COMPREHENSIVE MOLECULAR AND CLINICOPATHOLOGICAL CHARACTERIZATION OF BREAST CANCER METASTASIS - EMPHASIS ON INVASIVE LOBULAR CARCINOMA

by

Ahmed Mohammed Basudan

MS, Biochemistry and Molecular Biology, Georgetown University, 2011

BSc, Clinical Laboratory Sciences, King Saud University, Saudi Arabia, 2008

Submitted to the Graduate Faculty of

Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Ahmed Mohammed Basudan

It was defended on

April 3, 2018

and approved by

Dissertation Advisor:
Steffi Oesterreich, PhD
Professor
Department of Pharmacology and Chemical Biology
School of Medicine
University of Pittsburgh

Committee Members:

Candace M. Kammerer, PhD
Associate Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Adrian V. Lee, PhD
Professor
Department of Pharmacology and Chemical Biology
School of Medicine
University of Pittsburgh

Ryan L. Minster, PhD, MSIS
Assistant Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Steffi Oesterreich, PhD

COMPREHENSIVE MOLECULAR AND CLINICOPATHOLOGICAL CHARACTERIZATION OF BREAST CANCER METASTASIS - EMPHASIS ON INVASIVE LOBULAR CARCINOMA

Ahmed Mohammed Basudan, PhD

University of Pittsburgh, 2018

ABSTRACT

Metastatic breast cancer (MBC) has been the main cause of death in breast cancer patients, demonstrating a major public health burden. Estrogen receptor-alpha (ER; *ESR1*) is expressed in nearly 70% of breast tumors and has been a key target for endocrine therapy. Yet, 20-40% of patients eventually develop relapse. Invasive lobular breast cancer (ILC) is the second most common histological subtype of breast cancer, characterized by near universal expression of ER, and by frequent late recurrences. The goal of my studies was to identify genetic changes that might cause endocrine resistance and metastases, with emphasis on ILC.

While many studies have identified *ESR1* mutations in recurrent tumors, less is known about changes in *ESR1* DNA copy number (CN) changes and their clinical relevance. First, sensitive nanoString-technology was used to comprehensively investigate the role of these alterations in MBC. Our analysis identified substantial rates of *ESR1* gains and amplifications in MBC that were of metastatic-site tropism and showed significant association with poor overall survival. Additionally, mutually exclusive amplifications of *CCND1* and deletions of *CDKN1B* and *CDKN2A* were identified, potentially defining a subset of patients with improved response to CDK4/6 inhibition. I also identified frequent *ESR1* amplifications in ILC, and they were significantly enriched in tumors with recurrence and showed association with recurrence-free survival. CN analysis also discovered a unique group of tumors negative for HER2 by IHC

characterized by HER2 DNA amplifications, which correlated with high mRNA and protein expression, and enrichment of molecular HER2 signature. Lastly, in addition to clinicopathological evaluation, I utilized RNA-seq approach to understand transcriptomic changes involved in unique ILC metastasis to the ovaries. Our analyses revealed unique transcriptomic alterations in WNT, glutamate/calcium receptors, and MAPK/ERK pathways. Analysis of clinically actionable genes identified *MYCN* as potential mediator of endocrine resistance. Using targeted sequencing, I further validated previously reported *PIK3CA* and *FOXA1* mutations, and identified novel *NCOR1* mutations enriched in the metastases. Collectively, these studies address distinct molecular changes involved in endocrine resistance and metastasis. Such findings are of public health significance, as they can serve as novel therapeutic targets for the leading cause of death in women.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

As they always say, every wonderful journey must come to an end. Throughout the years in Dr. Oesterreich's lab, I had the honor to work with many intelligent and altruistic members. I would like to start by thanking my advisor, Dr. Steffi Oesterreich for being remarkable mentor. From her, I have learned patience, perseverance, punctuality, and many other great skills. She not only has been a fantastic mentor, but she has taught me how to mentor other students. She also taught me how to be a good philanthropist, by talking directly to breast cancer patients, hearing from them, and increasing awareness. I am grateful to her for opening my eyes into new stages of opportunity and strengths. I count myself lucky for getting the chance to work with her, and I look forward for continuing collaboration in future. I must also thank my academic advisor, Dr. Candace Kammarer. I thank her for all the guidance, encouragement, and advice through my graduate studies. She always welcomes me with open heart. Being an international student, I needed more help from her but she never let me down. I endlessly thank her for her encouragement and support.

Next, I would like to thank my committee member Dr. Adrian Lee. Being closely involved in my projects, he has been instrumental with his valuable advice and productive discussions. I next want to thank my committee member, Dr. Ryan Minster for his extended support and critique to improve my work. I had my first encounter with R in his NGS class, and this was a huge shift towards improving my bioinformatics skills.

I show my genuine appreciation for my lab mates: Ali, Amir, Becky, Beth, David, Emily, Julie, Kevin, Lan, Maddie, Nadine, Nick, Nilgun, Rekha, Sreeja, Susan, Tanya, Tiffany, Zeynep. I specially thank Emily as she was among those who helped me get on my feet when I first joined the lab, and my R buddy Kevin for his continuous help solving codes problems, and Nolan for

dealing with bioinformatics hassles and contribution to my studies, and Becky for help in the CN study. I extend my gratitude to Dr. Ryan Hartmaier for his critical help in my projects and sequencing analysis. I also thank John Skoko for his support and education.

Of course, I also thank our collaborators as they were essential to my projects. Dr. Peter Lucas has been the main pathologist involved in our studies with such amazing skills. Dr. Esther Elishaev was instrumental to establish the collection of ovarian metastases.

I wouldn't also forget to thank my previous lab members: former mentor Dr. Denise O'keefe, Dr. Dean Bacich, Dr. Allison Madigan, Dr. Katy Sobek, and Jessica Cummings for spending the time to improve my lab skills and help me through my early PhD life.

I would like to thank my government and my work place, King Saud University for the extended support with my graduated studies and scholarship.

Lastly and most importantly, I endlessly thank my parents Mohammed and Safiah, my siblings Amani, Amira, Abdulrahman, Abdulaqader, and Abdulaziz. We always take family for granted but they are the reason behind our success. My mom Safiah, I dedicate this work to you. I know no matter what I do I won't be able to repay you. Thank you for patience with your son being away all these years, I am endlessly grateful.

LIST OF ABBREVIATIONS

aCGH	array comparative genomic hybridization
AF-1	activation function 1
AF-2	activation function 2
AI	Aromatase Inhibitor
CCLE	the cell cancer cell line encyclopedia
CN	copy number
DBD	DNA-binding domain
ddPCR	digital droplet polymerase chain reaction
DGIdb	drug gene interaction database
E2	17beta-estradiol
ER	estrogen receptor
ERE	estrogen response element
FDA	Federal Drug Administration
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
FISH	fluorescence in situ hybridization
IDC	invasive ductal breast cancer
IHC	immunohistochemistry
ILC	invasive lobular breast cancer
LBD	ligand binding domain

LTED	long-term estrogen-deprived
MBC	metastatic breast cancer
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
OS	overall survival
PDX	patient derived xenograft
PR	progesterone receptor
qPCR	quantitative polymerase chain reaction
RFS	recurrence-free survival
SERM	Selective Estrogen Receptor Modulator
SERD	Selective Estrogen Receptor Downregulator/degrader
SPM	survival post metastasis
TCGA	The Cancer Genome Atlas
HER2	human epidermal growth factor receptor 2
E-cad	ecadherin
Dx	diagnosis
Adj Tx	adjuvant therapy
DFS	disease-free survival
MFS	metastasis-free survival

1.0 INTRODUCTION

Breast cancer can be characterized as abnormal proliferation of malignant cells in mammary epithelial tissue. Over the years, breast cancer research has evolved tremendously, giving new insights into the disease's diagnosis, development, progression, metastasis, and treatment. Despite remarkable advances in our understanding of the disease in the last half century, it is still a major public health burden worldwide and poses significant challenge.

1.1 HISTOLOGICAL SUBTYPES OF BREAST CANCER AND UNIQUE FEATURES OF INVASIVE LOBULAR CARCINOMA (ILC)

Breast cancer is heterogeneous disease with numerous clinical behaviors. Breast tumors can be classified into multiple grades based on degree of differentiation and proliferative activity, which are usually used as markers for aggressiveness [1,2]. On the other hand, these tumors can be categorized based on their cytological and morphological growth patterns into multiple histopathological types. Invasive breast cancer can be sub-classified into invasive ductal (IDC), lobular (ILC), ductal/lobular, mucinous (colloid), tubular, medullary, and papillary carcinomas. The two commonly seen subtypes are IDC and ILC, which represent 80% and 10-15% of invasive breast carcinomas, respectively [3–7]. The growth pattern of ILC is usually characterized by small regular uniform round cells that infiltrate the stroma in single file cellular growth, while IDC tend to grow as clusters of cells together (Figure 1) [8–10]. In general, ILC tumors are more often

mammaographically occult, estrogen receptor (ER) positive, and of lower proliferative capacity compared to IDC [11,12] (Table 1).

Table 1: Unique characteristics of ILC and IDC

	IDC	ILC
% of breast cancer	80-90% of the cases	10-15%
Growth pattern	Large mass (bulky growth) "easily detected"	small, round cells that infiltrate the stroma in single file
Incidence	Relatively constant	Has been increasing (especially among postmenopausal women)
Metastasis	Bone, brain, & liver. More significantly to lung than ILC	Bone, brain, & liver. More significantly to ovaries and GIT than IDC
ER status	~70% ER+	~90% ER+
Survival	Although ILC tend to be more ER+, Her2-, less Ki67, ILC survival not necessarily better than IDC	

Another hallmark of ILC growth features is E-cadherin inactivation in about 95% of the cases, which is considered a defining characteristic aid in the diagnosis of ILC [13–15]. In a recent multidimensional work characterizing the molecular portrait of ILC, our group's principal investigator and my mentor, Dr. Oesterreich along with the TCGA breast cancer ILC group have shown that about 63% of E-cadherin inactivation can be explained by missense and truncating mutations in the *CDH1* gene (chromosome 16q22) which is usually accompanied by heterozygous loss the other allele (in 89% of the cases) [16]. Other possible mechanisms for E-cadherin loss include transcriptional repression by epithelial to mesenchymal transition proteins, epigenetic silencing, and mutations in the *CDH1* regulatory region, which require further elucidation [17]. E-cadherin normally forms adherens junctions (AJ) intracellularly with actin cytoskeleton (group of catenins)

and intercellularly with E-cadherins on other cells (Figure 2) [18]. The inactivation of E-cadherin in ILC results in the loss of catenins, where p120-catenin becomes up-regulated and re-localized to the cytoplasm [14,19–24].

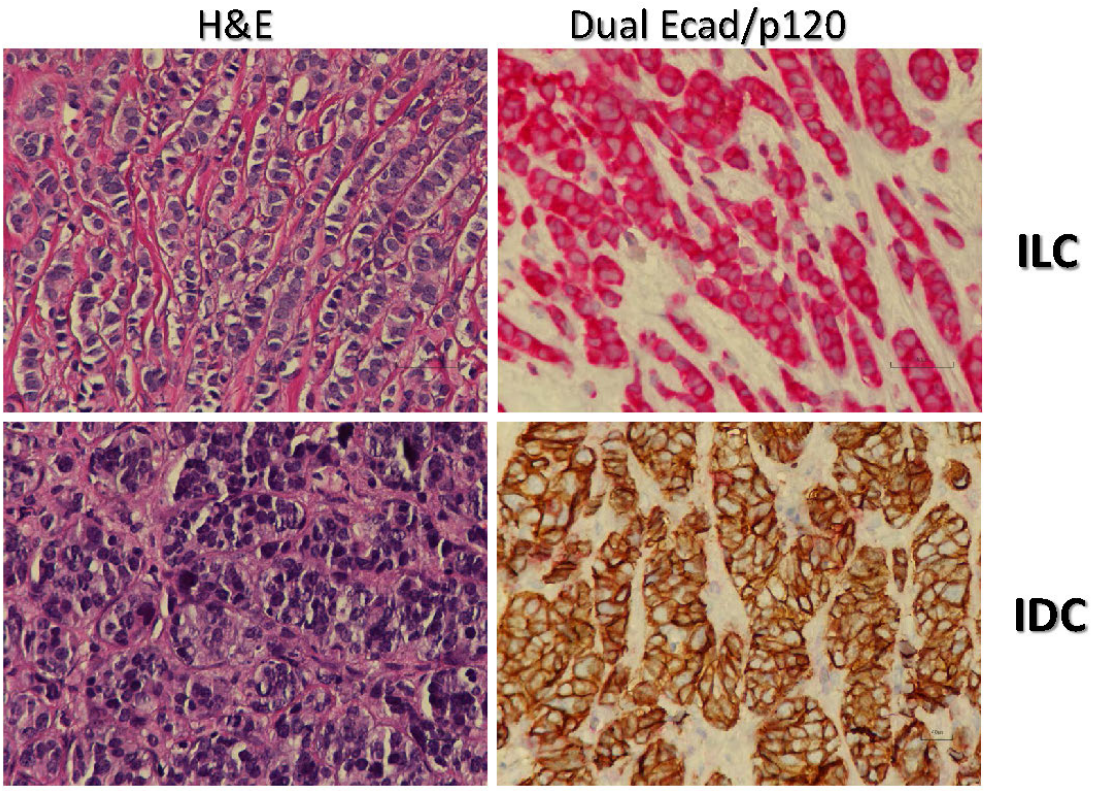


Figure 1: ILC vs IDC growth characteristics
Left; Hematoxylin and Eosin (H&E) staining shows the single file growth patterns in ILC (top) vs IDC (bottom). Right; dual E-cadherin/p120 Immunohistochemistry (IHC) staining validates the loss of E-cad in ILC (top) vs IDC (bottom).

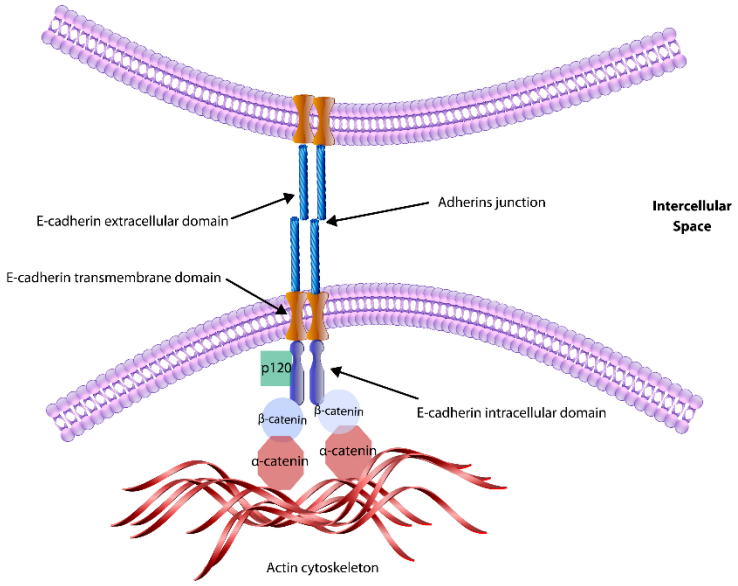


Figure 2: E-cadherin and formation of adherins junctions
Schematic diagram of E-cadherin interactions with other E-cadherins (intercellularly) to form adherins junctions, and with actin cytoskeleton via catenins (intracellularly).

Outside of the loss of E-cadherin and unique growth patterns, not much is known about distinguishing histological features between ILC and other subtypes such as IDC. Given that clinical presentation of this breast cancer subtype is not frequently uncommon, finding additional subtype-specific markers is crucial for promoting diagnosis, treatment, and thus patient outcome.

1.2 MOLECULAR SUBTYPES OF BREAST CANCER

In addition to being clinically heterogeneous disease, breast cancer is complex disorder with very heterogeneous biology. One of the earliest observations that defined a large subset of breast tumors was the association with estrogen receptor [ER] expression as a distinct molecular feature [25]. Subsequent studies showed growing evidence that ER is a putative mechanism for estrogen action, the link that eventually led to antiestrogen therapy as an option for treatment in the clinic [26–28].

With the observation that some tumors grow and spread rapidly while others do not, it has become obvious that the existence of other distinct molecular target (s) is potential. In mid-1980s, Slamon *et al.* discovered that the human epidermal growth factor receptor 2 (HER2) gene *ERBB2* is highly expressed in 25-30% of breast cancers, linked to high rate of tumor growth, and poor overall patient survival [29,30].

Extensive molecular breast cancer research and advances in molecular technologies allowed for better understanding of molecular subtypes of breast cancer. With the idea that phenotypic diversity in breast cancer can be associated with diversity in gene expression patterns, hierarchical clustering of gene expression microarray data was used to uncover similarities and differences between tumors [31]. Besides the expected clusters for ER+ and HER2+ tumors, additional intrinsic clusters for luminal/ER+ and basal epithelial cells emerged. Further follow up by Sorlie *et al.* identified two subgroups within the luminal cluster to be luminal A and B which showed different clinical outcome [32]. To develop a clinical test, further work has improved and standardized this classification using 50-gene set called PAM50 (Prediction Analysis of Microarray) [33]. The intrinsic subtypes gene expression profiling has provided us with prognostic information beyond standard clinical assessment. For example, the 21-gene OncotypeDx assay™ can be used to predict recurrence of node-negative early-stage ER+ breast cancer [34,35]. In addition to ER+, the 70-gene MammaPrint™ assay has shown prognostic significance in ER- node-negative early-stage breast cancer [36,37]. Recently, an FDA-approved PAM50-based test (Prosigna™) was developed to be used for defining a category of risk of recurrent metastasis [38].

The current era of large scale genomics and big data has shown explosion of multi-dimensional studies such as exome/whole genome sequencing, transcriptomics, and proteomics. This in turn allowed for further molecular characterization of breast cancer. Multiple inter-

institutional comprehensive and coordinated effort emerged to accelerate our understanding of the molecular basis of cancer, including The International Cancer Genome Consortium (ICGC), The Cancer Genome Atlas (TCGA), and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) [39–41]. These studies revealed frequent *PIK3CA*, *TP53*, and *GATA3* mutations in breast cancer patients (> 10%), and enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype suggesting new targets for response classification and therapy.

It should be noted that most of these studies were focused mainly on primary tumors while patients usually struggle with metastasis. Metastatic breast cancer remains understudied and is of distinct molecular features compared to primary tumors. Building comprehensive molecular portraits of breast cancer metastasis to address clinical implications deserve further investigation.

1.3 THERAPEUTIC STRATEGIES IN BREAST CANCER

Breast cancer treatment can be divided into two main categories; local and systemic therapies. The local option usually involve surgery (lumpectomy or mastectomy) and can be followed by radiation therapy. Depending on the type of breast cancer, different systemic treatment options can be used including chemotherapy, endocrine therapy, and targeted therapy.

1.3.1 Chemotherapy

Although usually associated with high toxicity, chemotherapy is commonly the treatment option in cases which are both endocrine receptor-negative and HER2-negative [42]. This type of treatment

can also extend to patients with node-positive cancer, larger tumors (>1 cm), or progressive disease [43]. Anthracyclines (e.g. Doxorubicin and Epirubicin) and Taxanes (e.g. Docetaxel and Paclitaxel) are commonly accepted regimens in breast cancer, and many cases can be administered in combinations with other chemotherapeutic agents and /or endocrine/hormone therapy [43,44]. The choice of chemotherapy combinations is variable between patients and depends on patient's tolerance and response. However, a new strategy of response-guided treatment (where there is a switch to another chemotherapy in case of early resistance or increase in cycle number in case of early response) showed significantly improved disease-free survival (DFS) and overall survival (OS) [45]. Of note, this was mainly observed in hormone receptor positive patients

1.3.2 Endocrine (hormonal) therapy

Estrogen receptor [ER-alpha] expression is a distinct and prevalent molecular feature in about 60-70% of all breast cancers[46,47]. ER+ tumors utilize endogenous estrogen to activate ER-responsive genes and variety of signaling pathways involved in growth and survival [48]. Therefore, blocking ER activation and/or exposure to estrogen represents an effective strategy in improving survival in patients with ER+ breast cancer. Several classes of endocrine therapy have evolved through the years and these mainly include Selective Estrogen Receptor Modulators (SERMs), Selective Estrogen Receptor Downregulators (SERDs), and Aromatase Inhibitors (AIs). The choice of treatment option usually depends on menopause status, tolerance, and stage of the disease. The first and most commonly used example of SERMs in clinics is Tamoxifen, which was discovered in the late 1960s [49,50]. Tamoxifen is considered a 'modulator' because of its tissue specific activity. While it competes with estrogen in breast cancer cells, it can act as estrogen in other tissues, including bone (osteoporosis) and uterus (endometrial hyperplasia) [51,52]. In contrast to Tamoxifen,

Fulvestrant, which was introduced to clinical practice in 2002 for advanced-stage breast cancer, is categorized as a SERD, and is anti-estrogenic in all tissues. It is usually a useful option in the treatment of advanced ER+ breast cancer that has progressed on SERMs. Unlike Tamoxifen, the binding affinity of Fulvestrant is 100 times stronger to ER and it induces rapid degradation of the receptor [53–55]. Instead of modulating binding to ER, AIs class of treatment acts by reduction of estrogen production through inhibiting the activity of estrogen synthase, the aromatase enzyme responsible for androgens conversion into estrogen [56,57]. The non-steroidal sub-class of AIs (e.g. Anastrozole and Letrozole) has reversible ability to inhibit aromatase, while steroidal AIs (e.g. Exemestane) bind to aromatase irreversibly. AIs are indicated only for premenopausal women where the main source of limited estrogen is not the ovaries. In premenopausal women, lower serum levels of estrogen lead to upregulation of aromatase enzymes in the ovaries as compensatory mechanism [58]. However, AIs might be considered for premenopausal women after ovarian ablation either permanently by surgical removal (oophorectomy) or temporarily by using gonadotropin or luteinizing releasing hormone (GnRH or LHRH) analogs [59,60].

