovich, M. E. *et al.* Aberrant expression of LMO4 induces centrosome amplification and mitotic spindle abnormalities in breast cancer cells. *J. Pathol.* **222**, 271-281, doi:10.1002/path.2762 (2010).
- 37 Carroll, P. E. *et al.* Centrosome hyperamplification in human cancer: chromosome instability induced by p53 mutation and/or Mdm2 overexpression. *Oncogene* **18**, 1935-1944, doi:10.1038/sj.onc.1202515 (1999).
- 38 Slack, A. D., Chen, Z., Ludwig, A. D., Hicks, J. & Shohet, J. M. MYCN-Directed Centrosome Amplification Requires MDM2-Mediated Suppression of p53 Activity in Neuroblastoma Cells. *Cancer Res.* **67**, 2448-2455, doi:10.1158/0008-5472.can-06-1661 (2007).
- 39 Croessmann, S. *et al.* NDRG1 links p53 with proliferation-mediated centrosome homeostasis and genome stability. *Proc. Natl. Acad. Sci. USA* **112**, 11583-11588, doi:10.1073/pnas.1503683112 (2015).
- 40 Harrison Pitner, M. K. & Saavedra, H. I. Cdk4 and nek2 signal binucleation and centrosome amplification in a her2⁺ breast cancer model. *PLoS One* **8**, e65971, doi:10.1371/journal.pone.0065971 (2013).
- 41 Suizu, F., Ryo, A., Wulf, G., Lim, J. & Lu, K. P. Pin1 Regulates Centrosome Duplication, and Its Overexpression Induces Centrosome Amplification, Chromosome

- Instability, and Oncogenesis. *Mol. Cell. Biol.* **26**, 1463-1479, doi:10.1128/MCB.26.4.1463-1479.2006 (2006).
- 42 Liu, X. & Erikson, R. L. Activation of Cdc2/cyclin B and inhibition of centrosome amplification in cells depleted of Plk1 by siRNA. *Proc. Natl. Acad. Sci. USA* **99**, 8672-8676, doi:10.1073/pnas.132269599 (2002).
- 43 Lončarek, J., Hergert, P. & Khodjakov, A. Centriole reduplication during prolonged interphase requires procentriole maturation governed by Plk1: Plk1 in procentriole maturation. *Curr. Biol.* **20**, 1277-1282, doi:10.1016/j.cub.2010.05.050 (2010).
- 44 Habedanck, R., Stierhof, Y. D., Wilkinson, C. J. & Nigg, E. A. The Polo kinase Plk4 functions in centriole duplication. *Nat. Cell Biol.* **7**, 1140-1146, doi:10.1038/ncb1320 (2005).
- 45 Shinmura, K. *et al.* SASS6 overexpression is associated with mitotic chromosomal abnormalities and a poor prognosis in patients with colorectal cancer. *Oncol. Rep.* **34**, 727-738, doi:10.3892/or.2015.4014 (2015).
- 46 Tang, C.-J. C. *et al.* The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. *EMBO J.* **30**, 4790-4804, doi:10.1038/emboj.2011.378 (2011).
- 47 Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The Ensembl Regulatory Build. *Genome Biology* **16**, 56, doi:10.1186/s13059-015-0621-5 (2015).
- 48 Budczies, J. *et al.* Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization. *PLoS ONE* **7**, e51862, doi:10.1371/journal.pone.0051862 (2012).

Table 2.1 CA20 genes

Gene	Experimental manipulation	Centrosomal phenotype	Possible mechanism	References
AURKA	Overexpression of AURKA in mouse NIH 3T3 cells and human MCF10A breast cells; overexpression of AURKA in synchronized HeLa cells	Increased proportion of cells with >2 centrosomes	Tetraploidization	Zhou et al., 1998; Meraldi et al., 2002
CCNA2	Knockdown of CCNA2 during DNA damage-induced G2 arrest in murine embryonic fibroblasts	Abrogation of centrosome overduplication (cells with >2 centrosomes)	Centrosome overduplication	Hanashiro et al., 2008
CCND1	Overexpression of CCND1 in hepatocytes	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Nelsen et al., 2005
CCNE2	Overexpression of CCNE2 and siRNA of TP53 in bladder cancer cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Kawamura et al., 2004; Nelsen et al., 2005
CDK1	Overexpression of CDK1 in osteosarcoma cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Loffler et al., 2011
CEP63	Overexpression of CEP63 in osteosarcoma cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Loffler et al., 2011
CEP152	Overexpression of CEP152 in osteosarcoma cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Dzhindzhev et al., 2010
E2F1	Overexpression of E2F1 in mammary epithelial cells	Increased proportion of cells with >2 centrosomes	Binucleation	Lee et al., 2014
E2F2	Overexpression of E2F2 in mammary epithelial cells	Increased proportion of cells with >2 centrosomes	Binucleation	Lee et al., 2014
LMO4	Overexpression of LMO4 in breast cancer cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Montañez-Wiscovich et al., 2010
MDM2	Overexpression of MDM2 in murine fibroblasts; pharmacologic inhibition of	Increased proportion of cells with >2 centrosomes;	Centrosome overduplication	Carroll et al., 1999; Slack et al., 2007

	MDM2-p53 interaction in MYCN-overexpressing cells subjected to ionizing radiation	decreased proportion of cells with >2 centrosomes		
MYCN	Overexpression of MYCN in neuroblastoma and p53 ^{+/+} colon cancer cells and subjection to ionizing radiation	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Slack et al., 2007
NDRG1	Knockdown of NDRG1 in breast epithelial and colorectal cancer cells	Increased proportion of cells with >2 centrosomes	Suppression of centrosome overduplication	Croessmann et al., 2015
NEK2	Overexpression in breast cancer cells	Increased proportion of cells with >2 centrosomes	Binucleation	Harrison Pitner et al., 2013
PIN1	Overexpression of PIN1 and induction of S-phase arrest in murine embryonic fibroblasts and overexpression of PIN1 in mammary epithelial cells (transgenic mice)	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Suizu et al., 2006
PLK1	Knockdown of PLK1 in S phase-arrested osteosarcoma cells; constitutive activation in S phase-arrested osteosarcoma cells	Increased proportion of cells with >2 centrosomes or >4 centrioles	Centrosome/centriole overduplication	Liu et al., 2002; Loncarek et al., 2010
PLK4	Overexpression of PLK4 in osteosarcoma cells	Increased proportion of cells with >2 centrosomes or >4 centrioles	Centriole/centrosome overduplication	Habedanck et al., 2005; Dzhindzhev et al., 2010
SASS6	Overexpression of SASS6 in colorectal cancer cells	Increased proportion of cells with >2 centrosomes	Centrosome overduplication	Shinmura et al., 2015
STIL	Overexpression of STIL in osteosarcoma cells	Increased proportion of cells with >4 centrioles	Centriole overduplication	Tang et al., 2011
TUBG1	Most abundant centrosome structural protein	Marker for centrosome amplification	Excessive accumulation or increased copy number of centrosomes	Bauer et al., 2016

Table 2.2 Descriptive statistics for the METABRIC breast discovery and validation sets

Variable	Level (categorical variables) or Statistic (continuous variables)	Discovery	Validation
Nottingham grade	1	80	86
	2	404	363
	3	464	486
	Missing	37	51
Tumor stage	0	230	262
	I	188	174
	II	299	274
	III	45	41
	IV	4	6
Chemotherapy	Missing	219	229
	No	764	790
	Yes	221	196
Hormone therapy	Missing		
	No	378	375
	Yes	607	611
Radiotherapy	Missing	0	0
	No	406	399
	Yes	579	587
Subtype	Missing	0	0
	Luminal	647	623
	TNBC	126	124
	HER2	207	228
CA20 group (average)	Missing	5	11
	Low	234	243
	High	751	743
CIN25 group (average)	Missing	0	0
	Low	337	329
	High	648	657
CA20 group (optimal)	Missing	0	0
	Low	274	278
	High	711	708
CIN25 group (optimal)	Missing	0	0
	Low	405	402
	High	580	584
	Mean	60.8	61.5

	Median	61.28	62.56
	Standard Deviation	13.01	12.92
Age at diagnosis (years)	Minimum	21.93	26.36
	Maximum	92.14	96.29
	Number Missing	0	0
	Mean	30.46	30.33
	Median	30.08	29.86
CA20	Standard Deviation	5.74	5.83
	Minimum	17.27	17.7
	Maximum	49.25	48.45
	Number Missing	0	0
	Mean	53.3	53.05
	Median	52.84	52.9
CIN25	Standard Deviation	13.08	13.45
	Minimum	22.16	20.9
	Maximum	90.67	88.46
	Number Missing	0	0

Table 2.3 Descriptive statistics for the TCGA breast dataset

Variable	Level (categorical variables) or Statistic (continuous variables)	Value
AJCC stage	I	49
	II	241
	III	92
	IV	13
	Missing	129
CA20 group (average)	Low	240
	High	284
	Missing	0
CA20 group (optimal)	Low	238
	High	286
	Missing	0
CIN25 group (average)	Low	256
	High	268
	Missing	0
CIN25 group (optimal)	Low	514
	High	10
	Missing	0
Age at diagnosis (years)	Mean	58.11
	Median	59
	Standard Deviation	13.19
	Minimum	26
	Maximum	90
	Number missing	74
	Mean	23.66
Median	24.57	
CA20 score	Standard Deviation	8.91
	Minimum	0
	Maximum	45.7
	Number missing	0
	Mean	45.37
CIN25 score	Median	46.07
	Standard Deviation	16.71
	Minimum	0
	Maximum	88.52
	Number missing	0

Table 2.4 Final multivariable Cox proportional-hazards models of breast cancer-specific survival including CA20 and CIN25 (based on average scores) in full models using METABRIC data

Covariates	Discovery set				Validation set			
	p-value	HR	95% CI for HR		p-value	HR	95% CI for HR	
			Lower	Upper			Lower	Upper
CA20 (high)	0.006	2.13	1.24	3.68	0.028	1.82	1.07	3.12
CIN25 (high)	0.053	1.49	0.99	2.22	0.29	1.27	0.81	1.97
Stage 0	<0.001				<0.001			
Stage I	0.001	0.48	0.31	0.75	0.005	0.53	0.34	0.82
Stage II	0.003	0.58	0.41	0.83	0.78	0.95	0.68	1.34
Stage III/IV	0.002	2.18	1.32	3.58	0.001	2.27	1.39	3.68
Chemotherapy	<0.001	2.24	1.57	3.19	0.002	1.73	1.22	2.44
Hormone therapy	0.04	1.38	1.01	1.88	0.71	0.94	0.7	1.27
Radiotherapy	0.059	0.73	0.53	1.01	0.51	1.11	0.81	1.51

Table 2.5 Final multivariable Cox proportional-hazards models of breast cancer-specific survival including CA20 and CIN25 (based on optimal thresholds) in full models using METABRIC data

Covariates	Discovery set				Validation set			
	p-value	HR	95% CI for HR		p-value	HR	95% CI for HR	
			Lower	Upper			Lower	Upper
CA20 (high)	0.006	2.13	1.24	3.68	0.028	1.82	1.07	3.12
CIN25 (high)	0.053	1.49	0.99	2.22	0.29	1.27	0.81	1.97
Stage 0	<0.001				<0.001			
Stage I	0.001	0.48	0.31	0.75	0.005	0.53	0.34	0.82
Stage II	0.003	0.58	0.41	0.83	0.78	0.95	0.68	1.34
Stage III/IV	0.002	2.18	1.32	3.58	0.001	2.27	1.39	3.68
Chemotherapy	<0.001	2.24	1.57	3.19	0.002	1.73	1.22	2.44
Hormone therapy	0.04	1.38	1.01	1.88	0.71	0.94	0.7	1.27
Radiotherapy	0.059	0.73	0.53	1.01	0.51	1.11	0.81	1.51

Table 2.6 Cox models of overall survival including CA20 and CIN25 (based on average or optimal thresholds) and AJCC stage in full models using TCGA data

Model	Covariates	HR	p-value	95% CI for HR		Bootstrap p-value	Bootstrap 95% CI for HR	
				Lower	Upper		Lower	Upper
CA20 and CIN25 (average)	CA20 (high)	2.8	0.008	1.31	5.99	0.01	1.31	7.43
	Stage III/IV	2.68	0.006	1.33	5.41	0.011	1.13	6.5
CA20 and CIN25 (optimal)	CA20 (high)	2.55	0.009	1.27	5.15	0.01	1.34	7.29
	Stage III/IV	2.87	0.008	1.32	6.25	0.012	1.14	6.16

Table 2.7 Final multivariable Cox proportional-hazards models of overall survival including CA20 and CIN25 (based on average or optimal thresholds), AJCC stage, and age at diagnosis in full models using TCGA data

Model	Covariates	HR	p-value	95% CI for HR		Bootstrap p-value	Bootstrap 95% CI for HR	
				Lower	Upper		Lower	Upper
CA20 and CIN25 (average)	CA20 (high)	3.82	0.001	1.69	8.67	0.007	1.7	13.11
	Age at diagnosis	1.04	0.01	1.01	1.07	0.028	1	1.08
	Stage III/IV	2.74	0.005	1.35	5.55	0.003	1.22	6.85
CA20 and CIN25 (optimal)	CA20 (high)	3.67	0.002	1.63	8.26	0.003	1.6	12.83
	Age at diagnosis	1.04	0.015	1.01	1.07	0.032	1	1.07
	Stage III/IV	2.5	0.011	1.24	5.05	0.009	1.14	6.05

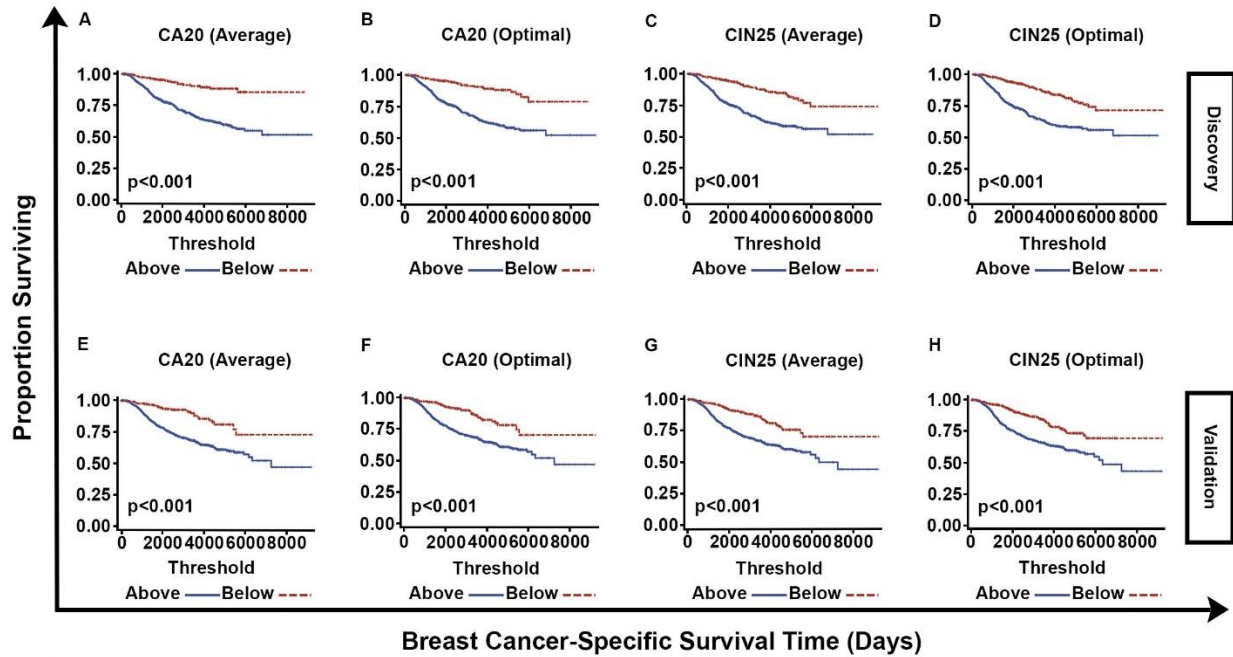


Figure 2.1 Plots of Kaplan-Meier product limit estimates of breast cancer-specific survival of patients in METABRIC discovery and validation sets Stratified by (A, E) CA20 (optimal threshold), (B, F) CA20 (average value), (C, G) CIN25 (optimal threshold), and (D, H) CIN25 (average value), respectively. Optimal thresholds and average values were determined using the discovery set.

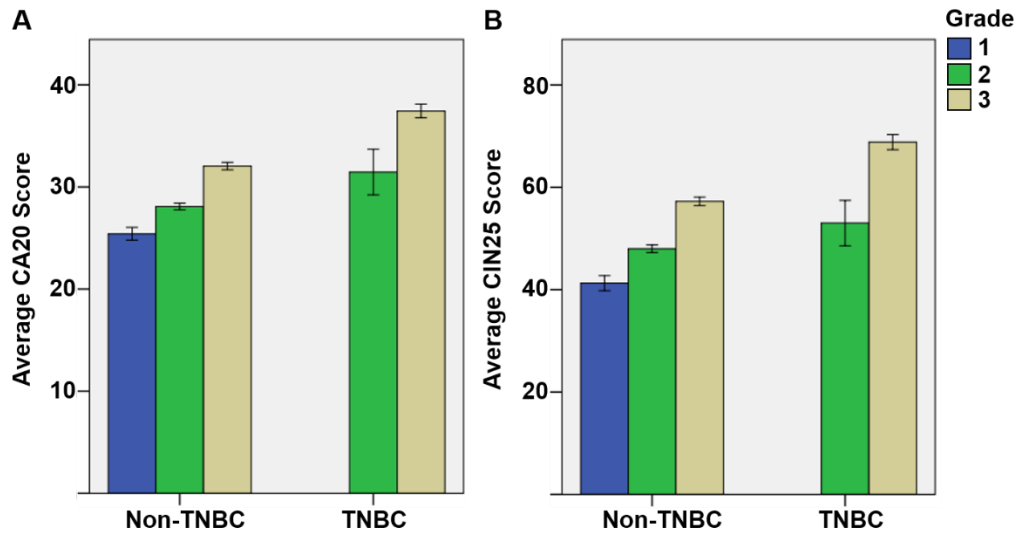


Figure 2.2 Grade-wise comparisons of (A) average CA20 score and (B) average CIN25 score in non-TNBCs vs. TNBCs in the METABRIC dataset

Scores between grades were significantly different at the $p < 0.001$ level. Error bars represent 95% confidence intervals.

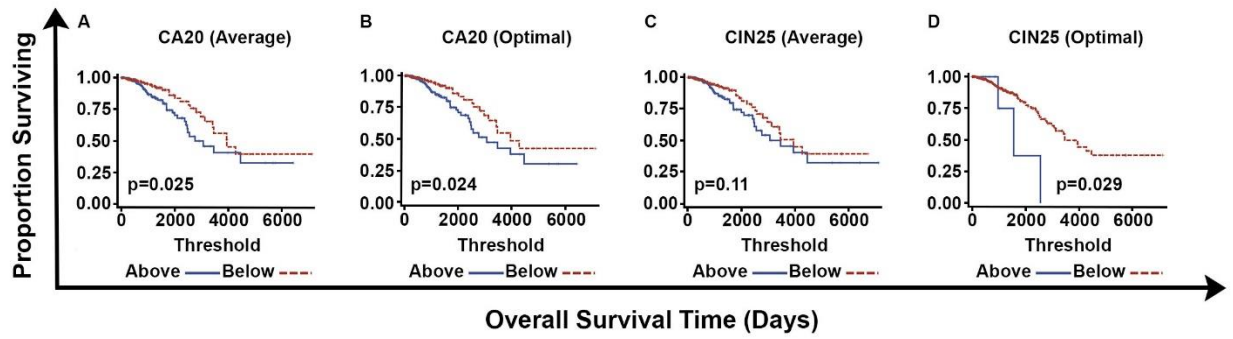


Figure 2.3 Plots of Kaplan-Meier product limit estimates of overall survival of patients in TCGA dataset

Stratified by (A) CA20 (optimal threshold), (B) CA20 (average value), (C) CIN25 (optimal threshold), and (D) CIN25 (average value) determined using the entire dataset.

3 MULTI-INSTITUTIONAL STUDY OF NUCLEAR KIFC1 AS A BIOMARKER OF POOR PROGNOSIS IN AFRICAN AMERICAN WOMEN WITH TRIPLE-NEGATIVE BREAST CANCER³

3.1 Abstract

Nuclear KIFC1 (nKIFC1) predicts worse outcomes in breast cancer, but its prognostic value within racially distinct triple-negative breast cancer (TNBC) patients is unknown. Thus, nKIFC1 expression was assessed by immunohistochemistry in 163 African American (AA) and 144 White TNBC tissue microarrays (TMAs) pooled from four hospitals. nKIFC1 correlated significantly with Ki67 in White TNBCs but not in AA TNBCs, suggesting that nKIFC1 is not merely a surrogate for proliferation in AA TNBCs. High nKIFC1 weighted index (WI) was associated with significantly worse overall survival (OS), progression-free survival (PFS), and distant metastasis-free survival (DMFS) (Hazard Ratios [HRs] = 3.5, 3.1, and 3.8, respectively P=0.01, 0.009, and 0.007, respectively) in multivariable Cox models in AA TNBCs but not White TNBCs. Furthermore, KIFC1 knockdown more severely impaired migration in AA TNBC cells than White TNBC cells. Collectively, these data suggest that nKIFC1 WI an independent biomarker of poor prognosis in AA TNBC patients, potentially due to the necessity of KIFC1 for migration in AA TNBC cells

3.2 Introduction

KIFC1, a minus end-directed microtubule motor belonging to the C-terminal kinesin subfamily [1], crosslinks and slides microtubules in mammalian meiotic and mitotic spindles,

³Parts of this chapter have been published verbatim in *Scientific Reports* 2017 Feb 20;7:42289. doi: 10.1038/srep42289 as “Multi-institutional study of nuclear KIFC1 as a biomarker of poor prognosis in African American women with triple-negative breast cancer.”

facilitating tight pole focusing [2]. This kinesin can also bind to and traffic early endocytic vesicles [3] and DNA oligonucleotides [4] along microtubules. Recently, attention has focused on KIFC1 due to its association with malignancy. High KIFC1 transcript levels in lung tumors predict increased risk of metastatic dissemination to the brain [5] Primary breast tumors also overexpress KIFC1 as compared with matched normal breast tissue [6], and TNBCs express higher KIFC1 than non-TNBCs [7]. nKIFC1 expression correlates with advanced tumor grade and worse overall and progression-free survival in breast cancer [6]. KIFC1-overexpressing MDA-MB-231 and MDA-MB-468 TNBC cells exhibit enhanced survival compared with vector controls [8]. KIFC1 may contribute to apoptosis reluctance in TNBC because KIFC1 overexpression stabilizes survivin by decreasing its polyubiquitination in MDA-MB-231 cells [6]. KIFC1 promotes the survival of TNBC cells with supernumerary centrosomes, which rely on KIFC1 for clustering of supernumerary centrosomes, which facilitates chromosomal instability [9, 10]. Indeed, *KIFC1* was identified as the top hit in a genome-wide *Drosophila* screen of centrosome clustering genes, and it is essential for clustering of supernumerary centrosomes and thereby suppression of multipolar division of human cancer cell lines [9].

AA women with breast cancer experience a more aggressive clinical course than White women with breast cancer partly due to the ~2-3-fold increased risk AA women have of developing TNBC, a subtype with a high risk of distant relapse and mortality; nonetheless, racial disparity may exist even within this subtype [11, 12]. A critical need exists for validation of novel biomarkers to risk-stratify AA breast cancer patients because they experience higher breast cancer mortality than any other racial group, which might indicate that high-risk AA patients are not being identified as such using standard clinical prognostic tools and are thus not being

prescribed sufficiently aggressive treatment. TNBC remains defined by what biomarkers it lacks, whereas non-TNBCs are defined by expression of hormone receptors and/or HER2, which can be targeted with specific inhibitors. Treatment guidelines in the US stratify TNBC patients for systemic adjuvant treatment based primarily on tumor staging. The association of nKIFC1 with triple-negative receptor status and worse clinical outcomes in breast cancer suggests that nKIFC1 drives aggressive breast cancer disease course and may potentially serve as a prognostic biomarker in TNBC, although its utility in specific racial groups is unclear. Given that AA women with breast cancer suffer a more aggressive disease course than White women, we hypothesized that KIFC1 would hold greater value as a prognostic biomarker in AA women with TNBC. Given the association between KIFC1 and brain metastases in lung cancer, we also hypothesized that KIFC1 is critical for TNBC cell migration. Herein, we analyze nKIFC1 expression by immunohistochemistry in race-annotated TNBC specimens to test its association with race, standard clinical prognostic factors, and clinical outcomes within racial groups and determine the effect of KIFC1 knockdown on migration and proliferation in White and AA TNBC cell lines.

3.3 Results

We found that nKIFC1 WI was significantly higher in AA than White TNBCs (154.66 vs. 133.74, respectively, $p=0.036$) (**Figure 3.1**, representative staining). In Whites, nKIFC1 WI was significantly higher in grade 3 than grade 1 and 2 TNBCs ($p=0.035$ and $p=0.001$, respectively, per post-hoc comparison), whereas nKIFC1 WI did not significantly differ by grade in AA TNBCs (**Table 3.1**). nKIFC1 WI was significantly higher in high-stage AA TNBCs, whereas nKIFC1 WI did not significantly differ by stage in White TNBCs. In both races,

nKIFC1 WI did not significantly differ by lymph node status, and nKIFC1 WI was not significantly correlated with tumor size or age at diagnosis.

Among a panel of 21 potential biomarkers, Ki67 WI alone was significantly correlated with nKIFC1 WI in White TNBCs ($\rho=0.65$, $p=0.00076$), whereas no significant correlations were found in AA TNBCs, suggesting that nKIFC1 is not merely a surrogate of any of these biomarkers in AA TNBCs.

nKIFC1 was not significantly associated with any survival outcome in simple Cox models (**Table 3.2**) or Kaplan-Meier analysis (**Figures 3.3-3.5**). However, in multivariable analysis adjusting for tumor stage, age at diagnosis, chemotherapy, and hospital, high nKIFC1 WI was associated with >3 times worse OS, PFS, and DMFS in AA TNBCs (bootstrap $p<0.05$ for all), whereas it did not significantly impact any survival outcome in White TNBCs.

To gain insights into why KIFC1 is prognostic in AA TNBCs but seemingly not White TNBCs, we performed proliferation and wound-healing assays with 2 AA and 2 White TNBC cell lines in which KIFC1 was transiently knocked down with siRNA (**Figure 3.2A**). We found that KIFC1 knockdown did not significantly impact proliferation in any of the cell lines (**Figure 3.2B**). By contrast, KIFC1 knockdown significantly impaired wound healing in one of the White TNBC cell lines and both AA TNBC cell lines; however, the percentage wound closure was more drastically reduced in AA TNBC cell lines ($p<0.001$) (**Figure 3.2C,D**). These results suggest that AA TNBC cells more heavily rely on KIFC1 for migration than white TNBC cells,

which may partly underlie the differential prognostic importance of KIFC1 by race as we found in clinical TNBC specimens.

3.4 Discussion

AA breast cancer patients are known to suffer increased cancer-related morbidity and mortality compared to all other racial groups even after adjusting for socioeconomic and lifestyle factors [13-15]. In particular, AA TNBC patients have a particularly dire prognosis, especially those with high-stage tumors. Given that AA women afflicted with breast cancer typically suffer from a much more aggressive disease course, it is critical to identify and validate biomarkers that could indicate differences in tumor biology between racial groups and serve as risk predictors that would help to stratify AA patients better and mitigate health disparity in disease outcomes.

In the clinic, establishment of triple-negative receptor status is a crucial factor determining treatment choice for patients because the absence of therapeutic targets in TNBC patients (a large fraction of whom are AA) leaves clinicians with no option but to resort to conventional chemotherapeutics that are non-discriminating and highly toxic. AA patients have also not benefited from the advances in breast cancer therapy as much as other patient groups, owing in large part to the high incidence of TNBC in this racial group. In our study, nKIFC1 was found to be a strong predictor of significantly worse survival outcomes in multivariable analysis in AA TNBCs. Our finding is valuable because it positions nKIFC1 not only as a critical racial disparity biomarker in TNBC but also as great potential therapeutic target. For the first time, AA patients whose tumors are triple-negative can be categorized based on what their tumors express, rather than what they lack (hormone receptors and Her2). Importantly, in AA TNBCs nKIFC1

WI does not significantly correlate with WIs of members a large panel of potential cancer biomarkers, suggesting it is not merely a surrogate for them and can offer novel prognostic information. It is remarkable that nKIFC1 significantly predicts survival outcomes in AA but not White TNBCs considering the extensive overlap in nKIFC1 scores between the two groups, since nKIFC1 was only slightly (albeit significantly) higher in AA than White TNBCs. This finding suggests that the differential prognostic value of nKIFC1 between these groups may arise from inherently different biology, possibly rooted in genetic ancestry, rather than substantially higher nKIFC1 levels in AA than White TNBCs. Our finding that AA TNBC cell lines are more reliant on KIFC1 for migration suggests provides possible mechanistic insights into why nKIFC1 levels are prognostic in AA but not White TNBCs. AA TNBC cells with high nKIFC1 levels may more efficiently migrate and thereby metastasize distantly, conferring worse prognosis. Our failure to reject the null hypothesis that there is no difference between nKIFC1-low and high groups among White TNBC patients aligns with our finding that KIFC1 is not entirely necessary for the migration of White TNBC cells; however, we may have failed to reject the null not because no true difference exists between nKIFC1-low and high White TNBC groups, but rather because the difference is very small or perhaps because of sampling or other errors, given that one White TNBC cell line relied to a significant extent on KIFC1 for migration. Larger future studies may be able to lend more confidence to our findings regarding a seeming lack of prognostic value of nKIFC1 among White TNBCs.