Endocrine therapy has been one of the most successful forms of treatments in breast cancer and has led to substantial improvements in patients' outcomes. However, tumors can acquire resistance over time, limiting the efficacy of endocrine therapy [61]. The intrinsic resistance to endocrine therapy can be attributed to changes in the ER pathway, alterations in cell cycle and survival molecules, and activation of other escape pathways. The loss of ER expression and *ESR1* mutations are known mechanisms of anti-estrogen therapy resistance. ER expression is lost in 15-20% of refractory tumors while the estimated rate of *ESR1* mutations in advanced breast cancer is 22% [62–66]. ER co-regulators can also influence response to endocrine therapy. For example, it has been shown that overexpression of ER coactivator NCOA3 and downregulation of the corepressor

NCoR are associated with tamoxifen resistance [67,68]. Another mechanism for endocrine resistance includes cell cycle molecules. Upregulation of positive cell cycles regulators (e.g. *MYC, CCNE1, CCND1*) results in therapy resistance by activation of cyclin-dependent kinases (CDKs) and inhibition of negative regulators p21, p27, and Rb [69–72]. Alternatively, tumors can trigger activation of other pathways to circumvent the inhibitory effect of endocrine therapy. Growth factor pathways such as HER tyrosine kinase family, IGFIR, FGF, and Akt have been all implicated as ER-independent survival pathways [73–76]. In addition, pathway activation through downstream targets, such as PI3K activating mutations or loss of *PTEN* tumor suppressor gene have been reported [77].

1.3.3 Targeted therapy

1.3.3.1 HER2 Targeted therapy

Unlike chemotherapy, targeted therapy is very focused towards specific genes and proteins that contribute to cancer progression and survival. As mentioned earlier in this chapter, the human epidermal growth factor receptor 2 (HER2) is highly expressed in about 1 in 5 women with breast cancer, and associated with tumor aggressiveness and poor survival. HER2 belongs to the family of receptor tyrosine kinases, such as the epidermal growth factor receptor (EGFR), HER3, and HER4. In most cases, the signal is generated by ligand binding to one of the members and dimerization with HER2 to regulate cell behavior [78]. The massive amplifications and overexpression of HER2 in breast tumors leads to chaotic constitutive signal activation to drive tumorigenesis [79]. Discovery of the oncogenic role of HER2 in breast cancer has culminated in the development of the anti-HER2 antibody Trastuzumab for HER2+ tumors [80,81]. The molecular mechanism of Trasstuzumab is very complicated but can be grouped into three parts; 1) As an

antibody, Trastuzumab targets and blocks HER2, which attracts the immune system to tumor cells for destruction by immune system antibody-dependent cellular cytotoxicity (ADCC) [82,83], 2) Binding of Trastuzumab to HER2 can recruit c-Cbl leading to HER2 ubiquitination [84], and 3) By interfering with HER2 dimerization, Trastuzumab inhibits the MAPK and PI3K/Akt pathways, leading to cell cycle arrest [85–87]. The continuous interest in finding options for HER2 inhibition has led to the development of another humanized antibody, Pertuzumab. While Trastuzumab binds close to the transmembrane domain, Pertuzumab binds directly to the dimerization domain of HER2, showing efficacious inhibition [88]. Unsurprisingly, dual HER2 blockade with both antibodies has been associated with prolonged both progression-free survival and overall survival [89–91].

In addition to HER2-homing drugs (e.g. Trastuzumab), HER2 signaling can be inhibited using kinase inhibitors that compete with adenosine triphosphate (ATP) and inhibit its catalytic function [92]. Examples of drugs from this class include the first tyrosine kinase inhibitor approved for HER2 Lapatinib, and the newer irreversible HER2 inhibitor Neratinib [93,94]. In advanced breast cancer, Lapatinib has shown significant benefit after resistance to Trasuzumab, and the combination of both showed higher response rate [95,96].

Trastuzumab has tremendously improved outcome in HER+ breast cancer patients. However, less than 35% of the patients respond initially, and about 70% of those who respond to treatment progress to metastasis, suggesting both inherited and acquired resistance [97]. It has been proposed that the resistance to Trastuzumab can be due to HER2 structural mutations, alternative upregulation of other tyrosine kinase receptors, or alterations in HER2 downstream signaling. Some HER2 mutations can lead to proteolysis of its extracellular domain which inhibit binding to Trastuzumab. A well-known example of such process is the generation of truncated p95HER2

isoform with constitutively active kinase activity [98–100]. Other studies suggested that *ERBB2* L869R is functional mutation that leads to increased sensitivity to Neratinib, but upon progression tumors acquired *ERBB2* T798I mutation that conferred resistance [101]. similarly, *ERBB2* L755S has shown to be a mechanism of induced acquired resistance to Lapatinib in cell lines models [102]. Another mechanism for resistance to HER2 targeted therapy can be achieved by overexpression of other tyrosine kinase receptors. For example, overexpression of HER3 can overcome inhibition of HER2 by Trastuzumab, which subsequently leads to PI3K downstream activation [103,104]. Similar examples include insulin-like growth factor 1 receptor (IGF1R) and c-Met whose overexpression is involved in treatment-acquired resistance to Trastuzumab [105,106]. The last proposed mechanism for HER2 therapy resistance involves alternative activation of downstream mediators. The most notable examples in this category are PTEN inactivation and/or PI3K/Akt activation. It has been reported that 36% of HER2+ primary breast tumors show loss of PTEN, which was associated with remarkable resistance to Trastuzumab [85]. In addition, PI3KCA activating mutations were reported in 25% of patients resistant to Trastuzumab and were associated with shorter progression-free survival [107,108].

1.3.3.2 Other targeted therapies

Over the last decade, our understanding of cancer cell progression has advanced, leading to identification of more targets for cancer therapy. Examples of such promising targets include CDK4/6 and PI3K/Akt/mTOR targeted therapies.

Estrogen can facilitate G1 to S cell cycle transition by activation of CCND1 dimerization with CDK4/CDK6 [109–111]. Amplifications of *CCND1* and overexpression of CDK4/6 have been reported as frequent occurrences in breast cancer [112]. In addition, CCND1 has been shown to upregulate estrogen regulated genes by activation of ER as another survival mechanism in absence

of estrogen [113]. Thus, dysregulation of this pathway has been reported to be implicated in breast cancer, representing unique opportunity for intervention. This has led to the development of Palbociclib, Ribociclib as potent CDK4/6 inhibitors. These inhibitors block the ability of CDK4/6 to phosphorylate retinoblastoma (Rb), a necessary step for growth progression [114]. Currently, these inhibitors are FDA-approved only for ER+ HER2- in combination with either AIs or Fulvestrant, and showed improved outcome versus endocrine therapy alone [115].

Being downstream mediators to HER2 and ER, PI3K/Akt/mTOR became principal targets for developing new breast cancer regimens. This pathway is altered in more than 70% of breast cancers, with the alterations mainly include PIK3CA and AKT mutations (which affects the downstream target mTOR), and PTEN loss [116–119]. Multiple PI3K inhibitors have been developed but unfortunately discontinued due to poor toxicity. The mTOR inhibitor Everolimus has shown promising results in combination with AIs in ER+ HER2 postmenopausal women who didn't benefit from AIs only [120]. This class is relatively new and multiple drugs are being tested in early phase clinical trials targeting the PI3K/Akt/mTOR pathway hoping for better outcome.

Breast cancer therapy has revolutionized in the last few decades. The breast cancer community around the world are working together find better ways for breast cancer prevention and treatment. For example, immunotherapy is emerging as a new therapeutic option, especially the use of PD-1 and PD-L1 inhibitors in triple negative breast cancer [121]. Overall, the key will be defining the most responsive subset of patients, and predicting the most efficacious drug combinations.

1.4 BREAST CANCER METASTASIS

Despite major advancements in early detection and treatment, nearly 30% of early-stage breast cancers develop recurrence eventually [122]. Unfortunately, patients with metastatic breast cancer (MBC) generally have poor 5-year survival rate of about 25% [123,124]. Almost all deaths from breast cancer do not result from tumors that are confined to the breast but rather from metastases. MBC remains challenging to treat, and most patients develop resistance to therapy eventually. This is largely due to the patient-specific tumor heterogeneity where cancer cells go through unique evolutionary processes to confer resistance. Our understanding of breast cancer metastasis is still not complete and needs to be extended to improve long-term control of the disease progression.

Metastasis is a multi-step and complex biological process that allows for survival and spread advantages to cancer cells. One of the postulated theories for metastasis is called the 'seed' and 'soil' mechanism that was introduced by Stephen Paget in the 1980s [125]. This means that successful growth of the cancer cell 'seed' depends on the compatibility with the secondary organ 'soil'. Successful metastasis comprises sequential molecular 'cascade' through which the primary tumor can spread to distant organs [126]. The first step involves the development of new blood vessels to the growing tumor (angiogenesis) for oxygen and nutrients supply [127]. The next step is the ability of tumor cells to escape tumor mass. In doing so, tumor cells undergo epithelial to mesenchymal transition (EMT) which allow them to diminish their cell to cell adhesion properties. This is usually mediated by molecules such as TGFß, MAPK, Wnt, Notch, Hedgehog, and the transcription factors Twist and Snail [128]. Once escaped, these cells enter the bloodstream (intravasation) migrating toward distant target organs. The next obstacle is how to leave blood circulation and break the vascular walls (extravasation), and invade the target organ. It has been shown that in some organs (e.g. bone and liver) the microvessels walls are weak and highly permeable for cancer

cells [129]. However, that is not the case in many other organs, where the endothelial lining is very tight and not permeable. Therefore, it has been established that tumor cells can form complexes with platelets and leucocytes, and interact with chemokines to facilitate metastasis to target organs [130]. Once in a new target organ, mesenchymal to epithelial (MET) process take place, and the tumor cells must adapt to their new environment.

Cancer cells exert many physiological changes on the colonized organ stroma, such as inflammation, immune response suppression, angiogenesis, and release of growth and survival factors [131]. There is increasing evidence that these modifications occur prior to tumor seeding to the target organ, where in response to factors released by the primary tumor, bone marrow derived cells create adequate microenvironment in the target organ (pre-metastatic niche) [132]. This seems to be supported by the 'organotropism' phenomenon frequently observed in breast cancer metastasis. For example, HER2+ tumors tend to preferentially form metastasis in the brain, where heregulin (HER2 ligand) is highly expressed [133]. In addition, breast cancer tumors highly express the chemokine receptor 4(CXCR4), while its ligand is highly expressed in the most common metastatic sites lung, liver, and bone [134].

One challenging aspect of cancer metastasis is the 'tumor cell dormancy", as evidenced by late relapses after years or even decades of surgical removal of tumor lesions. Some populations of tumor disseminated cells are with low proliferative and invasion capabilities and are unable to develop tumors. However, these cells remain viable and eventually they can develop tumors by interaction with surrounding microenvironment and acquisition of genetic modifications [135].

Many mechanisms in the extremely complex process of breast cancer metastasis have been identified. However, many mechanisms still need to be elucidated, as MBC is a great clinical challenge to treat in oncology. It's still unclear when and how metastatic cells appear within the

primary tumor. Also, with the field of liquid biopsy emerging, identifying circulating tumor cells with metastatic properties might provide effective way to predict metastasis at early stage, thus adaptation of novel therapeutic approaches.

1.5 ESR1 AND BREAST CANCER

The effect of estrogen is mediated in target tissue through binding with two members of nuclear receptor superfamily, estrogen receptor alpha (ERα) and beta (ERβ), encoded by *ESR1* and *ESR2* genes, respectively [136,137]. ERα is the predominant estrogen receptor in estrogen-induce breast cancer, thus it will be referred to as ER moving forward. Like other nuclear receptors, ER structure consist of A/B domain at the N-terminus with ligand-independent activation function (AF-1), DNA-binding domain (DBD, C domain), a hinge (D domain), followed by E/F domain at the C-terminus, which encompasses ligand-dependent activation function (AF-2) and ligand-binding domain (LBD) (Figure 3) [138].

Figure 3: Schematic representation of *ESR1* genomic and functional structure
The numbers represent amino acid sequence from N-terminus (left) to C-terminus (right). The region A/B contains the activation function 1 (AF-1) domain followed by DNA-binding domain (DBD) and a hinge in regions C and D, respectively. At the C-terminus, the region E/F contains the activation function 2 (AF-2) and ligand-binding (LBD) domains.

Once activated by ligand (e.g. estrogen), ER can dimerize and interact with estrogen response elements (EREs) in the promoter of wide range target genes to initiate transcription, which is known as the 'classic' genomic nuclear pathway of ER [139–141]. Alternatively, the estrogen-bound ER can interact with other direct DNA-binding transcription factors (e.g. AP1, SP1, CREB, and STAT5), leading to activation of target genes through non-classical signaling [142,143]. In addition, ER can elicit transcriptional activity in absence of ligand. Activation of receptor tyrosine kinases and growth factors (e.g. EGFR, HER2, IGF1R, and FGFR) in response to growth factors can activate ERK and PI3K/AKT, which can phosphorylate and activate ER [144].

It has been well established that activation of ER plays a key role in ER+ breast cancer initiation and progression. Nearly 70% of breast tumors express ER, and changes in its expression have been associated with clinical outcome [32]. As discussed in previous sections, ER is the principal marker for targeted endocrine therapy in breast cancer. The majority of ER+ endocrine resistant tumors retain ER expression, suggesting involvement in therapy resistance [145]. Many mechanisms of in ER activation have been elucidated and these include *ESR1* activating mutations, increased promoter activity, gene amplifications, and loss of protein degradation. *ESR1* gene mutations have been detected at low frequency in primary breast cancers [146–149]. Subsequent studies on metastatic lesions showed substantial enrichment of *ESR1* mutations (e.g. hotspot mutations Tyr537 and Asp538) in breast cancer metastasis (10-50%) in the LBD, suggesting a mechanism of resistance to endocrine therapy [65,145,150–152]. Unlike point mutations, the frequency of *ESR1* copy number amplifications is very controversial in literature. Importantly, its frequency in metastatic breast cancer remain to be elucidated. This subject constitutes large component of this dissertation and will be discussed in detail in the subsequent chapters.

1.6 PUBLIC HEALTH SIGNIFICANCE

According to the American Cancer Society, breast cancer alone is expected to account for 30% (266,120 cases) of all new cancer diagnoses in women in 2018 in US [153]. Breast cancer is the leading cause of cancer death in women aged 20 to 59 years, and second leading after lung cancer in women aged 60 years or older [154]. Despite major advances in breast cancer early detection and treatment, 20-45% of the patients relapse years or decades later [155]. Although the 5-year survival (5YS) of women living with distant metastatic breast cancer (MBC) has improved from 18% (1992-1994) to 36% (2005-2012), it is still incomparable to the 5YS of localized disease (>95%) [156,157]. Currently, we lose 40,000 every year within the US alone due to late stage MBC, thus it is not considered curable disease. With MBC remains challenging to treat, clearly identification of pathways involved in metastatic spread and therapy resistance is essential.

One breast cancer subtype under the focus in this dissertation is invasive lobular carcinoma (ILC). ILC is considered the second most common breast cancer subtype, representing 5-15% of newly diagnosed breast cancers (25,000-30,000 new) [4,5]. If considered alone ILC ranks as the 6th most common cancer in women [11]. Given that the focus in breast cancer research has been on the invasive ductal carcinoma (IDC), the incidence of this subtype has been relatively constant in the last two decades. In contrast, ILC is an understudied subtype and has been showing significant increase in the number of cases [4]. In women age ≥ 60, the percent increase in the rate of incidence in ILC was more than 2-fold greater than the increase in IDC from 1977-1980 to 1993-1995. In addition, despite ILC shows better prognostic and predictive markers (less proliferative and more ER positive), this subtype does not necessarily show better survival and outcome compared to IDC [158,159]. Better understanding of etiology and unique biology of the ILC subtype has become a necessity and deserve further investigation. Molecular characterization of ILC will aid in

identification of unique mechanisms involved in metastasis and endocrine therapy resistance, a key for finding therapeutic targets.

1.7 HYPOTHESIS

The use of endocrine therapy has substantially improved outcomes and quality of life in patients with breast cancer. However, its effectiveness is somewhat limited as about 20-40% of patients eventually develop relapse [160]. Although ER is the major target for such therapy, the conflicting evidence about *ESR1* amplifications and gains in literature leaves an unanswered question about the frequency and clinical significance of *ESR1* amplifications in breast cancer. Moreover, these alterations remain to be characterized in the context of breast cancer metastasis. I hypothesize that *ESR1* amplifications are involved in tumor progression and/or endocrine resistance in metastatic breast cancer.

Despite expressing biomarkers predictive of good prognosis (e.g. being more often ER+), ILC patients do not have improved survival compared IDC [159,161]. While more than 90% of ILC are ER+, a retrospective analysis from the Breast International Group (BIG) 1-98 trial suggests that subset of ILC patients does not benefits as much as IDC patients from tamoxifen treatment [162]. Previous characterization of endocrine response in ILC cell lines by our lab identified unique transcriptomic regulation in response to E2 [163]. In that study, the ILC cell line MDA-MB-134 exhibited resistance to tamoxifen and partial growth induction in response to its metabolite 4-hydroxytamoxefin (4-OHT). The results address the necessity of further understanding of endocrine resistance in ILC. In line with these findings, I hypothesize that ILC has distinctive copy

number profile of *ESR1* and other genes that are involved in modulation of response to endocrine therapy.

There is increasing evidence that, in addition to the metastasis to common sites, ILC can also spread to unique sites such as the urogenital and GI tracts compared to IDC [161,164–166]. This indicates unique biological differences in between ILC and IDC that yet to be elucidated. With ILC being strongly influenced by hormones, I hypothesize that the unique ovary microenvironment (including high E2 levels) provides an attractive niche for the unique metastatic spread of ILC.

2.0 FREQUENT ESR1 AND CDK PATHWAY COPY NUMBER ALTERATIONS IN METASTATIC BREAST CANCERS

2.1 INTRODUCTION

Breast cancer is a genetic disease that is driven by accumulations of single nucleotide mutations and structural alterations [167]. The latter include gene fusions, deletions, tandem duplications, and copy number (CN) amplifications such as *ERBB2* (HER2), *CCND1*, and *MYC*. Over the years, there have been several studies showing varying levels of estrogen receptor (ERalpha) gene (*ESR1)* CN amplifications in breast cancer. In 2007, *Holst et al* identified *ESR1* CN amplifications and gains in 20.6% and 15.3% of primary breast tumors respectively, using fluorescence in situ hybridization (FISH) [168]. Subsequently, multiple groups reported *ESR1* CN amplifications at much lower frequency (0-10%), using FISH and array comparative genomic hybridization (aCGH) [169–174]. Analysis of The Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) datasets using cBioPortal [40,175,176] revealed that *ESR1* CN amplifications are present in only 2.5% and 2.3% of breast cancer samples, respectively. More recently, *Desmedt et al* showed *ESR1* CN gains and corresponding increased *ESR1* mRNA levels in 25% of primary invasive lobular breast cancers (ILC) [177].

There is less data about structural variants (especially gene fusions) in metastatic breast cancer (MBC). There have been reports showing gene arrangements, and *ESR1*, *FOXA1*, *CYP19A1*, and *ERBB2* CN amplifications in endocrine resistant breast cancer [178–181]. Li et al [150] originally described an *ESR1* fusion arising in a PDX model. We have recently identified nine additional *ESR1* fusions in advanced disease. Importantly we were able to show protein expression

of the fusion for some cases [182]. The junction between the N-terminal *ESR1* gene, and the C-terminal fusion partner was located between *ESR1* exons 6 and 7. Gene fusion events are often associated with CN alterations adjacent to the fusion junction and imbalance in DNA CN of the exons around the break [183,184]. We therefore designed a novel copyshift algorithm that identifies CN imbalances, and applying this algorithm to *ESR1*, we identified frequent CN imbalances, significantly enriched in ER+ metastatic lesions compared to primary breast tumors [182].

Despite overwhelming evidence for a critical role of *ESR1* mutations and increasing evidence for *ESR1* fusions in endocrine resistant breast cancer, a comprehensive study of *ESR1* CN variations in metastatic disease is lacking. Here, I set out to characterize *ESR1* copy number alterations in a set of well curated metastatic lesions, and in a subset of those, their patient-matched primary tumors. I utilized a highly sensitive nanoString-based approach, resulting in high resolution data. In addition, I also studied CN alterations in 66 druggable candidate genes with known roles in breast cancer progression and metastasis.

2.2 MATERIALS AND METHODS

2.2.1 Sample collection

141 formalin-fixed paraffin-embedded (FFPE) and frozen sections were obtained from eligible metastatic cases from the University of Pittsburgh School of Medicine (SOM) Health Sciences Tissue Bank (HSTB). Collection and analysis of specimens were approved under the University of Pittsburgh IRB and Charite Universitätsmedizin Berlin guidelines. When available, patient-matched primary tumors were included in the study (n=26). Seven samples were excluded from

the study due to unknown ER status. In total, I had 134 samples from 100 patients that met the study criteria (82 ER+ and 52 ER-), including 26 primary tumors (15 ER+ and 11 ER-) and 108 metastases (67 ER+ and 41 ER-). In some cases, multiple primary tumors or metastases were from the same patient. More than 85% of the samples had tumor cellularity > 40%. Clinicopathological characteristics are provided in Data Supplement 1: Table S1. Information on ER status, source of tissue, site of metastasis, histologic subtype were collected.

2.2.2 Immunohistochemistry (IHC)

Hematoxylin and eosin (H&E) and ERα staining procedures were performed at the Histology and Micro-imaging Core (HMC) facility at Magee-Women's Research Institute (MWRI). ER IHC was performed using the mouse monoclonal antibody: Novacastra Leica, Catalog No: NCL-L-ER-6F11. Analysis of histological classification was performed by two pathologists (Drs. Peter Lucas and Esther Elishaev) (Data Supplement 1: Table S1).

2.2.3 DNA isolation

DNA was isolated from 4-6 formalin-fixed paraffin-embedded (FFPE) macrodisected sections (10 µm) per sample depending on tumor size and cellularity using the Qiagen AllPrep FFPE kit (cat# 80234) as per manufacturer's instructions. DNA from frozen samples was extracted using the Qiagen DNeasy kit (cat# 69506) according to the manufacturer's instructions. All DNA quantifications were done using Qubit dsDNA HS/BR assay kits (ThermoFisher).

2.2.4 Control samples for nanoString and Digital droplet PCR (ddPCR) comparison

I isolated DNA from MCF-7 long-term estrogen-deprived (LTED) cells, BT-474, and MCF10A cells for validation of CN calls. MCF-7 LTED cells have *ESR1* amplifications [150], and BT-474 cells have been described to have a heterozygous *ESR1* deletion (Cancer Cell Line Encyclopedia (CCLE)) [185]. MCF10A is a non-tumorigenic epithelial cell line with normal (2N) *ESR1* CN [186].

2.2.5 Digital droplet PCR (ddPCR)

Primers and probes were designed and ordered through Integrated DNA Technologies (IDT) for *ESR1* and Bio-Rad for two reference genes recommended by the manufacturer (*EIF2C* and *AP3B1*) (Table 9). 60 ng of control samples were processed for ddPCR analysis as previously described [187]. Briefly, DNA samples were combined with primers, probes, and supermix, and then added to cartridge. Droplets were generated using Biorad QX100 Droplet Generator and transferred into a 96-well plate for PCR amplification, and a droplet reader (Biorad QX100 Droplet Reader) was used to count PCR+ve and PCR-ve droplets. Data were analyzed using QuantaSoft™ software (Bio-Rad) where target concentration was normalized to reference concentration and multiplied by the number of reference loci in the genome (ideally 2) to generate copy number calls.

2.2.6 nanoString

In collaboration with the manufacturer, I designed 100bp DNA hybridization probes for a total of 67 genes (Data Supplement 1: Table S2), including *ESR1* (n=10 probes) (Figure 24), and 66 genes known to be frequently altered in breast cancer (n=3 probes per gene). For the latter, we queried

TCGA breast cancer dataset and previous breast cancer metastasis studies [152,188–190], with a focus on potentially druggable genes as identified in drug gene interaction database (DGIdb) [191]. Although there is a limitation in this approach by decreasing the chance of finding more novel alterations, it is still focused , well curated, and represent a comprehensive panel for potential cancer genes. Processing of CN data was performed as per manufacturer recommendations. Briefly, DNA samples were fragmented at 37 ºC using Alu1 restriction enzyme, denatured at 95 ºC, and hybridized overnight with target probes. Post-hybridization sample processing was done using the automated nCounter Prep Station. Raw counts were then collected from the nCounter Digital Analyzer and transferred to the nSolver™ software (v 2.5) for data analysis [192]. Raw counts were normalized to 10 invariant reference probes (count normalization) and to reference samples pools of normal breast FFPE (n=13) and frozen (n=4) DNA to generate CN calls (Data Supplement: Tables S2-S4). Average CN estimate value was calculated per gene based on all probes for that gene relative to the CN estimate in the normal sample pool. The 67 genes of the nanoString code set for comprehensive quality control assessment of copy number calls between FFPE and cell pellet DNA.

2.2.7 Preparation and sequencing of RNA-seq libraries

RNA sequencing data was available for a subset of 66 samples (46 ER+), and was used to associate correlation between *ESR1* CN and mRNA expression. TruSeq RNA Access library preparation (illumina) and sequencing was performed at the sequencing core facility at Children's Hospital of University of Pittsburgh Medical Center (UPMC) using the NextSeq500 platform that produced paired-end reads (2X75bp). Gene expression values for *ESR1* were calculated using counts per

million normalized to trimmed-means of M-values (TMM-normalized CPM) using DESeq2 and edgeR packages in R [193,194].

2.2.8 *ESR1* validation datasets

Patients data from the American Association for Cancer Research (AACR) GENIE and Memorial Sloan Kettering Cancer Center (MSKCC) was available through cBioPortal [195,196]. Access to Foundation Medicine (FM) data was provided through collaboration with Drs. Ryan Hartmaier and Ethan Sokol. Statistical analyses with p value significance of 0.05 were performed in R and GraphPad. Figures were plotted using ggplot2.

2.2.9 Statistical and bioinformatics analysis

R environment was used for statistical computing and graphics [197]. Oncoprints visualizing multiple genomic alterations were generated using ComplexHeatmap package [198]. CN increase by 35% (CN ≥2.7) and decrease by 50% (CN ≤ 1) were considered gains/amplifications and deletions, respectively. Copy number calls above ≥10 were considered high amplifications. For other genes in the panel, I used one copy number cutoff of ≥5 as amplification. Plots and heatmap of actual CN imbalance and shifts were generated using ggplot2 and heatmap.3 packages, respectively [199,200]. For CN imbalance, I reported only shifts of ≥ 30% difference in CN (ratio of ≥ 1.3) between *ESR1* 5' and 3' exons, which is similar to the 35% CN increase used to define gains. Multiple correlations between the different *ESR1* exons and mRNA expression were clustered by first principal component (FPC) scores. Spearman and Pearson correlations, Fisher's exact, and

Wilcoxon rank sum tests were performed in R and GraphPad Prism with significance cutoff of 0.05.

2.3 RESULTS

2.3.1 Sensitive and efficient measurement of DNA CN changes using nanoString technology

First, I set out to determine, in our hands, which platform (ddPCR or nanoString) is best suited for quantifying CN alterations. I determined *ESR1* CN status in MCF7-LTED, BT-474, and MCF10A cells by ddPCR, using *EIF2C* and *AP3B1* reference probes. Our CN calls were consistent between the two probes for the *ESR1* CN normal (MCF10A) and deletion (BT-474) models, but not MCF7-LTED cells, that have a known *ESR1* amplification (Figure 4A and Data Supplement: Table S5). These data suggested that use of two references probes was not sufficient for accurate measurement of CN. Next, *ESR1* CN was measured using nanoString technology, where ten reference probes were included in the library. This analysis correctly identified CN in the three control cell lines. I then performed extensive quality control experiments to test sensitivity and reproducibility of the CN calls within and between the nanoString runs using DNA from fresh cell pellets and from processed FFPE sections. There was a high correlation between DNA isolated from fresh and fixed samples (Figure 4B-C, and Data Supplement: Table S7; rho > 0.9, $p < 2.2e\text{-}16$) and excellent reproducibility between different nanoString runs using the same DNA (Figure 4D-E and Table S8; rho > 0.99, $p < 2.2e\text{-}16$). I therefore proceeded using nanoString technology for the characterization of CN alterations in a large set of well curated breast tumors.

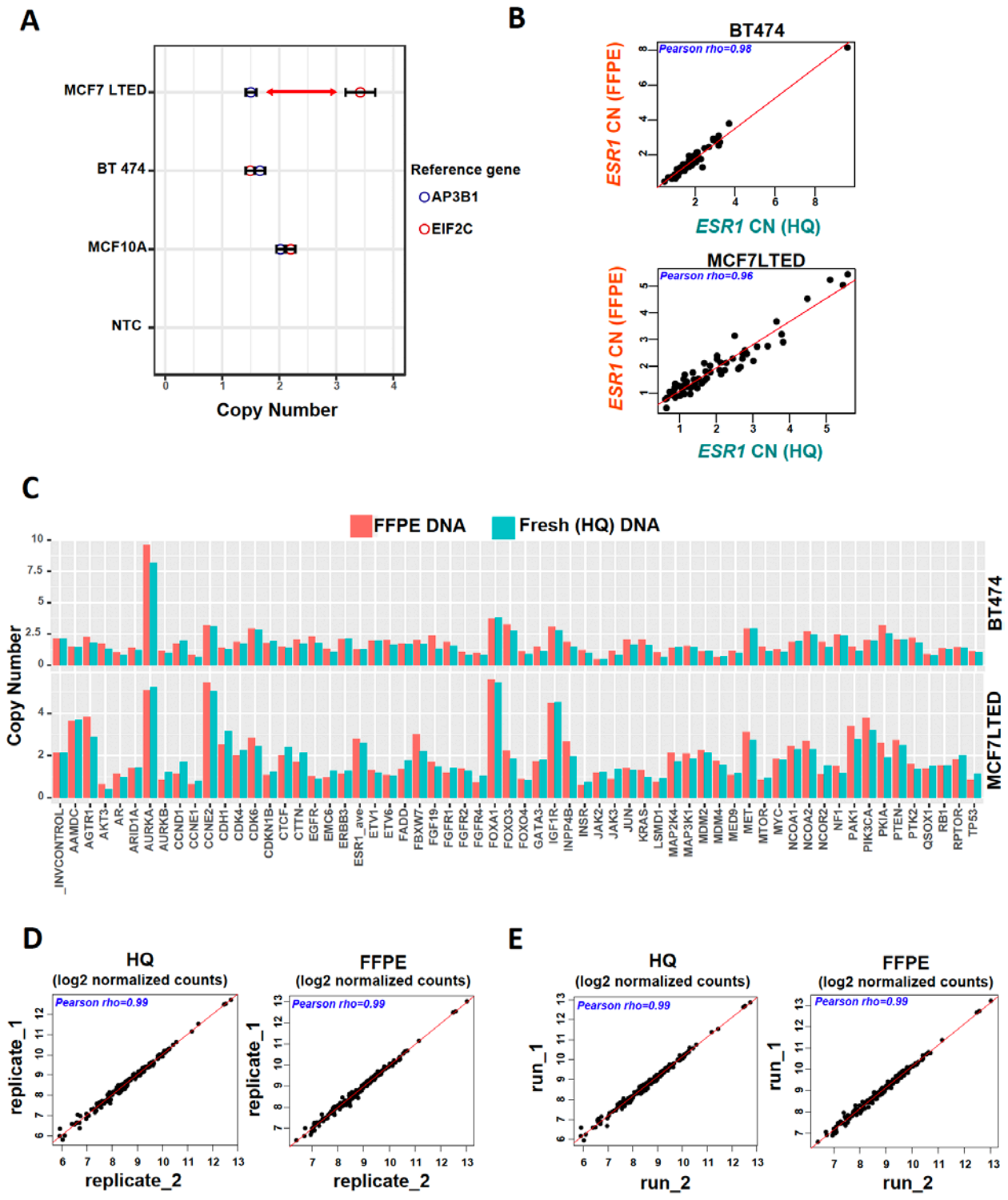


Figure 4: nanoString and ddPCR CN analysis in control cell lines

A. ddPCR *ESR1* copy number analysis of MCF7-LTED, BT-474, and MCF10A cell lines that represent amplification, deletion, and normal copy number models, respectively. The X-axis represents CN status for the *ESR1* gene that was normalized separately using two reference genes normalizers (*AP3B1* and *EIF2C*). The red bidirectional arrow indicates inconsistent *ESR1* CN calls for MCF7 LTED. Error bars of the droplet

Poisson distribution for the 95% Confidence Interval are indicated for each data point. NTC; non-template control. **B-C.** Comparison of nanoString CN calls from FFPE vs immediately processed high quality (HQ) samples of the same cell lines (top; BT474, bottom; MCF7LTED) showed very high correlation (Pearson's rho > 0.9, p < 2.2e-16). Data points at extreme ends (i.e. very low or very high expressed genes, n=6) are not included in the graph for better visualization of data (addition/exclusion didn't change rho value significantly). Plots of 61 genes and invariant controls shows highly correlated absolute CN calls for HQ and FFPE DNA of each cell line (top; BT474, bottom; MCF7LTED). Y-axis represent CN calls for each gene. **D-E.** MCF7LTED HQ and FFPE DNA was analyzed twice within the same run (D) and in two separate runs (E) to assess reproducibility. Reproducibility was very high within and between the runs (Pearson's rho > 0.99, p < 2.2e-16).

2.3.2 Frequent *ESR1* amplifications in metastatic breast cancer

Our probe design targeted *ESR1* untranslated exons (E1 and E2), promoter region, and coding exons (E3-E10) (Figure 24). *ESR1* CN was measured in a total of 134 tumor samples, defining CN ≥ 2.7 as gain, and CN ≥ 10 as amplification (Figure 5). When averaging CN calls from all *ESR1* probes, I detected *ESR1* amplifications in eleven (13.5%) out of 82 ER+ tumor samples (67 metastases and 15 primary) (Figure 5A and Table 2). Specifically, eight (9.8%) samples had CN gains and three (3.7%) samples had amplifications. There was a trend for enrichment of *ESR1* amplifications in metastatic samples (14.9%) compared to primary samples (6.6%). Bone metastases showed significant enrichment for amplifications vs primary and ovaries (Figure 5B and Table 2; $p < 0.05$).

In ER- tumors (41 metastases and 11 primary), I did not detect any sample with amplification, and found only one primary tumor with a CN gain (Figure 5A and Table 10). *ESR1* amplifications were significantly enriched in ER+ vs ER- tumors (Table 3; $p = 0.0192$).

Since our cohort included 13 ER+ patient-matched primary-metastatic tumor pairs, I was able to explore if any metastatic lesions demonstrated an increase in CN as compared to their matched primary lesions. Indeed, I observed this for one pair, where the ER+ primary tumor BP51

already had a gain in *ESR1* (CN 3.9), and the patient-matched brain metastases BM51 showed high level amplifications (CN 10.9) (Figure 5C and Figure 26).

We have recently described a novel algorithm that might aid in identification of gene fusions from hybrid capture NGS based on the frequent finding of DNA CN imbalance of the exons at a breakpoint associated with fusion [182]. The high resolution of our approach allowed us to adapt this approach to determine *ESR1* CN imbalance by calculating the ratio of CN of 5' exons (3-6) to 3' exons (7-10). Our analysis revealed that five out of 11 amplified samples (45.5%) exhibited ≥ 30% increase in CN (ratio of ≥ 1.3) in the 5` exons (Figure 5D and Table 12, *p= 0.0024*), while there was no sample in which the CN of 3' exons were increased relative to 5' exons.

In summary, I observed enrichment of *ESR1* amplifications in ER+ and recurrent samples, with indications for metastatic site tropism. I also identified frequent 5'-3' exonic imbalances indicating potential breakpoints in the *ESR1* gene.

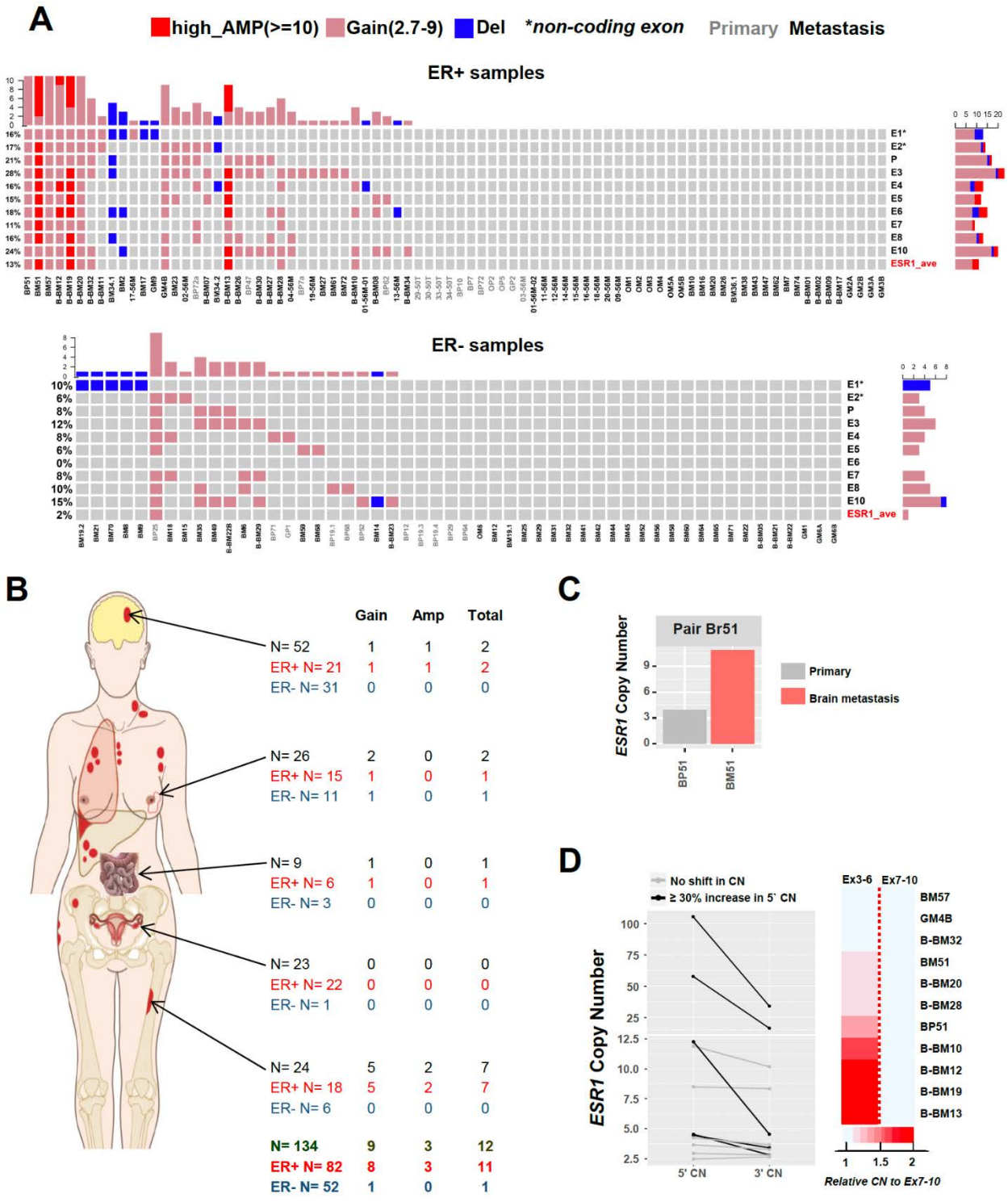


Figure 5: Frequent ESR1 amplifications in metastatic breast cancer

A. Oncoprint visualization of *ESR1* copy number alterations in ER+ (top) and ER- (bottom) samples. Levels of amplification and deletion are color coded. Each column represents a single sample (sample IDs labelled as: grey; primary tumor, black; metastasis). Each row indicates copy number call of the correspondent single exon probe. Untranslated exons (E1 and E2) are annotated with '*' symbol. P; Promoter probe. *ESR1*_ave: average copy number call of all probes. **B.** Distribution of *ESR1* copy number amplifications by site (primary,

brain, bone, GI, and ovaries as indicated by arrows) and ER status (color coded). Bone metastases showed significant enrichment for amplifications vs primary and ovaries (Fisher's exact test, $p < 0.05$). **C.** Single pair where the brain metastasis had higher level *ESR1* amplifications compared to its primary tumor. **D.** Graphical representation (left; actual copy number) and heatmap (right; ratio) of 5'-3' copy number imbalance among the *ESR1* amplified samples. The dark red color indicates higher copy number amplification toward the 5' side of *ESR1* (Wilcoxon matched pairs signed-rank test $p = 0.0024$).

Table 2: *ESR1* copy number alterations by site in ER+ samples

Site	N	n (%)		
		AMP	Gain	Total
Primary	15	0.0	1 (6.7)	1 (6.7)
Brain	21	1 (4.8)	1 (4.8)	2 (9.5)
Bone	18	2 (11.1)	5 (27.8)	7 (38.9)*
Ovaries	22	0.0	0.0	0.0
GI	6	0.0	1 (16.7)	1 (16.7)
Total	82	3 (3.7)	8 (9.8)	11 (13.4)

*significant enrichment for amplifications vs Primaries and ovaries (Fisher's exact test, $p < 0.05$)
N= number of samples included in the cohort; n and (%)= count and percentage of samples with alterations for the indicated site, respectively.

Table 3: *ESR1* copy number alterations by ER status

ER status	N	n (%)		
		AMP	Gain	Total
ER+	82	3 (3.7)	8 (9.8)	11 (13.4)*
ER-	52	0.00	1 (1.9)	1 (1.9)
Total	134	3 (2.2)	9 (6.7)	12 (9)

*Fisher's exact test, $p = 0.0192$
N= number of samples included in the cohort; n and (%)= count and percentage of samples with alterations for the indicated site, respectively.

2.3.3 Correlation of *ESR1* CN with ER mRNA and protein expression

To determine whether *ESR1* CN gains and amplifications were correlated with ER mRNA expressions, I utilized expression data from our RNAseq analysis. As expected, I observed higher *ESR1* mRNA expression in samples with *ESR1* gains/amplifications (Figure 6A), although this association did not reach statistical significance, likely due to the limited number of samples in the CN gains/amplification group for which RNAseq data was available (n=3). There was however a correlation between CN and expression with the group of *ESR1* amplified samples (Figure 25). The first translated *ESR1* exon E3 was most predictive for mRNA expression (Figure 6B and Data Supplement: Table S9; rho=0.34, $p= 0.0219$).