Our study, which has unraveled nuclear accumulation of the kinesin motor protein KIFC1 as a novel biomarker for cancer's unequal burden within the AA TNBC population, and which also demonstrates that KIFC1 is more important for the migration of AA than White TNBC cells,

opens up several important avenues of further investigation. An important next step is to determine whether nuclear KIFC1 accumulation is a cause or consequence of tumor aggressiveness. Several studies have implicated KIFC1 in disease aggressiveness and metastases, particularly to the brain.[5] Our data, which establish a tight correlation between KIFC1 nuclear expression and disease outcomes in AA TNBCs, affirms previous reports alluding to a connection between KIFC1 expression and disease aggressiveness [6]. Consequently, our findings make a strong case for examining whether a more direct causative link may exist between nKIFC1 and metastasis, especially given that we uncovered a critical role for KIFC1 in TNBC cell migration. Establishment of a causative role for nKIFC1 in metastasis would position nuclear KIFC1 as a biomarker that prognosticates distant metastasis risk in AA TNBC patients and anti-nKIFC1 targeted therapies as suppressors of further cancer cell dissemination.

nKIFC1 may have distinct functions in different phases of the cell cycle and at different subcellular locations. In mitosis, KIFC1 plays an essential role in the clustering of supernumerary centrosomes in cancer cells [9]. In doing so, KIFC1 not only ensures the survival of cancer cells with amplified centrosomes but may also function as a driver of tumor evolution, as low-grade chromosome missegregation and consequent generation of karyotypic heterogeneity may occur during clustering of amplified centrosomes [16, 17]. Based on these lines of evidence, the notion that KIFC1 overexpression is responsible for driving evolution of more aggressive phenotypes (e.g., within the group of AA TNBC patients) is plausible. KIFC1 also aids in the focusing of acentrosomal microtubule-organizing centers during the construction of a bipolar mitotic spindle in cancer cells [10, 18]. Since TNBCs display rampant centrosome amplification [19], targeting KIFC1 in TNBCs could provide tremendous therapeutic benefits by

inducing lethal levels of spindle multipolarity [20]. Most importantly, since KIFC1 is essential for tumor cell viability but is dispensable for the survival of non-cancerous cells [9], targeting KIFC1 would be an invaluable cancer cell-selective and non-toxic chemotherapeutic approach.

Regarding interphase activities, KIFC1 participates in vesicle transport from the Golgi to the ER in NIH3T3 murine embryonic fibroblasts [21] along with early endocytic vesicle transport and fission in murine hepatocytes [3]. Our immunohistochemistry data clearly show nuclear accumulation of KIFC1 as a negative prognostic marker in AA TNBCs, suggesting a possible interphase-specific nuclear role for this protein, although perhaps only within this specific patient population. In previous studies, KIFC1 has been implicated in the transport of DNA along microtubules [4], has been shown to bind nuclear transport factors such as β -importin [22] and to harbor a specific affinity for nucleoporins NUP50 and NUP153 [4] and a complex containing NUP62 [23]. However, the primary sequence of this fascinating protein does not reveal the presence of any recognizable DNA-binding domains. Thus, it is currently unclear what function, if any, KIFC1 serves in the interphase nucleus that might contribute to its ability to promote metastases.

Given the recent progress made in the development of KIFC1 inhibitors [18], it is also important to identify the patient population that might most benefit from KIFC1-targeted therapy. Pre-clinical development of KIFC1 inhibitors may now benefit from consideration of race and triple-negative status when choosing a disease model. For instance, initial mechanistic study of the first KIFC1 inhibitor implemented a TNBC cell line (BT-549), but it was derived

from a White patient. Future analyses of drug efficacy may find other models more favorable (e.g., MDA-MB-468 cells, which are both triple-negative and derived from an AA patient).

3.5 Materials and Methods

3.5.1 Specimen procurement and datasets

We pooled 163 AA and 144 White TNBC TMAs from multiple institutions, which has been shown to synergistically augment the ability of biomarkers to classify patients into risk groups [24]. Patients were treated between the years 2003-2008 at Grady Memorial and Emory University Hospitals and 2005-2013 at Northside Hospital (all in Atlanta, GA) and 2001-2012 at Baylor Scott & White Medical Center (Taylor, TX), whose respective Institutional Review Boards approved all aspects of the study. Patient consent was not required due to the archival nature of the de-identified samples. ER, PR, and HER2 staining and scoring were performed as to the prevalent ASCO/CAP guidelines at the time of sample collection. Specimen procurement was also commensurate with the guidelines in effect at the time of collection. Pathologic characteristics were reviewed by the pathologists of the respective hospitals. Descriptive statistics for continuous and categorical variables, including patient and pathologic characteristics, are available in **Tables 3.3 and 3.4**, respectively. The median follow up-time was 4.1 years.

3.5.2 Immunohistochemistry and specimen scoring

TMAs were deparaffinized, rehydrated in serial ethanol baths (100%, 90%, 75%, 50%) and then placed in citrate buffer (pH 6.0) in a pressure cooker at 15 psi for 30 min for antigen retrieval. nKIFC1 was immunostained at 1:1000 dilution in the corresponding author's lab using

rabbit polyclonal antibody, which was graciously provided by Dr. Claire Walczak (Indiana University) [25]. Enzymatic detection of the antibody was performed using the Universal LSAB2 Kit/HRP, Rabbit/Mouse (Dako, CA, US). A panel of potential breast cancer biomarkers (Androgen Receptor; Cytokeratins 5, 7, 8, 14, 18, and 19; CD44, c-Kit; Epidermal Growth Factor Receptor; Insulin-like Growth Factor Receptor; Ki67; Melatonin Receptor; p53; p63; P-cadherin; Survivin [nuclear]; Topoisomerase; Urokinase Receptor [nuclear]; Vimentin; and Zeb) had been previously immunostained [26]. Biomarkers were centrally reviewed and scored by the same two independent pathologists who were blind to clinical annotation. Scoring was performed for both the intensity of staining (0=none, 1=low, 2=moderate, 3=high) and the percentage of cells with any positivity (i.e., staining of 1+), and the average of the two pathologists' scores were taken. Weighted indices were calculated as the product of the staining intensity and percentage of positive cells. Representative images of nKIFC1 staining are provided in **Figure 3.1**.

3.5.3 Statistical analysis of clinical data

Because the distribution of nKIFC1 was strongly right-skewed for both races, non-parametric tests were employed: the Wilcoxon-Mann-Whitney or Kruskal-Wallis test for categorical covariates (with post-hoc pairwise comparisons made using the Dunn-Bonferroni approach) and Spearman correlation for numerical covariates. Monte-Carlo simulations (10,000 samples, 99% confidence interval [CI]) were performed to more robustly estimate all population parameters. Simple and multivariable Cox models of OS, PFS, and DMFS were also fit with bootstrapping (1,000 samples, 95% CI). Plots of partial residuals against rank time were approximately horizontal and centered near zero, indicating satisfaction of the proportional

hazards assumption. OS, PFS, and DMFS were defined, respectively, as the number of days from diagnosis to death, death or progression, and death or distant metastasis, whichever occurred first, or last follow-up if no event was recorded. The mean and +/- 1 standard deviation (SD) from the mean were tested as cutpoints to stratify patients based on OS. 1 SD above the mean demonstrated a non-significant trend among AA TNBCs, so that cutpoint was used. No trends were noted among White TNBCs, so the AA cutpoint was used for the purpose of comparison. For multivariable Cox models, covariates included AJCC stage, age at diagnosis, adjuvant chemotherapy, and hospital. Nottingham grade, which was not significant in full models, was not included due to multicollinearity with other covariates. The Bonferroni method was used to correct P-values following multiple hypothesis testing. Kaplan-Meier curves were also fit. Statistical analyses were performed using SPSS Version 21 (IBM). All relevant tests were 2-tailed and $p < 0.05$ was considered significant.

3.5.4 Statistical analysis of experimental data

Statistical bar graphs with the mean and error bars were plotted using Prism (GraphPad Software Inc., La Jolla, CA, USA). Experimental groups were compared with controls using unpaired two-tailed Student's t-test, and p-values were calculated using Prism. $p < 0.05$ was considered statistically significant.

3.5.5 Cell culture

White (HCC1143 and MDA-MB-231) and African American (HCC1906 and MDA-MB-468) cell lines were purchased from ATCC. MDA-MB-231 and MDA-MB-468 cells were cultured in Leibovitz's L-15 medium with 10% fetal bovine serum (ATCC) at 37 °C in 100% air.

HCC1806 and HCC1143 cell lines were cultured in Roswell Park Memorial Institute medium with 10% fetal bovine serum (ATCC) at 37 °C and 5% CO₂.

3.5.6 *siRNA*

RNAi knockdown of KIFC1 was performed using ON-TARGET plus KIFC1 siRNA. Non-targeting control *siRNAs* were used as controls. siRNA pools were transfected using 5 µg siRNA and 5 µl Dharmafect were used per well of a 6-well plate (resulting in a final siRNA concentration of 25 nM/well) according to the manufacturer's protocol. All reagents were purchased from Dharmacon (Lafayette, CO, USA). Efficiency of RNAi knockdown was analyzed after 36 h of transfection by Western blot analysis using antibodies specific to KIFC1. The knockdown efficiency was >90%.

3.5.7 *Western blotting and antibodies*

36 h after transfection, cells were scraped from culture plates and sonicated in lysate buffer with 1X protease inhibitor cocktail (Invitrogen). Polyacrylamide gel electrophoresis was performed for protein resolution, and then transfer was made to a polyvinylidene difluoride membrane. The membrane was incubated in 1:1,000 mouse monoclonal antibody specific for KIFC1 (M-63: sc-100947) from Santa Cruz Biotechnology (Dallas, TX, USA) with 1:10,000 goat anti-mouse secondary antibody (Abcam; Cambridge, MA, USA). B-actin was used as an internal loading control and was detected using rabbit monoclonal anti-B-actin antibody with goat anti-rabbit secondary antibody (both from Abcam). The Pierce ECL detection kit (Thermo Scientific; Waltham, MA, USA) was used to visualize bands.

3.5.8 *Cell proliferation assay*

Cell proliferation assay was performed using the BrdU Cell Proliferation Kit (#2750, EMD Millipore). Equal quantities of cells were seeded into 96 well plates. Following attachment, cells were incubated in BrdU for 4 hrs, and BrdU incorporation was measured spectrophotometrically as per kit instructions using TMB as a substrate. Absorbance was recorded at 450 nm. Each experiment was performed at least in triplicate.

3.5.9 *Wound healing assay*

Cells were seeded in 6-well plates and cultured until confluent, then they were serum starved for 2 h prior to the assay. Using a 100 ul pipette tip, a scratch was made keeping the pipette tip under an angle of around 30 degrees to keep the scratch width limited, simulating a wound. After scratching, the cells were washed in 1X Dulbecco's phosphate-buffered saline, serum-containing medium was added, and cells were incubated for 24 h in 5% CO₂ at 37 °C. Wound edges were imaged using a 10X objective. All the experiments were performed in triplicate and two values were measured from each scratched region.

3.6 **References**

1. DeLuca, J.G., et al., *Purification and characterization of native conventional kinesin, HSET, and CENP-E from mitotic hela cells*. J Biol Chem, 2001. **276**(30): p. 28014-21.
2. Mountain, V., et al., *The kinesin-related protein, HSET, opposes the activity of Eg5 and cross-links microtubules in the mammalian mitotic spindle*. J Cell Biol, 1999. **147**(2): p. 351-66.
3. Nath, S., et al., *Kif5B and Kifc1 interact and are required for motility and fission of early endocytic vesicles in mouse liver*. Mol Biol Cell, 2007. **18**(5): p. 1839-49.
4. Farina, F., et al., *Kinesin KIFC1 actively transports bare double-stranded DNA*. Nucleic Acids Res, 2013. **41**(9): p. 4926-37.

5. Grinberg-Rashi, H., et al., *The expression of three genes in primary non-small cell lung cancer is associated with metastatic spread to the brain*. Clin Cancer Res, 2009. **15**(5): p. 1755-61.
6. Pannu, V., et al., *HSET overexpression fuels tumor progression via centrosome clustering-independent mechanisms in breast cancer patients*. Oncotarget, 2015. **6**(8): p. 6076-91.
7. Wu, J., et al., *Discovery and mechanistic study of a small molecule inhibitor for motor protein KIF1C*. ACS Chem Biol, 2013. **8**(10): p. 2201-8.
8. De, S., et al., *Overexpression of kinesins mediates docetaxel resistance in breast cancer cells*. Cancer Res, 2009. **69**(20): p. 8035-42.
9. Kwon, M., et al., *Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes*. Genes Dev, 2008. **22**(16): p. 2189-203.
10. Kleylein-Sohn, J., et al., *Acentrosomal spindle organization renders cancer cells dependent on the kinesin HSET*. J Cell Sci, 2012. **125**(Pt 22): p. 5391-402.
11. Dietze, E.C., et al., *Triple-negative breast cancer in African-American women: disparities versus biology*. Nat Rev Cancer, 2015. **15**(4): p. 248-54.
12. Brewster, A.M., M. Chavez-MacGregor, and P. Brown, *Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry*. The Lancet. Oncology, 2014. **15**(13): p. e625-e634.
13. Newman, L.A., et al., *Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status*. J Clin Oncol, 2006. **24**(9): p. 1342-9.
14. Hershman, D.L., et al., *Treatment quality and outcomes of African American versus white breast cancer patients: retrospective analysis of Southwest Oncology studies S8814/S8897*. J Clin Oncol, 2009. **27**(13): p. 2157-62.
15. Chlebowski, R.T., et al., *Ethnicity and breast cancer: factors influencing differences in incidence and outcome*. J Natl Cancer Inst, 2005. **97**(6): p. 439-48.
16. Ganem, N.J., S.A. Godinho, and D. Pellman, *A mechanism linking extra centrosomes to chromosomal instability*. Nature, 2009. **460**(7252): p. 278-82.
17. Silkworth, W.T., et al., *Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells*. PLoS One, 2009. **4**(8): p. e6564.
18. Wu, J., et al., *Discovery and Mechanistic Study of a Small Molecule Inhibitor for Motor Protein KIF1C*. ACS Chem Biol, 2013.
19. Pannu, V., et al., *Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers*. Oncotarget, 2015. **6**(12): p. 10487-97.
20. Ogden, A., P.C. Rida, and R. Aneja, *Let's huddle to prevent a muddle: centrosome declustering as an attractive anticancer strategy*. Cell Death Differ, 2012. **19**(8): p. 1255-67.
21. Dorner, C., et al., *Characterization of KIF1C, a new kinesin-like protein involved in vesicle transport from the Golgi apparatus to the endoplasmic reticulum*. J Biol Chem, 1998. **273**(32): p. 20267-75.
22. Yang, W.X. and A.O. Sperry, *C-terminal kinesin motor KIF1C participates in acrosome biogenesis and vesicle transport*. Biol Reprod, 2003. **69**(5): p. 1719-29.

23. Yang, W.X., H. Jefferson, and A.O. Sperry, *The molecular motor KIFC1 associates with a complex containing nucleoporin NUP62 that is regulated during development and by the small GTPase RAN*. Biol Reprod, 2006. **74**(4): p. 684-90.
24. van Vliet, M.H., et al., *Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability*. BMC Genomics, 2008. **9**: p. 375.
25. Cai, S., et al., *Kinesin-14 family proteins HSET/XCTK2 control spindle length by cross-linking and sliding microtubules*. Mol Biol Cell, 2009. **20**(5): p. 1348-59.
26. Oprea-Ilies, G., et al., *Expression of melatonin receptors in triple negative breast cancer (TNBC) in African American and Caucasian women: relation to survival*. Breast Cancer Research and Treatment, 2013. **137**(3): p. 677-687.

Table 3.1 Relationships between nuclear KIFC1 weighted index and patient and clinicopathological factors

Categorical Variable	Level	N	nKIFC1 WI Mean Rank	p-value
Grade				
White	1	3	17.17	<0.001
	2	25	41.36	
	3	99	71.14	
AA	1	2	78.5	0.21
	2	24	65.46	
	3	135	83.8	
AJCC Stage				
White	I/II	115	62.5	1
	III/IV	9	62.44	
AA	I/II	109	75.3	0.02
	III/IV	52	92.94	
Nodes				
White	Negative	95	61.86	0.82
	Positive	27	60.22	
AA	Negative	96	75.31	0.11
	Positive	63	87.14	
Continuous Variable	N	Correlation with nKIFC1 WI (ρ)		p-value
Tumor size				
White	127	0.15		0.09
AA	161	0.13		0.09
Age at diagnosis				
White	144	-0.1		0.25
AA	161	-0.08		0.33

Table 3.2 Impact of Nuclear KIFC1 weighted index on survival outcomes

The impact of nuclear KIFC1 weighted index (~1 standard deviation above the mean in both racial groups) on overall survival (OS), progression-free survival (PFS), and distant metastasis-free survival (DMFS) in simple and multivariable Cox models adjusted for AJCC stage, age at diagnosis, adjuvant chemotherapy, and hospital at which treatment was received. Bonferroni-corrected p-values are displayed for simple Cox models of OS, for which multiple hypotheses were tested.

Outcome	Race	HR	p-value	95% CI	
				Lower	Upper
Simple Cox model					
OS	White	0.56	1	0.04	1.97
	AA	1.82	0.3	0.77	3.73
PFS	White	0.85	0.75	0.18	2.16
	AA	1.46	0.29	0.67	2.9
DMFS	White	1.17	0.81	0.04	3.63
	AA	1.55	0.22	0.73	3.02
Multivariable Cox model					
OS	White	0.28	0.15	8.97	1.31
	AA	3.45	0.01	1.4	20.8
PFS	White	0.65	0.5	0	2.08
	AA	3.14	0.009	1.07	9.18
DMFS	White	0.73	0.55	0	2.76
	AA	3.83	0.007	1.45	16.32

Table 3.3 Descriptive statistics for continuous variables by race.

Race	Nuclear HSET Group		Positive Lymph Nodes		Nottingham Grade			TNM Stage				Adjuvant Chemotherapy	
	Low	High	No	Yes	1	2	3	I	II	III	IV	No	Yes
White	122	22	95	27	5	27	110	81	47	9	4	41	102
AA	139	24	96	63	2	24	135	30	79	15	37	16	95

Table 3.4 Descriptive statistics for categorical variables by race.

Race	Nuclear HSET WI			Age at Diagnosis (years)			
	Median	Range	Missing	Mean (SD)	Median	Range	Missing
White	60	0-300	0	59 (13)	59	26-93	0
AA	60	0-210	0	54 (13)	54	25-90	2

A. White

B. AA

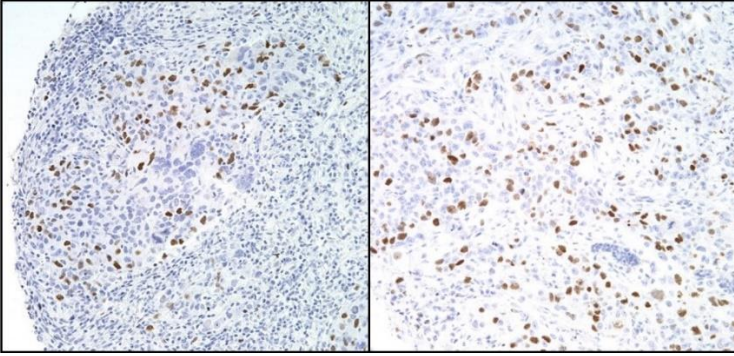


Figure 3.1 Representative images of nKIFC1 staining
(A) White and (B) African American (AA) triple-negative breast tumors. 20X objective.

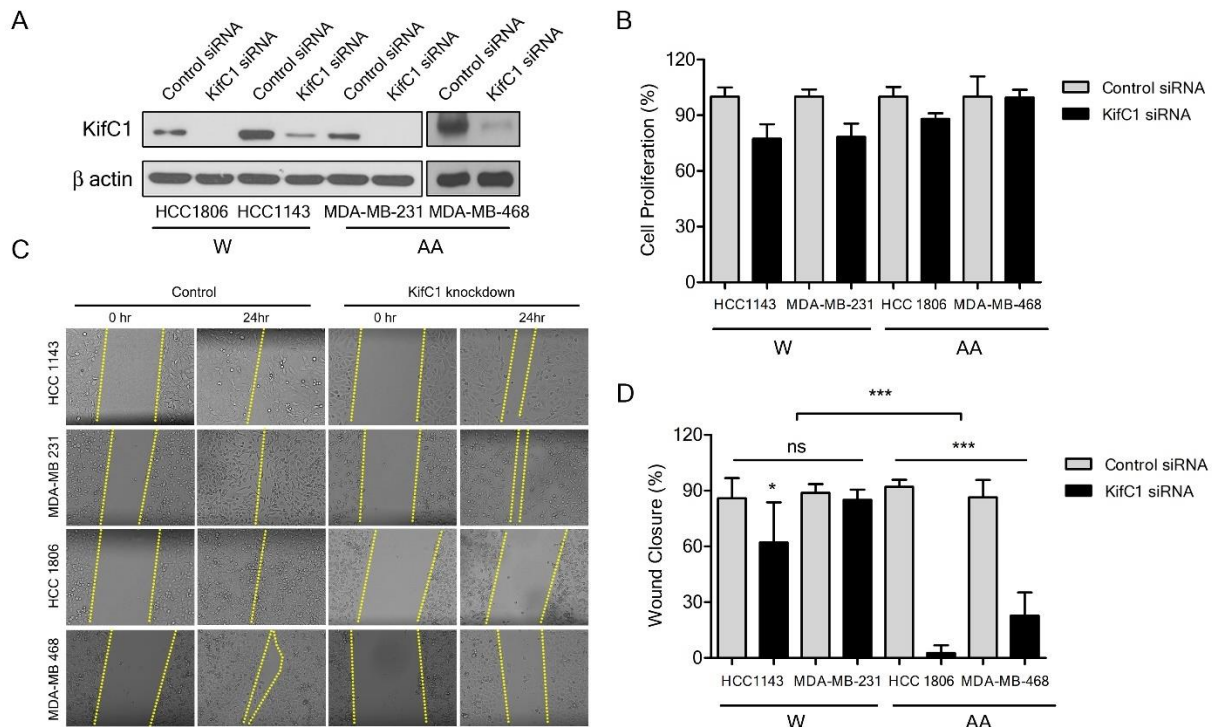


Figure 3.2 Impact of KIFC1 knockdown on proliferation and migration of White (W) (HCC1143 and MDA-MB-231) and African American (AA) (HCC1906 and MDA-MB-468) triple-negative breast cancer (TNBC) cell lines.

(A) Western blots of KIFC1 from White and AA TNBC cell lines treated with control or KIFC1 siRNA. (B) Proliferation of White and AA TNBC cell lines following treatment with control or KIFC1 siRNA. (C) Micrographs from wound healing assays of White and AA TNBC cell lines treated with control or KIFC1 siRNA. (D) Percentage wound closure in White and AA TNBC cell lines following treatment with control or KIFC1 siRNA. * $p < 0.01$; *** $p < 0.0001$; ns=not significant.

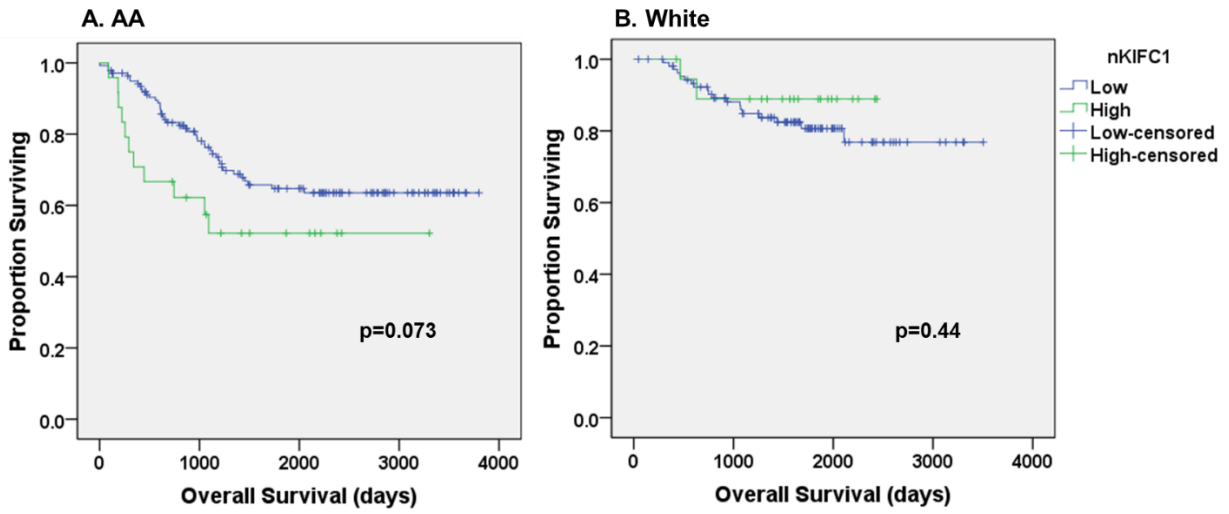


Figure 3.3 Kaplan-Meier curves of overall survival based on nKIFC1 weighted index (stratified by 1 standard deviation above the mean)

A. African American (AA) and **B.** White triple-negative breast cancer patients

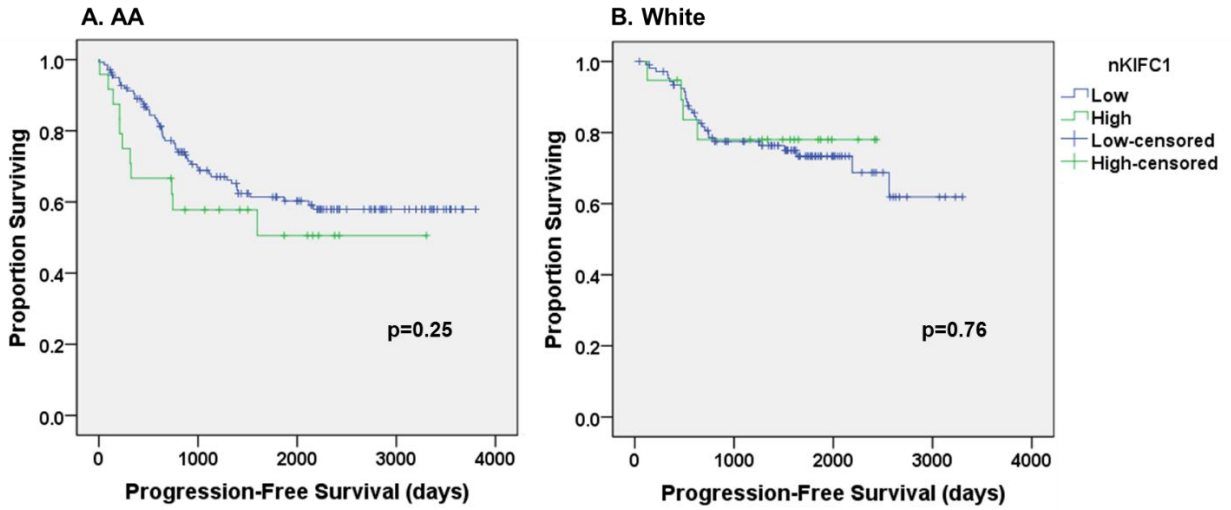


Figure 3.4 Kaplan-Meier curves of progression-free survival based on *nKIFC1* weighted index (stratified by 1 standard deviation above the mean)

A. African American (AA) and **B.** White triple-negative breast cancer patients

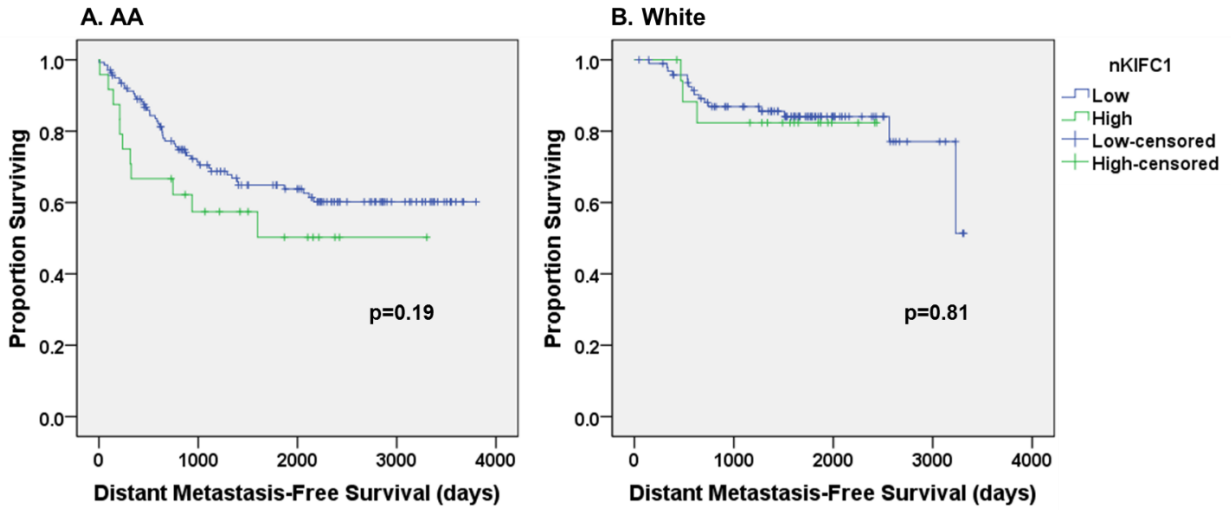


Figure 3.5 Kaplan-Meier curves of distant metastasis-free survival based on nKIFC1 weighted index (stratified by 1 standard deviation above the mean)

A. African American (AA) and **B.** White triple-negative breast cancer patients

4 QUANTITATIVE MULTI-PARAMETRIC EVALUATION OF CENTROSOME DECLUSTERING DRUGS: CENTROSOME AMPLIFICATION, MITOTIC PHENOTYPE, CELL CYCLE AND DEATH⁴

4.1 Abstract

Solid and hematological cancers often exhibit numerical centrosome amplification (CA), the presence of more than one centrosome before S phase or two thereafter, a rarity in normal human tissues. CA correlates positively with malignancy, suggesting that CA benefits cancer cells [1]. Indeed, CA induced tumor formation and metastasis in a *Drosophila* larval brain transplantation assay [2]. Moreover, CA may promote metastasis by enhancing directional migration and invasion [3]. Nonetheless, CA also poses a liability to cancer cells. CA may result in potentially lethal mitotic spindle MP. MP can induce death following prolonged mitotic arrest (MA) or cause multipolar mitosis, resulting in daughter cells sustaining an intolerable degree of aneuploidy [4, 5]. In order to escape these calamitous fates, cancer cells employ centrosome clustering mechanisms to assemble a pseudo-bipolar mitotic spindle [6-8]. Centrosome clustering not only circumvents death but also promotes low-grade chromosome missegregation, which may drive tumor evolution [9, 10].