Subsequently, I evaluated the effect of *ESR1* amplifications on ER protein expression. IHC staining for ER (3 amplified vs 3 normal) revealed very strong and homogeneous ER expression in the amplified samples (Figure 6C). Some samples showed high ER protein expression despite normal CN (Figure 27), which is likely the result of either sub-threshold amplifications or transcriptional/posttranscriptional regulation of ER expression.

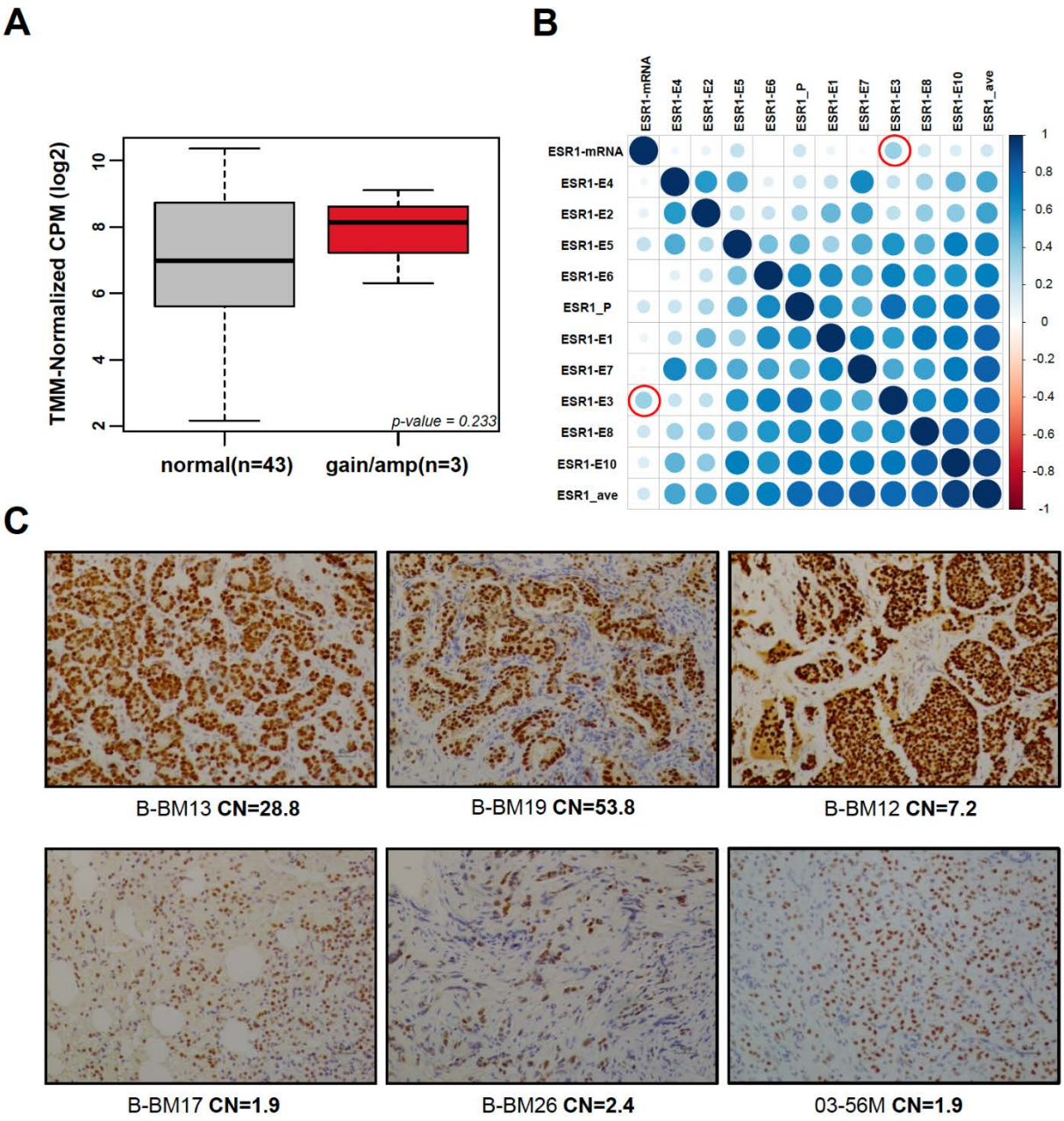


Figure 6: Correlation of ESR1 copy number with mRNA expression and immunohistochemistry (IHC)

A. Box plot comparison of *ESR1* mRNA expression between normal vs gain/amplification groups (Wilcoxon rank-sum test, p=0.233). Y-axis: TMM-normalized log2 of count per million (CPM). **B.** Correlation matrix plot of *ESR1* copy number calls by different exons and mRNA expression. Multiple correlations were clustered by first principal component (FPC) scores. Bigger and darker blue circles represent higher correlation. Red outline represents significant p-value for correlation with mRNA expression ($p < 0.05$). **C.** IHC staining of *ESR1* amplified (top row) and normal samples (bottom row). Correspondent *ESR1* DNA copy number calls are indicated below each image.

2.3.4 Validation of *ESR1* CN amplifications enrichment in additional cohorts

Driven by the enrichment of *ESR1* CN amplifications in our metastatic cohort, I sought to validate these findings in three additional datasets (AACR_GENIE, MSKCC, and FM) that used NGS approaches [201–204]. Our analysis revealed significant enrichment for *ESR1* amplifications in metastatic samples versus primary tumors (Figure 7A). The frequency of the amplifications in the metastatic cohorts were 1.65% (n=27/1637)), 2.28% (n=19/835), and 1.84% (n=122/6629) in AACR_GENIE, MSKCC, and FM respectively. On note, these alterations are probably underestimated given that I did not limit the study to ER+ tumors due to lack of ER status iformation in the AACR_GENIE and MSKCC. Indeed, correction of FM for ER status increased the frequency to 3.71%. Furthermore, analysis of survival data from MSKCC showed significantly worse overall survival (OS) even after performing metastasis-only analysis (p value <0.05, Figure 7B).

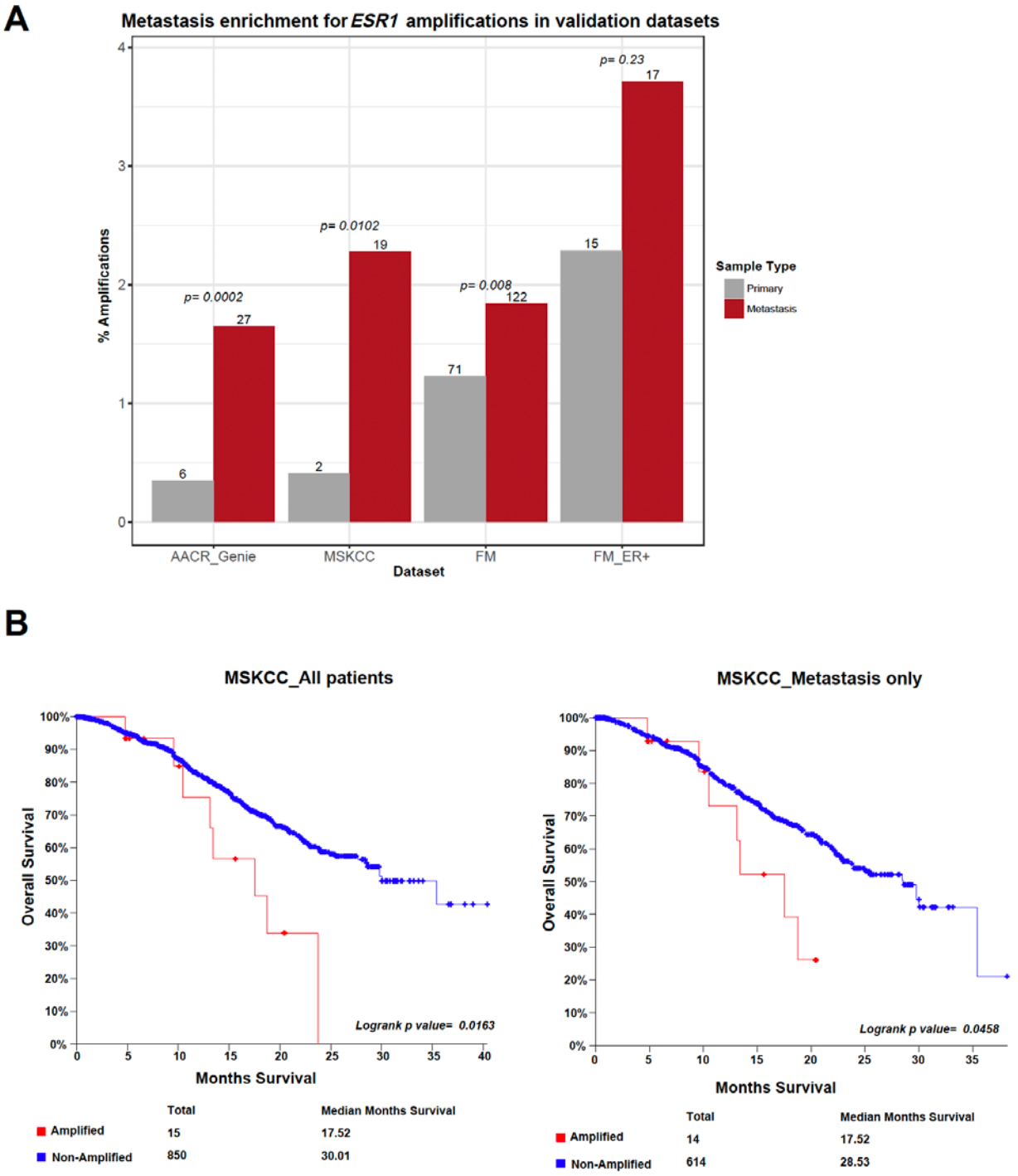


Figure 7: Metastasis enrichment for *ESR1* amplifications in validation datasets
A. Barplot for the percentage of *ESR1* amplifications in three different genomic datasets (FM was also plotted by ER+ only). Sample sites are color coded (grey, primary; red, metastasis). Actual counts for amplified samples are indicated on the top of each bar, with p values for metastatic enrichment. AACR, American Association for Cancer Research; MSKCC, Memorial Sloan Kettering Cancer Center; FM, Foundation Medicine. MSKCC is also a major contributor to the AACR_GENIE dataset. **B.** Kaplan-Meier curves for MSKCC survival data grouped by *ESR1* CN status (red, amplified; blue, not amplified) for both primary and metastatic (left) and metastatic cases only (right).

2.3.5 CN alterations of known and potential breast cancer driver genes other than *ESR1*

In addition to *ESR1*, I extended our CN analysis to 66 genes with described roles in breast cancer progression and metastasis. The genes with most frequent CN amplifications are shown in Table 4 and Figure 28. The most amplified genes were *ERBB2* and *GRB7* (35%), showing amplifications of 44% and 21% in brain and bone metastases, respectively. Previous studies have reported co-amplifications of *ERBB2* and *GRB7* on amplicon 17q12 [205–207]. This was confirmed by our eSNP-karyotyping analysis using RNAseq data from a subset of our samples that predicts the amplification to cover a broader region rather than being limited to a focal event (Figure 29).

The analysis of CN alterations also revealed numerous deletions. The most frequently deleted gene was *TP53* (10%) and this loss was mainly observed in brain metastases (15.4%) (Table 4 Figure 28). Metastases site-unique or enriched CN alterations were seen for a number of genes, including higher rates of *FADD* amplifications (17%) in bone metastasis compared to other metastatic sites, and higher rates of *PTK2* and *PKIA* amplifications (~10-20%) in brain and GI metastases. Additionally, comparison of ER+ vs ER- brain metastases showed significant enrichment of *FGFR1* amplifications in the ER+ group (Figure 30; p= 0.0221). The comparative analysis between ER+ and ER- tumors was only possible in the brain metastases cohort where I had balanced distribution of ER+ vs ER- tumors.

Multiple co-occurrence analysis identified multiple gene combinations that are enriched in ER+ tumors (Figure 8A, top panel; $p < 0.05$). For example, I confirmed previously reported co-amplifications of genes (*CCND1, CTTN, FADD, PAK1, AAMDC, and FGF19*) at the 11q13 amplicon [208–213] , of *MYC (8q24)* and *ERBB2 (17q12)* [214–216], and of *NCOA3 (20q13*) and *MDM2 (12q15)* [217] (Figure 8B). I also observed co-amplification of *MDM2* and *ERBB3 (12q13)*, which had not been previously described. Further investigation of this co-amplification in TCGA dataset

suggested that these amplifications originate from two separate amplicons (Figure 31). Finally, there was an enrichment of *MYC* (45%) and *CCND1* (36%) in tumors with *ESR1* amplifications (Figure 8C; $p < 0.05$). I identified only one event of co-occurrence in ER- tumors, involving CN amplifications of *ERBB2* and *GRB7* at the 17q12 amplicon (Figure 8A; bottom panel).

Among the most frequent recurrent deletions, I identified mutually exclusive deletions of *CDKN2A*, *CDKN2B*, and *CDKN1B* with amplifications of *CCND1* and *CDK4/6* (Figure 8D; $p < 0.0001$). Expression analysis from RNAseq data showed significant correlation with these CN alterations (Figure 8D).

Table 4: Most frequent copy number alterations by site

Alteration	Gene	n (%)					
		Primary	All Mets	Brain	Bone	Ovary	GI
Amplifications	ERBB2	9 (34.6)	31 (28.7)	23 (44.2)	5 (20.8)	2 (8.6)	1 (11.1)
	GRB7	9 (34.6)	29 (26.8)	24 (46.1)	2 (8.3)	2 (8.6)	1 (11.1)
	CCND1	1 (3.8)	13 (12)	6 (11.5)	5 (20.8)	2 (8.6)	0 (0)
	MYC	4 (15.3)	13 (12)	8 (15.3)	4 (16.6)	0 (0)	1 (11.1)
	FGF19	3 (11.5)	10 (9.2)	6 (11.5)	2 (8.3)	2 (8.6)	0 (0)
	CTTN	0 (0)	9 (8.3)	5 (9.6)	2 (8.3)	2 (8.6)	0 (0)
	FADD	1 (3.8)	8 (7.4)	3 (5.7)	4 (16.6)	1 (4.3)	0 (0)
	PTK2	0 (0)	8 (7.4)	6 (11.5)	0 (0)	0 (0)	2 (22.2)
	PKIA	0 (0)	7 (6.4)	5 (9.6)	1 (4.1)	0 (0)	1 (11.1)
	FGFR1	4 (15.3)	6 (5.5)	4 (7.6)	0 (0)	2 (8.6)	0 (0)
Deletions	TP53	0 (0)	11 (10.1)	8 (15.3)	1 (4.1)	1 (4.3)	1 (11.1)
	AURKB	0 (0)	9 (8.3)	7 (13.4)	2 (8.3)	0 (0)	0 (0)
	CDKN1B	3 (11.5)	9 (8.3)	9 (17.3)	0 (0)	0 (0)	0 (0)
	CDKN2B	3 (11.5)	9 (8.3)	6 (11.5)	1 (4.1)	2 (8.6)	0 (0)
	CDKN2A	4 (15.3)	7 (6.4)	5 (9.6)	1 (4.1)	1 (4.3)	0 (0)
	MAP2K4	0 (0)	7 (6.4)	7 (13.4)	0 (0)	0 (0)	0 (0)
	JUN	0 (0)	7 (6.4)	6 (11.5)	0 (0)	1 (4.3)	0 (0)
	LSMD1	1 (3.8)	7 (6.4)	4 (7.6)	0 (0)	3 (13)	0 (0)
	PTEN	2 (7.6)	6 (5.5)	6 (11.5)	0 (0)	0 (0)	0 (0)

n and (%)= count and percentage of samples with alterations for the indicated site, respectively.

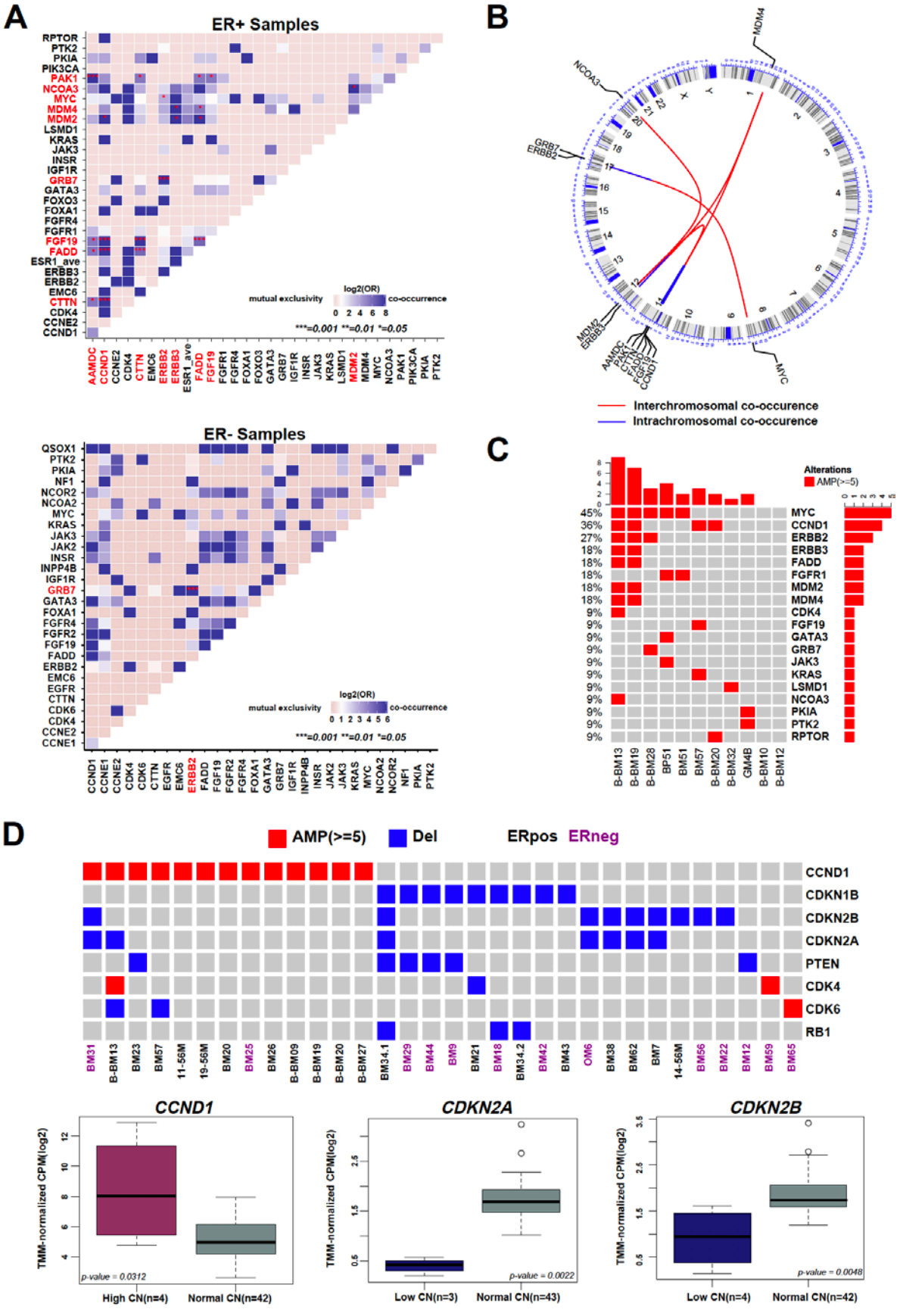

A
ER+ Samples
mutual exclusivity
log2(OR)
co-occurrence
***=0.001 **=0.01 *=0.05
ER- Samples
B
Interchromosomal co-occurence
Intrachromosomal co-occurence
C
Alterations
AMP(>=5)
D
AMP(>=5)
Del
ERpos
ERneg
CCND1
CDKN1B
CDKN2B
CDKN2A
PTEN
CDK4
CDK6
RB1
CCND1
CDKN2A
CDKN2B
TMM-normalized CPM(log2)
p-value = 0.0312
High CN(n=4)
Normal CN(n=42)
p-value = 0.0022
Low CN(n=3)
Normal CN(n=43)
p-value = 0.0048
Low CN(n=4)
Normal CN(n=42)

Figure 8: Co-occurrence and mutual exclusivity of CN alterations in metastatic breast cancer
A. Tile plots of ER+ (top) and ER- (bottom) samples for copy number amplifications co-occurrences. Gene pairs with significant co-occurrence are colored in red and marked with '*' in the plot (Multiple Fisher's exact test with FDR adjusted p <0.05). **B.** Circos plot for genomic co-occurrence events. Circular tracks from outside to inside: genome positions by chromosomes (black lines are cytobands); Inside arcs connects genes with co-occurrence (red; inter-chromosomal, blue; intra-chromosomal). **C.** Oncoprint shows enrichment of *MYC* (45%) and *CCND1* (36%) amplifications in tumors with *ESR1* amplifications (Fisher's exact test; *MYC* p= 0.0083, *CCND1*, p= 0.0365). **D.** Top: Oncoprint of samples with CN alterations in the CDKs pathway and its inhibitors. Amplifications of the activators *CCND1* and *CDK4/6* of the pathway were mutually exclusive to the deletions of the inhibitory members *CDKN1B*, *CDKN2A*, and *CDKN2B* (Fisher's exact test p < 0.0001). Bottom: RNAseq data for a subset of samples (n=46) shows significant correlation of mRNA expression with copy number status for genes of interest (p indicated for Wilcoxon rank-sum test).

2.4 DISCUSSION

Breast cancer is increasingly being recognized as a disease of structural aberrations, in addition to well described single base pair mutations. The majority of CN studies have been performed in primary tumors, at least in part due to difficulties in accessing metastatic tissue. However, this has become easier due to the realization of gene expression changes in treatment targets, such as ER and HER2, and thus increasing biopsy rates of metastases by clinicians. There have been some conflicting reports on frequencies of *ESR1* amplification, ranging from 0 to 30% [169–174]. CN has been explored using FISH, aCGH, and qPCR, and the NGS is currently evolving as promising approach. Due to the potential clinical relevance of this question, I sat out to use a sensitive and specific nanoString methodology to measure *ESR1* CN in a unique cohort of metastatic, and when available patient-matched primary tissues.

In total, I detected gains or amplifications of the *ESR1* gene in 14% of ER+ breast tumors. Given the sensitivity and specificity of our method, and careful sampling of tumors with cellularity of at least 40%, I am confident that this is a representative number for ER+ tumors. Previously reported wide ranges in detected CN alterations are due, at least in part, to the use of technologies with limited sensitivities and specificity such as aCGH, FISH, and qPCR. Our study clearly shows

PCR approaches using a limited number of reference genes need to be interpreted with caution. Additional reasons for variability include tumor cellularity and threshold for calling gain and amplification, as previously discussed [218].

I observed significant enrichment of *ESR1* gains and amplification in the ER+ metastatic lesions, compared to the primary tumors, and I was able to validate this finding three independent datasets. In one out of 15 ER+ paired samples, *ESR1* CN increased from 3.9 in the primary tumor to 10.9 in the patient-matched brain metastasis. This enrichment suggest for selection of the CN increase under endocrine treatment, similar to what has been described for the *ESR1* mutations [152,188], and HER2 amplifications in brain metastases recently described by us [181]. Although somewhat limited by small numbers, I did observe a trend towards a correlation between *ESR1* amplifications and *ESR1* mRNA and protein expression, a finding that has previously been reported by others [168,219]. Also, the significant enrichment of *ESR1* CN gains and amplification in bone metastases, an environment that is known to support hormone-dependent tumors, provides further evidence for functionality of these CN changes. However, there are reports in which *ESR1* amplification do not correlate with expression level [171,218,220], and *ESR1* amplifications have been previously detected in ER- tumors with poor prognosis [221]. Of note, in our study, I only observed an *ESR1* gain (CN= 2.7) in one ER- sample. Additional studies are necessary to decipher the functional consequences of low level *ESR1* amplifications.

We [182] and others[150] have recently identified *ESR1* gene fusions in MBC, with the fusions generally maintaining DNA binding and the N-terminal transcriptional activation domains, but deleting the ligand-binding domain. We recently described a copyshift algorithm that determines imbalance in DNA CN of exons that are 5' or 3' to the break [182]. The high resolution of our assay design allowed us to detect CN imbalances at the exon level, and I found that five out of 11

amplified samples (45.5%) showed at least 30% increase in CN at the 5` side of *ESR1*. The lack of RNAseq data for four out of these five cases didn't allow us to characterize the fusions extensively. However, these imbalances are indicative of genetic rearrangements, therefor our future studies will test whether these tumors harbor *ESR1* fusion genes, and if so, which ones might be non-functional vs drivers of endocrine resistance.

The clinical relevance of *ESR1* CN is unclear at this point in time. *ESR1* amplifications have been associated with improved [168,218,219] as well as worse outcome [222] in patients treated with endocrine therapy. In endocrine resistant cell line models [223] and in a patient-derived xenografts (PDX) model [150] with *ESR1* amplifications, estradiol treatment resulted in tumor regression. We recently described a metastatic breast cancer case with *ESR1* amplifications that showed sustained partial response to high dose estradiol treatment as measured by CA 27-29 level and by decrease in liver metastasis burden [224]. These findings might deserve further exploration in a clinical trial setting including prospective measurement of *ESR1* CN.

Analysis of 66 other - mainly druggable - genes revealed frequent *ERBB2* and *GRB7* amplifications in different metastatic sites, with an enrichment in brain metastases. These results support the increasingly growing need of testing for HER2 in metastatic settings [225,226]. Amplifications of genes at the 11q13 locus has been reported in in about 15% of primary breast cancer cases, and is associated with poor prognosis [208–213]. In our analysis, amplifications of multiple genes at this locus (*CCND1, CTTN, FADD, PAK1, AAMDC, and FGF19*) were also frequent in multiple metastatic sites. Intriguingly, I observed co-amplification of *ERBB3* and *MDM2*, which has not been previously described. Given prior evidence for functional interaction between *ERBB3* signaling and *MDM2* complex formation, co-amplification might be selected for in some tumors. Similar to *ESR1*, some amplifications also showed organ tropism. For example,

brain and GI metastases showed higher *PTK2* and *PKIA* amplifications (~10-20%), while FADD amplifications were more frequent in bone metastases (17%). In our patient-matched samples, most of the CN alterations were maintained in the pairs except for two pairs where there was slight increase in CN for *PKIA*, *PTK2*, and *FGFR4* in the metastases (Figure 32). Moreover, *FGFR1* amplifications were more enriched in ER+ vs ER- brain metastases.

Cyclin-dependent kinases (CDKs), which control transitions through the different stages of cell cycle, have been considered as promising targets for cancer therapy. Amplifications/overexpression of *CCND1* (*CDK4/6* activator) and loss of *CDKN2A* (p16, *CDK4/6* inhibitor) have been described in primary breast cancer patients [111,227–232] . Preclinical and in vitro data across multiple cancers support that loss of *CDK4/6* inhibitory members may serve as biomarkers for *CDK4/6* inhibition sensitivity [114,233–237]. On the other hand, clinical evidence in breast cancer trials performed thus far showed that *CCND1* amplifications or p16 loss were unlikely to predict treatment benefit [238]. However, these trials assessing p16 as a biomarker were earlier phase with small sample size, and didn't show if concurrent alterations in other genes existed, which can alter response to therapy. In our analysis, 32 out of 108 metastases (29%) showed aberrations in the CDKs pathway in general and 24 samples (22%) had alterations in the *CCND1-CDK4/6* axis specifically. Among those, deletions of the *CDK4/6* inhibitors *CDKN2A/B* were significantly mutually exclusive to *CCND1* amplifications, and correlated with mRNA expression, suggesting functional consequences for the alterations. Those deletions were maintained in the paired metastases, which supports the concept of using *CDK4/6* inhibitors in metastatic settings as previously done. Moreover, it is important to note that these deletions also occur in ER- tumors (some of them might also have HER2 amplifications). Most clinical trials have been conducted in patients with ER+/HER2- breast cancer. Our data suggests that testing of

CDK4/6 inhibitors in other subgroups (e.g. ER-) deserve further investigation. Collectively, these data define a subset of metastatic tumors that can be more sensitive to *CDK4/6* inhibition therapy.

3.0 ENRICHMENT OF ESR1 COPY NUMBER AMPLIFICATIONS IN ENDOCRINE RESISTANT INVASIVE LOBULAR CARCINOMA

3.1 INTRODUCTION

As discussed in Chapter 1, ILC is a different histological disease than IDC. Despite clinical and pathological differences, ILC patients are treated as those with IDC, due to the lack of knowledge of targetable pathways underlying the observed differences. Several studies have dedicated efforts to mutational and copy number characterization of breast cancer [239]. However, ILC was clearly underrepresented, which limits our understanding of the genomic alterations that drive this unique subtype of breast cancer. Growing evidence suggests that ILC is enriched for mutations targeting *PTEN, TBX3* and *FOXA1,* in addition to the best known ILC genetic hallmark *CDH1* [16]. Concurrent Loss of *PTEN* and AKT phosphorylation have also been noted as driving pathways. In addition, *ERBB2* and *ERBB3* mutations have been reported in about 3-5% of ILC patients, suggesting a role of the human epidermal growth factor in driving ILC progression [177].

ESR1 gene mutations have been detected at low frequency in primary breast cancers [146–149]. Although ER has been the principal marker for targeted endocrine therapy in breast cancer, copy number alterations of *ESR1* have been understudied, especially in metastatic settings and in unique primary tumor subtypes such as ILC. Comprehensive portrayal of these alterations in metastatic samples has been analyzed and discussed in depth in the previous chapter. This chapter characterize *ESR1* copy number (CN) alterations and their clinical relevance in primary ILC. It also explores copy number changes among other candidate driver genes potentially involved in endocrine resistance.

3.2 MATERIALS AND METHODS

3.2.1 Sample acquisition

Formalin-fixed paraffin-embedded (FFPE) sections were obtain from 71 primary ILC cases (including 18 with recurrences) diagnosed between 1990 and 2011 through University of Pittsburgh Health Sciences Tissue Bank (HSTB). The study was reviewed and approved Institutional Review Board (IRB) at the University of Pittsburgh. All samples were ER+ and diagnosed as primary invasive lobular carcinoma. Samples with synchronous bilateral and/or metastatic breast cancer at the time of diagnosis were excluded. In addition, patients with a history of non-breast malignancy within 5 years prior to ILC were excluded.

3.2.2 DNA isolation

4-6 FFPE sections were utilized for DNA isolation. Guided by H&E slides, those sections were macrodisected to enrich for tumor region and keep cellularity more than 40%. DNA isolations were performed using either Qiagen AllPrep FFPE kit (cat# 80234) or the QIAamp DNA FFPE kit (cat#56404) as per manufacturer's instructions. Isolated DNA was quantified using the ThermoFisher Qubit dsDNA HS/BR kits. 150ng DNA was used in subsequent nanoString analysis.

3.2.3 nanoString copy number analysis

A target panel of 206 hybridization probes covering 67 genes including *ESR1* were used in this analysis. An average of 3 probes per gene were used, except for *ESR1* where I used 10 probes

covering promoter, coding and non-coding exons. Detailed description of this approach has been described in depth in Chapter 2 (section 2.2.6). Raw counts were normalized to 10 invariant reference probes and to reference sample pool of 13 normal breast FFPE samples, using the nSolver™ software. Average CN estimate value was calculated per gene based on all probes for that gene relative to the CN estimate in the normal sample pool. Detailed copy number calls for all genes and samples are illustrated in Data Supplement 2: Table S1.

3.2.4 ESR1 mRNA expression and H-scores

ESR1 mRNA expression data was available through previous nanoString expression study by Sikora *et al*, that comprised 695 genes. Raw counts were normalized to multiple housekeeping genes *(ACTB, RPLP0, GAPDH,* and *GUSB)* using the nSolver™ software. ER histology scores (H-scores) were obtained through curation of patients' clinical data from cancer registry.

3.2.5 Survival analysis

Recurrence free survival (RFS) data was available for 68 out of the 71 collected samples, and was considered the primary end-point (median follow-up of 15.9 years). RFS was measured as the time between date of diagnosis and date of local, regional, or distant recurrence. Samples were grouped based on *ESR1* amplified vs unamplified groups. Log-rank test was implemented to assess significant differences in RFS between the two groups.

3.2.6 HER2 and Ecad/p120 immunohistochemistry

To validate *ERBB2* amplifications findings, 5-um FFPE sections were subject to HER2 and dual Ecad/p120 staining. The staining was performed at the Histology and Micro-imaging Core (HMC) facility at Magee-Women's Research Institute (MWRI) using standard methods. The Ecad/p120 mouse antibodies used were Ventana cat.no: 790-4497 and BD Biosciences cat.no: 610134, respectively. The HER2 antibody was rabbit monoclonal, Ventana, REF #790-2991.

3.2.7 Independent validation of HER2 amplifications and mRNA expression in TCGA data

Datasets of HER2 mRNA expression (n=804) and Reverse phase protein array of 804 and 654 TCGA primary breast tumors were downloaded through cBio portal (Data Supplement 2: Table S2) [175]. Both data were correlated with GISTIC HER2 copy number and IHC calls forming 6 groups per dataset (HER2/AMP, HER2-/AMP, HER2+/Gain, HER2-, HER2+/Diploid, and HER2+/Del). The GISTIC algorithm groups copy number status into four levels with -2 as deep loss, -1= shallow loss, 0=diploid, 1= gain, and 2= high-level amplification. In addition, these groups were investigated for enrichment of HER2 PAM50 subtype calls generated from RNA-seq data using the *genefu* package (Data Supplement 2: Table S3) [240].

3.2.8 Statistical and bioinformatics analysis

ComplexHeatmap R package was used to generate oncoprints for copy number alterations [198]. Chi square test with p value significance cutoff of 0.05 was used to test the enrichment of *ESR1*

amplifications in samples with recurrence. Wilcoxon rank sum test was performed for boxplots comparisons with p value significance cutoff of 0.05. CN increase by 35% (CN ≥2.7) and decrease by 50% (CN ≤ 1) were considered gains/amplifications and deletions, respectively. Copy number calls above ≥10 were considered high amplifications. Giving lower resolution for copy number amplifications of genes other than *ESR1,* I used one copy number cutoff of ≥5 as amplification.

3.3 RESULTS

3.3.1 Frequent *ESR1* amplifications in ILC

For high resolution of copy number detection, *ESR1* was targeted with 10 probes covering non-coding exons, promoter, and coding exons. I measured *ESR1* CN alterations in 71 primary ILC samples, and reported the average copy number call of all probes. About 24% (n= 17) of primary ILC samples showed gains and amplifications. Particularly, ten samples (14%) had copy number gains, while seven samples (10%) showed high amplifications (Figure 9A). There were no *ESR1* deletions in any of the samples.

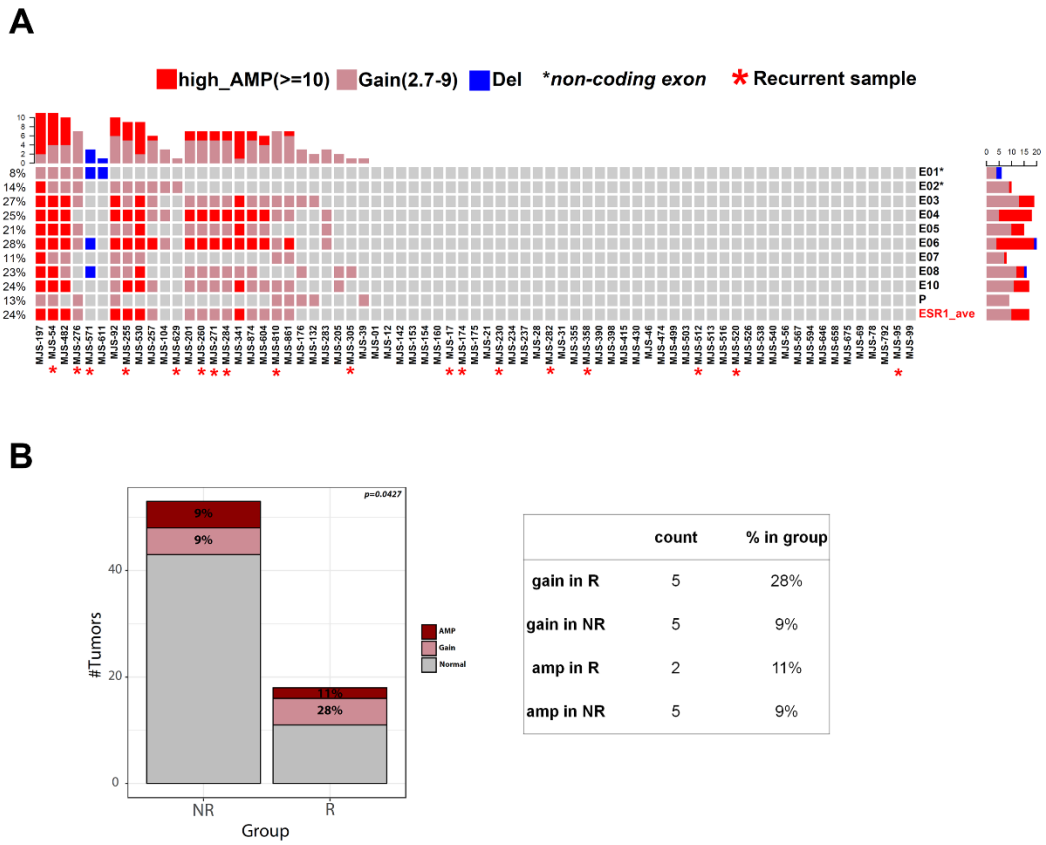


Figure 9: Frequent *ESR1* gains and amplifications in endocrine resistant ILC
A. Oncoprint visualization of *ESR1* copy number alterations in primary ILC samples. Levels of amplification and deletion are color coded. Each column represents a single sample. Each row indicates copy number call of the correspondent single exon probe. Samples with recurrence are annotated with '*' red symbol. P; Promoter probe. ESR1_ave: average copy number call of all probes. **B.** Frequency of *ESR1* copy number amplifications for the non-recurrent (NR) and recurrent groups (Chi-square test 0.0427).

Since 18 out the 71 primary ILCs showed either local (n=6) or distant (n=12) recurrence, I was able to test if the *ESR1* amplifications are enriched in patients with recurrence. 39% (n=7) of the group with recurrence showed amplifications while only 19% (n=10) of non-recurrent group presented with amplifications (Figure 9B, p value= 0.0427). Overall, the results demonstrate frequent *ESR1* amplifications in primary ILC especially in samples with recurrence, suggesting a role in endocrine resistance.

3.3.2 Association of *ESR1* amplifications with clinical outcome

Given that I detected significant enrichment of *ESR1* amplifications in patients with recurrence (Chapter 2), I wanted to see if those patients confer worse clinical outcome. Recurrence-free survival (RFS) data was available for 68 out of the 71 patients. As expected, the patients with copy number amplifications conferred worse RFS outcomes, although didn't reach statistical significance (Figure 10A, p value= 0.2). This likely due to the limited number of samples in this analysis, especially in the amplifications group.

3.3.3 Correlation of *ESR1* CNAs with mRNA expression and H-score

Giving that copy number amplifications might potentially have an impact on gene expression, I tested if the *ESR1* amplifications are correlated with higher mRNA expression. nanoString expression data from previous study (Tasdemir, Sikora *et al*, manuscript in preparation) was available for all samples. Indeed, patients with *ESR1* amplifications showed significantly higher mRNA expression compared to the group without amplifications (Figure 10B, p value= 0.001).

Next, I sought to assess if *ESR1* amplifications are also associated with higher protein expression. ER histological scores (H-scores) was available for a subset of 53 samples. Although H-scores were slight higher in the group with amplifications, the association did not reach statistical significance (Figure 10B). Of note, the median H-score for all 53 samples was 244, suggesting involvement of other mechanisms in regulation of ER protein expression.

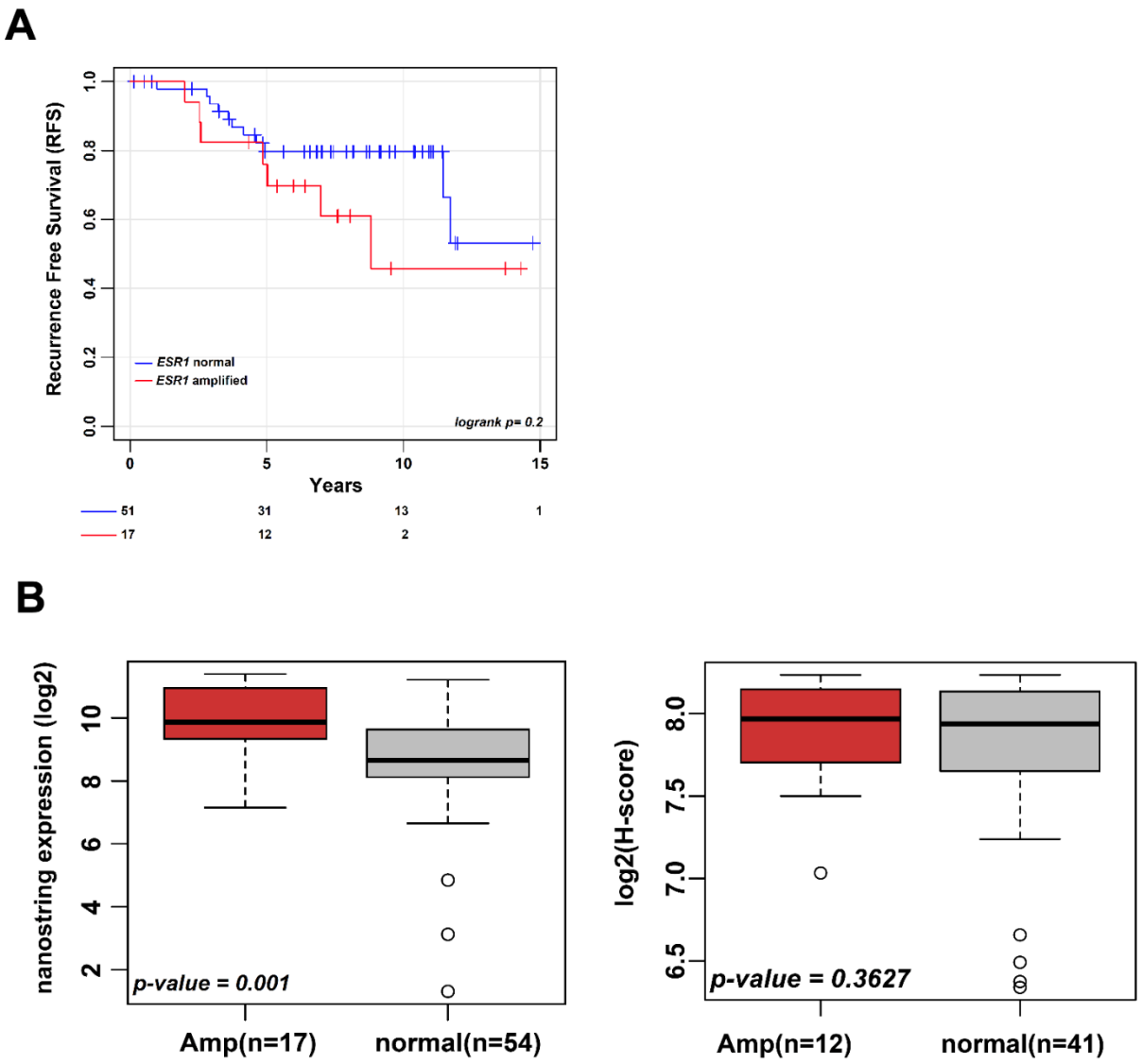


Figure 10: correlation of *ESR1* amplifications with RNA and protein expression, and clinical outcome
A. Kaplan–Meier curves of recurrence-free survival (RFS) analysis for the *ESR1* amplified (red) vs unamplified (blue) groups (p value indicated for log-rank test). **B.** Box plot comparison of *ESR1* mRNA (left) and protein (right) expression between normal vs gain/amplification groups (Wilcoxon rank-sum test).