Various factors are critical for centrosome clustering, such as cortical actin, the spindle assembly checkpoint (SAC), cell adhesion and polarity regulators, the Ndc80 microtubule-kinetochore attachment complex, augmin complex members, and microtubule motors (e.g., dynein and HSET) [11-14]. Inhibition of the kinesin-14 HSET has attracted attention lately because it causes death selectively in cancer cells with supernumerary centrosomes [11]. Based

⁴Parts of this chapter have been published verbatim in *Cell Death and Disease*, 2014, 5:e1204. doi: 10.1038/cddis.2014.16 as “Quantitative multi-parametric evaluation of centrosome declustering drugs: centrosome amplification, mitotic phenotype, cell cycle and death.”

on these promising findings, two HSET inhibitors have been developed recently [15, 16]. Inhibitors of centrosome clustering are selective for cancer cells because most healthy adult human cells do not exhibit CA, precluding their dependence on clustering mechanisms [6, 17]. Consequently, putative centrosome declustering agents have emerged as promising anticancer drugs. These agents include griseofulvin (GF), noscapine (Nos), and Nos derivatives (e.g., brominated noscapine (BN) and reduced BN (RBN)), all of which modulate microtubule dynamicity, as well as the phenanthrene-derived poly(ADP-ribose) polymerase inhibitor, PJ-34 (PJ).

Like traditional spindle poisons, declustering drugs are known to cause G2/M phase arrest (Nos [18], BN [18], RBN [19], GF [20], PJ [21]). Spindle poisons induce MA by perturbing microtubule attachment to kinetochores, thereby activating the SAC [22]. The SAC prevents cyclin B1 degradation, resulting in sustained MA [23]. The fates of spindle poison-treated, mitotically arrested cells are thought to be governed by two competing pathways: one that induces apoptosis in mitosis (via caspase-dependent mechanisms) and another that induces mitotic exit (via progressive degradation of cyclin B1) [22, 23]. Mitotic exit may culminate in apoptosis in the subsequent interphase or may permit cell survival [24]. Nevertheless, controversy surrounds the role of MA in inducing apoptosis. Specifically, it remains unclear which factors – the extent, onset, and/or duration of MA – are critical for inducing apoptosis [22, 24].

Despite their common ability to induce MA, declustering agents differ in their clinical indications and molecular targets. GF is an antifungal and Nos is an antitussive, both with different binding sites on microtubules, whereas PJ is a phenanthrene-derived poly(ADP-ribose) polymerase inhibitor with no known microtubule-binding capacity. To date, no study has quantitatively compared the declustering abilities of these diverse agents, which may contrast given their pharmacologic differences. Furthermore, these agents paradoxically are effective in cancer lines, such as HeLa, PC3, and DLD-1 cells, which lack significant CA²⁶ (e.g., Nos and BN [18], GF [20, 25], PJ [26], and RBN [27]). Unless these drugs also induce CA (such as occurs with BN [27] and RBN [19]), MP is perhaps being induced in such cell lines in a centrosome declustering-independent manner. It was recently demonstrated that inhibition of HSET can induce MP via acentrosomal pole formation in cancer cells irrespective of centrosome number [28]. Thus, 'acentrosomal pole amplification' represents a potential mechanism by which declustering drugs might induce spindle MP in cancer cells without CA.

Herein, we seek to quantitatively determine which of these phenotypes – MA, CA, declustering, acentrosomal pole amplification, and/or spindle MP – promote cancer cell death. This information can provide a conceptual framework to aid rational design of novel centrosome declustering drugs.

4.2 Results

4.2.1 Characterization of MA induced by centrosome declustering drugs

To evaluate the impact of putative declustering drugs on cell cycle progression and hypodiploidy (<2N DNA content, which may indicate apoptotic cells), MDA-MB-231 (231), PC3, and HeLa cells were treated with different concentrations of declustering drugs, stained with propidium iodide, labeled with anti-MPM2 antibody, and then assessed by flow cytometry at multiple time points over 48 h. The chosen cell lines displayed different levels of endogenous CA. 231 cells (mutant p53) exhibit high levels of CA (~20–45%) [11, 28] compared with PC3 (p53 null) and HeLa (wild-type but E6-inactivated p53), which have low basal levels of CA. Consistent with previous reports, our data showed that all drugs induced sustained MA (at least 2X mitotic cells compared with untreated control cultures) at the concentrations indicated. The duration, highest degree, and rapidity of onset of MA varied between drugs, drug concentrations, and cell lines (**Figures 4.1A and B**). In general, the maximum MA achieved was less pronounced in Nos- and PJ-treated cells (**Figures 4.1A and B**). Drug-induced onset of MA was corroborated by substantial increases in cyclin B1 levels in all cell lines (**Figure 4.1C**). For most cases, prolonged MA (~24 h in duration) was followed by a substantial increase in the subG1 population fraction (**Figures 4.1A and B**). In all cases, we observed significant increases in cleaved caspase-3 over controls (**Figure 4.1C**), suggesting apoptosis. Instances wherein the subG1 fraction was elevated without cleaved caspase-3 may either represent caspase-independent cell death or the presence of hypodiploid cells whose fate is unclear. In general, we found no consistent associations between the extent, duration, or timing of MA within drugs or across cell lines. Altogether, although centrosome declustering drugs induced MA, significant differences existed in the (i) extents and durations of MA, (ii) the size of the subG1 population,

(iii) the rapidity of the onset of MA and hypodiploidy, and (iv) the extent to which hypodiploidy was accompanied by caspase-dependent apoptosis (**Figure 4.1A-C**) even within a given cell line.

4.2.2 Declustering drugs induce CA in cancer cell lines

As RBN increases the expression of Plk4 [19], a mediator of CA, we investigated whether other declustering drugs affect the expression of PLK4 along with two other mediators of CA, Cyclin E and Aurora A. Interestingly, all the drugs we studied increased PLK4, Cyclin E, and Aurora A compared with untreated cultures (**Figure 4.2A**). Consequently, we assessed CA in cultures treated with different concentrations of declustering drugs for 6, 12, 18, or 24 h and untreated controls via microscopy. Centrosomes were identified by γ -tubulin and centrin-2 colocalization at discrete foci. Interestingly, all drugs tested induced CA in a statistically significant manner in at least one cell type and drug concentration (10 or 25 μ M for all drugs except GF, which was used at 25 and 50 μ M). The average percentages of CA over 24 h and the associated fold increases over controls is shown in **Figure 4.3**, respectively. Representative confocal micrographs of CA in interphase and mitotic cells, both control and drug treated, are depicted in **Figure 4.4**. We did not find significant correlations between the degree of CA (**Figure 4.3A**) and the expression levels of PLK4, Cyclin E, and Aurora A (**Figure 4.2**). Although declustering drugs might induce CA through pathways independent of these proteins, we conjecture that the levels of these proteins required to induce CA may simply vary between cell lines. Furthermore, many of the cells in which CA was induced may have died, precluding their quantitation. To determine whether the CA-inducing activity of declustering drugs is restricted to cancer cells, we treated two non-malignant cell lines, mammary fibrocystic (MCF10A) cells and adult human dermal fibroblasts with these drugs. We found that RBN, GF,

and PJ did induce CA and cell death in these cell lines to varying extents. In summary, all the centrosome declustering drugs in our study were also centrosome-amplifying drugs, depending on the cell line and drug concentration.

4.2.3 Effect of putative declustering agents on centrosome declustering and spindle MP

Having identified that all the declustering drugs in our study induce CA, we were interested to quantitatively evaluate the extent to which they induce MP and declustering. MP was considered low grade if there were only 3 or 4 spindle poles and high grade if there were ≥ 5 poles. We found that all declustering drugs, at one or both concentrations, induced spindle MP in at least one cell type above control levels (**Figure 4.5A**). Several of the drugs induced acentrosomal or 'acentriolar' poles (wherein at least one spindle pole stained positively for γ -tubulin but not centrin-2; **Figure 4.5A**), a phenotype not previously reported for these particular drugs. This phenotype has been reported following knockdown of HSET [28]. We found that acentriolar poles were more readily induced in HeLa than in PC3 or 231 cells (**Figure 4.5A**). The mechanism undergirding this phenotype is presently unknown. Some of the forces that tether together supernumerary centrosomes may also preserve spindle pole integrity [29], and our observations support that hypothesis.

Next, we evaluated the extents to which these drugs induced declustering. We found that the extent of total declustering (the percentage of cells with amplified centrosomes in which no centrosomes were clustered) induced by all the drug regimens was the lowest in 231 cells, which have higher endogenous CA (**Figure 4.5B**). By contrast, in HeLa and PC3 cells, which have comparatively low levels of CA, a majority of the amplified centrosomes were found to be

totally declustered (**Figure 4.5B**). For comparison, we assessed drug-induced MP, declustering, and acentrosomal pole formation in non-malignant cell lines. We found that RBN, GF, and PJ significantly induced MP over control levels, and the supernumerary centrosomes induced tended to be declustered.

Ultimately, it appears that the drugs tested largely induce spindle MP in a declustering-independent manner, at least in the cancer cell lines tested here. Declustering drugs may therefore prove effective in cancers regardless of the extent of CA present.

4.2.4 Cross talk between drug-induced spindle MP, declustering, and drug efficacy

Next, we probed the associations between drug-induced spindle MP, centrosome declustering, and drug efficacy (subG1 extent) in order to identify the phenotypes that contributed most to cell death. We used beta regression (a statistical methodology more appropriate for proportions data than linear regression when very low or high percentages are observed) to analyze correlates of peak subG1. For this technique, pseudo R^2 (the squared correlation of linear predictor and link-transformed response) is reported rather than R^2 as in linear regression, and it indicates the goodness-of-fit of the model.

We discovered that across all drugs and cell lines, peak MP significantly correlated with peak subG1 ($P=0.00840$, pseudo $R^2=0.321$), suggesting that generation of spindle MP is a shared mechanism whereby declustering drugs trigger cell death. Importantly, we found no significant

associations between CA and spindle MP, which corroborates our finding that declustering drugs appear to induce spindle MP by disrupting spindle pole and/or centrosome integrity, which in some cases may also decluster centrosomes if an excess is present. Within 231 cells, we found an even stronger, positive correlation with a very good fit between peak high-grade MP and peak subG1 (**Figure 4.6Ai**; $P=0.006$; $\text{pseudo}R^2=0.833$), underscoring that a desirable attribute for declustering drugs is the ability to induce high-grade rather than low-grade MP. We also found that a model including both peak high-grade and low-grade MP together was better in predicting peak subG1 ($P=0.001$; $\text{pseudo}R^2=0.860$) (**Figure 4.6Aii**). Specifically, within this model, the prediction of peak subG1 using peak high-grade MP was very highly statistically significant ($P<0.00001$) and the beta coefficient was positive, indicating a positive correlation between peak high-grade MP and subG1 generation. The prediction of peak subG1 using peak low-grade MP was very highly statistically significant ($P=0.00001$), and the beta coefficient was negative, indicating a negative correlation between peak low-grade MP and peak subG1. This finding is consistent with the notion that high-grade MP engenders intolerably severe aneuploidy that is likely to culminate in cell death, whereas low-grade MP is more likely to be survivable and perhaps advantageous to cancer cells. Clear trends were not uncovered for centrosome declustering and subG1 across drugs, although we cannot rule out its importance within individual drugs, as the number of data points for peak subG1 was limiting.

In HeLa cells, peak MP (any grade) positively correlated with peak subG1 ($P=0.0055$; $\text{pseudo}R^2=0.575$; **Figure 4.6Ci**). Also, peak high-grade MP positively correlated somewhat with peak subG1 ($P=0.028$; $\text{pseudo}R^2=0.271$; **Figure 4.6Cii**). Notably, the peak acentriolar pole percentage positively correlated with peak subG1 ($P=0.0023$; $\text{pseudo}R^2=0.600$; **Figure 4.6Ciii**),

so daughter HeLa cells without centrosomes may be inviable. Indeed, based on the pseudo R^2 value, peak acentriolar pole formation was superior to all other variables in predicting peak subG1. Peak total declustering also positively correlated with peak subG1 ($P=0.020$; pseudo $R^2=0.424$; **Figure 4.6Civ**), strengthening the idea that more extensive declustering kills more cancer cells.

In PC3 cells, we did not find an association between peak MP and peak subG1 across drugs. However, when we analyzed the correlation between the average fold increase in CA induction with peak subG1 percent, we found an interesting trend. In PC3 cells, the proportion variable (peak subG1) always lay within the 30–70% range and the other variable (fold increase in CA) was continuous, so we implemented linear regression for analysis. We found that the average fold increase in CA in interphase positively correlated with peak subG1 ($P=0.057$; $R^2=0.619$; **Figure 4.6b**), suggesting that an increase in CA may promote cell death.

We also analyzed the impact of treatment with declustering drugs on spindle MP and subG1 induction in non-transformed cells. In both MCF10A cells and human dermal fibroblasts, peak MP positively correlated with peak subG1 ($R^2=0.82$ with $P=0.003$ and $R^2=0.89$ with $P<0.001$, respectively), suggesting that MP is also toxic to normal cells.

4.3 Discussion

Declustering of supernumerary centrosomes is a promising chemotherapeutic approach that has recently come to light. Centrosome declustering kills cancer cells with supernumerary centrosomes while sparing normal cells, which rarely have supernumerary centrosomes. To date, drugs that decluster centrosomes have been grouped together given that they share this ability; however, the precise mechanisms by which they decluster centrosomes may differ and are unknown. It is also unclear whether centrosome declustering is their sole, or even primary, cellular activity. Ours is the first study to assess the differential abilities of various declustering drugs to induce MA, CA, MP, declustering, acentrosomal pole amplification, and cell death in different malignant and non-malignant cell lines. It would be valuable to have a framework to quantitatively evaluate novel declustering drugs to improve their efficacy via rational design. Our study lays the foundation for defining such a framework, which could guide development of a next generation of declustering drugs that are even more effective in killing cancer cells and even less toxic to normal cells.

Centrosome declustering drugs induce MA followed by death [19, 25, 30-32], and the duration of MA induced by some microtubule-targeting drugs determines cell death after mitotic exit [33]. In concordance, we found that declustering drugs induce MA, which is followed by a peak in the subG1 fraction. However, we also found great variability in the extent of MA, how rapidly its onset occurs, its duration, and the extent of hypodiploidy generated between drugs and cell lines. Differences were also apparent in the extent to which MA was accompanied by the appearance of cleaved caspase-3, an early marker of apoptosis. There were no clear associations between peak MA, the duration of MA, and the timing of the onset of MA with hypodiploidy or

the various centrosome- and spindle-related phenotypes we observed. These findings convey the complexity of the cellular responses to these drugs and suggest that the mechanisms of actions of these drugs probably differ significantly.

A central finding of our study is that all the declustering drugs upregulate PLK4, Cyclin E, and Aurora A and induce CA (**Figure 4.2**). This finding helps to clarify the seeming contradiction that centrosome declustering drugs are effective in cell lines without much CA. Although it has been previously shown that BN[27]- and RBN [19] induce CA, no other studies to our knowledge have demonstrated induction of CA by Nos, GF, or PJ. It currently does not appear that the CA induced by these drugs is coupled in any way to their declustering activity, although further studies are needed to establish conclusively whether these two activities are truly independent. Another key finding from our study is that declustering drugs can trigger 'spindle pole amplification' via two means, acting separately or in concert: (i) genesis of acentriolar poles and (ii) amplification and declustering of centrosomes, both of which precipitate spindle MP.

SubG1 fraction (i.e., cell death) typically exceeded the extent of CA, often by a wide margin (**Figures 4.1 and 4.2**). Thus, CA and declustering alone cannot explain the efficacy of these agents in the cell lines examined. Many reports have demonstrated that MP can be induced without centrosome declustering, such as through spindle pole fragmentation, loss of microtubule anchoring at the centrosome, PCM fragmentation, or centriole separation in mitosis [34-37]. Consequently, it is conceivable that 'declustering drugs' are inducing MP in a

centrosome declustering-independent manner, such as by 'declustering' (defocusing) spindle poles. For instance, treatment of HeLa cells with 50 μM GF ($\text{IC}_{50}=35 \mu\text{M}$ [25]) for 24 h induced MP in 72% of cells, even though we found that $\leq 4\%$ of untreated controls had CA, as others have found [38]. We also discovered that 50 μM GF induced only a small increase in the extent of CA in HeLa cells. Although the minority of multipolar cells with CA did exhibit centrosome declustering in the presence of GF, the majority of multipolar cells exhibited acentrosomal poles declustered. Nos, likewise, is effective in cells without much endogenous CA, such as HeLa ($\text{IC}_{50}=25 \mu\text{M}$ [31]), and Nos frequently induced MP by generating acentrosomal poles (**Figure 4.5A**). Nos also induced further CA in HeLa and PC3 cells and achieved nearly complete declustering of these amplified centrosomes. Intriguingly, although we confirmed that BN induced CA that closely paralleled MP across cell lines (i.e., cells with CA were generally the ones that were also MP; **Figure 4.5A**), we discovered that the subG1 population exceeded the percentage of cells with CA and/or MP. This finding suggests that additional mechanisms may contribute to this drug's anticancer activity, even though CA and declustering appear to operate as well. We also verified that RBN substantially increases CA across cell lines but it often produces MP spindles with one or more acentrosomal poles. Consequently, like all the other putative centrosome declustering agents that we studied, it does not seem that RBN is a 'pure' declustering drug but may additionally disrupt spindle pole integrity or otherwise cause spindle pole amplification, a downstream consequence of which is centrosome declustering. Altogether, it seems that MP itself, however it is generated (whether via bona fide centrosome declustering or acentrosomal spindle pole amplification), is a key driver of cell death in 231 and HeLa cells, although CA may instead be mechanistically involved in cell death in PC3 cells.

Ultimately, the diversity of previously unrecognized or under-appreciated phenotypes uncovered by our study, illustrated in **Figure 4.7**, raises a host of questions that represent ample opportunities for future investigations. We have discerned certain phenotypes that contribute to cell death, which may guide rational development of these drugs or formulation of synergistic drug combinations to enhance chemotherapeutic efficacy. Altogether, our data highlight the potential utility of centrosome declustering drugs in combatting malignancy given that we found cancer cells are often more susceptible to these agents than normal cells. Interestingly, we uncovered that centrosome declustering drugs may act partially or primarily through spindle pole amplification. Declustering may either operate in addition to spindle pole amplification or may simply be a corollary of that process. Future studies are awaited to further unravel the precise mechanisms by which currently available declustering drugs operate to guide rational design of novel members of this promising class.

4.4 Materials and Methods

4.4.1 Cell lines, cell culture, and drugs

HeLa, MDA-MB-231 (231), and PC3 cells were grown in medium (Dulbecco's modified Eagle's medium for HeLa and 231 cells, Roswell Park Memorial Institute medium for PC-3 cells; Invitrogen, Carlsbad, CA, USA) supplemented with 10% HyClone fetal bovine serum (Thermo Scientific, Waltham, MA, USA) and maintained in a 95% CO₂ atmosphere at 37 °C. BN and RBN were prepared as described previously [19, 39]. Nos, GF, and PJ were purchased from Sigma (St. Louis, MO, USA). Cells were grown to ~70% confluence and treated with drug or

vehicle (0.1% dimethyl sulfoxide) followed by processing for flow cytometry, immunofluorescence confocal microscopy, or immunoblotting as previously described [40].

4.4.2 Immunofluorescence microscopy and western blotting

Primary antibodies against γ -tubulin, α -tubulin, and centrin-2 (Abcam, Cambridge, MA, USA), Alexa 488-, 555-, and 647-conjugated secondary antibodies (Invitrogen), and DAPI (Invitrogen) were diluted in 2% BSA/1 \times PBS 1:2000. Confocal microscopy was performed using the Zeiss LSC 700 microscope (Oberkochen, Germany) with a 1.4 NA oil-immersion lens, and image processing was performed with Zen software (Oberkochen, Germany). For each drug at a particular concentration (10, 25, or 50 μ M) in a specific cell line (HeLa, 231, or PC3 cells), 200 cells in randomly selected fields were assessed for each time point (0, 12, 18, and 24 h). If the number of mitotic cells out of 200 did not reach $n=50$, additional randomly selected mitotic cells were assessed until $n=50$. The number of interphase cells was never less than $n=133$. Mitotic index was calculated for each time point, and for both interphase and mitotic cells, the number of γ -tubulin/centrin-positive foci (i.e., centrosomes) in each cell was recorded. For mitotic cells, the number of spindle poles was also recorded, along with the number of centrosomes present, if any, at each pole. Western blotting was performed as previously described [40] using lysates from cells treated with 25 μ M drug for 0, 4, 8, 12, 16, 18, 20, 24, 28, 36, or 48 h.

4.4.3 Cell cycle analysis

The flow cytometric evaluation of cell cycle status was done using FlowJo software (Ashland, OR, USA). After the treatment, cells were harvested at different time intervals, washed twice with ice-cold PBS, and fixed in 70% ethanol for at least 24 h. Cell pellets were then washed with PBS and added with 0.5 ml of RNase A (2 mg/ml) and stained with MPM-2 primary mouse antibody followed by Alexa-488 secondary antibody. In addition, 0.5 ml of propidium iodide (0.1% with 0.6% Triton X-100 in PBS) was added for 45 min in the dark followed by analysis on a FACS Cantor flow cytometer (BD Canto, Franklin Lakes, NJ, USA). Experiments were repeated in triplicate.

4.4.4 Statistical analysis

To compare proportions between samples, Fisher's exact test was performed using SPSS. Sample size was always $n \geq 50$. The only exception occurred with calculation of the proportion of cells exhibiting centrosome declustering (the percent of cells with declustered centrosomes out of all cells with CA at each time point), as the number of cells with CA was typically < 50 . In all cases, results were considered significant if $P < 0.05$. To perform regression using proportions, beta regression was used as described by Ferrari and Cribari-Neto [41] using MATLAB.

4.5 References

1. Chan, J.Y., *A Clinical Overview of Centrosome Amplification in Human Cancers*. International Journal of Biological Sciences, 2011. **7**(8): p. 1122-1144.
2. Basto, R., et al., *Centrosome Amplification Can Initiate Tumorigenesis in Flies*. Cell, 2008. **133**(6): p. 1032-1042.

3. Ogden, A., P.C.G. Rida, and R. Aneja, *Heading off with the herd: How cancer cells might maneuver supernumerary centrosomes for directional migration*. Cancer metastasis reviews, 2013. **32**(0): p. 269-287.
4. Vitale, I., et al., *Mitotic catastrophe: a mechanism for avoiding genomic instability*. Nat Rev Mol Cell Biol, 2011. **12**(6): p. 385-92.
5. Holland, A.J. and D.W. Cleveland, *Boveri revisited: Chromosomal instability, aneuploidy and tumorigenesis*. Nature reviews. Molecular cell biology, 2009. **10**(7): p. 478-487.
6. Kramer, A., B. Maier, and J. Bartek, *Centrosome clustering and chromosomal (in)stability: a matter of life and death*. Mol Oncol, 2011. **5**(4): p. 324-35.
7. Ogden, A., P.C. Rida, and R. Aneja, *Let's huddle to prevent a muddle: centrosome declustering as an attractive anticancer strategy*. Cell Death Differ, 2012. **19**(8): p. 1255-67.
8. Godinho, S.A., M. Kwon, and D. Pellman, *Centrosomes and cancer: how cancer cells divide with too many centrosomes*. Cancer Metastasis Rev, 2009. **28**(1-2): p. 85-98.
9. Ganem, N.J., S.A. Godunov, and D. Pellman, *A mechanism linking extra centrosomes to chromosomal instability*. Nature, 2009. **460**.
10. Silkworth, W.T., et al., *Multipolar Spindle Pole Coalescence Is a Major Source of Kinetochore Mis-Attachment and Chromosome Mis-Segregation in Cancer Cells*. PLoS ONE, 2009. **4**(8): p. e6564.
11. Kwon, M., et al., *Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes*. Genes Dev, 2008. **22**.
12. Leber, B., et al., *Proteins required for centrosome clustering in cancer cells*. Sci Transl Med, 2010. **2**(33): p. 33ra38.
13. Quintyne, N.J., et al., *Spindle multipolarity is prevented by centrosomal clustering*. Science, 2005. **307**(5706): p. 127-9.
14. Fielding, A.B., et al., *A critical role of integrin-linked kinase, ch-TOG and TACC3 in centrosome clustering in cancer cells*. Oncogene, 2011. **30**(5): p. 521-34.
15. Wu, J., et al., *Discovery and mechanistic study of a small molecule inhibitor for motor protein KIF1C*. ACS Chem Biol, 2013. **8**(10): p. 2201-8.
16. Watts, Ciorsdaidh A., et al., *Design, Synthesis, and Biological Evaluation of an Allosteric Inhibitor of HSET that Targets Cancer Cells with Supernumerary Centrosomes*. Chemistry & Biology, 2013. **20**(11): p. 1399-1410.
17. Korzeniewski, N., M. Hohenfellner, and S. Duensing, *The centrosome as potential target for cancer therapy and prevention*. Expert Opin Ther Targets, 2013. **17**(1): p. 43-52.
18. Aneja, R., et al., *Synthesis of microtubule-interfering halogenated noscapine analogs that perturb mitosis in cancer cells followed by cell death*. Biochem Pharmacol, 2006. **72**(4): p. 415-26.
19. Pannu, V., et al., *Induction of robust de novo centrosome amplification, high-grade spindle multipolarity and metaphase catastrophe: a novel chemotherapeutic approach*. Cell Death Dis, 2012. **3**: p. e346.
20. Rebacz, B., et al., *Identification of griseofulvin as an inhibitor of centrosomal clustering in a phenotype-based screen*. Cancer Res, 2007. **67**(13): p. 6342-50.
21. Inbar-Rozensal, D., et al., *A selective eradication of human nonhereditary breast cancer cells by phenanthridine-derived polyADP-ribose polymerase inhibitors*. Breast Cancer Res, 2009. **11**(6): p. R78.

22. Matson, D.R. and P.T. Stukenberg, *Spindle poisons and cell fate: a tale of two pathways*. Mol Interv, 2011. **11**(2): p. 141-50.
23. Sakurikar, N., J.M. Eichhorn, and T.C. Chambers, *Cyclin-dependent kinase-1 (Cdk1)/cyclin B1 dictates cell fate after mitotic arrest via phosphoregulation of antiapoptotic Bcl-2 proteins*. J Biol Chem, 2012. **287**(46): p. 39193-204.
24. Machado, E., M. Guillaumot, and M. Malumbres, *Killing cells by targeting mitosis*. Cell Death Differ, 2012. **19**(3): p. 369-77.
25. Rathinasamy, K., et al., *Griseofulvin stabilizes microtubule dynamics, activates p53 and inhibits the proliferation of MCF-7 cells synergistically with vinblastine*. BMC Cancer, 2010. **10**: p. 213.
26. Castiel, A., et al., *A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells*. BMC Cancer, 2011. **11**: p. 412.
27. Karna, P., et al., *A novel microtubule-modulating noscapinoid triggers apoptosis by inducing spindle multipolarity via centrosome amplification and declustering*. Cell Death Differ, 2011. **18**(4): p. 632-44.
28. Kleylein-Sohn, J., et al., *Acentrosomal spindle organization renders cancer cells dependent on the kinesin HSET*. J Cell Sci, 2012. **125**(Pt 22): p. 5391-402.
29. Godinho, S.A., M. Kwon, and D. Pellman, *Centrosomes and cancer: how cancer cells divide with too many centrosomes*. Canc Met Rev, 2009. **28**.
30. Castiel, A., et al., *A phenanthrene derived PARP inhibitor is an extra-centrosomes de-clustering agent exclusively eradicating human cancer cells*. BMC Cancer, 2011. **11**(1): p. 412.
31. Ye, K., et al., *Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induces apoptosis in dividing cells*. Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1601-6.
32. Jaiswal, A.S., et al., *9-bromonoscapine-induced mitotic arrest of cigarette smoke condensate-transformed breast epithelial cells*. J Cell Biochem, 2009. **106**(6): p. 1146-56.
33. Bekier, M.E., et al., *Length of mitotic arrest induced by microtubule-stabilizing drugs determines cell death after mitotic exit*. Mol Cancer Ther, 2009. **8**(6): p. 1646-54.
34. Logarinho, E., et al., *CLASPs prevent irreversible multipolarity by ensuring spindle-pole resistance to traction forces during chromosome alignment*. Nat Cell Biol, 2012. **14**(3): p. 295-303.
35. Gergely, F., V.M. Draviam, and J.W. Raff, *The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells*. Genes Dev, 2003. **17**(3): p. 336-41.
36. Wu, Q., et al., *Cep57, a NEDD1-binding pericentriolar material component, is essential for spindle pole integrity*. Cell Res, 2012. **22**(9): p. 1390-401.
37. Torres, J.Z., et al., *The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly*. Cell, 2011. **147**(6): p. 1309-23.
38. Starita, L.M., et al., *BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number*. Mol Cell Biol, 2004. **24**(19): p. 8457-66.
39. Zhou, J., et al., *Brominated derivatives of noscapine are potent microtubule-interfering agents that perturb mitosis and inhibit cell proliferation*. Mol Pharmacol, 2003. **63**(4): p. 799-807.

40. Karna, P., et al., *Induction of reactive oxygen species-mediated autophagy by a novel microtubule-modulating agent*. J Biol Chem, 2010. **285**(24): p. 18737-48.
41. Ferrari, S. and F. Cribari-Neto, *Beta Regression for Modelling Rates and Proportions*. Journal of Applied Statistics, 2004. **31**(7): p. 799-815.

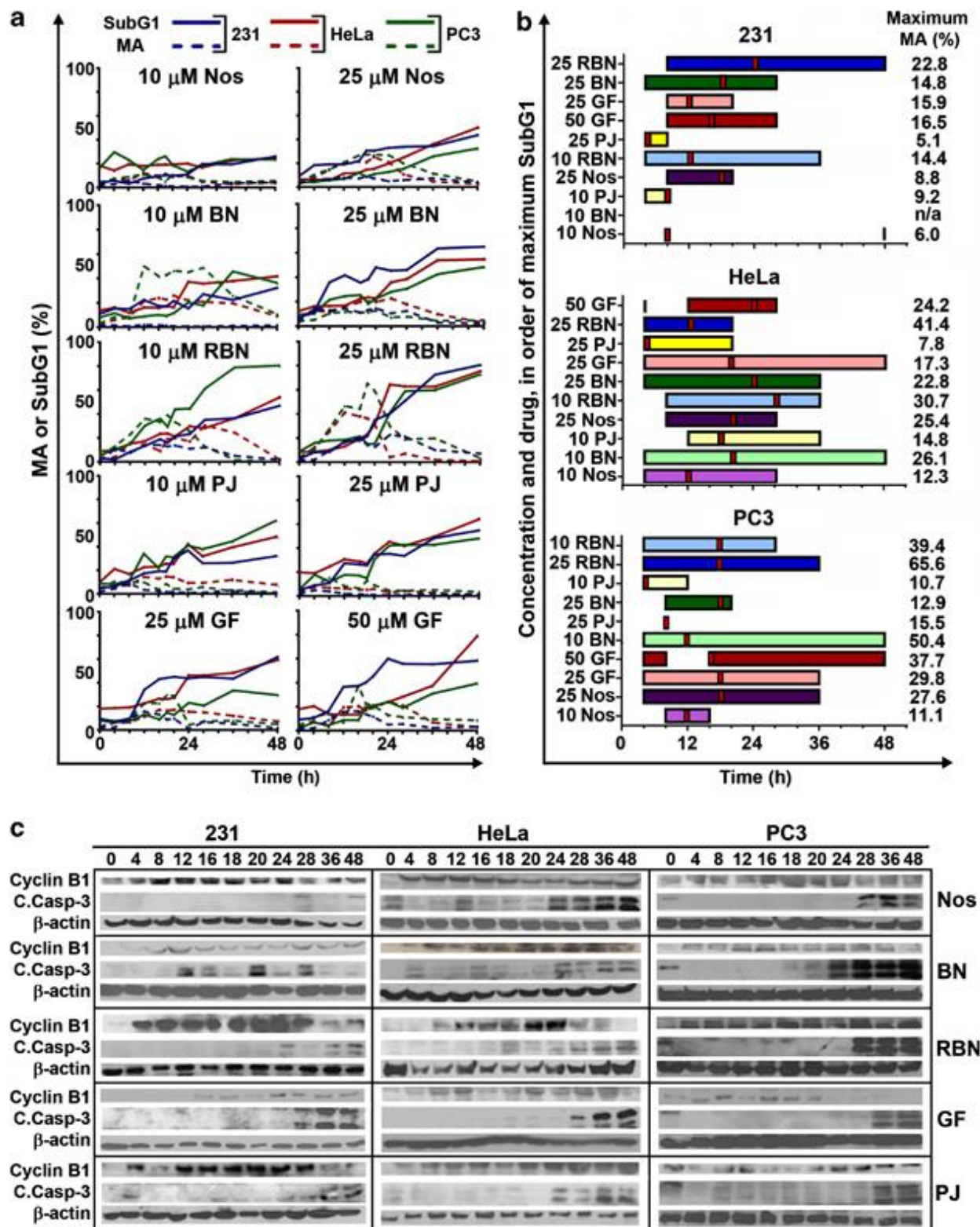


Figure 4.1 Mitotic arrest (MA) phenotypes observed upon treatment with putative centrosome declustering drugs.

(a) SubG1 and mitotically arrested cell population fractions with respect to time post-treatment with various putative declustering drugs. Declustering drugs included Nos, BN, RBN, PJ, and GF, all at 10 and 25 μM except GF, which was used at 25 and 50 μM , and cell lines included 231, PC3, and HeLa. These cell lines demonstrated differential susceptibility to various agents depending on drug concentration over the 48 h time period. In general, MA increased from 0 h to a peak near 24 h, followed by a decline in MA that coincided with increases in subG1 fractions. Results are representative of three independent experiments. (b) Duration of MA and peak MA by maximum subG1 fraction. Drugs are ranked in order of increasing peak subG1 from bottom to top along the y axis. The duration of MA (defined as the duration for which the mitotic population in drug-treated cells was greater than two times that in control cells) is plotted along the x axis. The time at which peak MA occurred is illustrated as a red bar and the value of peak MA is listed to the right of the graph. In 231 cells, 10 BN did not cause any MA; therefore, no bar is plotted. For 10 Nos in 231 cells and 25 PJ in PC3 cells, MA was observed at only one time point and is depicted using a single red bar. Some drugs produced a MA that then subsided and ultimately recurred, resulting in two bars being plotted, namely 50 GF in HeLa and PC3 cells. (c) Western blotting of cell cycle-related proteins and caspase-3, a marker for apoptosis. To assess cell cycle progression following treatment with different declustering drugs (all at 25 μM), cell lysates were obtained at multiple time points over 48 h and immunoblotted for Cyclins E and B1. Increased levels of both cyclins compared with controls (0 h) were detected across cell lines with variable expression patterns depending on the drug and cell line. To evaluate apoptosis, cleaved caspase-3 (C. Caspase-3) was immunoblotted and eventual increases over controls were universally detected, typically by 24 h

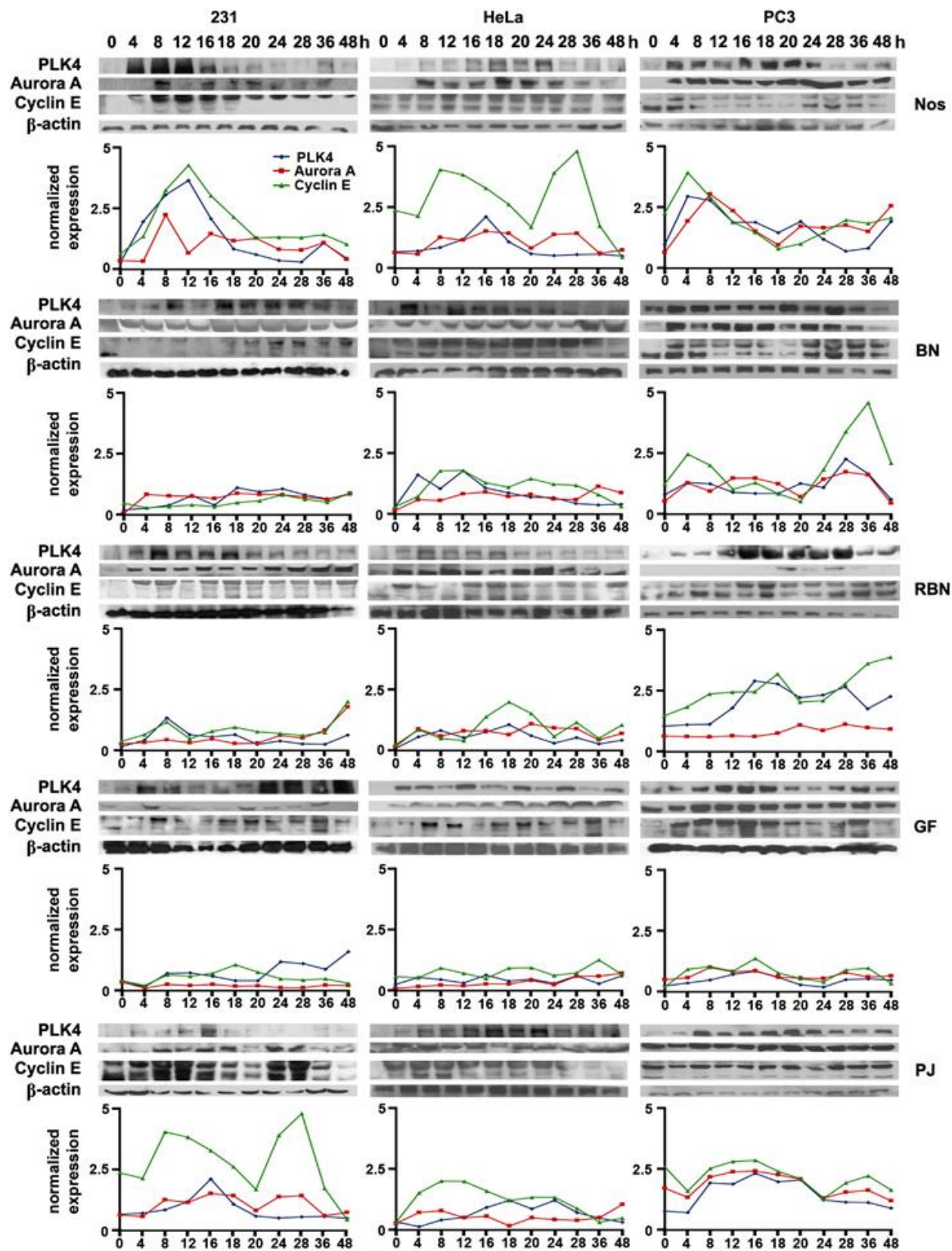


Figure 4.2 Centrosome declustering drug-induced changes in expression levels of markers of centrosome amplification.

Centrosome declustering drug-induced changes in expression levels of markers of centrosome amplification. To evaluate the levels of some of the well-established markers of CA upon treatment with declustering drugs at a concentration of 25 μ M, the levels of PLK4, Cyclin E, and Aurora A were assessed by western blotting, revealing eventual increases over untreated controls across cell lines. Increase in expression levels of PLK4 and Aurora A was generally rapid, often appearing by 4 h. Levels tended to vary thereafter depending on the drug and cell line.

Densitometry was performed to quantitate the changes in levels of CA markers relative to β -actin over time, and the changes in actin-normalized expression levels over the time-course of the experiment are depicted graphically beneath each sets of blots. As the Cyclin E blots revealed two closely placed bands (49 and 43 kDa) corresponding to the two spliced forms, the Cyclin E band intensity was generated as a sum of the two band intensities

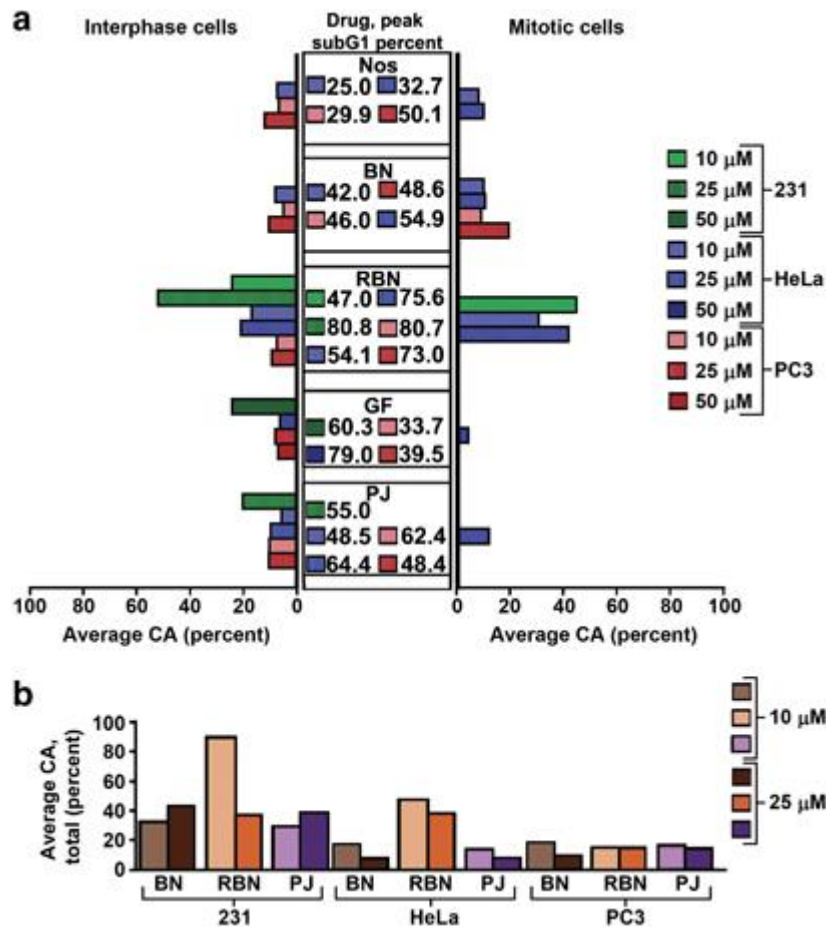


Figure 4.3 Average CA observed over 24 h and its relationship with peak subG1 for each drug treatment regimen

(a) Displayed are only statistically significant ($P < 0.05$) increases in average CA over controls. To calculate average CA, the sum of percentage of (interphase or mitotic) cells showing CA at the 6, 12, 18, and 24 h time points was divided by 4. (b) Depiction of the sum of average CA (interphase plus mitotic) observed when 231 cells were treated with RBN, BN, and PJ, compared with the treatment of HeLa and PC3 cells with the same three drugs.

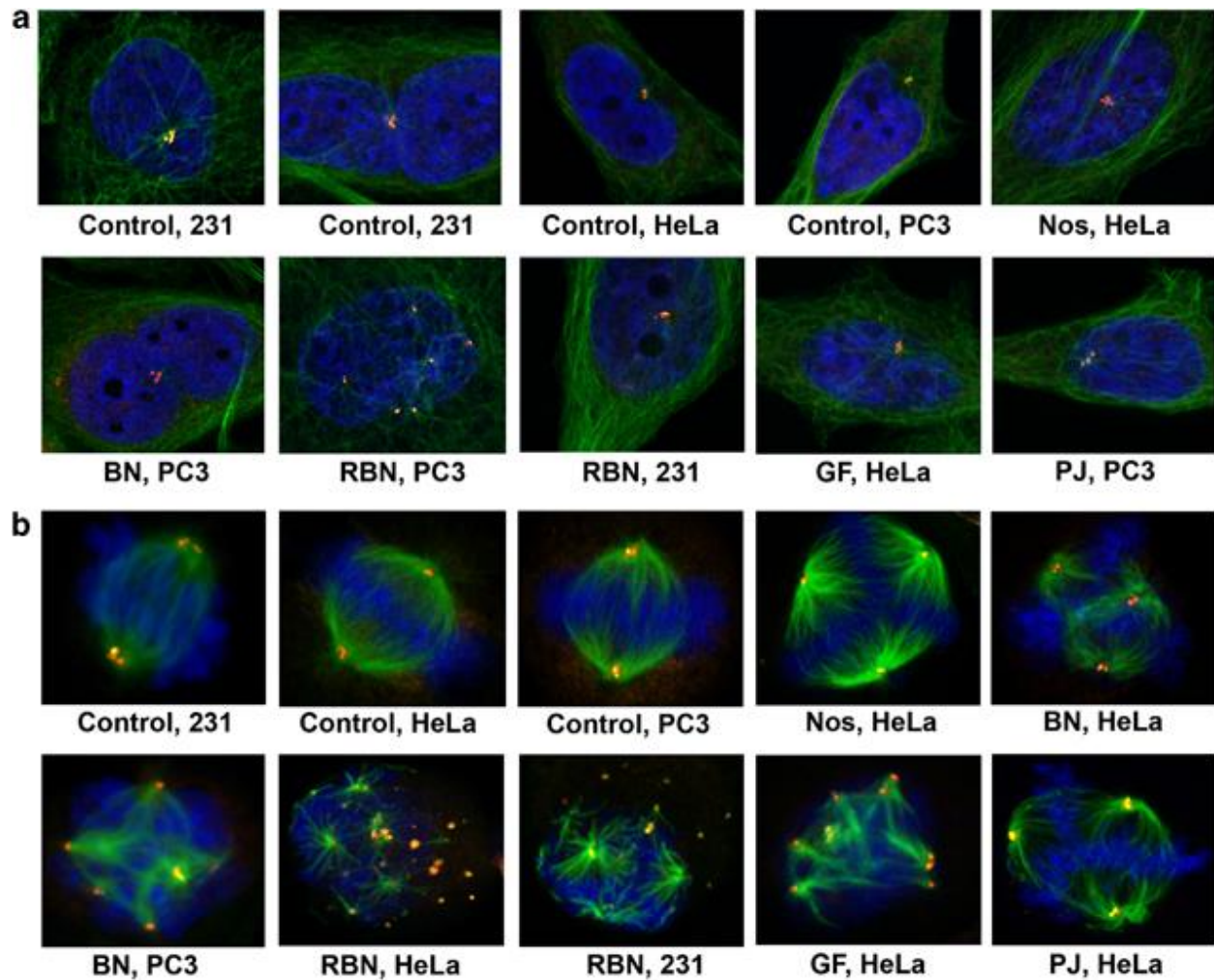


Figure 4.4 Representative confocal micrographs depicting centrosome amplification in controls and drug-treated cancer cells

(a) Interphase and (b) mitotic cells. Only displayed are drug/cell line combinations in which statistically significant ($P < 0.05$) fold increases in CA were found in mitotic cells, although all controls are shown regardless of the extent of endogenous CA. Blue=DNA, green=microtubules, red= γ -tubulin, and orange=centrin-2

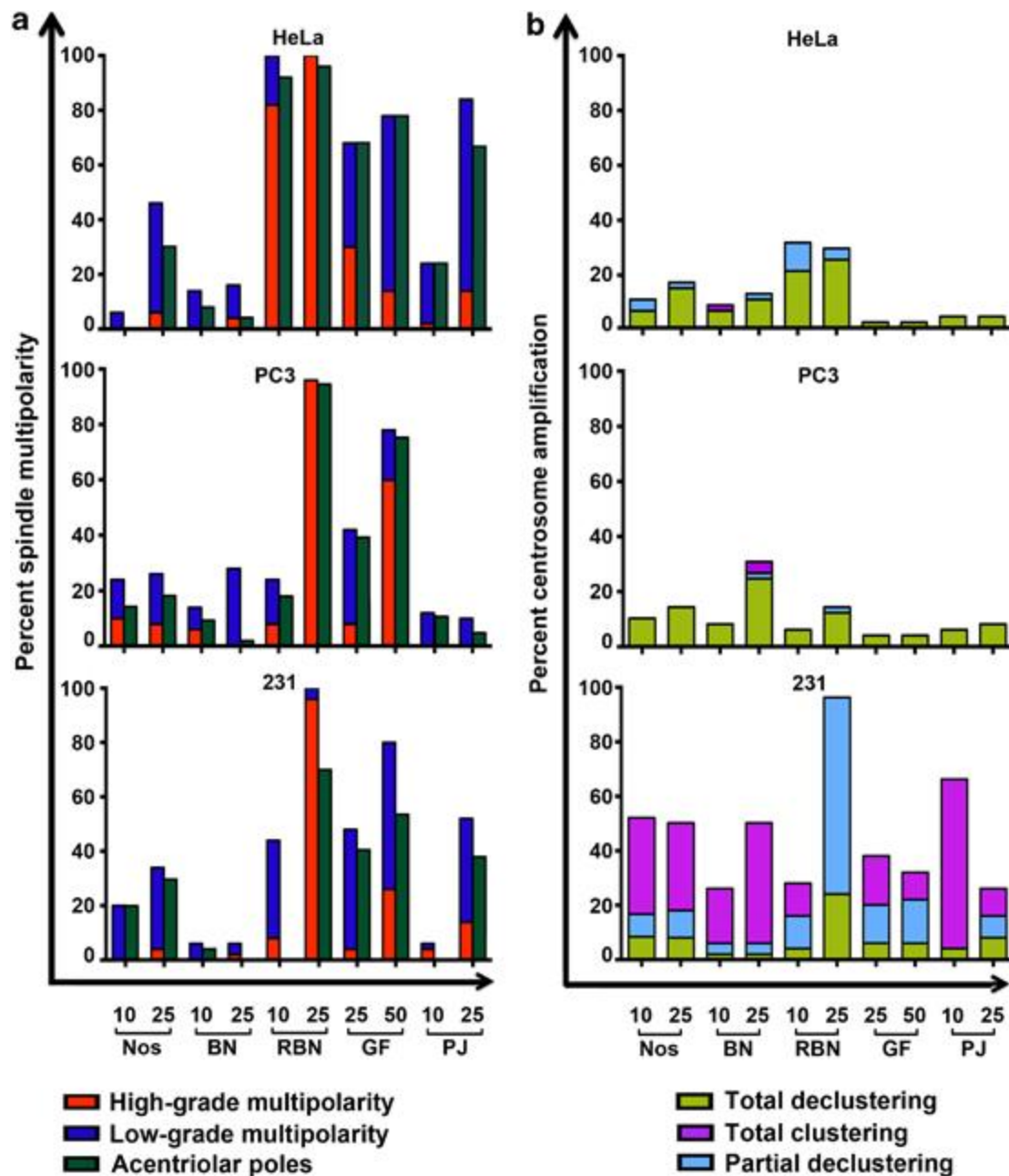


Figure 4.5 Phenotypes induced by declustering drugs at different concentrations.

(a) Peak MP and peak acentriolar pole formation induced by different declustering drugs in 231, HeLa, and PC3 cells. The maximum extents of MP induction of high grades (5+ poles) and low grades (3–4 poles) and acentriolar pole formation (at least one pole without centrioles) across a 24-h period are given for all drugs. (b) Peak CA and declustering of amplified centrosomes induced in 231, HeLa, and PC3 cells. The maximum extent of CA in mitosis over 24 h is depicted by the height of the bar. The extent of total clustering (all centrosomes clustered at two poles), total declustering (all centrosomes separated to different poles), and partial declustering (one or more poles with 2+ centrosomes) are given for that same time point

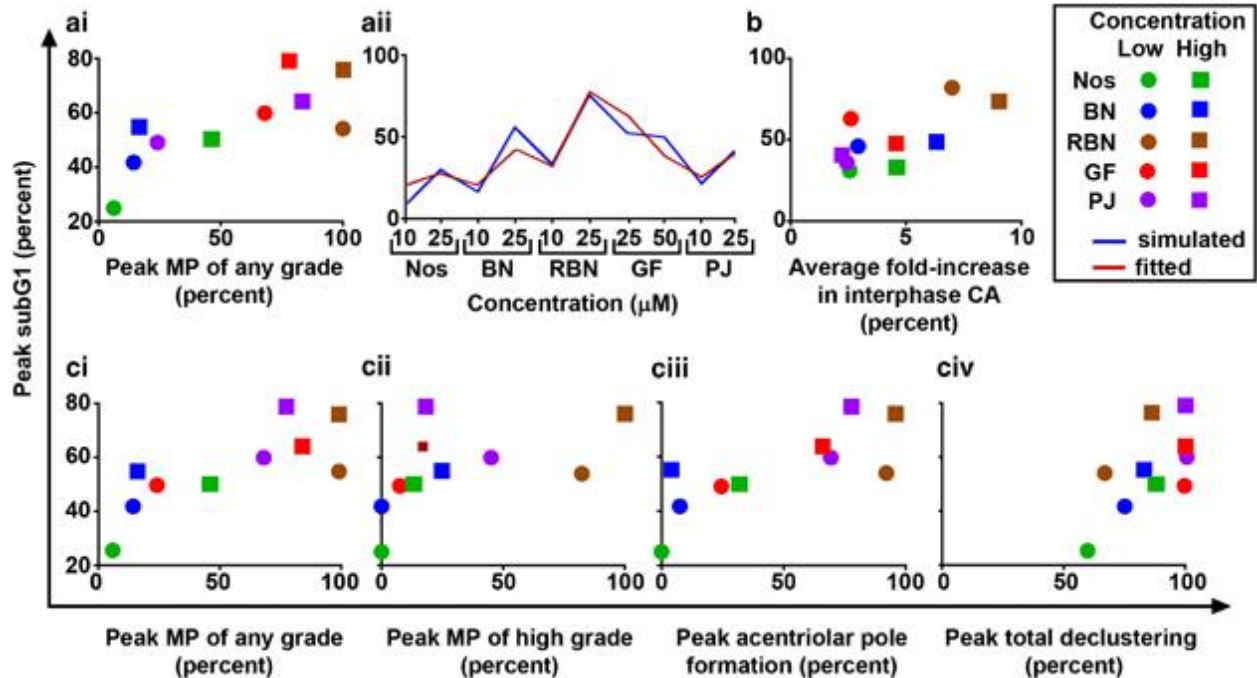


Figure 4.6 Correlates of peak subG1 percent within cell lines

(a) 231 cells, beta regression. (ai) In 231 cells, a clear trend was found for increasing peak MP of any grade and peak subG1, which was highly statistically significant ($P=0.006$; $\text{pseudoR}^2=0.833$). (aii) Furthermore, multiple regression using peak MP (high grade) and peak MP (low grade) produced an even better, statistically significant fit (red line) compared with simulated values ($P=0.001$; $\text{pseudoR}^2=0.860$). Within this model, both variables were very highly statistically significant ($P<0.0001$), with peak high-grade MP showing a positive correlation and peak low-grade MP showing a negative correlation with peak subG1 (based on the sign of the beta coefficients). (b) PC3 cells, linear regression. In these cells, the average fold increase in interphase CA shows some association with peak subG1, which almost reached statistical significance and which produced a good fit ($P=0.057$; $R^2=0.619$). (c) HeLa cells, beta regression. (ci) Increasing peak MP of any grade was associated with peak subG1 ($P=0.0055$; $\text{pseudoR}^2=0.575$), as was (cii). Increasing peak MP of high grade ($P=0.028$; $\text{pseudoR}^2=0.271$). (ciii) Increasing peak acentriolar pole formation ($P=0.0023$; $\text{pseudoR}^2=0.600$), and (civ) peak total declustering ($P=0.020$; $\text{pseudoR}^2=0.424$)

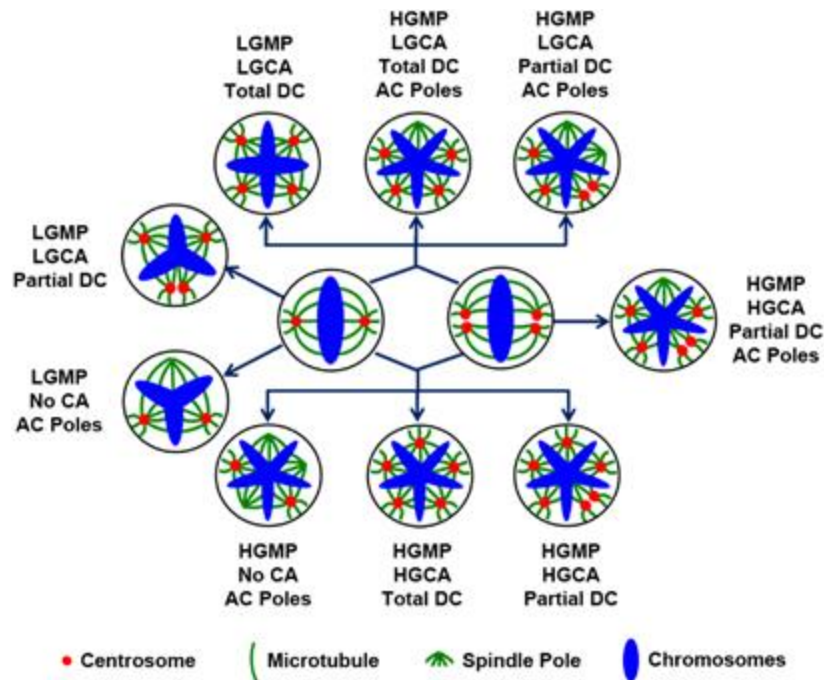


Figure 4.7 Diversity of phenotypes produced by putative centrosome declustering drugs

One of the two central cells with bipolar spindles show normal centrosome number (left cell) and the other one shows amplified centrosomes (right cell). Both types of cells are susceptible to the declustering agents. For the cell without CA, acentriolar pole formation (i.e., pole amplification or pole declustering) and MP (low- or high-grade) may be induced. Alternatively, these agents may induce CA and permit bona fide centrosome declustering to occur, partially or in total, with or without acentriolar pole formation. For the cell with CA, genuine centrosome declustering may occur, partially or in total, with or without acentriolar pole formation, and with or without further CA. AC, acentriolar; DC, declustering; HGCA, high-grade CA; HGMP, high-grade MP; LGCA, low-grade CA; LGMP, low-grade MP

5 MULTI-INSTITUTIONAL STUDY OF TRIPLE NEGATIVE BREAST CANCER STRATIFICATION BY A NOVEL BIOMARKER OF CELL CYCLING KINETICS

5.1 Abstract

Breast cancer prognostication includes assessment of tumor grade, which in the Nottingham Grading System involves assessment of tubule formation, nuclear pleomorphism, and mitotic index (MI). It may also include assessment of proliferation, most commonly via immunohistochemical analysis of Ki67 antigen, which yields the Ki67 index (KI). MI and KI are inextricably interrelated, as mitosis is a sub-phase of proliferation, so a more rational approach would be to integrate these two measures to gain insights into the cycling kinetics of the tumor, which could inform risk prediction. We hypothesized that a larger fraction of mitotic cells among proliferating cells (i.e., higher MI/KI, which we term the Ki67-Adjusted MI Score, or KAMS) would predict a greater risk of poor outcomes among chemotherapy-treated breast cancer patients. We evaluated KAMS in three unique cohorts and found that it was a superior biomarker to both KI and MI, with higher KAMS predicting worse survival specifically in triple-negative breast cancer (TNBC) patients (n=244 across all three cohorts). KAMS-stratified TNBCs exhibited differential expression of immunohistochemical biomarkers and RNA-sequenced molecular pathways, suggesting fundamentally different underlying tumor biology that could be targeted therapeutically. Specifically, the poor-prognosis high-KAMS group could potentially benefit by targeting DNA damage response, EGFR, or circadian rhythm pathways (which are upregulated) or boosting immune regulatory pathways (which are suppressed).

5.2 Introduction

In current diagnostic practice, breast tumor grading via the Nottingham Grading System involves microscopic evaluation of three histological parameters: tubule formation, nuclear pleomorphism, and mitotic index (i.e., mitoses/10 high-power fields [HPFs]) [1, 2]. In addition to tumor grading, pathologists often quantify the proliferating cell population within tumors using Ki67 staining. Extensive research suggests that, generally speaking, the percentage of Ki67-positive nuclei (or KI) is a prognostic biomarker in breast cancer that also predicts response to neoadjuvant chemotherapy [3]. However, breast cancer is a remarkably heterogeneous disease, and the value of KI in distinct breast cancer subtypes is not firmly established. In TNBC, an aggressive, highly proliferative subtype, KI has been less well studied, and the evidence regarding its prognostic and predictive value is mixed [4-7].

While MI and KI are considered independent biomarkers, in reality, mitosis is a sub-phase of the proliferative cell cycle, so a more rational approach would be to consider these related measures in an integrated manner rather than treating them as separate entities. In so doing, it may be possible to capture information about the cycling kinetics of the tumor, which could enhance risk prediction. We hypothesize that the higher the proportion of mitotic cells in the proliferative tumor cell population, the greater is the risk to the patient. This elevated risk conceivably arises from more frequent error-prone mitoses that generate diverse and aggressive karyotypes. Unlike the *independent* determination of KI and MI, a quantitation of the proportion of mitotic cells *among* the proliferating cells may harness kinetic information regarding how fast the proliferative cells are actually cycling and thereby generating intratumor heterogeneity. Because rapid tumor growth is a sign of aggressive disease, cycling kinetics can serve as a

powerful beacon of risk that quantifies a fundamental aspect of a tumor's biology – its mitotic propensity – and may be a stratifying risk factor for cancer types such as breast that exhibit heterogeneity in cycling kinetics. The primary objective of this study was to develop a novel risk-predictive metric that rationally integrates KI and MI using retrospective clinical data from breast cancer patients treated with chemotherapy at three different hospitals and to evaluate the ability of this biomarker to risk-stratify different breast cancer subtypes.