3.3.4 CN alterations of other breast cancer driver genes

Our probe design was extended to include additional 66 genes that are previously reported in breast cancer progression (Data Supplement 2: Table S1). The most frequently altered genes are

illustrated in Figure 11A. Most amplified genes included *CCND1(32%), ERBB2 (18%), MDM4 (17%), and MYC (17%)*. mRNA expression was strongly correlated with these amplifications except for *MYC* (Figure 11B).

Analysis of copy number alterations revealed infrequent deletions (Figure 34). The most frequently deleted gene was *NCOR2* in about 7% (n=5) of samples. All patients with this deletion were not recurrent, which align with the previously reported role as independent indicator of poor outcome [241].

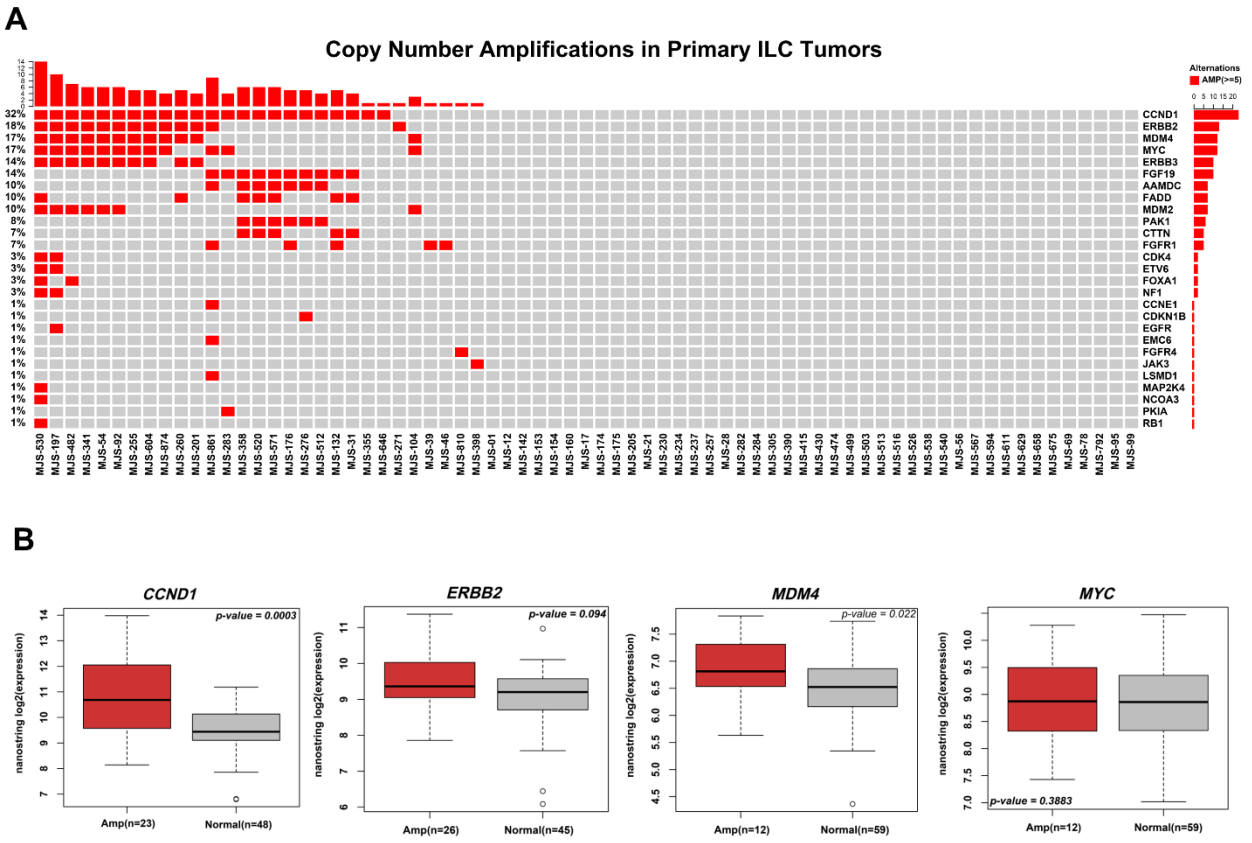


Figure 11: Most frequently amplified genes in primary ILC samples
A. Oncoprint visualization of copy number amplifications by genes (rows) and samples (columns). Frequencies of the amplifications are indicated to the left side of each row. Top and right-side bars show counts for alteration events within a sample and by gene, respectively. **B.** Correlation of most frequently amplified genes in the ILC cohort. Correlation of most frequently amplified genes in the ILC cohort. Y-axis: nanoString mRNA expression (log2).

Unexpectedly, the frequency of *ERBB2* amplifications was higher than previously reported, suggesting a novel underestimated role of the ERBB pathway in ILC progression. Captivated by this finding, I did HER2 IHC staining for six samples to evaluate protein expression. Interestingly, these samples showed small subpopulations of HER+ cells. I further stained these samples with Ecad/p120 to confirm that these are epithelial cells. Indeed, the staining of these markers overlapped perfectly (Figures 12 and 32). Of note, these tumors would have been called HER2- clinically.

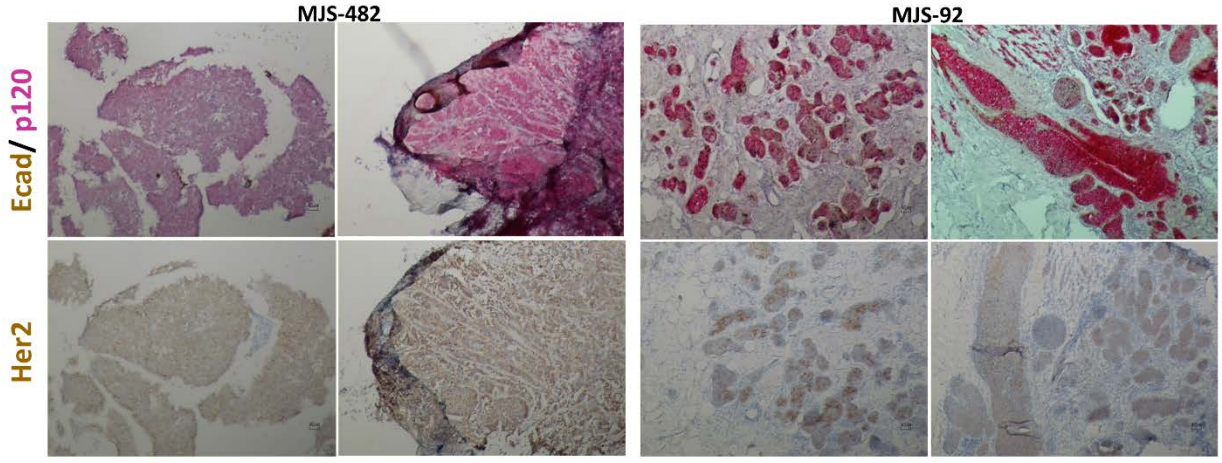


Figure 12: IHC validation for HER2 amplified ILC samples
Staining characterization of two ILC cases with Ecad/p120 (top panel) and HER2 (bottom panel) immunohistochemistry. Two fields per sample are shown.

3.3.5 Validation of underrepresented HER2 amplifications in TCGA dataset

Intrigued by the HER2 findings, I sought to investigate further how common patients with *ERBB2* amplifications can be HER2- by IHC using TCGA data. Surprisingly, about 25% (n=26) of the *ERBB2* amplified samples were not identified as HER2+ by IHC (Figure 13A). Correlation with RNA-seq and RPPA data showed higher mRNA and protein expression for this group (IHC HER2-

but *ERBB2* amplified) compared to samples without amplifications (Figure 13B). I further interrogated the RNA-seq data to assess the possibility of molecular enrichment for the HER+ PAM50 subtype. Our analysis revealed that the HER+ signature is enriched in about 15% of the HER2-/AMP group (Figure 13C). As expected, the HER2+/AMP group showed the highest enrichment (41%) while the HER2 unamplified groups didn't show any enrichment.

In summary, these results propose a subset of HER+ tumors with functional expression that are usually missed by conventional techniques.

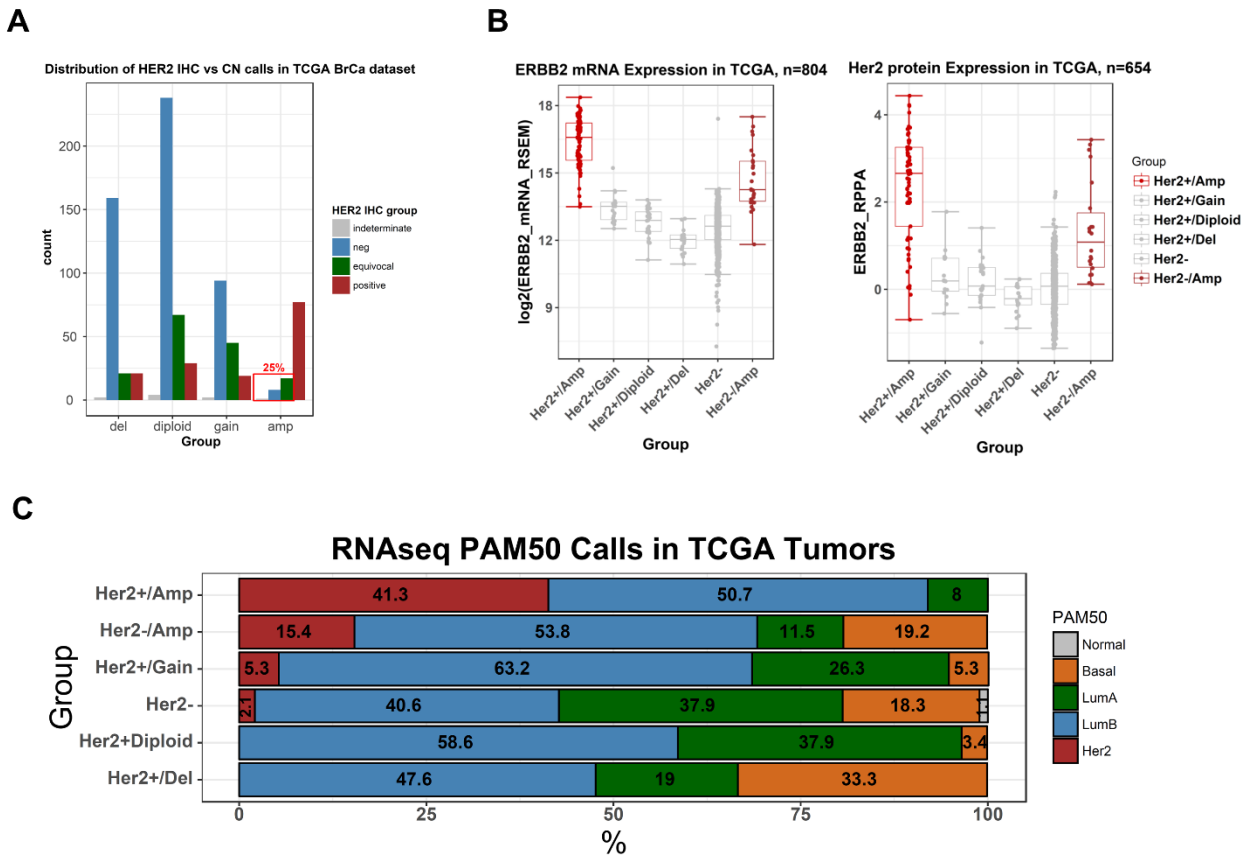


Figure 13: validation of unique HER2 amplifications in TCGA dataset
A. Distribution of HER2 IHC scores (color coded) grouped by DNA copy number status in TCGA breast cancer samples. The group with interest (IHC-/AMP+) is highlighted with red square. **B.** Correlation of each HER2 IHC/copy number group with mRNA (left) and protein (right) expression, from RNA-seq and RPPA data, respectively. **C.** PAM50 molecular predictions (RNA-seq data) for each HER2 IHC/copy number group. Each PAM50 subtype is color coded and the percentages for enrichments are indicated. Additional details are illustrated in Data Supplement 2.

3.4 DISCUSSION

ILC is a unique disease associated with some unique pathological, clinical, and molecular features compared to IDC. Although it represents a significant proportion of breast cancer patients, there is no special treatment recommendations for this subtype. The molecular characterization of ILC has been increasing in the last few years, with the goal to identify unique pathways that could drive development of new therapeutic interventions for patients with ILC. These analyses suggested that although there is big overlap with IDC at the molecular level, ILC has fairly a different mutational and genomic landscape than that of IDC, as evidenced by high prevalence of alterations in the PI3K pathway and unique *FOXA1* mutations [16,177].

Although *ESR1* has been an excellent endocrine therapeutic target for decades, there is no clear agreement on the clinical relevance of copy number (CN) changes associated with expression of this gene, and with response to endocrine treatment. Factors such as sample processing, tumor heterogeneity, and sensitivity of the technology used have contributed to inconsistency in the frequency of *ESR1* amplifications reported in literature. In addition, most of the previous CN studies were performed on primary tumors that are enriched for the IDC subtype. As a result, characterization of CN changes in unique subtypes such us ILC has been understudied.

Our CN characterization of ILC revealed frequent *ESR1* amplifications in about one-fourth of the cases (24%). As expected, these amplifications were associated with significantly higher mRNA expression. More importantly, *ESR1* amplifications were significantly enriched in samples from patients with tumor recurrences, and showed strong trend towards association with worse recurrent-free survival, supporting the association with worse overall survival discussed in Chapter 2. Of note, our findings are in agreement with the recently reported *ESR1* amplifications in 25% of ILC, association with mRNA expression, and enrichment in aggressive ILC tumors [177]. In that

study, the alterations were significantly enriched in ILC vs IDC, suggesting a unique role in modulating response to E2 and endocrine therapy. Collectively, these findings support further clinical investigation in the context of endocrine therapy, such as increased sensitivity to estradiol therapy described by us and others [224,242].

In addition to *ESR1*, our analysis validated previously reported ILC amplifications in 11q13 (*CCND1*, 32%) and 8q24 (*MYC*, 17%) [16,177,243]. Furthermore, I observed frequent *ERBB2* and *MDM4* amplifications in 18% and 17% of the cases, respectively. To our knowledge, these represent novel findings and have not been reported with such high frequencies. Interestingly, our IHC evaluation of 6 samples (clinically HER2-) with *ERBB2* amplifications revealed small subpopulations with HER2+ staining.

Determination of the best biomarker for response to HER2 directed therapy is a controversial topic that continues to evolve. For example, although a sub-analysis of IHC HER2- cases in the NSABP B31 trial showed benefit to HER2 treatment, there was no benefit for this group in the more recent comprehensive NSABP B47 trial that ruled out samples with DNA amplifications [244,245]. IHC evaluation of HER2 became routinely in use by the late 1990s and its scoring method kept evolving since then [246]. Although it is considered a good predictive marker, IHC alone does not address the benefit of HER2 treatment in HER2 IHC- patients but with *ERBB2* amplifications, as they will be called HER-. Thus, ILC alone might not be enough to identify HER+ samples. Indeed, it has been shown that patients with *ERBB2* amplifications, despite being IHC-, had worse overall survival compared to IHC- patients without amplifications [247]. Our analysis of TCGA data showed frequent occurrences of HER2 IHC- but *ERBB2* amplified cases, representing 25% of all amplified cases. This subgroup clearly showed higher mRNA and protein expressions compared to the groups without amplifications. In addition, the same group showed molecular

enrichment for the HER2 PAM50 predictions compared to the unamplified groups. It is possible that factors such as sample processing, poor fixation, variation in antigen retrieval techniques, antibody sensitivity and specificity, and variation in scoring methods can contribute to weak or false negative HER2 signals [248]. Our results support further exploration for the benefit of HER2 therapy in this subgroup.

4.0 COMPREHENSIVE MOLECULAR AND CLINICOPATHLOGICAL LANDSCAPE OF UNIQUE INVASIVE LOBULAR CARCINOMA METASTSIS TO THE OVARIES

4.1 INTRODUCTION

Invasive lobular carcinoma (ILC) is the second most common breast cancer histological subtype, accounting for approximately 10-15% of the cases [4,5]. Although it accounts for 10% of breast cancer, ILC incidence is twice that of invasive cervical cancer and equivalent to that of ovarian cancer [249]. Both ILC and invasive ductal carcinoma (the more common histological subtype) can metastasize to common sites, such as liver, bone, lung, and brain. In 1984, however, Harris *et al* suggested that ILC behaves differently from IDC with regard to metastatic spread [164]. In their observation, it was clear that ILC tend to metastasize more frequently to the gastrointestinal tract, peritoneum, and urogenital organs (e.g. ovaries). Although these differences well indicate important biological differences between the two subtypes, these findings were not investigated further until two decades later. In 2004, Arpino *et al* undertook comprehensive metastasis and clinical outcome comparison of around 4000 ILC and 45,000 IDC patients with median follow-up period of 87 months [161]. In their analysis, ILC was three times more likely to spread to ovaries, peritoneum, and gastrointestinal system (6.7% vs 1.8% in IDC). Subsequently, two more studies by Ferlicot *et al* and He *et al* validated and confirmed these findings, supporting a different molecular biology of ILC [165,166]. Moreover, our analysis of Magee-Women's Hospital (MWH) of UPMC registries from 1994-2014 revealed significant ILC metastasis to ovaries when compared to IDC (Table 5). These findings, in addition to accessibility to tissue materials through The Health

Science Tissue Bank (HSTB) at UPMC, have motivated us to investigate and characterize further the unique pattern of ILC metastasis to the ovaries.

Table 5: Breast cancer metastasis to the ovaries and peritoneum (MWH database from 1990-2014).

Subtype	No. of primaries	% met to peritoneum	% met to ovaries	% met to ovary and peritoneum
ILC	1384	1.16*	2.96*	4.12*
IDC/ILC	822	0.36	2.43	2.8*
IDC	12121	0.17	1.23	1.4

*p-value vs IDC < 0.0001

ILC, invasive lobular carcinoma; IDC, invasive ductal carcinoma

4.2 MATERIALS AND METHODS

4.2.1 Collection of clinical samples

Collection of 41 samples (13 primary, 28 metastases, 6 normal breast, 4 normal ovaries) was performed in collaboration with Health Sciences Tissue Bank (HSTB) at the pathology department at MWH under Institutional Review Board (IRB) guidelines. 14 ovarian metastases were matched with their 13 primary tumors (two metastases shared the same primary), while 14 samples were unmatched ("orphan" samples). Clincopathological features of the patients/tumors are illustrated in Table 13 and Data Supplement 3: Table S1. In addition, 6 normal breast and 4 normal ovarian samples were collected for comparison and quality control in the subsequent analyses. Analysis of histological classification was performed by two pathologists (PL and EE).

4.2.2 FFPE processing and nucleic acid isolations

Overview of the workflow of formalin-fixed paraffin-embedded (FFPE) sample processing is illustrated in Figure 35. RNA and DNA were isolated from 3-6 FFPE macrodisected sections per sample depending on tumor size and cellularity using the Qiagen AllPrep FFPE kit as per manufacturer's instructions. RNA and DNA concentrations were determined with the Qubit 3.0 Fluorometer (ThermoFisher Scientific). RNA fragment sizes distributions (DV200 metrics) was obtained utilizing either the Agilent 2100 Bioanalyzer or the Agilent 4200 TapeStation. Distribution of the samples based on nucleic acid yields and tumor cellularity is shown in Figure 36 and Table 14, respectively.

4.2.3 *In vitro* migration assay to the ovaries

Collection of mice ovaries from FVB/N mice strain (10 weeks of age) was performed in collaboration with Tanya Minteer, Susan Farabaugh PhD, and Rebecca Watters PhD from the Lee and Oesterreich groups. Cell lines were subject to overnight serum starvation prior to running the assay. 500 μL of media containing 10% fetal bovine serum or desired attractant (e.g. minced ovaries) were added to the lower well of the migration plate. 300 μL of the cell suspension solution (150,000-300,000 cells in serum-free media) were added to the inside of each insert. Cells were incubated at 37Cº with 5% CO2 for 72 hours. Inserts were then washed 2X with PBS, air dried, and stained with 400 μL 0.1% crystal violet for 15min at RT. Inserts were washed 3X with water, dried on bench, and destained using 300 μL 10% acetic acid for 10min. In the final step, 200 μL of the stained cells were transferred to 96-well plate and read in spectrophotometer at 590nm. Workflow of the migration assay is illustrated in Figure 37.