5.3 Materials and Methods

5.3.1 Description of datasets

We compiled clinical data for 4,587 breast carcinoma patients from Northside Hospital, Atlanta, GA, US diagnosed between years 2000-2013 and 142 TNBCs from Emory University Hospital, Atlanta, GA, US (**Table 1**, descriptive statistics), both of which were racially diverse, along with 124 chemotherapy-treated TNBC patients in Nottingham University Hospital, Nottingham, UK diagnosed between 1987-1998 (**Table 2**, descriptive statistics), a predominantly white cohort. Descriptive statistics for a subset of 99 breast tumors of different Nottingham grades from the Northside cohort used to determine the average number of mitoses for MI categories as used in the Nottingham Grading System are given in **Table 3**. In addition, 72 chemotherapy-treated TNBC patients from Stavanger University Hospital (all white women) and 48 chemotherapy-treated TNBC patients from Olabisi Onabanjo University, Nigeria (all black women), were compiled (**Table 1**). TNBC was defined as ER/PR/HER2-; Luminal A as ER and/or PR+, HER2-, KI<15%; Luminal B as ER and/or PR+ and either HER2- with KI \geq 15% or HER2+; and HER2-enriched as ER and/or PR-, HER2+. Patients missing information pertaining to KI, MI, or Nottingham grade were excluded from analyses. Patients additionally missing

information pertaining to chemotherapy or outcomes (either breast cancer-specific survival or overall survival, BCSS and OS, respectively) were excluded from survival analyses.

5.3.2 Formulation of KAMS

MI data for all hospitals except for Stavanger were categorical in nature, with Nottingham Grading System mitotic scores of 1, 2 and 3 representing 0-7, 8-14 and >15 mitoses/10 HPFs, respectively; however, continuous MI scores were desired to rationally integrate them with continuous KI scores (which are recorded as percentages). Towards this end, we sought to estimate the average number of mitoses for each mitotic score category. We obtained 99 hematoxylin and eosin-stained slides of breast carcinomas from Northside Hospital and 3 tissue microarrays of 142 TNBCs from Emory University Hospital, and the number of mitoses/10 consecutive HPFs at the most poorly differentiated area at the periphery of the tumor (40× objective, magnification ×400, field area 0.196 mm²) were manually and independently scored by two pathologists who were blind to clinical annotation. We obtained mean counts of 2.94, 11.12, and 32.62 mitotic cells/10 HPFs for Nottingham Grading System mitotic scores of 1, 2, and 3, respectively. These counts were then converted into percent MI after taking into account the cellularity of the samples. We were thus able to derive a rationally integrated measure of cycling kinetics, KAMS, for each patient as the percent mitotic cells:percent Ki67-positive cells. For the Stavanger cohort, mitotic index was recorded as a continuous variable (calculated according to the stringent Multicenter Morphometric Mammary Carcinoma Project protocol[8, 9]) and was thus used to calculate KAMS.

5.3.3 *Statistical analyses*

Average MI and KI values were calculated and compared between Nottingham grades using Analysis of Variance (ANOVA). Comparison of the slopes of the plots of mitotic cells versus Ki67-positive cells across grades was performed using Analysis of Covariance (ANCOVA). Survival curves were plotted using the Kaplan-Meier method, with equality of survival distributions tested using the log-rank test. Survival time was measured from the initial diagnosis to either an event (breast cancer-specific death for the Nottingham and Stavanger cohorts or death from any cause for the Nigeria cohort, for which cause of death was not available) or to the final follow-up (censor) and was thus an indicator of BCSS or OS in the respective cohorts. Average BCSS time was 201.3 months with n=35 events (Nottingham cohort) and 201.1 months with n=23 events (Stavanger cohort) (median BCSS could not be calculated due to >50% cumulative survival), and average and median survival times of 56.1 days and 35.0 days, respectively, with 43 events (Nigeria cohort). Simple and multivariable Cox proportional hazard models were also fit. Satisfaction of the proportional hazards assumption was verified by plotting partial residuals against rank-time and assuring that the slope was approximately horizontal. Multivariable models were fit using backward stepwise elimination of covariates with $p > 0.10$. Covariates included age at diagnosis (years), tumor size (cm), Nottingham grade, and either lymph node stage per AJCC guidelines (Nottingham and Stavanger cohorts) or lymph node status (positive or negative, Nigeria cohort). For categorical variables, the lowest risk group was used as the reference for calculation of hazard ratios (HRs). CutoffFinder [10] was used to discern optimal cutpoints in KAMS and KI based on BCSS or OS since there are no established cutpoints for these biomarkers. Because of differences between datasets in terms of patient characteristics and tissue processing/imaging protocols, optimal

cutpoints were determined for each individual dataset. Standard cutpoints in MI as per the Nottingham grading system (i.e., scores of 1, 2, and 3) were used. For survival analyses, results with $p < 0.05$ were considered statistically significant.

To compare immunohistochemical biomarkers between KAMS-low and high groups in the Nottingham cohort, multiple two-tailed t-tests were performed, guided by Levene's test for equality of variances. To set the false-discovery rate at 0.25, a p-value of $0.25 * R/m = 0.026$ (where R =number of rejected null hypotheses, in our case 30, and m =the number of tested null hypotheses, in our case 292) was considered statistically significant. The comprehensive biomarker panel included 105 unique biomarkers recorded as either percentages and/or H-scores, as appropriate for the particular biomarker, with nuclear, cytoplasmic, total (nuclear + cytoplasmic), and/or membranous levels (for cellular biomarkers) or intratumoral, stromal levels, and/or total levels (tissue biomarkers), resulting in 292 comparisons. SAS version 9.4 was used for ANOVA and ANCOVA, and IBM SPSS Statistics 21 was used for all other statistical tests.

5.3.4 RNA sequencing and analysis

FFPE TNBC primary tumor and lymph node sections with for which KAMS had been calculated ($n=84$ and $n=51$ in the discovery and validation cohorts, respectively, from 75 and 30 patients, respectively, due to some patients having multiple samples), from a mix of chemotherapy treated and untreated patients, were deparaffinized with xylene, and RNA was extracted using the Omega Mag-Bind XP FFPE RNA isolation kit (Omega, M2595-01) and KingFisher Flex magnetic particle separator (ThermoFisher). RNA concentration was measured with Nanodrop 2000c spectrophotometer (Thermo Scientific Inc., Waltham, MA, USA).

Integrity was assessed using Agilent 2200 TapeStation instrument (Agilent Technologies, Santa Clara, CA), and the percentages of fragments larger than 200 nucleotides (DV200) were calculated. First strand cDNA syntheses were performed on ~100 ng RNA at 25°C for 10 min, 42°C for 15 min, and 70°C for 15 min using random hexamers and ProtoScript II Reverse Transcriptase (New England BioLabs, Ipswich, MA). Second strand synthesis and RNA sequencing libraries were prepared using TruSeq RNA Access library kit according to the manufacturer's protocol (Illumina, Inc., San Diego, CA, USA). Briefly, the RNA templates were removed and a second replacement strand was generated by incorporation dUTP to generate double-stranded cDNA. The blunt-ended cDNA was cleaned up from the second strand reaction mix with Mag-Bind RxnPure Plus magnetic beads (Omega Bio-tek, Norcross, GA). The 3' ends of the cDNA were then adenylated and followed by the ligation of indexing adaptors. PCR (15 cycles of 98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec) to selectively enrich DNA fragments with adapter molecules on both ends. The library was qualified using an Agilent 2200 TapeStation and quantified using QuantiFluor dsDNA System (Promega, Madison, WI). A 4-plex pool of libraries was made by combining 200 ng of each DNA library. The pooled DNA libraries were then mixed with capture probes to targeted regions of interest. The hybridization was performed by 18 cycles of 1 min incubation, starting at 94°C, and then decreasing 2°C per cycle. Then streptavidin-coated magnetic beads were used to capture probes hybridized to the target regions. The enriched libraries were then eluted from the beads and subjected to a second round of hybridization. The enriched libraries were amplified by 10 cycles of PCR amplification (98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec) followed by RxnPure magnetic bead clean up. The final libraries were validated using Agilent High Sensitivity D1000 ScreenTape on

an Agilent 2200 TapeStation instrument. The size distribution of the library ranged from approximately 200 bp–1 kbp.

Libraries were normalized, pooled, and subjected to clustering, and pair-read sequencing was performed for 75 cycles on a HiSeq2500 instrument (Illumina, Inc. San Diego, CA, USA), according to the manufacturer's instructions. The fastq files were mapped to hg38 (GRCh38.P5) using STAR mapper. Transcript assembly and abundance quantification were performed using Cufflinks. In the discovery cohort 4 patients had multiple samples, and in the validation cohort 8 patients had multiple samples; for each of these patients, only the sample with maximum number of reads was retained in the analyses. Stratification of patients into high- and low-KAMS groups was based on the average KAMS score in the Nottingham TNBC cohort, with low KAMS defined as <1.0960236 (84% of patients). The DESeq2 approach [11], implemented in a Bioconductor package, was used for differential expression analyses. Association between binary KAMS and 220 signaling and metabolic KEGG pathways was assessed by the GAGE method [12]. Benjamini-Hochberg adjustment for multiplicity [13] was applied to p-values for differential expression but not KEGG pathway analysis results.

5.4 Results

5.4.1 KI and MI bear a highly variable relationship in breast tumors

We first evaluated how MI (the number of mitotic cells/10 HPFs) changes with KI in 241 H&E-stained breast tumors of differing Nottingham grades from the Northside and Emory cohorts. We observed that (a) all the slopes are <1 and (b) the slopes increase significantly grade-

wise (**Figure 1A**). Thus, with increasing tumor grade, KI increases at a faster rate than MI. A grade-wise comparison of mean MI and KI (**Figure 1B,C**) in the Nottingham and Northside cohorts also revealed a steeper increase in KI than MI (see y-axes in **Figure 1B,C**). Because KI and MI do not increase proportionately with each other, little information about cycling kinetics can be derived based upon independent consideration of these indices, suggesting a need for their rational integration, such as in metric KAMS (**Figure 1D**), which bears a more complex relationship with grade than MI or KI.

5.4.2 KAMS is a superior predictor of treatment response than both Ki67 and mitotic indices in three unique cohorts

Next, we analyzed the ability of KAMS to risk-stratify patients with different breast cancer subtypes based on BCSS (Nottingham and Stavanger cohorts) and OS (Nigeria cohort). We found that KAMS was unable to significantly stratify Luminal A, Luminal B, or HER2-overexpressing cases (data not shown). However, KAMS stratified TNBCs into subgroups with different survival prospects in three separate chemotherapy-treated cohorts ($p=0.015$, Nottingham; $p=0.073$, Stavanger; $p=0.006$, Nigeria) (**Figure 2**), with high-KAMS TNBCs having a worse prognosis than low-KAMS TNBCs. KAMS outperformed KI in all three cohorts in terms of the significance of the stratification (**Figure 2**). In all cohorts, KAMS also outperformed MI, which was far from significant in the Nottingham and Stavanger cohorts, whereas a non-significant trend was noted in the Nigeria cohort. Similar results were obtained in simple Cox models, wherein it was evident that model fit was better with KAMS than with KI per Akaike's Information Criterion (**Table 4**).

To determine whether KAMS is an independent predictor of poor outcomes, multivariable Cox models were fit. For all cohorts, age at diagnosis, tumor size, and Nottingham grade were available in clinical annotation and were included as covariates initially in full models. Lymph node stage was available for the Nottingham and Stavanger cohorts, but only lymph node status (positive or negative) was available for the Nigeria cohort, so these covariates (lymph node stage or status) were included in respective full models. Due to the presence of only a single Grade 1 tumor in the Nigerian cohort, Grades 1 and 2 were merged to obtain more precise estimates of the effect of grade on OS. KAMS outperformed KI and MI in all three cohorts and was a significant independent predictor of worse BCSS and OS in the Nottingham and Nigeria cohorts, respectively ($p=0.021$, $HR=2.20$ and $p=0.009$, $HR=2.46$); a non-significant trend was noted in the Stavanger cohort ($p=0.058$, $HR=4.19$) (**Table 5**). Thus, KAMS appears to be an independent predictor of poor survival, with high-KAMS groups exhibiting worse BCSS and OS. Remarkably, in the Nigeria cohort, KAMS was the only predictor in final models. By contrast, KI was only an independent predictor in the Nottingham cohort ($p=0.029$, $HR=0.47$), not appearing in final models of either the Stavanger or Nigeria cohorts, and MI was not a significant independent predictor in any cohort (data not shown).

5.4.3 KAMS groups are characterized by different levels of immunohistochemistry-based biomarkers

To derive insights into how KAMS high- and low-TNBCs differ biologically, we compared levels of 292 immunohistochemistry-based biomarkers between the two groups. Compared with the low-KAMS group, the high-KAMS group was enriched in nuclear UBC9 and SMC6L1 (H-scores) and nuclear wildtype ER β , γ -H2AX, and DNA-PK and membranous EGFR

(percentages). The high-KAMS group was deficient in cytoplasmic androgen receptor, nuclear VGLL1, total CD68, and intratumoral CD3, CD8, CD20, CD68, and FOXP3 (H-scores) and nuclear Ki67 and VGLL1 (percentages) (**Figure 3**). These data suggest that the KAMS groups exhibit significant differences in clinically meaningful immunohistochemical biomarker expression, such as the expression of DNA damage response proteins (e.g., nuclear SMC6L1, γ -H2AX, and DNA-PK) and immune cell markers (e.g., intratumoral CD3, CD8, CD20 CD68, FOXP3), which may underlie distinct and potentially actionable tumor biology.

5.4.4 KAMS groups are characterized by different molecular pathways

In the high KAMS group, 84 genes were overexpressed and 50 genes were underexpressed compared with the low KAMS group (**Tables 6 and 7**, respectively). Using the GAGE method, it was found that 31 KEGG pathways were differentially regulated according to KAMS status at the nominal $p \leq 0.05$ level, with 7 overexpressed and 24 underexpressed in the high KAMS group (**Table 8**). Many differentially regulated KEGG pathways belong under the umbrella of the immune system, with the high KAMS group exhibiting suppression of these pathways, which is consistent with the IHC biomarker analysis and may offer insights into the poor prognosis of this group. In addition, circadian rhythm and entrainment pathways, the disruption of which have been implicated in carcinogenesis [14], were upregulated in the high-KAMS group.

5.5 Discussion

We found that our cycling kinetics biomarker, KAMS, stratified patients with TNBC into groups with significantly different survival outcomes, but not patients with Luminal A, Luminal

B, or HER2-overexpressing breast cancers. In all three unique cohorts, KAMS outperformed KI as a predictor of clinical outcomes. KAMS groups were characterized by different immunohistochemistry-based biomarkers and RNA-sequenced molecular pathways, suggesting they constitute biologically distinct subgroups in TNBC that may be differentially susceptible to targeted therapies. Specifically, analysis of proteins implicated in cancer using immunohistochemistry revealed that the poor-prognosis, high-KAMS group exhibits upregulation of DNA damage response markers (e.g., SMC6L1, DNA-PK) and EGFR and downregulated immune cell markers (e.g., CD3, CD8, CD68) and immune-related molecular pathways (e.g., cytokine-cytokine receptor interaction and antigen processing and presentation), which could potentially be targeted with DNA damage response and EGFR antagonists and immunostimulatory molecules, respectively. The finding that the high-KAMS group exhibits immunosuppression was confirmed using RNAseq data, which revealed downregulation of immune-regulated KEGG pathways. An additional fascinating finding was that the high-KAMS group exhibits upregulation of pathways involved in circadian rhythm and entrainment. It has recently come to light that disruption of circadian rhythms at the cellular level is involved in multi-step breast tumorigenesis [14]. A broken “molecular clock” can lead to dysregulation of the cell cycle, suppress apoptosis, alter metabolism, promote epithelial-mesenchymal transition, thereby promoting tumorigenesis. Our findings here suggest that upregulation of circadian pathways in TNBC may be associated with aggressive disease features, such as a faster-running clock (high KAMS). TNBC is a highly heterogeneous disease in terms of gene expression profiles [15, 16], and KAMS may help to discriminate between clinically relevant TNBC subtypes in a relatively inexpensive and technically facile manner. Our study further reveals that the poor-prognosis high-KAMS group may benefit especially from certain targeted therapies,

like cetuximab (EGFR inhibitor) or immunostimulatory molecules, and is the first to lend evidence to the superior prognostic power of integrating KI and MI for chemotherapy-treated TNBC patients and novel TNBC subgroups that may be differentially susceptible to targeted therapies.

Our new integrated cell cycling kinetics metric KAMS revealed that, despite increases in KI and MI across tumor-grades, MI increases more slowly than KI, and the KI-MI relationship varies dynamically depending on Nottingham grade. This broken KI-MI relationship hints that mitosis may not always contribute the same proportion to the proliferative cell cycle. Despite the decades-long debate regarding the relative merits of KI and MI, it is indisputable that faster tumor growth is a sign of more aggressive disease. Faster tumor growth can result from two possible scenarios: (i) KI increases and MI increases proportionally with KI or (ii) both KI and MI increase but MI does not increase proportionally to the increase in KI. Our data show the latter possibility to be true. Moreover, recent studies have clearly divulged that the majority of cells within the proliferative cell population of a tumor are not actually dividing (i.e., are not in M-phase of the cell cycle) but are instead populating interphase [17-19]. The more rapidly cells transit through cell cycle, the higher is the proportion of mitotic cells observed in the proliferating population. Thus, KAMS should provide a measure of the mitotic propensity of the proliferative population.

It is well known that breast cancer molecular subtypes possess very different characteristics, and they may even be considered distinct diseases [20]. For example, they exhibit different proliferation rates (i.e., KIs), the highest of which is found among TNBCs [21]. The

literature is conflicting on the prognostic and predictive value of KI in TNBC [4-7]. We found that it was not an independent prognostic biomarker in chemotherapy-treated TNBC patients across three unique cohorts (British, Norwegian, and Nigerian women), nor was MI. However, rationally integrating these two indices in the cell cycling kinetics measure, KAMS, resulted in improved risk stratification in all cohorts. High KAMS was a significant independent predictor of poor survival in the Nottingham and Nigeria cohorts, and it nearly reached statistical significance in the Stavanger cohort, where it was a more significant biomarker than tumor size in final multivariable Cox models. A high KAMS value conceivably represents a tumor characterized by fast cell cycling, with proliferating cells rapidly progressing through interphase to enter M phase, resulting in aggressive tumor growth and, based on the higher frequency of mitoses (which are more error prone in cancer), higher intratumor heterogeneity, both of which contribute to increased risk of death.

TNBCs constitute an aggressive breast cancer subtype defined by the lack of ER and PR receptors and HER2 overexpression and characterized by high rates of recurrence and death, especially in the first 5 years of follow-up. TNBCs tend to be of a higher histological grade and clinical stage at diagnosis, and almost all currently available genomic prognostic signatures tend to assign poor prognostic risk status to TNBCs. Since TNBCs lack the most common breast cancer drug targets, the only therapeutic options currently available to treat them are cytotoxic chemotherapy, surgery, and radiation. TNBCs display marked heterogeneity in clinical behavior, spurring a search for stratifying biomarkers that allow discernment of inherent and potentially actionable differences in tumor biology among TNBCs that can inform clinical management. Although gene expression profiles have revealed the presence of distinct molecular subtypes

among TNBCs, the cost of TNBC subtyping by gene expression profiling is very steep, provides little predictive information, and is currently impractical for routine use. Thus, there is a paucity of relatively inexpensive and technically facile risk-predictive biomarkers that can risk-stratify TNBCs, reveal differences in their biology, and indicate to what drugs they may be sensitive, thereby enabling individualization of treatment.

It should be emphasized that the present study investigated KI and MI as assessed in separate microscopic fields, not always as continuous variables (e.g., in the Nottingham and Nigeria cohorts). In current clinical practice, KI and MI are normally determined by pathologists in different tissue sections and evaluated on disparate scales, which (i) overlooks the fact that mitotic cells comprise a subset of cycling cells, (ii) makes a direct cell-matched comparison of KI and MI impossible, and (iii) precludes evaluation of mitotic propensity and cell-cycling kinetics of the proliferative population in a tumor. Through our novel metric, KAMS, we attempted to “recapture” information about cell cycling kinetics, although it required us to make assumptions about MI based on empirical evidence to convert it from a categorical score to a percentage. Our promising results herein pave the way for future studies to investigate the value of extracting KI and MI from the same microscopic field, both as continuous variables, which we expect to yield superior risk-predictive information.

5.6 References

1. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up*. *Histopathology*, 1991. **19**(5): p. 403-10.
2. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up*. *Histopathology*, 2002. **41**(3A): p. 154-61.

3. Denkert, C., et al., *Strategies for developing Ki67 as a useful biomarker in breast cancer*. *The Breast*, 2015. **24**, **Supplement 2**: p. S67-S72.
4. Wang, J., et al., *Value of Breast Cancer Molecular Subtypes and Ki67 Expression for the Prediction of Efficacy and Prognosis of Neoadjuvant Chemotherapy in a Chinese Population*. *Medicine (Baltimore)*, 2016. **95**(18): p. e3518.
5. Wang, W., et al., *Prognostic and predictive value of Ki-67 in triple-negative breast cancer*. *Oncotarget*, 2016. **7**(21): p. 31079-87.
6. Hao, S., et al., *New insights into the prognostic value of Ki-67 labeling index in patients with triple-negative breast cancer*. *Oncotarget*, 2016. **7**(17): p. 24824-31.
7. Kashiwagi, S., et al., *Advantages of adjuvant chemotherapy for patients with triple-negative breast cancer at Stage II: usefulness of prognostic markers E-cadherin and Ki67*. *Breast Cancer Res*, 2011. **13**(6): p. R122.
8. van Diest, P.J., et al., *Reproducibility of mitosis counting in 2,469 breast cancer specimens: Results from the Multicenter Morphometric Mammary Carcinoma Project*. *Human Pathology*. **23**(6): p. 603-607.
9. Baak, J.P.A., et al., *Prospective Multicenter Validation of the Independent Prognostic Value of the Mitotic Activity Index in Lymph Node–Negative Breast Cancer Patients Younger Than 55 Years*. *Journal of Clinical Oncology*, 2005. **23**(25): p. 5993-6001.
10. Budczies, J., et al., *Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization*. *PLoS ONE*, 2012. **7**(12): p. e51862.
11. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biol*, 2014. **15**(12): p. 550.
12. Luo, W., et al., *GAGE: generally applicable gene set enrichment for pathway analysis*. *BMC Bioinformatics*, 2009. **10**(1): p. 161.
13. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995. **57**(1): p. 289-300.
14. Blakeman, V., et al., *Circadian clocks and breast cancer*. *Breast Cancer Research*, 2016. **18**(1): p. 89.
15. Burstein, M.D., et al., *Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer*. *Clin Cancer Res*, 2015. **21**(7): p. 1688-98.
16. Lehmann, B.D., et al., *Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection*. *PLoS ONE*, 2016. **11**(6): p. e0157368.
17. Komlodi-Pasztor, E., et al., *Mitosis is not a key target of microtubule agents in patient tumors*. *Nat Rev Clin Oncol*, 2011. **8**(4): p. 244-50.
18. Komlodi-Pasztor, E., D.L. Sackett, and A.T. Fojo, *Inhibitors targeting mitosis: tales of how great drugs against a promising target were brought down by a flawed rationale*. *Clin Cancer Res*, 2012. **18**(1): p. 51-63.
19. Ogden, A., et al., *Interphase microtubules: chief casualties in the war on cancer?* *Drug Discov Today*, 2014. **19**(7): p. 824-9.
20. Sorlie, T., *Molecular portraits of breast cancer: tumour subtypes as distinct disease entities*. *Eur J Cancer*, 2004. **40**(18): p. 2667-75.

21. Nishimura, R., et al., *Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer*. *Exp Ther Med*, 2010. **1**(5): p. 747-754.

Table 5.1 Patient and clinicopathologic characteristics of breast carcinoma patients in the Northside and Emory cohorts.

*Level: only applicable for categorical variables. **Missing: percentage calculated as the missing count out of the total number of cases (10,504 and 142 for the Northside and Emory cohorts, respectively), unlike other variables, which are calculated out of the valid (non-missing) cases in the column.

Variable	Level*	Statistic	Cohort		
			Northside	Emory	
Age at diagnosis (years)		Mean	57	56	
		Median	57	56	
		Minimum	21	23	
		Maximum	97	86	
		Missing	0	0	
Nottingham grade	1	Count	1638	2	
		Percent	35.7%	1.4%	
	2	Count	1804	26	
		Percent	39.3%	18.3%	
	3	Count	1145	114	
		Percent	25.0%	80.3%	
	Missing**	Count	0	0	
		Percent	0.0%	0.0%	
	Stage	0	Count	49	0
			Percent	1.1%	0.0%
I		Count	2495	56	
		Percent	54.6%	39.7%	
II		Count	1537	67	
		Percent	33.6%	47.5%	
III		Count	409	18	
		Percent	8.9%	12.8%	
IV		Count	81	0	
		Percent	1.8%	0.0%	
Missing**		Count	16	1	
		Percent	0.3%	0.7%	
Subtype		Luminal A	Count	1609	0
			Percent	35.1%	0.0%
	Luminal B	Count	1545	0	
		Percent	33.7%	0.0%	
	HER2+	Count	215	0	
		Percent	4.7%	0.0%	
	TNBC	Count	390	142	
		Percent	8.5%	100.0%	
	Missing**	Count	828	0	

Race	African American	Percent	18.1%	0.0%	
		Count	679	93	
	European American	Percent	14.8%	67.4%	
		Count	3630	40	
	Other	Percent	79.1%	29.0%	
		Count	160	5	
	Missing**	Percent	3.5%	3.6%	
		Count	118	4	
	Mitotic index (categorical)	1	Percent	2.6%	2.8%
			Count	2541	23
2		Percent	55.4%	16.2%	
		Count	1206	23	
3		Percent	26.3%	16.2%	
		Count	840	96	
Missing**		Percent	18.3%	67.6%	
		Count	0	0	
Ki67 index (Northside) or mitotic index (Emory)		Percent	0.0%	0.0%	
		Mean	23	25	
	Median	14	20		
	Minimum	0	0		
	Maximum	100	105		
	Missing	0	0		

Table 5.2 Patient and clinicopathologic characteristics of triple-negative breast cancer patients in the Nottingham, Stavanger, and Nigeria cohorts

*Level: only applicable for categorical variables. **Missing: percentage calculated as the missing count out of the total number of cases (10,504), unlike other variables, which are calculated out of the valid (non-missing) cases in the column. ***In the Stavanger cohort, mitotic index was calculated as number of mitoses/10 high-power fields; in the Nottingham and Nigeria cohorts, it was calculated based on the average number of mitotic cells/10-high-power fields for each Nottingham Grading System mitotic score (1, 2, or 3), as determined from the 241 total cases in the Northside subset and Emory cohort (n=99 and n=142, respectively). – Nodal stage was not available for the Nigeria cohort.

Variable	Level*	Statistic	Cohort			
			Nigeria	Nottingham	Stavanger	
Age at diagnosis (years)		Mean	49	44	51	
		Median	47	44	51	
		Minimum	26	25	22	
		Maximum	77	60	84	
		Number missing	0	0	0	
			Mean	3.4	2.6	2.4
Tumor size (cm)		Median	3	2.2	2.3	
		Minimum	1	0.6	0.2	
		Maximum	8	8	10	
		Number missing	0	0	2	
	Negative		Count	4	56	48
			Percent	8.30%	45.20%	66.70%
Positive		Count	44	68	24	
		Percent	91.70%	54.80%	33.30%	
Missing**		Count	0	0	0	
		Percent	0.00%	0.00%	0.00%	
Nodal stage	0	Count	-	56	48	
		Percent	-	45.20%	66.70%	
	1	Count	-	54	19	
		Percent	-	43.50%	26.40%	
	2	Count	-	14	5	
		Percent	-	11.30%	6.90%	
	Missing**	Count	48	0	0	
		Percent	100.00%	0.00%	0.00%	
	Nottingham grade	1	Count	1	0	1
			Percent	2.10%	0.00%	1.40%
2		Count	22	2	11	
		Percent				

		Percent	45.80%	1.60%	15.30%
	3	Count	25	122	60
		Percent	52.10%	98.40%	83.30%
	Missing**	Count	0	0	0
		Percent	0.00%	0.00%	0.00%
		Mean	11.5	31.6	24
		Median	2.9	32.6	23
Mitotic index (continuous)***		Minimum	0	2.9	0
		Maximum	32.6	32.6	105
		Number missing	0	0	0
	1	Count	22	2	8
		Percent	45.80%	1.60%	11.10%
	2	Count	12	3	7
Mitotic index (categorical)		Percent	25.00%	2.40%	9.70%
	3	Count	14	119	57
		Percent	29.20%	96.00%	79.20%
	Missing**	Count	0	0	0
		Percent	0.00%	0.00%	0.00%
		Mean	28	70	47
		Median	20	80	50
Ki67 index		Minimum	0	3	1
		Maximum	80	100	99
		Number missing	0	0	0
	Low	Count	35	35	14
		Percent	72.90%	28.20%	19.40%
Ki67 group	High	Count	13	89	58
		Percent	27.10%	71.80%	80.60%
	Missing**	Count	0	0	0
		Percent	0.00%	0.00%	0.00%
		Mean	1.5	0.8	1.3
		Median	0.4	0.4	0.5
KAMS		Minimum	0	0	0
		Maximum	9.8	10.9	14.5
		Number missing	0	0	0
	Low	Count	16	90	17
		Percent	33.30%	72.60%	23.60%
KAMS group	High	Count	32	34	55
		Percent	66.70%	27.40%	76.40%
	Missing**	Count	0	0	0
		Percent	0.00%	0.00%	0.00%

Table 5.3 Patient and clinicopathologic characteristics of the 99 breast carcinoma samples from the Northside cohort used to determine the average number of mitoses for mitotic index categories as used in the Nottingham Grading System

Variable	Level*	Statistic	Value
Age at diagnosis (years)		Mean	55
		Median	54
		Minimum	27
		Maximum	97
		Missing	0
Nottingham grade	1	Count	26
		Percent	26.30%
	2	Count	28
		Percent	28.30%
	3	Count	45
		Percent	45.50%
	Missing**	Count	0
		Percent	0.00%
	0	Count	0
		Percent	0.00%
AJCC stage	I	Count	36
		Percent	36.40%
	II	Count	39
		Percent	39.40%
	III	Count	18
		Percent	18.20%
	IV	Count	6
		Percent	6.10%
	Missing**	Count	0
		Percent	0.00%
Subtype	Luminal A	Count	17
		Percent	17.50%
	Luminal B	Count	41
		Percent	42.30%
	HER2+	Count	9
		Percent	9.30%
Race	TNBC	Count	30
		Percent	30.90%
	Missing**	Count	2
		Percent	2.00%
African American	Count	40	

		Percent	40.80%
	European American	Count	57
		Percent	58.20%
	Others	Count	1
		Percent	1.00%
	Missing**	Count	1
		Percent	1.00%
	No	Count	30
		Percent	33.30%
Chemotherapy	Yes	Count	60
		Percent	66.70%
	Missing**	Count	9
		Percent	9.10%
Mitotic index (categorical)	1	Count	31
		Percent	31.30%
	2	Count	40
		Percent	40.40%
	3	Count	28
		Percent	28.30%
Missing**	Count	0	
	Percent	0.00%	
Mitotic index (continuous)		Mean	17
		Median	10
		Minimum	0
		Maximum	129
		Missing	0
		Mean	33
Ki67 index		Median	28
		Minimum	0
		Maximum	96
		Missing	0

Table 5.4 Predictive value of KAMS and Ki67 index in simple Cox models of breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort)

Cohort	Variable	Omnibus Test of Model Coefficients			Variable in the equation			
		AIC	Overall (score)		HR	95% CI for HR		p-value
			X ²	p-value		Lower	Upper	
Nottingham	KAMS	320.2	5.93	0.015	2.25	1.15	4.4	0.018
	KI	320.8	5.24	0.022	0.47	0.24	0.91	0.025
Stavanger	KAMS	181.2	3.22	0.07	3.48	0.82	14.88	0.09
	KI	184.3	1.03	0.34	0.62	0.24	1.58	0.32
Nigeria	KAMS	268.1	7.08	0.008	2.46	1.25	4.86	0.009
	KI	274.9	0.49	0.49	1.01	0.99	1.02	0.49

Table 5.5 Multivariable Cox models of breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort) that include KAMS or KI in full models. Final model covariates, selected via backward stepwise elimination if $p > 0.10$, are shown. Model fit statistics for the final models are given. KI was excluded from final models in the Stavanger and Nigeria cohorts due to $p > 0.10$.