4.2.4 Immunohistochemistry (IHC)

Samples were subject hematoxylin and eosin (H&E), Ecad/p-120, and ER staining to confirm the characteristic growth patterns of the ILC and IDC subtypes. The staining was performed at the Histology and Micro-imaging Core (HMC) facility at Magee-Women's Research Institute (MWRI) using standard methods. Dual Ecad/p120 mouse antibodies used were Ventana cat.no: 790-4497 and BD Biosciences cat.no: 610134, respectively. ER IHC was performed using the mouse monoclonal antibody: Novacastra Leica, cat.no: NCL-L-ER-6F11. Characteristic growth and staining patterns for each subtype were evaluated by two pathologists (PL and EE).

4.2.5 Exome-capture library preparation and RNA sequencing

Library preparation using TruSeq RNA Access library preparation (Illumina) and sequencing was done at the Genomic Core Facility at Children's Hospital of UPMC. RNA samples were sonication-fragmented, followed by cDNA synthesis, adapters ligation, and PCR amplification. Lastly, biotinylated probes were hybridized to the target regions (exome), followed by capture step using streptavidin beads, then elution of the beads. NextSeq500 platform was used to produce 2X75bp paired-end reads with target reads of 40-50 million. General overview of subsequent RNA-seq analyses workflow is described in Figure 38.

4.2.6 Processing of raw data, gene expression quantification, and normalization

Gene transcripts quantification from RNA-seq data was performed from FASTQ files using the quasi-mapping-based mode of *Salmon v0.8.2* and Ensembl GRCh38 v82 transcripts annotation

GTF file [250]. Gene counts from Salmon were then converted to counts per million normalized using trimmed-means of M-values (TMM-normalized CPM) with *edgeR* package [194]. Log2 transformed CPM were used for plotting in subsequent analyses. To produce binary alignment map (BAM) files, two-pass *STAR v2.4* package was used in Linux-based environment [251]. FASTQ reads were aligned to the genome using locally generated genome indexes produce by the same package. Detailed RNA-seq information about read counts and mapping rates are available in Data Supplement 3: Table S2.

4.2.7 Quality assessment of RNA sequencing data, principal component analysis, and molecular validation of paired samples

Quality control assessment of RNA-seq data was performed using *QoRTs v1.1.8* package using gene mapping rates and reads as inputs [252]. Detailed RNA-seq QC metrics such as base-quality score, clipping rate, gene-body coverage are illustrated in Figure 39 and Data Supplement 3: Table S2. To broadly explore transcriptomic difference between the samples, principal component analysis (PCA) was performed using the top 5000 most variable genes defined by interquartile range (IQR) in R environment (Figure 40) [197]. Outlier samples were excluded from further differential expression analysis. Paired samples were molecularly validated from RNA-seq data using the tumorMatch custom algorithm developed by Priedigkeit *et al* then plotted using *corrplot* package in R (Figure 41) [253,254].

4.2.8 Intrinsic PAM50 molecular subtyping analysis

Log2 TMM-normalized CPM values were used for PAM50 predictions using the *genefu* package in R [240]. Giving the sensitivity of this algorithm to ER status (test-set bias), a cohort of 20 balanced ER+ and ER- samples as primary cohort, then query samples of unknown molecular subtype were added. Each sample was tested 20 times and the mode was used as the final discrete PAM50 subtype. The average probability of all 20 testes was used as the final probability score. Visualization of the resulted PAM50 calls and scores was performed using the *ggplot2* package in R.

4.2.9 Differential gene expression analysis

Gene counts produced by *Salmon* were used as input for differential expression analysis between the primary and metastatic samples, using the *DESeq2* package [193]. The analysis was done twice using both the paired (for matched primary-metastasis) and unpaired mode (for all samples) of the package. Then, each analysis was compared to differentially expressed genes between normal ovaries vs primary tumors to correct for noise by site-specific genes (Figure 42). Genes with log2 fold change below 0.5 in each comparison were considered potential ovary-specific and were excluded. I used a cutoff of 2-fold change and adjusted p-value <0.05 for genes to be significantly up or down regulated (Data Supplement 3: Table S3). A heatmap of hierarchical clustering of the differentially expressed genes was generated using the *heatmap.3* function (https://github.com/obigriffith/biostar-tutorials/blob/master/Heatmaps/heatmap.3.R). Distance measurements of 1- Pearson correlation and the 'average' clustering agglomeration method were also implemented.

4.2.10 ConsensusPathDB (CPDB) pathway analysis

Gene set pathway over-representation analysis was performed in CPDB using interactions of 32 databases [255]. Enriched pathway-based sets parameters included a minimum of two input genes and hypergeometric p-value cutoff of 0.01 for the overlap of each predefined pathway set with the input gene list. Significantly upregulated pathways were plotted using the ggplot2 package in R [199].

4.2.11 Analysis of clinically actionable transcriptomic alterations

Data set of 399 clinically actionable genes was obtained from the Drug-Gene Interaction Database (DGIdb) [191,256]. Only genes with CPM value above one in at least two samples were included. In addition, genes due to site-specific contamination were excluded by comparison to highly expressed genes in normal ovaries vs primary tumors. As a result, 73 genes made it to the final analysis (Data Supplement 3: Table S4). Only genes with at least 4-fold change difference between paired metastasis vs its primary tumor were reported. Genes with changes in at least two pairs were plotted using the *oncoprint* function in *ComplexHeatmap* package [198].

4.2.12 MYCN survival analysis

MYCN gene expression and disease-specific survival (DSS) were obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) through Synapse software platform (syn1688369; Sage Bionetworks, Seattle, WA, USA). ER+ Patients were classified base on MYCN expression as upper quartile (>q3) vs lower quartile (<q3) groups. Additional analysis

of ILC-only and IDC-only cohorts was also performed. Data acquisition, analysis, and visualization was completed in collaboration with Kevin Levine.

4.2.13 MammaSeq DNA Library Preparation and Ion Torrent Sequencing

Targeted sequencing was performed using the MammaSeq™ panel (Thermo Fisher Scientific) which comprises 78 genes, with 688 amplicons targeting 1398 mutations (Data Supplement 3: Table S5) (Smith *et al*, manuscript under review). 25-100 ng DNA was used per each amplicon pool for library preparation of FFPE samples. Library preparation was performed using the Ion AmpliSeq™ Library Kit 2.0 and Ion Xpress™ Barcode Adapters. Sequencing templates were prepared and enriched with Ion PI™ Hi-Q™ OT2 200 Kit on the Ion OneTouch 2 System. Sequencing was performed on the Ion Proton using the Ion PI™ Hi-Q™ Sequencing 200 Kit and Ion PI™ Chip Kit v3. FFPE DNA samples were sequenced into separate PI chips (60 million reads) at 1600x empirical depth.

4.2.14 Single nucleotide variant analysis

Raw fastq files were aligned to the hg19 reference genome using CLC Genomics Workbench 11 (https://www.qiagenbioinformatics.com/) to generate VCF files using the specified parameters: minimum coverage= 10, minimum variant read count= 2, min allele frequency= 5%, base quality filter (central =20, neighbor=15, neighborhood radius=5), remove pyro-error variants (yes, length= 3, frequency= 0.8), minimum base quality= 20, minimum F/R strand bias= 0.02. Further variants annotation was performed using the Cravat CHASM-v4.3 tool [257]. Removal of germline and common SNPs (ExAC, common dbSNP, and 100Genomes, and synonymous variants using CLC

Genomics Workbench 11 (https://www.qiagenbioinformatics.com/) [258–260]. Additional data filtering, organization, and visualization was generated in R environment using Maftools and ComplexHeatmap packages [198,261].

4.3 RESULTS

4.3.1 Clinicopathological characterization of primary tumors and ovarian metastasis

An overview of the number of clinical samples stratified by subtype and site is illustrated in Table 15. Ecad/p120 IHC staining confirmed the loss of E-cadherin in the ILC subtype and it was maintained in the metastatic samples (Figure 14). Interestingly, clinicopathological analysis of 12 matched primary-metastasis pairs revealed median age at diagnosis of 40.5 years, which was also confirmed by enrichment of premenopausal status in 75% of the patients (Table 6). Comparison with independent cohorts of patients with brain (n=21) and bone (n=11) metastases showed significantly younger age at the diagnosis in the ovarian cohort (Data Supplement 3: Tables S6-7, $p < 0.05$).

ER and PR were frequently expressed in 83% and 75% of the ovarian metastatic patients respectively, while HER2 was expressed in only one metastatic sample (8%). As expected, E-cad was lost in all ILC samples but expressed in all IDCs. Most the patients (66%) were with pathological stage II disease. The median survival post metastasis (SPM) and overall survival (OS) were 29 and 86 months, respectively. Additional clinical characteristics by patient are explained more in depth in Table 13 and Data Supplement 3: Table S1.

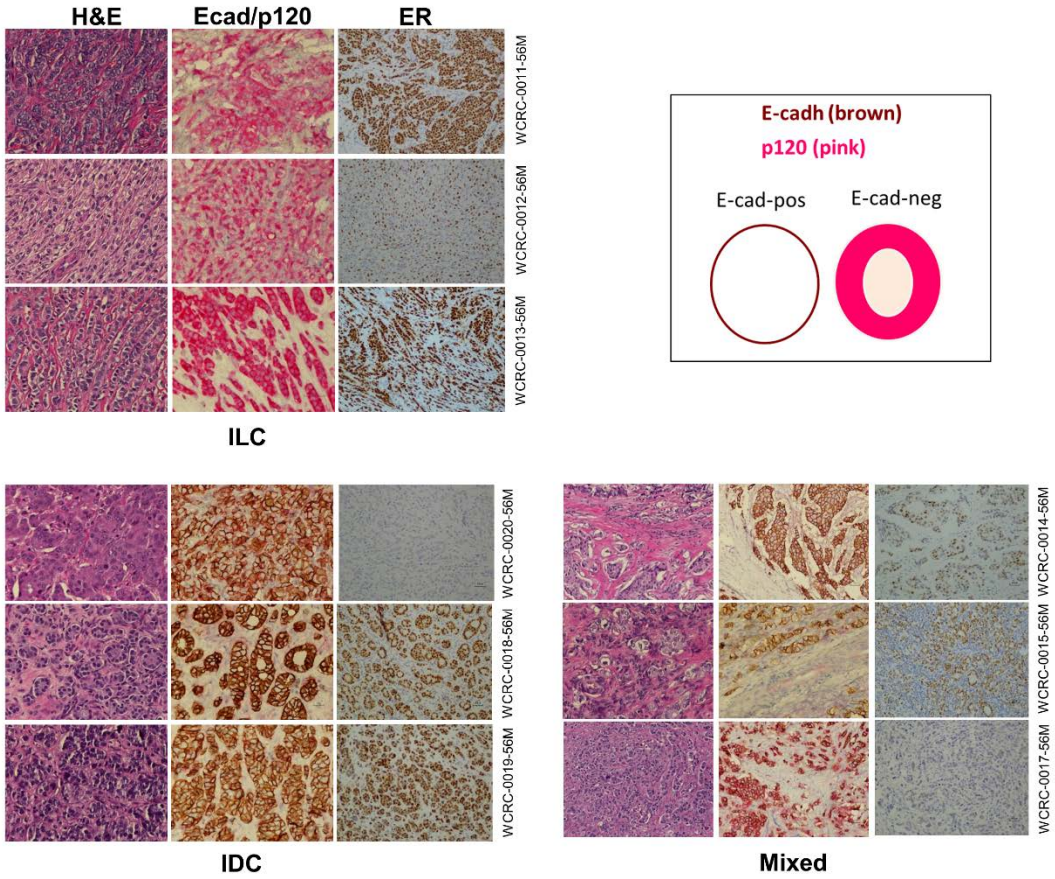


Figure 14: Immunohistochemical characterization of the ovarian metastases
H&E, dual Ecad/p-120, and ER staining for 9 cases (3 for each breast cancer subtype) are shown. ILC metastases maintain the E-cadherin loss and therefore infiltrate the breast stroma in single files. E-cadherin loss triggers relocalization of p-120 to the cytoplasm (pink color). IDC express E-cadherin and tend to form glandular structures.

Table 6: Clinicopathological features of the ovarian metastatic samples

Age		value			
	Median	40.5			
	Mean	42.9			
	Range	(31-60)			
Histological subtype		n	%		
	ILC	6	50		
	IDC	4	33		
	Mixed	2	17		
ER status		n	%		
	Pos	10	83		
	Neg	2	17		
PR status		n	%		
	Pos	9	75		
	Neg	3	25		
Her2 status		n	%		
	Pos	1	8		
	Neg	11	92		
E-cad status		n(ILC)	%(ILC)	n(IDC)	%(IDC)
	Pos	0	0	4	100
	Neg	6	100	0	0
Menopausal status at Dx		n	%		
	Pre	9	75		
	Post	3	25		
Pathological stage		n	%		
	IA	1	8		
	IB	1	8		
	IIA	4	33		
	IIB	4	33		
	IIIC	2	17		
Adj endocrine Tx		n	%		
	Yes	10	83		
	No	2	17		
Survival (months)		DFS	MFS	SPM	OS
	Median	26	49.5	29	86
	Range	(19-169)	(19-169)	(4-86)	(24-192

Abbreviations: ILC, invasive lobular carcinoma; IDC, invasive ductal carcinoma; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; E-cad, ecadherin; Dx, diagnosis; Adj Tx, adjuvant therapy; DFS, disease-free survival; MFS, metastasis-free survival; SPM, survival post metastasis; OS, overall survival.

4.3.2 *In vitro* mimicry of ILC metastasis to the ovaries

Mimicking ILC metastasis to the ovaries, I tested migration of the ILC cell lines MDA-MB-134 and SUM44PE to mice ovaries using Boyden chamber transwell approach. The ovaries used were from 10-week old mice, imitating the enrichment of premenopausal status in patients with metastasis to the ovaries. Interestingly, the results showed significant or trend of migration of the ILC cell lines MDA-MB-134 and SUM44PE toward the ovaries, respectively, while the IDC cell line MCF7 did not show that phenotype (Figure 15). This migration was also higher than the migration to FBS which was used as positive control.

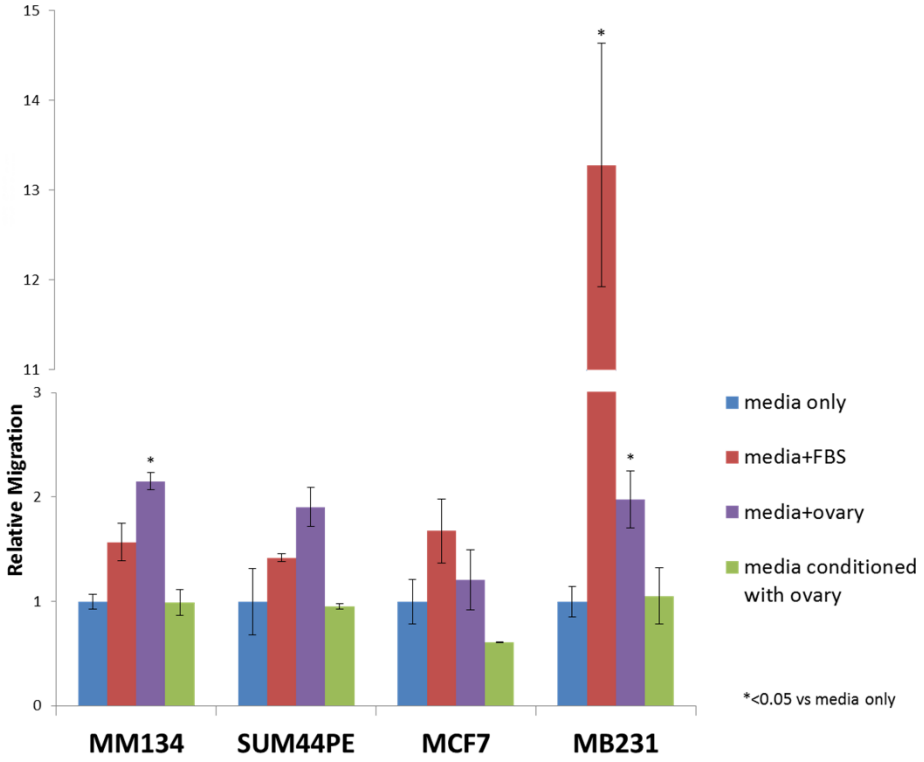


Figure 15: Migration of breast cancer cell lines to mice ovaries
Migration behavior of ILC cell lines (MDA-MB-134 and SUM44PE), IDC cell lines (MCF7), and the positive control MDA-MB-231 cell lines toward different chemoattractants (color coded). The Y-axis represents relative migration to migration media-only condition. The experiment was repeated 3 times but is not representative of all of them. This is because factors such as incubation time, use of fresh ovaries, and mice/age were changed each time to optimize the assay.

4.3.3 Intrinsic PAM50 molecular subtyping of the primary-metastasis matched pairs

Given potential shifts in the PAM50 classification between the primary tumors and their matched metastases during tumor progression, I wanted to explore such differences. 10 out of 13 (77%) pairs maintained similar identical subtype (Figure 16). Of these, nine (90%) pairs were LumA, while only one pair was with the Her2 subtype. Out of the three pairs with shifts, one pair switched from LumA to LumB, one from Basal to Normal, and one from LumB to LumA. However, a closer look at the probability scores within these pairs showed a noteworthy shared proportion of an identical subtype.

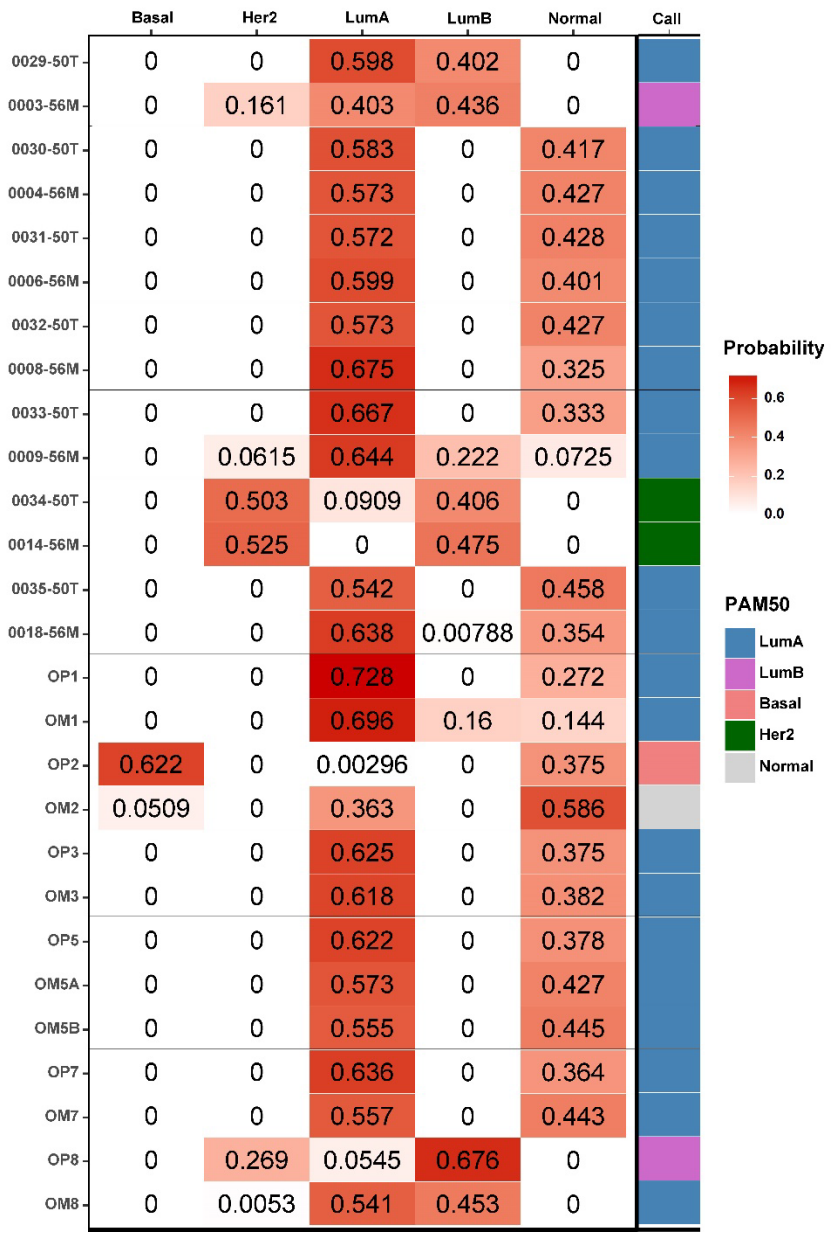


Figure 16: PAM50 predictions in primary-matched ovarian metastatic pairs
PAM50 intrinsic molecular subtype calls from RNA-seq data of primary-matched ovarian metastases. Each molecular subtype is color coded. Probabilities for each subtype (column) are indicated for each sample (row). Definitive calls are listed in the right side of the figure. Sample pairs are separated by grey horizontal lines.