Cohort	Variable	Omnibus Test of Model Coefficients			Variable in the equation			
		AIC	Overall (score)		HR	95% CI for HR		p-value
			X ²	p-value		Lower	Upper	
Nottingham	KAMS	320.2	5.93	0.015	2.25	1.15	4.4	0.018
	KI	320.8	5.24	0.022	0.47	0.24	0.91	0.025
Stavanger	KAMS	181.2	3.22	0.07	3.48	0.82	14.88	0.09
	KI	184.3	1.03	0.34	0.62	0.24	1.58	0.32
Nigeria	KAMS	268.1	7.08	0.008	2.46	1.25	4.86	0.009
	KI	274.9	0.49	0.49	1.01	0.99	1.02	0.49

Table 5.6 Genes overexpressed in the high KAMS group at $q < 0.05$

Ensembl ID	Symbol	Log2FC	p-value	q-value
ENSG00000183091	NEB	1.69	4.00E-07	0.0015
ENSG00000154310	TNIK	1.62	2.50E-07	0.0015
ENSG00000152969	JAKMIP1	1.89	1.90E-07	0.0015
ENSG00000049540	ELN	1.96	4.00E-07	0.0015
ENSG00000141232	TOB1	1.21	1.40E-07	0.0015
ENSG00000259581	TYRO3P	1.29	7.00E-07	0.0021
ENSG00000073605	GSDMB	1.21	1.10E-06	0.0028
ENSG00000241112	RPL29P14	1.76	2.60E-06	0.0043
ENSG00000100784	RPS6KA5	1.13	1.90E-06	0.0043
ENSG00000123562	MORF4L2	0.72	2.50E-06	0.0043
ENSG00000068976	PYGM	1.95	3.40E-06	0.0048
ENSG00000127129	EDN2	2.14	5.00E-06	0.0054
ENSG00000128655	PDE11A	1.97	4.40E-06	0.0054
ENSG00000128039	SRD5A3	1.46	4.80E-06	0.0054
ENSG00000109586	GALNT7	1.39	4.70E-06	0.0054
ENSG00000155657	TTN	1.51	6.80E-06	0.0068
ENSG00000102804	TSC22D1	1	7.80E-06	0.0071
ENSG00000185347	C14orf80	1.51	7.50E-06	0.0071
ENSG00000175264	CHST1	1.89	9.80E-06	0.0077
ENSG00000156689	GLYATL2	1.99	9.60E-06	0.0077
ENSG00000008952	SEC62	0.74	1.70E-05	0.011
ENSG00000134909	ARHGAP32	0.77	1.50E-05	0.011
ENSG00000108515	ENO3	1.38	1.50E-05	0.011
ENSG00000239305	RNF103	0.83	1.80E-05	0.012
ENSG00000167081	PBX3	1.16	2.10E-05	0.012
ENSG00000151090	THRБ	1.38	2.70E-05	0.014
ENSG00000145014	TMEM44	1.4	2.70E-05	0.014
ENSG00000254786	RP11-142C4.5	1.86	2.60E-05	0.014
ENSG00000239474	KLHL41	1.98	3.60E-05	0.016
ENSG00000196782	MAML3	1.16	3.40E-05	0.016
ENSG00000139219	COL2A1	2.07	3.40E-05	0.016
ENSG00000121769	FABP3	1.74	3.80E-05	0.017
ENSG00000212493	SNORD19	1.21	4.10E-05	0.017
ENSG00000122367	LDB3	1.76	4.90E-05	0.019
ENSG00000261324	RP11-174G6.5	1.5	5.20E-05	0.019
ENSG00000169330	KIAA1024	1.2	5.00E-05	0.019
ENSG00000183077	AFMID	1.22	4.70E-05	0.019
ENSG00000229992	HMGB3P9	1.53	6.80E-05	0.024
ENSG00000144802	NFKBIZ	1.18	7.10E-05	0.025

ENSG00000176428	VPS37D	1.96	7.70E-05	0.026
ENSG00000104221	BRF2	1.06	7.70E-05	0.026
ENSG00000204624	PTCHD2	1.84	9.10E-05	0.028
ENSG00000227295	ELL2P1	1.8	8.80E-05	0.028
ENSG00000077235	GTF3C1	0.56	0.00012	0.033
ENSG00000142599	RERE	0.53	0.00019	0.039
ENSG00000164049	FBXW12	1.9	0.00014	0.039
ENSG00000177694	NAALADL2	1.27	0.00019	0.039
ENSG00000245958	RP11-33B1.1	0.98	0.00017	0.039
ENSG00000164040	PGRMC2	0.84	0.00014	0.039
ENSG00000235109	ZSCAN31	1.23	0.00019	0.039
ENSG00000164951	PDP1	1.12	0.00015	0.039
ENSG00000198682	PAPSS2	1.42	0.00016	0.039
ENSG00000100596	SPTLC2	0.93	0.00016	0.039
ENSG00000140474	ULK3	0.65	0.00016	0.039
ENSG00000064270	ATP2C2	1.43	0.00016	0.039
ENSG00000260105	AOC4P	1.8	0.00016	0.039
ENSG00000184828	ZBTB7C	1.38	0.00018	0.039
ENSG00000128709	HOXD9	1.47	0.00021	0.04
ENSG00000173175	ADCY5	1.58	0.00021	0.04
ENSG00000196935	SRGAP1	0.8	0.00022	0.04
ENSG00000140993	TIGD7	1.01	2.00E-04	0.04
ENSG00000126351	THRA	0.88	0.00022	0.04
ENSG00000139722	VPS37B	0.73	0.00024	0.042
ENSG00000136436	CALCOCO2	0.78	0.00024	0.042
ENSG00000198740	ZNF652	0.87	0.00024	0.042
ENSG00000109089	CDR2L	1.24	0.00025	0.042
ENSG00000213139	CRYGS	1.37	0.00025	0.043
ENSG00000110484	SCGB2A2	1.78	0.00027	0.044
ENSG00000140987	ZSCAN32	0.82	0.00027	0.044
ENSG00000115561	CHMP3	1.17	0.00029	0.046
ENSG00000164776	PHKG1	1.45	0.00029	0.046
ENSG00000186104	CYP2R1	1.1	0.00028	0.046
ENSG00000170100	ZNF778	0.81	3.00E-04	0.046
ENSG00000256683	ZNF350	1.08	0.00031	0.046
ENSG00000054118	THRAP3	0.39	0.00034	0.048
ENSG00000127249	ATP13A4	1.65	0.00035	0.048
ENSG00000168916	ZNF608	0.97	0.00033	0.048
ENSG00000149016	TUT1	0.86	0.00035	0.048
ENSG00000139990	DCAF5	0.6	0.00034	0.048
ENSG00000166821	PEX11A	1.25	0.00033	0.048
ENSG00000244165	P2RY11	1.14	0.00034	0.048
ENSG00000142197	DOPEY2	0.76	0.00033	0.048

ENSG00000133138	TBC1D8B	0.86	0.00035	0.048
ENSG00000138030	KHK	1.27	0.00035	0.049

Table 5.7 Genes underexpressed in the high KAMS group at $q < 0.05$

Ensembl ID	Symbol	Log2FC	p-value	q-value
ENSG00000196116	TDRD7	-0.87	2.80E-06	0.0043
ENSG00000152778	IFIT5	-1.23	2.30E-06	0.0043
ENSG00000185885	IFITM1	-1.59	8.60E-06	0.0074
ENSG00000157601	MX1	-1.44	1.00E-05	0.0077
ENSG00000134321	RSAD2	-1.48	2.00E-05	0.012
ENSG00000181143	MUC16	-1.82	1.90E-05	0.012
ENSG00000138646	HERC5	-1.39	3.10E-05	0.015
ENSG00000165905	GYLTL1B	-1.33	2.80E-05	0.015
ENSG00000168394	TAP1	-1.24	3.60E-05	0.016
ENSG00000133256	PDE6B	-1.61	3.90E-05	0.017
ENSG00000125730	C3	-1.09	4.50E-05	0.018
ENSG00000073792	IGF2BP2	-1.58	6.40E-05	0.023
ENSG00000051523	CYBA	-1.08	6.20E-05	0.023
ENSG00000113273	ARSB	-0.77	7.90E-05	0.026
ENSG00000236496	GPS2P1	-1.87	8.90E-05	0.028
ENSG00000101311	FERMT1	-1.58	9.70E-05	0.03
ENSG00000187554	TLR5	-1.04	1.00E-04	0.032
ENSG00000107165	TYRP1	-1.92	0.00011	0.032
ENSG00000130303	BST2	-1.36	0.00012	0.033
ENSG00000155629	PIK3AP1	-0.96	0.00013	0.036
ENSG00000162552	WNT4	-1.67	0.00019	0.039
ENSG00000055732	MCOLN3	-1.51	0.00015	0.039
ENSG00000115271	GCA	-0.73	0.00019	0.039
ENSG00000196653	ZNF502	-1.32	0.00015	0.039
ENSG00000124780	KCNK17	-1.79	0.00018	0.039
ENSG00000215045	GRID2IP	-1.68	0.00017	0.039
ENSG00000179869	ABCA13	-1.76	0.00017	0.039
ENSG00000160584	SIK3	-0.41	0.00015	0.039
ENSG00000136048	DRAM1	-0.98	0.00015	0.039
ENSG00000135114	OASL	-1.53	0.00019	0.039
ENSG00000121281	ADCY7	-0.82	0.00019	0.039
ENSG00000160183	TMPRSS3	-1.43	0.00018	0.039
ENSG00000198715	GLMP	-0.78	0.00021	0.04
ENSG00000164362	TERT	-1.65	0.00021	0.04
ENSG00000120539	MASTL	-0.7	0.00021	0.04
ENSG00000139278	GLIPR1	-0.94	0.00021	0.04
ENSG00000187775	DNAH17	-1.65	2.00E-04	0.04
ENSG00000050344	NFE2L3	-1.11	0.00024	0.042
ENSG00000102445	KIAA0226L	-1.11	0.00024	0.042

ENSG00000115828	QPCT	-1.46	0.00025	0.043
ENSG00000106392	C1GALT1	-0.97	0.00026	0.043
ENSG00000165996	HACD1	-1.43	0.00025	0.043
ENSG00000137959	IFI44L	-1.26	0.00029	0.046
ENSG00000213064	SFT2D2	-0.89	3.00E-04	0.046
ENSG00000070214	SLC44A1	-0.62	3.00E-04	0.046
ENSG00000156587	UBE2L6	-1.09	3.00E-04	0.046
ENSG00000111331	OAS3	-1.11	0.00029	0.046
ENSG00000197536	C5orf56	-1.03	0.00035	0.048
ENSG00000135966	TGFBRAP1	-0.64	0.00036	0.049
ENSG00000233122	CTAGE7P	-1.74	0.00036	0.049

Table 5.8 Differentially regulated pathways according to KAMS

Pink and green shading indicate overexpression and underexpression, respectively, in the high KAMS group at the nominal $p \leq 0.05$ level.

Category	Subcategory	KEGG pathway	p
Metabolism	Energy metabolism	Oxidative phosphorylation	0.02
	Amino acid metabolism	Valine, leucine and isoleucine degradation	0.02
Genetic information processing	Translation	Ribosome	0.04
	Folding, sorting and degradation	Proteasome	0.04
Environmental information processing	Signal transduction	Wnt signaling pathway	0.05
		Jak-STAT signaling pathway	0.005
		NF-kappa B signaling pathway	0.0005
		TNF signaling pathway	0.01
		AMPK signaling pathway	0.03
	Signaling molecules and interaction	Neuroactive ligand-receptor interaction	0.05
		Cytokine-cytokine receptor interaction	3e-7
Cell adhesion molecules		0.003	
Cellular processes	Transport and catabolism	Phagosome	2e-6
		Lysosome	0.04
		Peroxisome	0.03
Organismal systems	Immune system	Hematopoietic cell lineage	8e-6
		Complement and coagulation cascades	0.003
		NOD-like receptor signaling pathway	0.004
		Cytosolic DNA-sensing pathway	0.02
		Natural killer cell mediated cytotoxicity	0.0005
		Antigen processing and presentation	1e-6
		T cell receptor signaling pathway	0.04
		B cell receptor signaling pathway	0.005
		Intestinal immune network for IgA production	0.001
	Chemokine signaling pathway	0.009	
	Endocrine system	Prolactin signaling pathway	0.04
	Circulatory system	Cardiac muscle contraction	0.04
	Digestive system	Mineral absorption	0.05
	Development	Osteoclast differentiation	0.01
	Environmental adaptation	Circadian rhythm	0.01
Circadian entrainment		0.03	

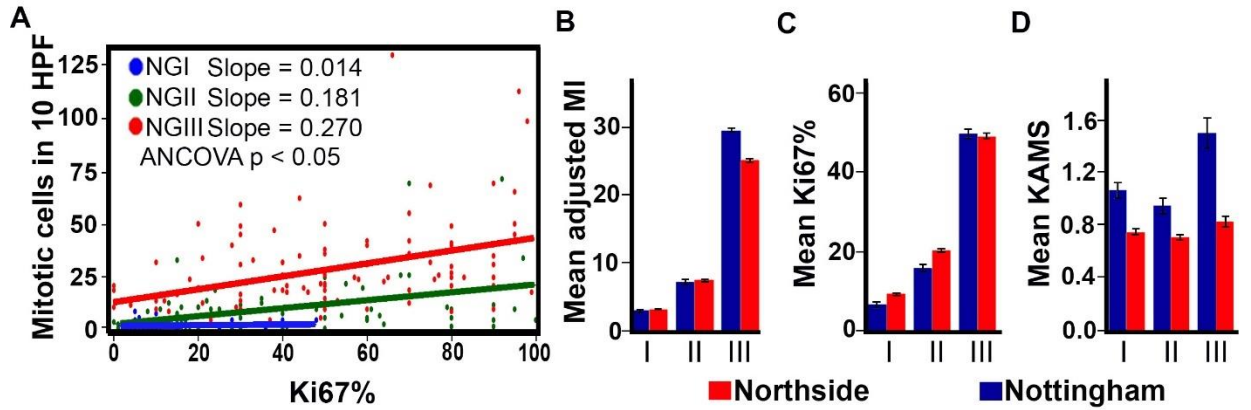


Figure 5.1 Analysis of grade-wise relationship between KI and MI

Comparison of mitotic counts for 241 total breast carcinoma samples from the Northside mixed-subtype and Emory TNBC cohorts (n=99 and 142, respectively), with the respective lines for each Nottingham grade (NG) representing significantly different fitted linear regressions per ANCOVA. Mean **B.** MI, **C.** KI, and **D.** KAMS of breast carcinoma samples from the Northside and Nottingham cohorts.

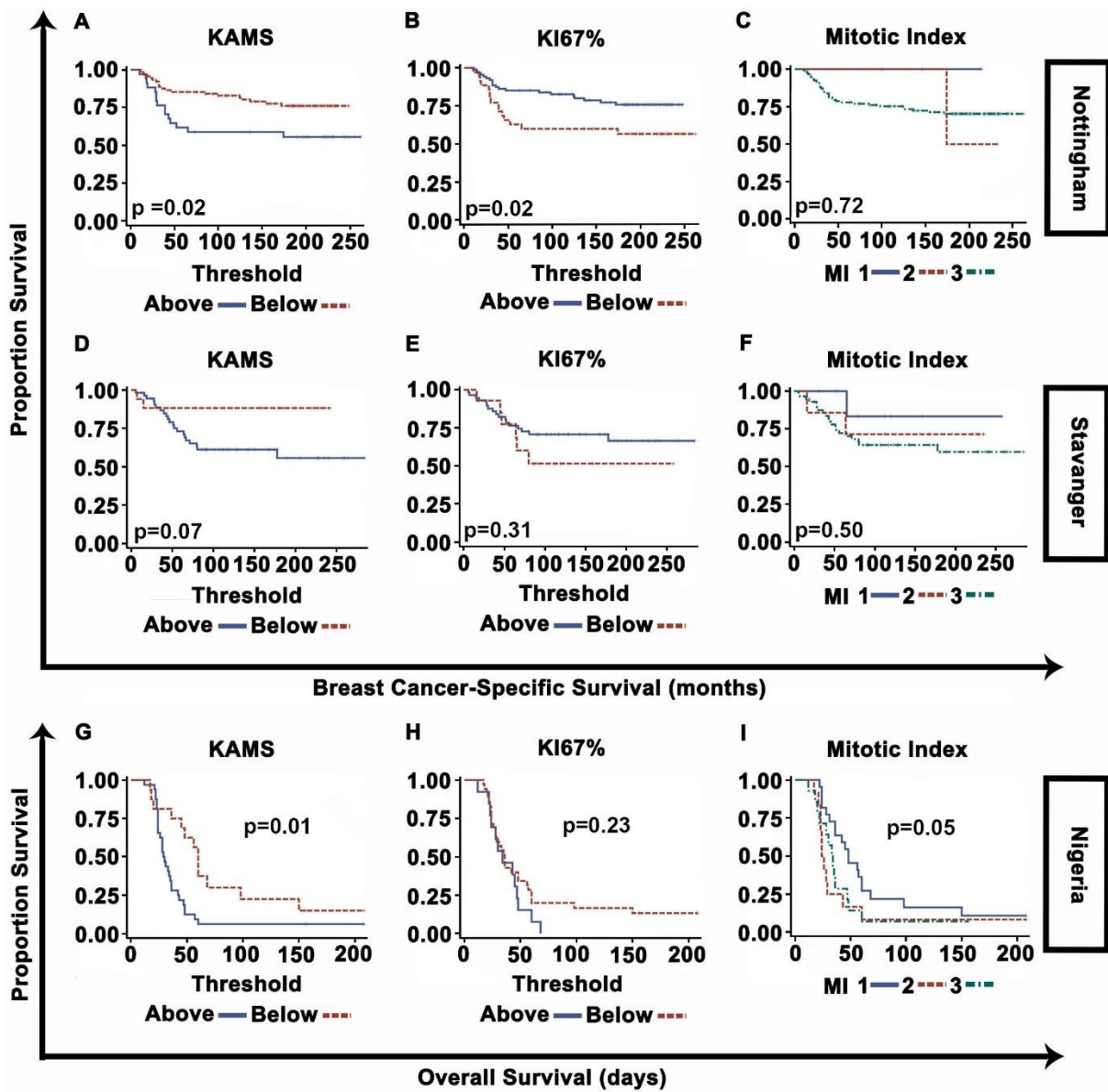


Figure 5.2 Stratification by KAMS, Ki67 index, and mitotic index according to breast cancer-specific survival (Nottingham and Stavanger cohorts) and overall survival (Nigeria cohort)

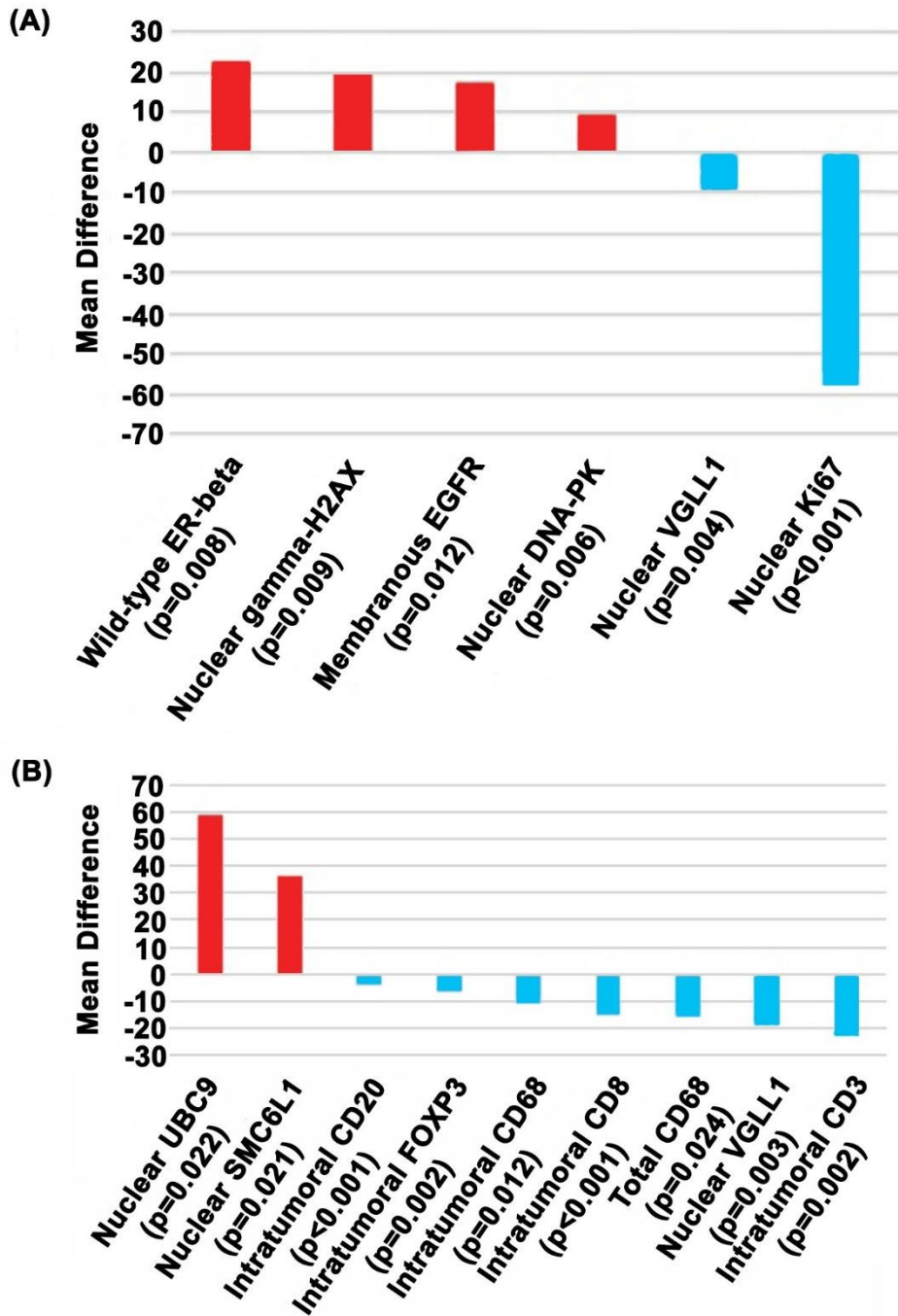


Figure 5.3 Comparison of significantly different immunohistochemistry-based biomarkers between high- and low-KAMS TNBCs.

A. H-scores. B. Percentages

6 HIGH HER3-EGFR SCORE PREDICTS AGGRESSIVE DISEASE COURSE IN TRIPLE-NEGATIVE BREAST CANCER

6.1 Abstract

Several proteins of the ErbB receptor tyrosine kinase family may have prognostic value in breast cancer. While the clinical utility of assaying levels of human epidermal growth factor receptor 2 (HER2) is well established, it is less apparent whether human epidermal growth factor receptor 3 (HER3) or epidermal growth factor receptor (EGFR) could guide clinical management of breast cancer. Biomarkers are urgently needed for triple-negative breast cancer (TNBC), which is defined by the biomarkers it lacks and for which no targeted therapies are available. There is limited evidence that HER3 and EGFR could serve as biomarkers in TNBC. However, HER3 is kinase impaired and must bind to other ErbB family members, like EGFR, to effect signaling, so it may be necessary to consider HER3 in this context. Thus, we developed a combined HER3-EGFR score (the sum of the individual HER3 and EGFR H-scores), which we found outperformed assessment of HER3 and EGFR H-scores in isolation in multivariable survival models. Specifically, in a multi-institutional study of n=510 TNBC patients, we found that an above-median HER3-EGFR score was associated with 2.30-times worse breast cancer-specific survival (BCSS) and 1.78-times worse distant metastasis-free survival (DMFS) ($p=0.006$ and $p=0.041$, respectively) in age- and stage-adjusted Cox models among adjuvant chemotherapy-treated (but not untreated) TNBC patients. Moreover, based on an analysis of 105 immunohistochemical biomarkers, we found that adjuvant chemotherapy-treated TNBCs with above-median HER3-EGFR scores exhibited upregulated expression of luminal cytokeratins, DNA damage-response proteins, and P-cadherin. Ingenuity Pathway Analysis (IPA) of

RNAsequenced TNBCs revealed that tumors with above-median HER3-EGFR scores were enriched in genes involved in hepatic fibrosis, which has been linked to distant metastasis. Altogether, the HER3-EGFR score may identify which chemotherapy-treated TNBC patients require more aggressive treatment and potential targets for therapy.

6.2 Introduction

Accruing evidence suggests that HER3 may be a valuable target in breast cancer, including TNBC, for which there are no approved targeted drugs to date. HER3 is kinase impaired, demonstrating only relatively weak autophosphorylation and no ability to phosphorylate other peptides [1]. It can, however, heterodimerize with other ErbB family members to effect downstream signaling. Targeted dual anti-HER3/EGFR therapy may be a useful adjunct to PI3K-AKT antagonism when treating TNBCs that exhibit PI3K-AKT pathway overdrive, a common phenotype in the disease, because PI3K-AKT inhibitors induce compensatory upregulation of receptor tyrosine kinases like HER3 and EGFR [2-4]. Indeed, antagonizing HER3 along with EGFR enhances the sensitivity of TNBC cells to PI3K-AKT pathway inhibitors [4]. In addition, HER3 may have value as a prognostic biomarker in TNBC, as HER3-positivity (defined as no HER3 immunostaining or <10% of cells with membranous immunostaining) is associated with shorter overall and disease-free survival in TNBC [5]. However, it is unknown whether HER3 is an independent predictor of poor outcomes in TNBC or whether it is merely a surrogate for advanced stage or other negative prognostic indicators, as multivariable analyses have not been explored. Furthermore, given that HER3 primarily relies on heterodimerization to signal [6], assessing HER3 scores in isolation as a prognostic measure may be inferior to assessing HER3 levels in combination with the scores of other ErbB family

members. For TNBC, which does not overexpress HER2, determining HER3 levels in the context of EGFR levels may prove superior to assessing HER3 levels alone. (HER3 may bind to HER4, but this interaction appears to be mitogenically unproductive [7, 8]). Thus, our goal herein was to evaluate whether a combined HER3-EGFR score (the sum of the individual HER3 and EGFR H-scores) could predict survival in TNBC patients and to compare its prognostic potential to individual HER3 and EGFR H-scores. Furthermore, we characterized differences in molecular pathways between HER3-EGFR groups using RNA-seq data and IPA to identify potential targets for therapy in TNBC.

6.3 Materials and methods

6.3.1 Datasets

TNBCs were identified as completely lacking ER and PR expression and having <10% membrane cells with any staining for HER2. Three TNBC datasets were included in the study: Nottingham (n=302), Norway (n=104), and Emory (n=104), consisting of adjuvant chemotherapy-treated and untreated breast cancer patients with stage I-III disease. Cases included patients from the well-characterized Nottingham Tenovus Primary Breast Carcioma Series [9, 10] diagnosed with primary operable invasive TNBC presenting between 1986 and 1998 for which staining of HER3 and EGFR has been previously described [11]. Descriptive patient and clinicopathologic statistics can be found in **Table 6.1**. The Nottingham Research Ethics Committee 2 approved all aspects of the study under the title “Development of a molecular genetic classification of breast cancer.” Samples were deidentified and collected before 2006, so informed consent was not required pursuant to the Human Tissue Act. Data

regarding mortality, disease progression, and distant metastases were maintained prospectively for patients presenting after 1989.

6.3.2 *Sample scoring*

For TNBC samples in the Norway and Emory cohorts, an experienced pathologist blind to clinical annotation scored membrane staining intensity (0-3+) and the percentage of cells with any evidence of membranous staining (0-100%), the product of which was the individual Histochemical (H)-score for the biomarker. Only cores with at least 15% of tumor cells were scored, and only invasive breast cancer cells were scored. The two cores were scored independently at 20X magnification and the average score was taken.

6.3.3 *Survival analyses*

Survival outcomes were defined as the time from the date of diagnosis to the time of breast cancer-specific death (BCSS) or distant metastasis (DMFS). Average BCSS and DMFS times were 224 months and 200 months, respectively, for the entire study and were 213 and 167 months, respectively, for the Nottingham cohort; 216 and 197 months, respectively, for the Norway cohort; and 123 and 117 months, respectively, for the Emory cohort. Continuous HER3, EGFR, and HER3-EGFR H-scores were converted to categorical variables based on the median. The impact of categorical HER3, EGFR, and HER3-EGFR H-scores on BCSS and DMFS in adjuvant chemotherapy-treated vs. untreated patients were tested using simple and multivariable Cox models, stratified by hospital. It was found that the proportional hazards assumption was satisfied by plotting partial residuals against rank-time and determining that the slope did not significantly differ from zero per ANOVA (data not shown). AJCC stage and age at diagnosis

were entered as covariates into multivariable models. Nottingham grade was not included because the paucity of grade 1 and 2 patients resulted in a great deal of uncertainty in hazard ratios (i.e., confidence intervals were excessively wide). Results of survival analyses were considered significant if $p < 0.05$. Analyses were conducted using IBM SPSS Statistics v. 21.

6.3.4 Biomarker comparisons

Differences in 105 cellular and tissue biomarkers between HER3-EGFR-high and low TNBCs among chemotherapy-treated patients from the Nottingham cohort were compared using 2-tailed Mann-Whitney U tests. Cellular biomarkers were recorded as intensities, percentages, and/or H-scores with nuclear, cytoplasmic, total (nuclear + cytoplasmic), and/or membranous levels. Tissue biomarkers were recorded as intratumoral, stromal, and/or total (intratumoral + stromal) levels. There were 314 comparisons, so to set the false-discovery rate at $q = 0.25$, we calculated $p = 0.25 * R/m$ (where R = number of rejected null hypotheses at $\alpha = 0.05$ and m = the number of tested null hypotheses) to determine statistical significance. As $R = 30$ and $m = 314$, $p = 0.018$ was considered statistically significant for this analysis. Analyses were conducted using IBM SPSS Statistics v. 21.

6.3.5 RNA sequencing

We performed RNA-seq on 78 TNBCs (pre-treatment formalin-fixed paraffin-embedded [FFPE] resection specimens) from the Nottingham cohort for which HER3 and EGFR scores were available. Following deparaffinization with xylene, samples were processed for RNA extraction using the Omega Mag-Bind XP FFPE RNA isolation kit (Omega, M2595-01) and KingFisher Flex magnetic particle separator (ThermoFisher). The Nanodrop 2000c

spectrophotometer (Thermo Scientific Inc., Waltham, MA, USA) was used to measure RNA concentration, and RNA integrity was analyzed with the Agilent 2200 TapeStation instrument (Agilent Technologies, Santa Clara, CA) by calculating the percentages of fragments larger than 200 nucleotides (DV200). First-strand cDNA syntheses were accomplished using ~100 ng RNA at 25°C for 10 min, 42°C for 15 min, and 70°C for 15 min by adding random hexamers and ProtoScript II Reverse Transcriptase (New England BioLabs, Ipswich, MA). The TruSeq RNA Access library kit (Illumina, Inc., San Diego, CA, USA) was used according to the manufacturer's instructions for second-strand synthesis and to prepare RNA sequencing libraries. RNA templates were eliminated and a second replacement strand was created using dUTP to synthesize double-stranded cDNA. Blunt-end cDNA was removed using Mag-Bind RxnPure Plus magnetic beads (Omega Bio-tek, Norcross, GA). The 3' ends were adenylated and then ligated with indexing adaptors. PCR (15 cycles of 98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec) was performed to enrich dual-end adaptor-ligated DNA fragments. Agilent 2200 TapeStation was used to qualify the library and the QuantiFluor dsDNA System (Promega, Madison, WI) to quantify it. A 4-plex pool of libraries was created by adding 200 ng of each DNA library to the pool, which was then mixed with capture probes. Hybridization was accomplished through 18 cycles of 1 min incubation, beginning at 94 °C, and then lowering the temperature by 2 °C per cycle. Probes hybridized to target regions were captured using streptavidin-coated magnetic beads. Enriched libraries were eluted and again subjected to hybridization followed by 10 cycles of PCR amplification (98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec). RxnPure magnetic beads were used for clean-up. Agilent High Sensitivity D1000 ScreenTape on an Agilent 2200 TapeStation instrument was used for final library validation. The final library size distribution was ~200 bp–1 kbp. Normalization, pooling, and

clustering was performed on the libraries, which were pair-red sequenced for 75 cycles on the HiSeq2500 (Illumina, Inc. San Diego, CA, USA), per the manufacturer's instructions.

6.3.6 RNA-seq data analyses

78 malignant samples from the Nottingham dataset were RNAsequenced. Differentially expressed transcripts between HER3-EGFR-high and low groups (stratified by the median) were determined using DESeq2 in GenePattern with default settings. 206 genes were differentially expressed at FDR<0.25, 178 genes of which were differentially expressed at least log₂-fold and which were entered into IPA for canonical pathway analysis with stratification based on the median HER3-EGFR score as determined in the multi-institutional n=510 TNBC sample. Default settings were used except the species was limited to *Homo sapiens*.

6.3.7 Oncomine analyses

Oncomine [12] breast cancer databases were interrogated by entering the terms, “HNF4A,” “breast cancer,” and “Stage Analysis” “Clinical Specimen,” and “mRNA.” All returned datasets had ≥95 patients. Early vs. advanced-stage breast cancers were compared. In addition, the database was interrogated by entering “Triple-Negative Breast Cancer” instead of “Stage Analysis,” and filtering for datasets with n≥50 patients. This returned 14 datasets, for which we compared HNF4A expression between TNBCs and non-TNBCs. We also compared copy-number variation between ductal carcinomas in situ and invasive ductal breast carcinomas in the Curtis Breast Dataset. We also conducted Cancer Outlier Profile Analysis (COPA) using Curtis Breast data, which identifies genes with striking overexpression in a subset of cases in a dataset [13], at the 90th percentile.

6.4 Results

We were interested to determine whether a combined HER3-EGFR score could predict survival in chemotherapy-treated or untreated patients. Although the treatment-predictive value of a biomarker can only be established in randomized controlled clinical trials, retrospective analysis of the impact of a biomarker on survival outcomes in chemotherapy-treated vs. untreated patients can provide insights into whether a biomarker may have predictive value. Furthermore, analyzing treatment subgroups is important because of the potentially confounding effect of chemotherapy on survival. Thus, we compared the impact of high vs. low HER3, EGFR, and HER3-EGFR H-scores categorized based on the median in chemotherapy-treated and untreated TNBC patients, stratified by hospital. In simple Cox models, high HER3-EGFR was associated with 2.50-fold increased risk of dying from breast cancer in chemotherapy-treated patients ($p=0.003$) but not untreated patients (**Table 6.2**). By contrast, HER3 and EGFR assessed as individual biomarkers did not have a significant effect on BCSS regardless of chemotherapy status. Similar results were obtained in analyses of DMFS. High HER3-EGFR was associated with 1.95-fold increased risk of distant metastasis in chemotherapy-treated patients ($p=0.022$) but not untreated patients (**Table 6.2**). Neither HER3 nor EGFR assessed as individual biomarkers significantly impacted DMFS in chemotherapy-treated or untreated patients.

In stage- and age-adjusted models, stratified by hospital, high HER3-EGFR remained a significant predictor of worse BCSS in chemotherapy-treated patients ($HR=2.30$, $p=0.006$) but not untreated patients (**Table 6.3**). Neither HER3 nor EGFR assessed as individual biomarkers significantly impacted BCSS in either treated or untreated patients in multivariable models. Similarly, in stage- and age-adjusted models, stratified by hospital, high HER3-EGFR predicted

worse DMFS in chemotherapy-treated patients (HR=1.78, p=0.041) but not untreated patients (**Table 6.3**). By contrast, neither HER3 nor EGFR assessed as individual biomarkers significantly impacted DMFS in multivariable models. Thus, consideration of HER3 and EGFR jointly independently predicts poor outcomes in adjuvant chemotherapy-treated patients, unlike assessment of HER3 or EGFR in isolation.

We were also interested in characterizing differences in 105 cellular and tissue immunohistochemical biomarkers implicated in cancer (recorded variously as intensities, percentages, and H-scores, as appropriate, for nuclear, cytoplasmic, total cell, and/or membranous levels for cellular biomarkers and as intratumoral, stromal, and/or total tissue levels for tissue biomarkers) between HER3-EGFR-stratified TNBC samples from adjuvant chemotherapy-treated patients. Among the 314 comparisons, only 8 had $q < 0.25$. These included luminal cytokeratins 7/8 and 18 (percentages and H-scores), DNA repair molecules SMC6L1 (percentage) and dicer (percentage and H-score), and placental cell-cell adhesion molecule P-cadherin (H-score), all of which were higher in the HER3-EGFR-high group of tumors from adjuvant chemotherapy-treated TNBC patients (**Table 6.4**).

Finally, we determined differentially expressed genes and differentially regulated pathways between HER3-EGFR-high and low TNBCs. 78 pre-treatment TNBC resections for which HER3-EGFR scores were available were subjected to RNA-seq. Differentially expressed genes were determined using DESeq2, and genes with log₂-fold differential expression and FDR < 0.25 were subjected to canonical pathway analysis using IPA. The top differentially regulated canonical pathway was Hepatic Fibrosis/Hepatic Stellate Cell Activation (p=0.008).

Consistent with this finding, the top toxicity function was early-stage liver cirrhosis ($p=9.4 \times 10^{-4}$). The top upstream regulator was Hepatocyte nuclear factor 4 alpha (HNF4A) ($p=0.012$), and the top disease/disorder was gastrointestinal cancer ($p=3.0 \times 10^{-9}$). A literature search revealed that HNF4A binds to ERBB3 (the gene encoding HER3), but not *EGFR*, in HepG2 cells [14]. Genes identified in the Hepatic Fibrosis pathway were distinct from those identified as being regulated by HNF4A, all of which were upregulated in the HER3-EGFR-high group.

Little is known about HNF4A in breast cancer, but because it is a key upstream regulator in the HER3-EGFR-high group, which is more likely to experience distant metastasis, we wondered whether HNF4A was associated with advanced nodal stage in breast cancer. To this end, we queried Oncomine by entering the terms, “HNF4A,” “breast cancer,” and “Stage Analysis” “Clinical Specimen,” and “mRNA,” which produced 7 microarray datasets for comparison, all of which had ≥ 95 patients. It was found that HNF4A was overexpressed in advanced-stage disease ($p=0.04$, median rank=1122.0). We also tested whether it was upregulated in TNBCs vs. non-TNBCs, so we repeated the query by entering “Triple-Negative Breast Cancer” instead of “Stage Analysis,” filtering for datasets with $n \geq 50$ patients, resulting in 14 datasets. It was found that HNF4A was not upregulated in TNBCs across datasets ($p=0.40$, median rank=5025.5). Thus, HNF4A is associated with advanced-stage breast cancer but it is not upregulated in TNBCs vs. non-TNBCs. We also assessed copy-number variation between ductal carcinomas in situ (DCIS; $n=10$) and invasive ductal breast carcinomas ($n=1,556$) in the Curtis Breast Dataset in Oncomine. We found that HNF4A copy number was significantly higher in malignant samples than DCIS samples ($p=0.019$, fold-change=1.039; top 6% of genes). In the entire $n=1,992$ -sample dataset, the Cancer Outlier Profile Analysis (COPA) score (which

identifies genes with striking overexpression in a subset of cases in a dataset[13]) at the 90th percentile is 2.437 (median rank=3,189); it was overexpressed in many datasets, and it was not underexpressed in any dataset. Thus, there appears to be a subset of breast tumors that overexpress HNF4A.

6.5 Discussion

The finding that a combined HER3-EGFR score, but not individual HER3 or EGFR H-scores, independently predicts worse BCSS and DMFS in adjuvant chemotherapy-treated TNBC patients suggests that membranous HER3 and EGFR levels do not have prognostic value in TNBC in isolation but, rather, must be considered jointly. Because HER3 is kinase impaired, it must heterodimerize with other ErbB family members such as EGFR. The tumors of chemotherapy-treated TNBC patients with high HER3-EGFR scores exhibited higher expression levels of luminal cytokeratins, DNA repair molecules, and P-cadherin than patients with low HER3-EGFR scores based on immunohistochemical analyses. It has been shown that TNBCs can be divided into four subtypes by immunohistochemistry, and the luminal subtype, which exhibits luminal cytokeratin expression, is associated with poor outcomes among TNBC patients treated with adjuvant cytotoxic chemotherapy [15]. Our results are in alignment with this finding, and so it is possible that the chemoresistant luminal-type TNBCs could be treated with a targeted dual HER3-EGFR inhibitor such as MEHD7945A. High levels of DNA repair molecules suggest that drugs targeting DNA repair pathways, such as PARP inhibitors, may also be useful in treating HER3-EGFR-high TNBCs, which appear to be less responsive to adjuvant cytotoxic chemotherapy than HER3-EGFR-low TNBCs. Because HER3-EGFR-high TNBCs exhibit relatively high P-cadherin immunostaining, they could also potentially be susceptible to

P-cadherin inhibitors, such as PCA062. A phase 1 multicenter Open-label Dose Escalation and Expansion Study of PCA062 is currently underway testing the efficacy of this drug in p-CAD-positive TNBCs.

The top differentially expressed canonical pathway – hepatic fibrosis – was upregulated in the HER3-EGFR-high group of TNBCs. Fibrosis has recently been connected to solid tumor metastasis [16]. During fibrosis of a primary tumor, tumor cells recruit and activate stromal cells, such as myofibroblasts, to produce desmoplasia, which has a composition similar to fibrotic non-malignant extracellular matrix and facilitates metastasis. Invasive breast cancers that metastasize exhibit higher expression of the myofibroblast marker α -smooth muscle actin than those that do not metastasize, and α -smooth muscle actin is an independent predictor of metastasis [17]. Thus, a high HER3-EGFR score may identify a subset of breast cancer patients whose tumors exhibit a transcriptional profile associated with fibrotic disease, which can drive distant metastasis, consistent with our finding that high HER3-EGFR score independently predicted worse DMFS in adjuvant chemotherapy-treated TNBC patients.

Seemingly in contrast, we identified the top differentially expressed upstream transcriptional regulator of differential gene expression in the HER3-EGFR high group of RNA-sequenced TNBCs as HNF4A. This transcription factor antagonizes hepatic fibrosis and promotes hepatocyte differentiation [18]. However, it is amplified in certain cancers, such as colon (in particular, a subtype associated with features of chromosomal instability), promotes gut tumorigenesis in mice, and it is a susceptibility gene for ulcerative colitis and other inflammatory bowel disorders [19-21]. It has been found that hepatocytes express only the “P1” isoforms of

HNF4A, whereas intestinal epithelial cells express both the “P1” and “P2” isoforms and hepatocellular carcinomas and murine embryonic liver preferentially express “P2” isoforms [22, 23]. P2 isoforms are a collection of three isoforms, $\alpha 7-9$, that result from the use of a more distant promoter, are shorter, and lack the cofactor-interacting domain present in group of six P1 isoforms, $\alpha 1-6$ [21]. P2 isoforms appear to promote proliferation of colon cancer cells, whereas P1 isoforms exhibit mixed effects [24]. A study of n=45 mucinous ovarian carcinomas found that about a third exhibit P1 staining and nearly all exhibit P1/P2 staining [25]. The involvement of P1 and P2 isoforms in breast carcinogenesis represents entirely unexplored territory. Although our Oncomine analysis did not differentiate between isoforms, we report here that HNF4A is associated with advanced-stage breast cancer and invasive breast cancer, may be overexpressed in a subset of breast tumors, and genes targeted by HNF4A are enriched in the poor-prognosis HER3-EGFR-high group. It appears as though a canonical pathway associated with hepatic fibrosis coexists with enrichment in HNF4A targets in the HER3-EGFR-high group; we speculate that P2 isoforms may be upregulated in the poor-prognosis HER3-EGFR-high group, resulting in inflammation that may promote desmoplasia, an exciting avenue for future research. This is consistent with the finding that a network involved in gastrointestinal inflammation was upregulated in the HER3-EGFR-high group and represents an exciting avenue for future studies.

6.6 References

1. Shi, F., et al., *ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(17): p. 7692-7697.
2. Serra, V., et al., *PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer*. Oncogene, 2011. **30**(22): p. 2547-57.
3. Sergina, N.V., et al., *Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3*. Nature, 2007. **445**(7126): p. 437-41.

4. Tao, J.J., et al., *Antagonism of EGFR and HER3 Enhances the Response to Inhibitors of the PI3K-Akt Pathway in Triple-Negative Breast Cancer*. *Science signaling*, 2014. **7**(318): p. ra29-ra29.
5. Bae, S.Y., et al., *HER3 status by immunohistochemistry is correlated with poor prognosis in hormone receptor-negative breast cancer patients*. *Breast Cancer Res Treat*, 2013. **139**(3): p. 741-50.
6. Littlefield, P., et al., *Structural analysis of the EGFR/HER3 heterodimer reveals the molecular basis for activating HER3 mutations*. *Sci Signal*, 2014. **7**(354): p. ra114.
7. Riese, D.J., et al., *The cellular response to neuregulins is governed by complex interactions of the erbB receptor family*. *Molecular and Cellular Biology*, 1995. **15**(10): p. 5770-5776.
8. Tzahar, E., et al., *A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor*. *Mol Cell Biol*, 1996. **16**(10): p. 5276-87.
9. Abd El-Rehim, D.M., et al., *Expression of luminal and basal cytokeratins in human breast carcinoma*. *J Pathol*, 2004. **203**(2): p. 661-71.
10. Rakha, E.A., et al., *Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation*. *J Pathol*, 2006. **208**(4): p. 495-506.
11. Green, A.R., et al., *Nottingham prognostic index plus (NPI+) predicts risk of distant metastases in primary breast cancer*. *Breast Cancer Res Treat*, 2016. **157**(1): p. 65-75.
12. Rhodes, D.R., et al., *ONCOMINE: a cancer microarray database and integrated data-mining platform*. *Neoplasia*, 2004. **6**(1): p. 1-6.
13. Rhodes, D.R., et al., *Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles*. *Neoplasia* (New York, N.Y.), 2007. **9**(2): p. 166-180.
14. Odom, D.T., et al., *Control of pancreas and liver gene expression by HNF transcription factors*. *Science*, 2004. **303**(5662): p. 1378-81.
15. Elsayaf, Z., et al., *Biological subtypes of triple-negative breast cancer are associated with distinct morphological changes and clinical behaviour*. *Breast*, 2013. **22**(5): p. 986-92.
16. Cox, T.R. and J.T. Erler, *Molecular Pathways: Connecting Fibrosis and Solid Tumor Metastasis*. *Clinical Cancer Research*, 2014. **20**(14): p. 3637-3643.
17. Yamashita, M., et al., *Role of stromal myofibroblasts in invasive breast cancer: stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome*. *Breast Cancer*, 2012. **19**(2): p. 170-6.
18. Yue, H.Y., et al., *Hepatocyte nuclear factor 4alpha attenuates hepatic fibrosis in rats*. *Gut*, 2010. **59**(2): p. 236-46.
19. Chellappa, K., G.R. Robertson, and F.M. Sladek, *HNF4a: a new biomarker in colon cancer?* *Biomarkers in medicine*, 2012. **6**(3): p. 297-300.
20. Darsigny, M., et al., *Hepatocyte nuclear factor-4alpha promotes gut neoplasia in mice and protects against the production of reactive oxygen species*. *Cancer Res*, 2010. **70**(22): p. 9423-33.
21. Babeu, J.P. and F. Boudreau, *Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks*. *World J Gastroenterol*, 2014. **20**(1): p. 22-30.

22. Babeu, J.-P. and F. Boudreau, *Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks*. World Journal of Gastroenterology : WJG, 2014. **20**(1): p. 22-30.
23. Dean, S., et al., *Developmental and tissue-specific regulation of hepatocyte nuclear factor 4-alpha (HNF4-alpha) isoforms in rodents*. Gene Expr, 2010. **14**(6): p. 337-44.
24. Zhang, B., et al., *Proteogenomic characterization of human colon and rectal cancer*. Nature, 2014. **513**(7518): p. 382-387.
25. Sugai, M., et al., *Expression of hepatocyte nuclear factor 4 alpha in primary ovarian mucinous tumors*. Pathol Int, 2008. **58**(11): p. 681-6.

Table 6.1 Clinicopathologic data for Nottingham, Norway, and Emory datasets

Variable	Statistic (continuous variables) or level (categorical variables)	Hospital		
		Nottingham	Norway	Emory
HER3 H-score	Median	146	50	5
	Mean	146	53	20
	Standard Deviation	75	46	38
	Minimum	0	0	0
	Maximum	300	180	185
	Missing	0	0	0
EGFR H-score	Median	0	0	9
	Mean	36	12	61
	Standard Deviation	62	38	86
	Minimum	0	0	0
	Maximum	300	300	300
	Missing	0	0	0
HER3-EGFR H-score	Median	170	60	33
	Mean	182	64	80
	Standard Deviation	106	65	93
	Minimum	0	0	0
	Maximum	600	430	330
	Missing	0	0	0
Age at diagnosis (years)	Median	50	52	56
	Mean	50	54	57
	Standard Deviation	11	14	12
	Minimum	27	22	23
	Maximum	71	85	85
	Missing	3	0	0
AJCC stage	I	95	30	40
	II	171	61	50
	III	35	11	14
	Missing	1	2	0
Chemotherapy	No	147	0	13
	Yes	131	80	77
	Missing	24	24	14

Table 6.2 Risk of worse breast cancer-specific or distant metastasis-free survival conferred by high levels of membranous HER3, EGFR, or HER3-EGFR H-scores in simple Cox models stratified by hospital

Biomarker	Chemo	Breast cancer-specific survival				Distant metastasis-free survival			
		p-value	HR	95% CI for HR		p-value	HR	95% CI for HR	
				Lower	Upper			Lower	Upper
HER3	No	1	1	0.52	1.92	0.81	0.93	0.51	1.69
	Yes	0.41	1.31	0.69	2.47	0.58	1.19	0.63	2.25
EGFR	No	0.91	1.03	0.58	1.84	0.8	1.08	0.62	1.86
	Yes	0.63	1.13	0.69	1.84	0.1	1.48	0.92	2.38
HER3-EGFR	No	0.97	1.01	0.55	1.84	0.89	0.96	0.55	1.68
	Yes	0.003	2.5	1.38	4.53	0.022	1.95	1.1	3.43

Table 6.3 Risk of worse breast cancer-specific or distant metastasis-free survival conferred by high levels of membranous HER3, EGFR, or HER3-EGFR H-scores, as defined by respective medians, in multivariable Cox models including each respective biomarker, adjusted for AJCC stage and age at diagnosis, and stratified by hospital

Model	Covariate	Chemo	Breast cancer-specific survival				Distant metastasis-free survival			
			P-value	HR	95% CI for HR		P-value	HR	95% CI for HR	
					Lower	Upper			Lower	Upper
HER3	HER3		0.58	1.21	0.62	2.36	0.77	1.1	0.59	2.06
	Age		0.96	1	0.97	1.03	0.99	1	0.97	1.03
	Stage	No	<0.001				<0.001			
	II vs. I		0.51	1.24	0.65	2.38	0.27	1.42	0.76	2.63
	III vs. I		<0.001	6.63	2.96	14.85	<0.001	6.09	2.67	13.92
	HER3		0.57	1.21	0.63	2.34	0.72	1.12	0.59	2.11
	Age		0.1	1.02	1	1.05	1	1	0.98	1.02
	Stage	Yes	0.01				0.02			
	II vs. I		0.79	1.1	0.57	2.12	0.97	1.01	0.56	1.85
	III vs. I		0.014	2.58	1.21	5.49	0.022	2.36	1.13	4.93
EGFR	EGFR		0.75	0.91	0.51	1.63	0.86	0.95	0.54	1.68
	Age		0.98	1	0.97	1.03	0.98	1	0.97	1.03
	Stage	No	<0.001				<0.001			
	II vs. I		0.56	1.21	0.64	2.3	0.28	1.4	0.76	2.59
	III vs. I		<0.001	6.42	2.89	14.27	<0.001	6.02	2.64	13.75
	EGFR		0.67	1.11	0.68	1.81	0.18	1.39	0.86	2.23
	Age		0.09	1.02	1	1.05	0.89	1	0.98	1.03
	Stage	Yes	0.008				0.024			
	II vs. I		0.81	1.09	0.56	2.1	0.98	1.01	0.55	1.84
	III vs. I		0.013	2.59	1.22	5.51	0.027	2.3	1.1	4.81
HER3-EGFR	HER3-EGFR		0.85	0.94	0.51	1.74	0.78	0.92	0.51	1.65
	Age		0.98	1	0.97	1.03	0.98	1	0.97	1.03
	Stage	No	<0.001				<0.001			
	II vs. I		0.58	1.2	0.63	2.29	0.3	1.39	0.75	2.56
	III vs. I		<0.001	6.33	2.88	13.96	<0.001	5.99	2.66	13.48
	HER3-EGFR		0.006	2.3	1.26	4.2	0.041	1.78	1.02	3.11
	Age		0.13	1.02	0.99	1.05	0.83	1	0.98	1.03
	Stage	Yes	0.016				0.007			
	II vs. I		0.73	1.12	0.58	2.17	0.83	1.07	0.58	1.95
	III vs. I		0.018	2.49	1.17	5.31	0.009	2.61	1.28	5.32

Table 6.4 Immunohistochemical biomarkers differentially expressed between HER3-EGFR-low and high groups at $q < 0.25$.

Unadjusted p-values are displayed.

Biomarker	HER3-EGFR	N	Mean rank	Sum of ranks	p-value
Cytokeratin 7/8 (percentage)	Low	38	54.43	2068.5	0.013
	High	93	70.73	6577.5	
Cytokeratin 7/8 (H-score)	Low	38	51.91	1972.5	0.006
	High	93	71.76	6673.5	
Cytokeratin 18 (percentage)	Low	38	52.49	1994.5	0.009
	High	91	70.23	6390.5	
Cytokeratin 18 (H-score)	Low	38	52.24	1985	0.007
	High	91	70.33	6400	
Cytoplasmic SMC6L1 (percentage)	Low	24	43.67	1048	0.017
	High	68	47.5	3230	
Dicer (percentage)	Low	27	38.87	1049.5	0.01
	High	68	51.63	3510.5	
Dicer (H-score)	Low	27	34.48	931	0.002
	High	68	53.37	3629	
P-cadherin (H-score)	Low	34	48.26	1641	0.015
	High	86	65.34	5619	

7 RETINOIC ACID RECEPTOR ALPHA PREDICTS GOOD PROGNOSIS IN TRIPLE-NEGATIVE BREAST CANCER

7.1 Abstract

Several studies have recently uncovered the potential therapeutic value of the retinoic acid receptor alpha (RARA) in breast cancer. Stimulation of RARA in hormone receptor-positive breast cancer cells induces growth arrest, differentiation, and apoptosis, and treatment with a specific RARA inhibitor suppresses mammary tumorigenesis and metastasis in an MMTV-Myc mouse model of breast cancer. However, RARA antagonists are not effective in suppressing the growth of triple-negative breast cancer (TNBC) cell lines because they tend to underexpress RARA. Whether a subset of tumors from TNBC patients exhibit upregulated RARA expression, and thus may be susceptible to RARA agonists, has not been established, nor has the potential prognostic value of RARA in TNBC been explored. We hypothesized that higher nuclear levels, representing the transcriptionally active form of RARA, would be associated with better outcomes in TNBC. We confirmed that TNBCs express lower levels of RARA mRNA than non-TNBCs in the METABRIC dataset (n=1975). We also tested the impact of nuclear RARA (nRARA) H-scores on breast cancer-specific survival (BCSS) and distant metastasis-free survival (DMFS) using 167 primary TNBC specimens immunolabeled for RARA. We found that high nRARA levels (based on an optimal cutpoint) were present in ~14% of tumors and independently predicted good prognosis in multivariable models (HR=0.26 for BCSS and HR=0.27 for DMFS, P<0.001 for both). Thus, determination of nRARA H-score by immunohistochemistry could provide prognostic information and guide clinical decision making, specifically revealing a subset of TNBC patients with a good prognosis. Future studies are

needed to evaluate whether this high-nRARA group could benefit from RARA agonists and potentially be spared harsher chemotherapeutic regimens.

7.2 Introduction

RARA, a nuclear hormone receptor, has emerged as a potential molecular target in certain types of breast cancer. For instance, stimulation of RARA with AM580, a RARA-specific retinoic acid analog, inhibits tumor growth and metastasis to the lung in an MMTV-Myc mouse model of breast cancer [1]. This is likely due to the ability of RARA to antagonize mitogen signaling and activate signaling pathways involved in tissue differentiation and induction of apoptosis. A subset of HER2-positive breast tumors exhibit coamplification of the gene encoding RARA, and co-targeting of HER2 and RARA with lapatinib and all-*trans* retinoic acid (ATRA), an active metabolite of Vitamin A, results in synergistic antagonism of proliferation and induction of differentiation and apoptosis [2]. Similarly, stimulation of RARA with AM580 or ATRA inhibits the proliferation of hormone receptor-positive breast cancer cells, an effect that is diminished by silencing of RARA [3]. On the other hand, RARA agonists do not affect the growth of triple-negative breast cancer (TNBC) cells, which express much lower RARA levels than hormone receptor-positive breast cancer cells. This finding is consistent with studies that found RARA expression is lower in basal-like breast cancer cells than breast cancer cells of other subtypes [4] and the RARA gene is hypermethylated in hormone receptor negative and basal-like breast tumors, but not hormone receptor positive and luminal breast tumors [5]. Thus, RARA agonists could be effective for hormone receptor-positive breast cancer, whereas induction of RARA may be necessary for such agents to be effective in triple-negative breast cancers.

Given its seeming role in suppressing breast tumorigenesis, RARA may serve as a positive prognostic biomarker, although this possibility has never been explored. While RARA levels are lower in TNBC than non-TNBC cells, it is possible that a subset of TNBC patients have higher levels and a more favorable prognosis, given how molecularly heterogeneous TNBCs are [6, 7]. Such biomarkers are urgently needed for TNBC to gain insight into a patient's likely disease course and the potential utility of systemic therapy, since to date no biomarkers have been approved to guide treatment decisions for TNBC per the latest ASCO guidelines due to a lack of evidence for their clinical utility [8, 9]. RARA could potentially fill this gap by serving as a positive prognostic biomarker in TNBC, although its potential in this regard has not been investigated. Towards this end, we explored the effect of nRARA levels on survival outcomes in TNBC patients. Nuclear levels were tested because they represent the transcriptionally active form of RARA [10]. Ultimately, we found that high RARA levels are associated with better breast cancer-specific and distant metastasis-free survival. Thus, a subset of TNBC patients with high intratumoral RARA could potentially be spared harsher chemotherapy regimens and may benefit from RARA agonism, whereas TNBC patients with low intratumoral RARA may require more aggressive treatment.

7.3 Materials and Methods

nRARA expression levels in breast carcinomas of different molecular subtypes were compared in the METABRIC dataset (n=1975; patient characteristics described in [11]). The impact of nRARA on the survival of 167 TNBC patients was also tested (continuous and categorical patient and clinicopathologic variables and nRARA scores are given in **Table 7.1**). Specifically, we immunolabeled primary TNBC specimens and calculated H-scores as the

product of the nuclear staining intensity (0-3) and the percent of cells with any nuclear staining (1+). The optimal, most-significant cut point based on BCSS was found using X-tile [12]. Tumors with scores below or equal to this cutpoint were classified as having a low nRARA score, and tumors with scores above this cutpoint were classified as having a high nRARA score. The impact of categorical nRARA (low vs. high) on BCSS and DMFS were assessed using the Kaplan-Meier method and Cox proportional hazards regression. The proportional hazards assumption was found to be satisfied because plots of partial residuals against rank time had nearly zero slope or y-intercept. Multivariable Cox models were fit using backward stepwise elimination of covariates with a p-value >0.10.

7.4 Results and Discussion

Using the METABRIC dataset, we found that RARA expression levels were approximately one third lower in TNBCs than non-TNBCs ($p=7.6E-92$) (**Figure 7.1**). We determined that the optimal cutpoint in nRARA H-score was 110 (≤ 110 =low, $n=21$; >110 =high, $n=146$). Based on Kaplan-Meier and univariate Cox analyses, high nRARA was associated with better BCSS and DMFS ($P<0.001$ for both, Kaplan-Meier analysis, **Figures 7.2 and 7.3**; and $HR=0.322$, $P<0.001$ and $HR=0.35$, $P=0.001$, respectively; simple Cox models). nRARA was found to be an independent predictor of better outcomes in multivariable analysis of BCSS and DMFS, where the covariates entered into the full model included age at diagnosis, tumor size, Nottingham grade, lymph node stage, basal phenotype, Ki67-labeling index, and receipt of adjuvant chemotherapy ($HR=0.26$ for BCSS and $HR=0.27$ for DMFS, $P<0.001$ for both; **Tables 7.2 and 7.3**, respectively). Only nRARA, Ki67, and tumor size (for BCSS) or tumor grade (for DMFS) remained in final models. Thus, our study corroborate previous reports that RARA is

underexpressed in TNBCs and basal-like breast cancers, the majority of which are TNBCs. The data also support our hypothesis that higher nRARA levels confer better survival outcomes in TNBC, and a small subset of TNBC patients have relatively high nRARA levels. Thus, certain TNBC patients may be good candidates for therapy with RARA agonists, such as ATRA.

7.5 References

1. Bosch, A., et al., *Reversal by RAR α agonist Am580 of c-Myc-induced imbalance in RAR α /RAR γ expression during MMTV-Myc tumorigenesis*. Breast Cancer Research : BCR, 2012. **14**(4): p. R121-R121.
2. Paroni, G., et al., *Synergistic antitumor activity of lapatinib and retinoids on a novel subtype of breast cancer with coamplification of ERBB2 and RARA*. Oncogene, 2012. **31**(29): p. 3431-43.
3. Terao, M., et al., *Induction of miR-21 by Retinoic Acid in Estrogen Receptor-positive Breast Carcinoma Cells: BIOLOGICAL CORRELATES AND MOLECULAR TARGETS*. Journal of Biological Chemistry, 2011. **286**(5): p. 4027-4042.
4. Centritto, F., et al., *Cellular and molecular determinants of all-trans retinoic acid sensitivity in breast cancer: Luminal phenotype and RAR α expression*. EMBO Molecular Medicine, 2015. **7**(7): p. 950-972.
5. Conway, K., et al., *DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival*. Breast Cancer Res, 2014. **16**(5): p. 450.
6. Lehmann, B.D., et al., *Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection*. PLoS ONE, 2016. **11**(6): p. e0157368.
7. Burstein, M.D., et al., *Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer*. Clin Cancer Res, 2015. **21**(7): p. 1688-98.
8. Harris, L.N., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. Journal of Clinical Oncology, 2016.
9. Van Poznak, C., et al., *Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. Journal of Clinical Oncology, 2015. **33**(24): p. 2695-2704.
10. Ablain, J. and H. de Thé, *Retinoic acid signaling in cancer: The parable of acute promyelocytic leukemia*. International Journal of Cancer, 2014. **135**(10): p. 2262-2272.
11. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups*. Nature, 2012. **486**(7403): p. 346-352.
12. Camp, R.L., M. Dolled-Filhart, and D.L. Rimm, *X-Tile. A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization*, 2004. **10**(21): p. 7252-7259.

Table 7.1 Descriptive statistics for continuous and categorical patient and clinicopathologic variables and nRARα scores

Continuous variable	Mean (SD)	Range	Missing
Age at diagnosis (years)	49 (11)	27-70	0
Tumor size (cm)	2.5 (1.2)	0.2-8.0	1
nRARα score	206 (71)	0-300	0
Categorical Variable	Level	Count	Column N %
Grade	2	9	5.40%
	3	157	94.60%
	Missing	1	
	Valid N	166	
LN stage	1	96	57.80%
	2	50	30.10%
	3	20	12.00%
	Missing	1	
	Valid N	166	
Adjuvant chemotherapy	None	70	45.50%
	Treatment Received	84	54.50%
	Missing	13	
Basal phenotype	Valid N	154	
	0	53	32.10%
	1	112	67.90%
	Missing	2	
	Valid N	165	

Table 7.2 Final multivariable Cox regression model of the impact of nRAR α on breast cancer-specific survival

Final Model Variables	p-value	HR	95.0% CI for HR	
			Lower	Upper
nRAR α	<0.001	0.255	0.12	0.539
Ki67	0.066	0.991	0.982	1.001
Tumor size (cm)	0.029	1.26	1.024	1.551

Table 7.3 Final multivariable Cox regression model of the impact of nRAR α on distant metastasis-free survival

Final Model Variables	p-value	HR	95.0% CI for HR	
			Lower	Upper
nRAR α	<0.001	0.265	0.129	0.542
Grade	0.179	0.25	0.033	1.89
Ki67	0.013	0.988	0.979	0.998

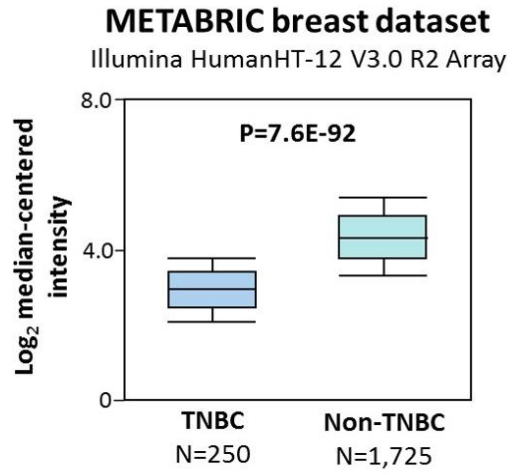


Figure 7.1 Normalized expression levels of *nRARA* in TNBCs vs. non-TNBCs in the METABRIC dataset.

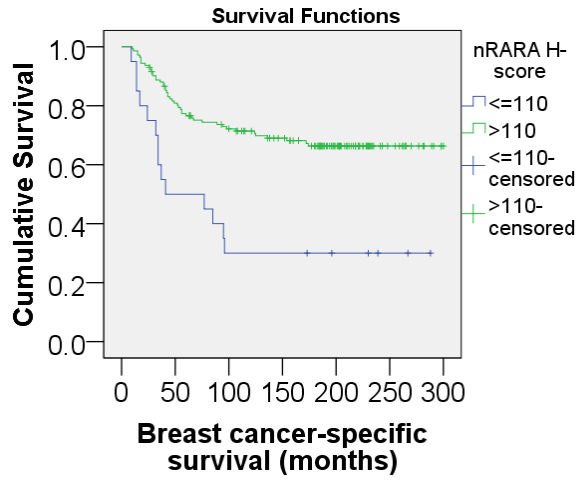


Figure 7.2 Kaplan-Meier survival plots of breast cancer specific-survival stratified by nRARA score.

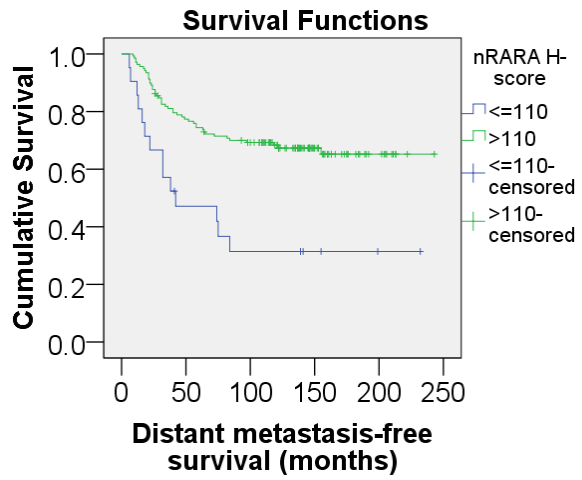


Figure 7.3 Kaplan-Meier survival plots of distant metastasis-free survival stratified by nRARA score.

8 CONCLUSIONS

Why are these findings clinically important?

The findings from the studies described in this dissertation serve an unmet need to identify novel and potentially actionable biomarkers that can predict the clinical course of breast tumors. Most of these biomarkers (nKIFC1, KAMS, HER3-EGFR, RARA) in particular risk-stratify TNBCs, for which no prognostic biomarkers or genomic tests are currently approved. Because the biomarkers explored herein offer independent prognostic information and the literature also reports that they actively drive breast tumorigenesis (overexpression in the cases of CA20, KIFC1, KAMS, and HER3-EGFR and underexpression in the case of nRARA), these biomarkers may also offer predictive information about responsiveness to specific drugs, including molecularly targeted agents (MTAs). For example, high CA20 score predicts poor survival, and CA has been shown to impart aggressive phenotypes, such as invasive behavior, enhanced directional migration, and chromosomal instability [1-3]. Thus, CA contributes to poor outcomes in breast cancer patients, rather than merely being correlated with aggressive disease course; as a consequence, high CA20 score may predict responsiveness to centrosome declustering drugs, such as the ones tested herein in TNBC cells, including the noscapinoids noscapine, bromonoscapine, and reduced bromonoscapine; the antifungal griseofulvin; and the phenanthrene-derived PARP inhibitor PJ-34. Similarly, AA TNBC patients with high nKIFC1 levels experience high mortality and distant metastasis rates, and I demonstrated that suppression of KIFC1 expression inhibits AA TNBC cell migration. Thus, treatment of AA TNBC patients whose tumors exhibit high nKIFC1 levels with a KIFC1 inhibitor may suppress the development of metastases and improve survival. TNBC patients whose tumors exhibit high KAMS may

respond well not only to cell cycle-active agents (such as taxanes) but perhaps also to immunostimulatory drugs because high KAMS was associated with suppression of immune pathways. KAMS may prove to be a biomarker with utility for racially diverse TNBC patients because it successfully stratified the predominantly white UK and Norwegian cohorts as well as the black Nigerian cohort. TNBC patients with high HER3-EGFR scores may exhibit responsiveness to targeted dual-HER3/EGFR or pan-HER family inhibitors, perhaps in conjunction with drugs that target the DNA damage response (e.g., PARP inhibitors) and P-cadherin, which are upregulated in the HER3-EGFR-high group. Furthermore, adding a PI3K-AKT inhibitor to a regimen including HER3 and EGFR inhibitors may be beneficial because antagonizing HER3 along with EGFR sensitizes TNBC cells to PI3K-AKT pathway antagonists [4]. Finally, I identified a good-prognosis subgroup of TNBC patients that overexpress nRARA; however, 5-year breast cancer-specific survival rates were still below 80%, indicating that these patients could benefit from some form of adjunctive therapy. Because RARA expression predicts responsiveness to RARA agonists, such as ATRA [5], TNBC patients with high nRARA levels may benefit from addition of a RARA agonists to their chemotherapeutic regimen. Furthermore, TNBC patients whose tumors underexpress nRARA may benefit from the addition of a drug that can upregulate nRARA expression (e.g., transcription factors for *RARA*), an interesting question for future studies. Another intriguing avenue for future research would be to test whether incorporating these biomarkers, which may reflect intratumor heterogeneity generated by chromosomal instability, along with others that encompass additional mechanisms that drive genome instability (e.g., abnormalities in pathways responsible for DNA repair pathways, maintenance of DNA methylation, and maintenance of chromosome structure) into a

comprehensive prognostic signature can improve risk prediction and, thus, quality of life and survival for breast cancer patients.

In addition to demonstrating the impact of these biomarkers on survival outcomes, this work also identifies correlates of in vitro efficacy for centrosome declustering drugs in TNBC cells, information that can be used to optimize these agents through rational drug design, which as mentioned may prove especially useful in TNBC patients with high CA20 scores since they likely exhibit augmented CA. Specifically, induction of high-grade spindle multipolarity by centrosome declustering drugs was found to eliminate TNBC cells with comparatively minimal toxicity to non-malignant breast cells, a mechanistic insight that can guide rational drug design. AA TNBC patients exhibit a more aggressive disease course than white TNBC patients, perhaps in part because their tumors display elevated intratumor heterogeneity [6], which promotes chemoresistance and disease relapse. The factors underpinning this elevated intratumor heterogeneity are not established, but because CA is a driver of CIN, which causes intratumor heterogeneity, it is possible that AA TNBCs exhibit higher levels of CA than TNBCs from patients of other races. Answering this question is an important focus for future studies because, if true, it would indicate that AA TNBCs may be more sensitive to centrosome declustering drugs. Furthermore, rational design of centrosome declustering drugs would then benefit from considerations about race in experimental design (e.g., the use of AA TNBC cell lines, such as HCC1806 and MDA-MB-468 TNBC cells, in preclinical testing of novel declustering drugs and the inclusion of AA TNBC patients in clinical trials of declustering drugs).

How do these findings align with the era of precision medicine?

Two decades ago, the era of precision medicine dawned in oncology with the arrival of MTAs, which disrupt specific processes that underpin tumorigenesis, such as apoptosis reluctance, angiogenesis, or invasiveness [7]. The National Cancer Institute defines precision or personalized 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 elaborates that “In cancer, personalized medicine uses specific information about a person’s tumor to help diagnose, plan treatment, find out how well treatment is working, or make a prognosis.” Thus, the biomarkers described herein constitute precision medicine because they can assist oncologists in prognostication and planning treatment. Furthermore, these biomarkers could potentially also predict responsiveness to specific MTAs, which I am eager to test in the future using in vitro and in vivo models. Thus, I am confident that these findings are congruent with the era of precision medicine and meaningfully contribute to the literature on the topic. Nevertheless, in the popular imagination, precision medicine may seem to include only the rather expensive, technically challenging, and analytically demanding next-generation “omics” tests in its purview, rather than these more straightforward methods to assess actionable tumor biology.

Next-generation tests are frequently touted as providing a comprehensive molecular portrait of the tumor landscape (whether genomic, transcriptomic, epigenomic, metabolomic, proteomic, or other “omic”), which can thereby yield potentially actionable information about molecular aberrations that can inform treatment decisions. Traditionally, breast cancer biomarkers (namely, ER, PR, and HER2) have been assayed primarily using immunohistochemistry (with HER2 also being assayed by in situ hybridization methods if

immunohistochemistry gives an equivocal result). These immunohistochemical biomarkers revolutionized clinical care for breast cancer patients because they offer information about the clinical course of the disease and which tumors may respond to treatment with hormone therapy (such as tamoxifen) and/or HER2 antagonists (such as trastuzumab). Today, immunohistochemistry-based biomarkers still constitute the cornerstone of precision oncology for breast cancer patients in everyday practice. For example, ER, PR, and HER2 are the only biomarkers indicated to guide the choice of specific treatment regimens per current ASCO recommendations regarding the use of biomarkers to guide decisions on systemic therapy for early-stage disease [8]. Moreover, for patients with late-stage breast cancer, it is recommended that ER, PR, and HER2 testing of accessible metastases be offered to the patient, and if the results of testing for the primary tumor and metastases are discordant, then the ER/PR/HER2 statuses of the metastases should direct therapy so long as this aligns with the clinical scenario and patient goals [9]. No other breast cancer biomarkers are currently recommended by ASCO to be used independently to guide treatment decisions in the metastatic setting. Therefore, immunohistochemical biomarkers form the foundation of precision oncology for breast cancer patients today.

By comparison, next-generation tests are expensive, technically challenging, and difficult to interpret, requiring complicated bioinformatic analysis and necessitating that the oncologist remains current on molecular therapeutics and open clinical trials for these tests to potentially translate their clinical benefits [10]. Next-generation tests are likely to reveal a multiplicity of molecular aberrations in the tumor, which may or may not be actionable, or whose actionability may be ambiguous. Furthermore, the complexity of the interactions of these aberrancies with

each other and in the context of the patient's unique genomic landscape (which is distinct for every single breast tumor cell [11]) is so tremendous that it is perhaps not possible to completely fathom, certainly not at this stage of the evolution of precision oncology. The hope for these tests is that some of these aberrations seem clinically actionable, such as an activating mutation in or high expression of a gene whose product is the target of an FDA-approved drug or a drug in an open cancer clinical trial for which the patient is eligible. Presently, clinical oncologists may not have background in molecular diagnostics advanced enough to confer expertise in interpreting test results. In an effort to fill this gap, cancer centers may have multidisciplinary "molecular tumor boards" or "sequencing tumor boards" (which may include oncologists, pathologists, surgeons, radiologists, and other clinicians; medical geneticists; basic and translational scientists; bioinformaticians, pathway analysts, and biostatisticians; social/behavioral scientists; and bioethicists). These boards assist in interpreting the next-generation reports and advising treating physicians about treatment options; however, these boards may not be accessible to practitioners working outside these centers, and research into the efficacy of such boards in translating next-generation reports into improved patient care is still inchoate, as very few studies have been published on the topic. A Pubmed search of "molecular tumor board" yields only 15 hits as of March 10, 2017, which filters to one result when "triple-negative breast cancer" is also included as a search term. This paper, published in 2017, describes a prospective trial at Johns Hopkins University School of Medicine that enrolled women with newly progressing metastatic TNBC, called the "Individualized Molecular Analyses Guide Efforts" or IMAGE trial. Next-generation sequencing of metastatic sites and plasma was conducted using a clinical cancer genomic profiling test that interrogates 287 cancer-related genes for base substitutions, indels, copy-number alterations, and fusions using massively parallel DNA sequencing [12]. The Johns

Hopkins Genomic Alterations In Tumors With Actionable Yields (GAIWAY) molecular profiling tumor board was tasked with interpreting genetic alterations found in the samples to discern potentially actionable biology and then making treatment recommendations to the patients' oncologists. Major real-world challenges ultimately resulted in early study termination because it met protocol-specified criteria for interim futility. These challenges included obtaining new or recent patient tissue suitable for next-generation sequencing in a timely fashion, DNA requirements for the next-generation tests, the necessity for patients to begin treatment if results were not returned in a clinically relevant clinical timeframe. Moreover, the actionability of the next-generation test results was often unclear or lacking. For example, *TP53* mutations were found in 95% of patients, but the GAIWAY board did not consider these to be actionable given the lack of FDA-approved drugs targeting mutant p53. In the end, only 20% of patients who received the next-generation test ultimately received a treatment based on GAIWAY recommendations even though 75% had potentially actionable aberrations, primarily due to technical and logistical challenges. Thus, for TNBC patients, significant hurdles may remain to translate the promise of next-generation tests in the clinic, not the least of which is the fact that there is no standard decision support framework to guide treatment decisions for TNBC patients based on the results of next-generation tests. Indeed, no next-generation test is indicated for TNBC, whether early stage or metastatic, as evidence for their clinical utility is deficient. Thus, for TNBC patients, the development of next-generation testing to guide treatment decisions is presently at a gestational stage.

Although it may be perceived that next-generation tests provide comprehensive portraits of a tumor (or at least cancer-relevant biology, such as in the test employed in the GAIWAY

trial), in reality these tests are usually based on a single biopsy and thus only capture a “snapshot” of the mutational landscape, which will inevitably change following drug treatment due to selective pressures, a sort of incarnation of Heisenberg’s uncertainty principle – by measuring the tumor, it is changed and no longer reflects its previous “position.” Indeed, the tumor and any circulating tumor cells and metastases will evolve even without treatment owing to internal selective pressures and the genomic instability of cancer cells. Serial tumor biopsies cannot usually be acquired due to their invasive nature, and less invasive liquid biopsies to detect circulating cell-free tumor DNA are not routinely collected, probably in large part due to the seriously limited sensitivity and specificity of current assays [13]. As a result, performing a test on a patient’s single biopsy to guide treatment selection is like trying to hit a moving target. Owing to the mostly random nature of mutation and the limits of current molecular modeling technologies, the direction in which a patient’s mutational landscape will drift simply cannot be predicted with reasonable accuracy. An additional spatial (vs. temporal) complication is the fact that the clones populating the biopsy (which are usually not multiregional in nature) may not represent the whole tumor owing to intratumor heterogeneity. Thus, the full range of driver mutations may be missed by a single biopsy. If all of these mutations are not targeted simultaneously, then the non-targeted subclones will be coaxed to evolve further and may become chemoresistant. Also, when multiple mutations are detected in the biopsy, it cannot be discerned which aberrations co-occur, and thus interact, in single cells (unless single-cell sequencing is performed). Co-occurrence in single cells can impact treatment sensitivity in ways that might not be known or predicted. Altogether, spatial and temporal intratumor heterogeneity obfuscate knowledge about the tumor landscape and thus the ability to successfully target driving aberrations.

This is not to say that cancer biology is of such Byzantine complexity as to represent a hopeless scenario. Rather, our expectations of next-generation methods should merely be tempered, as it is unlikely that a cost-effective, technically facile, and easily interpreted next-generation test that can reliably detect all the driver mutations (or other molecular aberrations) in a tumor and its metastases that must be targeted to cure the disease (as well as MTAs for all of these aberrations) will soon be clinically available. By assaying prognostic biomarkers, even from only a single biopsy, many patients with aggressive disease can be identified and administered an appropriately aggressive regimen of cytotoxic drugs, a sort of “shock and awe” approach that may be able to eliminate a substantial fraction of cancer cells regardless of their heterogeneity, as these less-specific drugs act on mechanisms that tend to be shared by cancer cells. To further personalize therapy while circumventing the manifold technical and logistical challenges intrinsic to next-generation methods, I believe that testing for potentially actionable breast cancer biomarkers, especially those responsible for chemoresistance, using immunohistochemistry holds great promise for improving clinical outcomes for TNBC patients in the very near future. Of course, the potential informativeness of immunohistochemical biomarkers is also circumscribed by the fact that they are based on a single biopsy in space and time, so potentially actionable information may be missed because a rare subclone goes undetected. However, if a biopsy detects drivers of ITH – such as CA, centrosome clustering, and mitotic propensity, the biomarkers studied in this work, which are relatively inexpensive and easy to interrogate – it brings to light that the tumor is likely to exhibit heterogeneity and chemoresistance, which may warrant more aggressive cytotoxic regimens while also suggesting potential MTAs that can inhibit further tumor evolution, which may need be given on a long-term basis, in line with the novel paradigm of treating cancer like a chronic (but manageable)

disease. For instance, I have great confidence in the ability of nKIFC1 to guide clinical decisions for AA TNBC patients. Because adjuvant-treated AA TNBC patients with high nKIFC1 levels experience worse outcomes, nKIFC1 may be an indicator of poor response to chemotherapy. In fact, KIFC1 overexpression causes resistance to docetaxel in TNBC cells [14], so TNBCs from AAs with high nKIFC1 levels may require a more aggressive cytotoxic chemotherapy regimen. Adjunctive treatment with a KIFC1 inhibitor may not only improve response to cytotoxic chemotherapy, it may also hinder further tumor evolution by selectively eliminating cancer cells with CA, a driver of CIN. An even greater degree of personalization may be possible to achieve for these patients if the proportion of African ancestry is also considered. AAs are a highly admixed population. On average, self-reported AAs have ~27% European ancestry with a range from 0-100% based on genome-wide estimates [10]. This contrasts strongly with whites, who are on average 99% European. Because African ancestry is an independent predictor of poor outcomes, percent African ancestry can confound survival analysis. Thus, in the nKIFC1 study, the AA cohort is likely to be substantially admixed, so it is necessary to determine whether there is a threshold in the percent African biogeographic ancestry above which nKIFC1 is maximally informative as a biomarker. Performing genetic ancestry testing to reveal proportion African ancestry and incorporating this information in statistical analyses as a covariate will shine a light on the potential utility of considering race at such fine-grained level in conjunction with assaying nKIFC1. For instance, it may reveal that AA TNBC patients whose percent African ancestry surpasses a certain threshold should be reflexively prescribed a test for nKIFC1, whereas AA patients below this threshold may not derive as great (or any) benefit. This exciting question is one I eagerly await testing in future work.

Finally, it bears mentioning that immunohistochemical methods are “information packed” because, generally speaking, proteins are the ultimate cellular effectors that impart phenotypes and thereby “dictate” health or pathology. Assessment of proteins via immunohistochemistry is also able to provide information of compartmentalization, post-translational modifications, and subcellular localization, which can more strongly impact phenotype than a whole-cell average gene expression score. For example, for AA TNBC patients it is specifically nuclear levels of KIFC1 that hold immense prognostic value to serve as a companion diagnostic, and which may inspire future studies investigating assaying nKIFC1 for AA patients with a certain threshold of African biogeographic ancestry. Likewise, it is specifically nuclear RARA and membranous HER3 and EGFR that are biologically active. This information is not provided by next-generation tests as routinely performed.

How do these findings relate to the next-generation of cancer hallmarks?

At the turn of the century, Hanahan and Weinberg defined six hallmarks of cancer that encompass molecular, biochemical, and cellular capabilities acquired in the course of their evolution (namely, sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis) [15]. More recently, they described their vision of the next-generation of cancer hallmarks, including delineation of two enabling characteristics of cancer that facilitate acquisition of the originally defined hallmark capabilities: tumor-promoting inflammation and genome instability [16]. Causes of the latter characteristic vary widely by tumor type; however, most solid tumors progress stochastically through non-clonal chromosomal abnormalities rather than through stepwise accrual of deleterious gene-level mutations [17]. Copy-number

abnormalities can cause overexpression of oncogenes or underexpression of tumor suppressors through alteration of gene dosages and thereby fuel tumorigenesis. For breast tumors, CA is a likely culprit because the vast majority of breast tumors exhibit aneuploidy coupled with this predominantly cancer-specific trait, which intensive research has revealed is a leading cause of CIN, a type of genome instability responsible for aneuploidy. Clustering of supernumerary centrosomes during mitosis causes chromosome missegregation; thus, CA, centrosome clustering, and mitotic propensity collectively fuel karyotypic diversification and, thus, ITH. Based on this paradigm, I envision that these drivers of ITH may serve as biomarkers of aggressive disease course in breast cancer and can identify patients with increased need for aggressive treatment. Assaying drivers of ITH may bring to light that the tumor is likely to exhibit heterogeneity and associated aggressive disease features, which may warrant more aggressive treatment.

Chromosomal instability drives tumor evolution, including the acquisition of metastatic potential, the most important clinical event to predict because it is the cause of cancer-related death in most cases and is generally incurable. Metastatic dissemination is an intricate cascade of events, each of which is a significant obstacle to metastasis. The cancer cell or cluster of cancer cells must at a minimum detach from adjacent cells, intravasate, circulate in the blood or lymph, adhere to a vessel wall, extravasate, and colonize the metastatic niche. Successful colonization may require that the cancer cell(s) have been reprogrammed to thrive in the specific niche, as for instance metabolic profiles of metastatic breast cancer cells differ substantially by organ site (viz., bone, liver, and lung) [18]. Determination of site-specific molecular profiles could inform rational development of biomarkers predicting metastasis, which is the next question I intend to

address in my post-doctoral research. Furthermore, I am also interested in investigating the processes by which metastatic cancer cells reprogram organ function, which, along with physical compression, ultimately compromises organ function. Biomarkers that could predict how a primary tumor is likely to disrupt organ function could guide clinicians in selecting preventative treatments most likely to prolong survival and resolve organ dysfunction, an exciting avenue I wish to explore.

1. Ganem, N.J., S.A. Godinho, and D. Pellman, *A Mechanism Linking Extra Centrosomes to Chromosomal Instability*. *Nature*, 2009. **460**(7252): p. 278-282.
2. Godinho, S.A., et al., *Oncogene-like induction of cellular invasion from centrosome amplification*. *Nature*, 2014. **510**(7503): p. 167-171.
3. Pannu, V., et al., *Rampant centrosome amplification underlies more aggressive disease course of triple negative breast cancers*. *Oncotarget*, 2015. **6**(12): p. 10487-97.
4. Tao, J.J., et al., *Antagonism of EGFR and HER3 Enhances the Response to Inhibitors of the PI3K-Akt Pathway in Triple-Negative Breast Cancer*. *Science signaling*, 2014. **7**(318): p. ra29-ra29.
5. Centritto, F., et al., *Cellular and molecular determinants of all-trans retinoic acid sensitivity in breast cancer: Luminal phenotype and RAR α expression*. *EMBO Molecular Medicine*, 2015. **7**(7): p. 950-972.
6. Keenan, T., et al., *Comparison of the Genomic Landscape Between Primary Breast Cancer in African American Versus White Women and the Association of Racial Differences With Tumor Recurrence*. *Journal of Clinical Oncology*, 2015.
7. Le Tourneau, C., et al., *Treatment Algorithms Based on Tumor Molecular Profiling: The Essence of Precision Medicine Trials*. *JNCI Journal of the National Cancer Institute*, 2016. **108**(4): p. djv362.
8. Harris, L.N., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. *Journal of Clinical Oncology*, 2016. **34**(10): p. 1134-1150.
9. Poznak, C.V., et al., *Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. *Journal of Clinical Oncology*, 2015. **33**(24): p. 2695-2704.
10. Meric-Bernstam, F., et al., *A Decision Support Framework for Genomically Informed Investigational Cancer Therapy*. *JNCI Journal of the National Cancer Institute*, 2015. **107**(7): p. djv098.
11. Wang, Y., et al., *Clonal Evolution in Breast Cancer Revealed by Single Nucleus Genome Sequencing*. *Nature*, 2014. **512**(7513): p. 155-160.

12. Frampton, G.M., et al., *Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing*. Nat Biotech, 2013. **31**(11): p. 1023-1031.
13. Alix-Panabieres, C. and K. Pantel, *Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy*. Cancer Discov, 2016. **6**(5): p. 479-91.
14. De, S., et al., *Overexpression of kinesins mediates docetaxel resistance in breast cancer cells*. Cancer Res, 2009. **69**(20): p. 8035-42.
15. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**(1): p. 57-70.
16. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation*. Cell. **144**(5): p. 646-674.
17. Heng, H.H., et al., *Chromosomal instability (CIN): what it is and why it is crucial to cancer evolution*. Cancer Metastasis Rev, 2013. **32**(3-4): p. 325-40.
18. Dupuy, F., et al., *PDK1-Dependent Metabolic Reprogramming Dictates Metastatic Potential in Breast Cancer*. Cell Metabolism. **22**(4): p. 577-589.