4.3.4 Differentially expressed genes and enriched pathways

To determine induced or downregulated genes in the metastases compared to primary tumors, I used a novel approach that takes into account paired status as well as potential impurity by site-specific genes (Figure 42). Our analysis revealed 235 differentially expressed genes with FDR-adjusted p value < 0.05 and fold-change ≥ 2 (Data Supplement 3: Table S3). Clustering by this gene list (71 with increased and 164 with decreased expression in metastasis) and separation between the primary and metastasis groups is illustrated in Figure 17A. The same list was also subject to pathway analysis using the ConsensusPathDB (CPDB) tool. Most notable upregulated pathways included WNT, metabotropic glutamate, and MAPK/ERK pathways (Figure 17B and Data Supplement 3: Table S8). These pathways appeared to be driven mainly by the calcium-sensing receptor *(CASR)* gene, Glutamate receptor metabotropic 7 and 8 *(GRM7 and GRM8),* and Wnt Family Member 5A *(WNT5A).* Expression analysis comparisons of these genes in 3 independent cohorts (brain, bone, and GI metastases) suggested that these changes are ovarian-metastasis-specific (Figure 18). Given that there is some literature evidence that suggest CASR is involved in WNT5A upregulation, I carried out multiple correlation testing of these two genes and other candidates. Indeed, in addition to being highly correlated together, the expression of the two was highly correlated with the expression of *GRM7*, *GRM8,* and the potential downstream effector of *WNTA5A*, β-catenin *(CTNNB1)* (Figures 19, 42, and Data Supplement 3: Table S10). On the other hand, the most notably downregulated pathways were generally driven by the Interleukin family member 6 (IL6) gene (Data Supplement 3: Table S9).

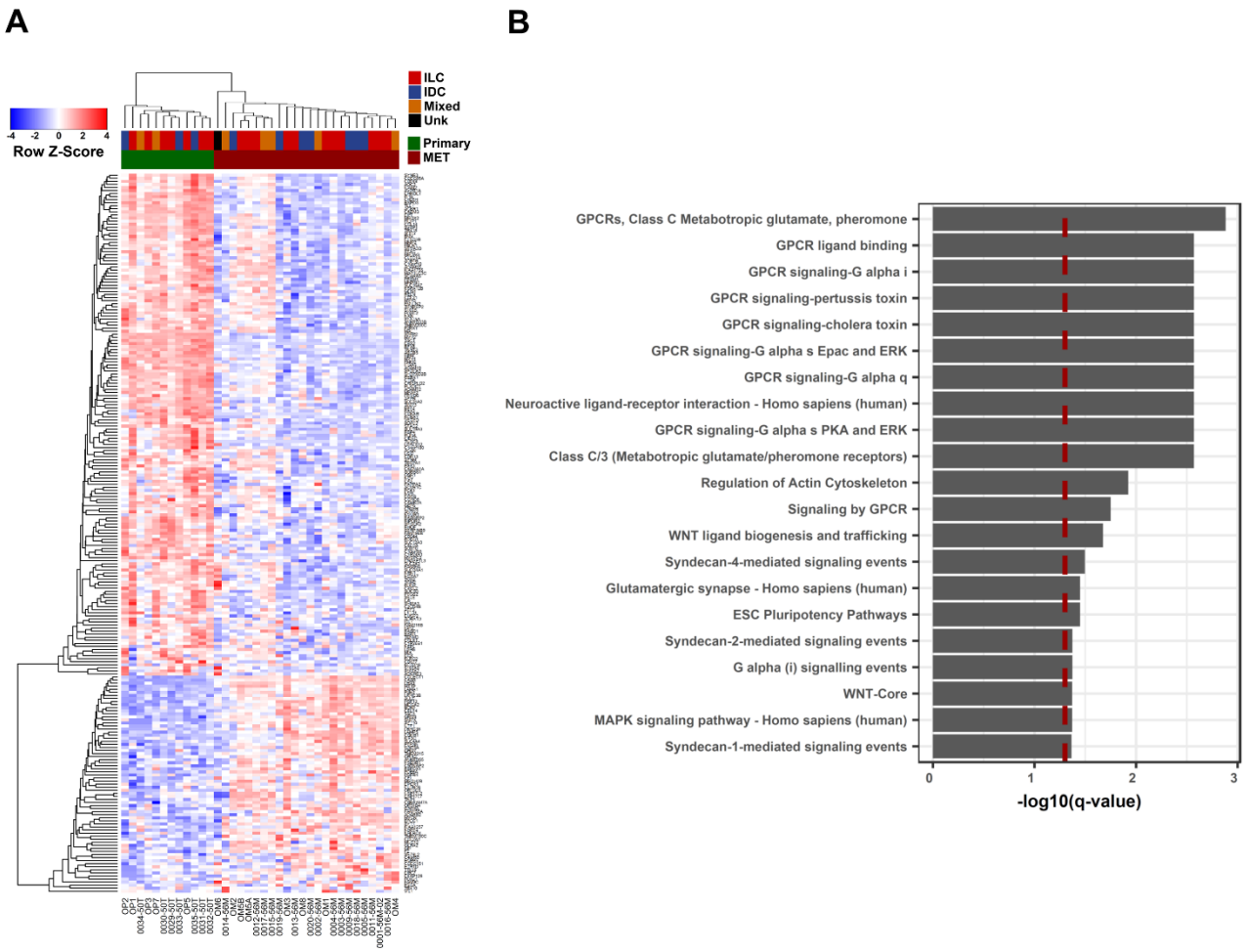


Figure 17: Differentially expressed genes and enriched pathways in the ovarian metastases
A. Unsupervised clustering heatmap of primary and ovarian metastases using the 235 differentially expressed genes. Site (primary/met) and histological subtypes (ILC, IDC, or mixed) are color coded for each sample (column). **B.** Significantly upregulated pathways in the ovarian metastases. Corrected p value (q value) is indicated for each pathway. The vertical dashed red line represents a q value cutoff of 0.05.

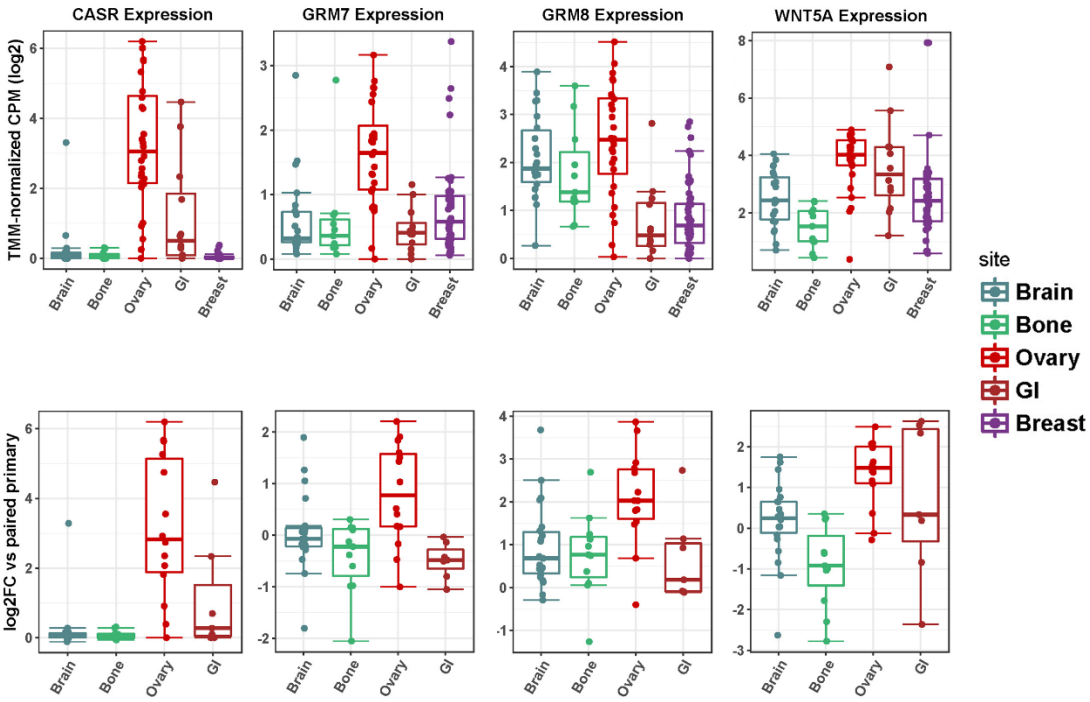


Figure 18: Comparison of the ovarian met upregulated genes among other metastatic cohorts
Expression analysis comparison of key genes (top upregulated pathways) in ovarian metastases with three independent cohorts (bone, brain, and GI). Top panel; expression status for each gene in each site cohort in addition to primary tumors (breast). Bottom panel; Fold-change increase in primary-matched metastases in each cohort (brain=22, bone=11, ovary=13, GI=7).

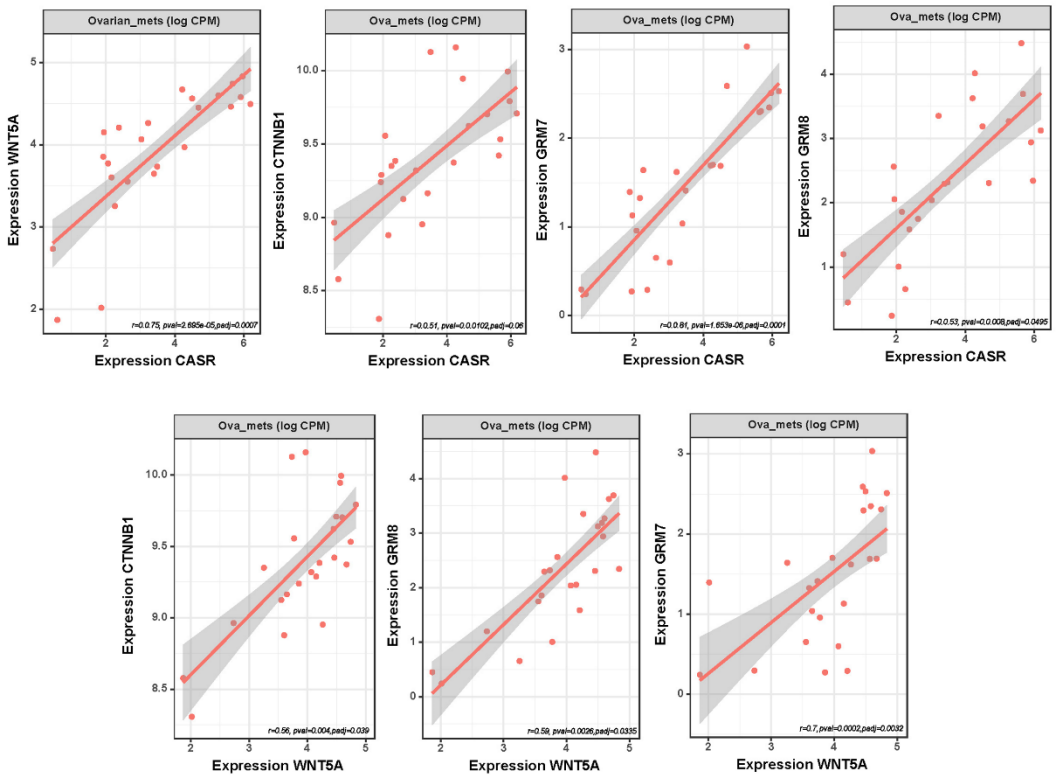


Figure 19: Multiple correlation testing between the upregulated genes in ovarian metastasis
Correlation of key genes involved in the ovarian metastasis upregulated pathways with *CASR* and *WNT5A* expression. Pearson's r and adjusted p values are indicated at the bottom of each plot.

4.3.5 Clinically actionable changes in ovarian metastases

Since metastasis can evolve and harbor distinct genetic alteration beyond its primary tumor, I sought to investigate the possibility of finding clinically informative alterations that are metastasis-enriched. Our analysis of 13 matched primary-metastatic pairs revealed enrichments of *FGFR4* and *MYCN* gains in 77% and 46% of the pairs, respectively (Figure 20 and Data Supplement 3: Table S4). To the contrary, the most notable expression losses included *SOX10*, *IGF2*, and *PDCD1* in 62%, 54%, and 54% of the tumors, respectively.

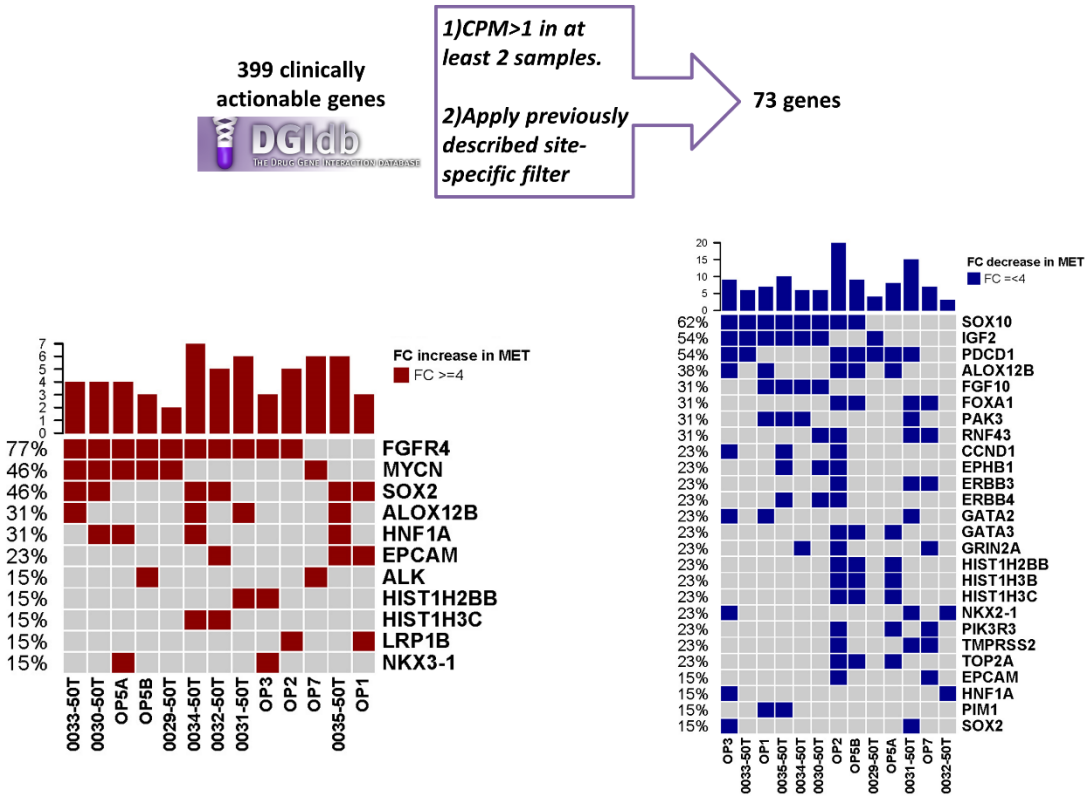


Figure 20: Changes in clinically actionable genes in ovarian met pairs
Recurrent expression gains (red; left) and losses (blue; right) ranked by frequency for primary-matched ovarian metastases. Top bars show counts for alteration events in each sample (column). Expression alterations with fold-change ≥ 4 are reported. Only genes with changes in at least two pairs (15%) are plotted.

Giving that *FGFR4* has been already a major focus in ILC research in our lab (a thesis research project of another student), I decided to investigate further *MYCN* as novel target for ILC progression and metastasis. Analysis of the TCGA and METABRIC breast cancer datasets showed significant expression of *MYCN* in ILC vs IDC (p values= 0.01 and 1e-04, respectively), and this enrichment was even further increased when the analysis was restricted to LumA-tumors (p value= 4.4e-06 and 8.6e-06, respectively) (Figure 21A). In addition, patients with higher *MYCN* expression showed worse disease-specific survival, and the association was more pronounced in ILC patients (Figure 21B). Interestingly, our ILC long-term estrogen-deprived (LTED) cell lines

models [262] showed higher expression of *MYCN* compared to their parental cell lines, further supporting a role in endocrine resistance (Figure 21C).

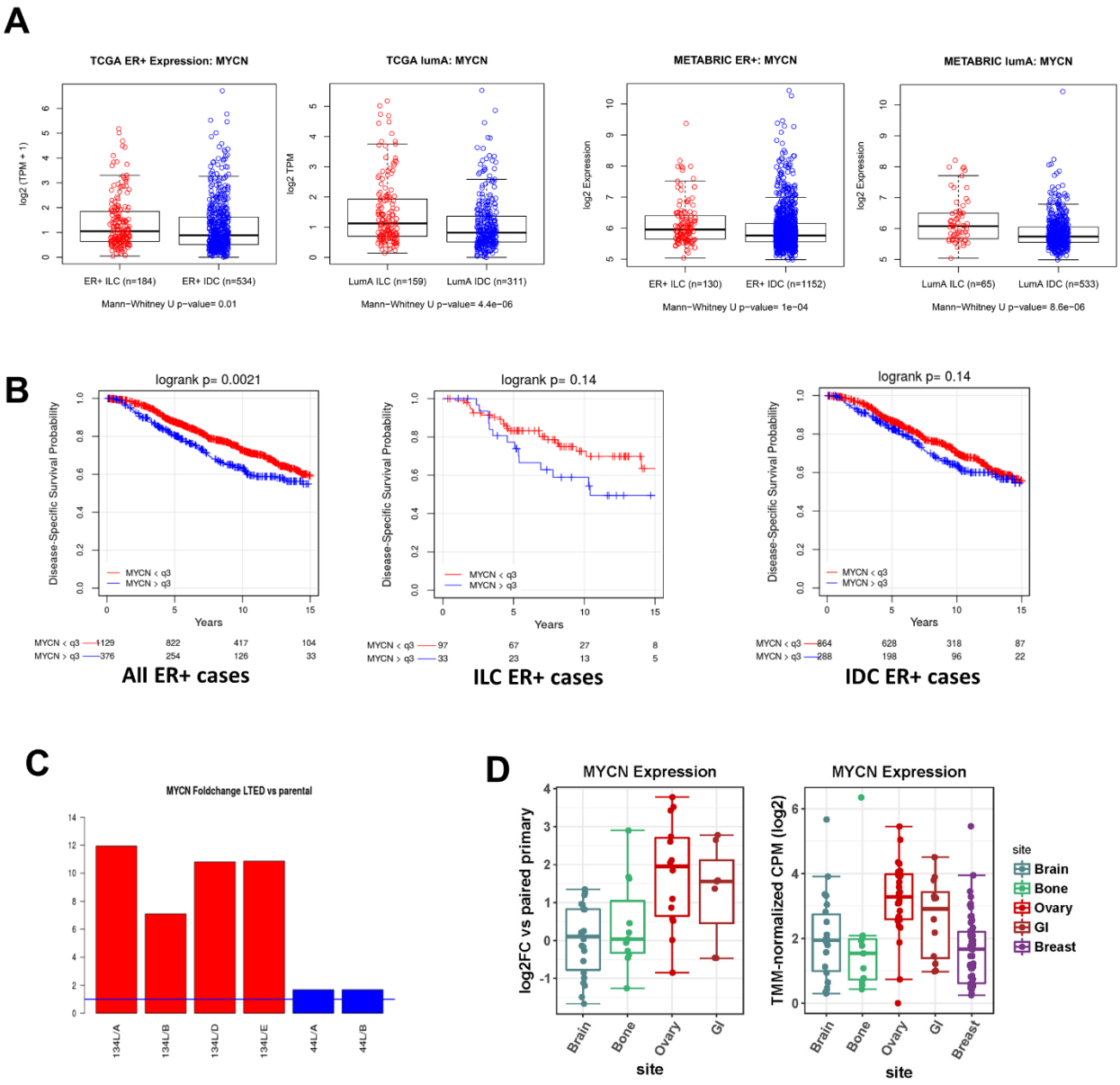


Figure 21: Characterization of MYCN in clinical and cell lines datasets
A. Comparison of *MYCN* mRNA expression for ILC versus IDC tumors in METABRIC and TCGA datasets. **B.** METABRIC Disease-specific survival (DSS) data for patients with high *MYCN* expression (upper quartile) vs patients with lower expression. Three Kaplan-Meier plots are shown for all ER+, ILC/ER+, and IDC/ER+ cases, respectively. P values indicated for log-rank test. **C.** Comparison of *MYCN* expression in long-term estrogen deprived ILC cell lines (MDA-MB-134 and SUM44PE) versus their parental (y-axis: fold-change vs parental). **D.** comparison of *MYCN* mRNA expression in three independent metastatic cohorts. Left; log2 fold-change vs matched primary. Right; overall expression.

4.3.6 MammaSeq™ targeted sequencing analysis

We recently developed an Ion Torrent NGS targeted panel comprised of 78 genes with 688 amplicons targeting 1398 mutations known to have role in cancer progression and endocrine resistance. Our preliminary analysis of 35 tumors (10 primary and 25 metastases) revealed frequent mutations in *CDH1 (43%), PIK3CA (40%),* and *FOXA1 (29%)* (Figure 22A and Data Supplement 3: Table s11). The most recurrent mutations were the previously reported hotspot mutations *PIK3CA E545K (17%) and H107R (17%),* corresponding to the helical and kinase domains, respectively (Figure 22B-C). All *FOXA1* mutations seem to cluster mainly in and around the fork-head DNA binding domain, as previously reported for ILC (Figure 22C). Comparison of primary vs metastatic tumors revealed enrichment for novel *NCOR1* mutations in the ovarian metastatic cohort (28% vs 0% in unpaired, and 40% vs 0% in paired analyses) (Tables 7-8). Further characterization of *CDH1* mutations in the paired samples showed limited maintenance of the same mutation in primary tumors and their matched metastases. A few maintained mutations were observed in two out of nine pairs (22%), and additional validation with ddPCR in one pair showed substantial drop in allele frequency (AF) from 28% in primary tumor to below 5% in metastasis (Figure 23).

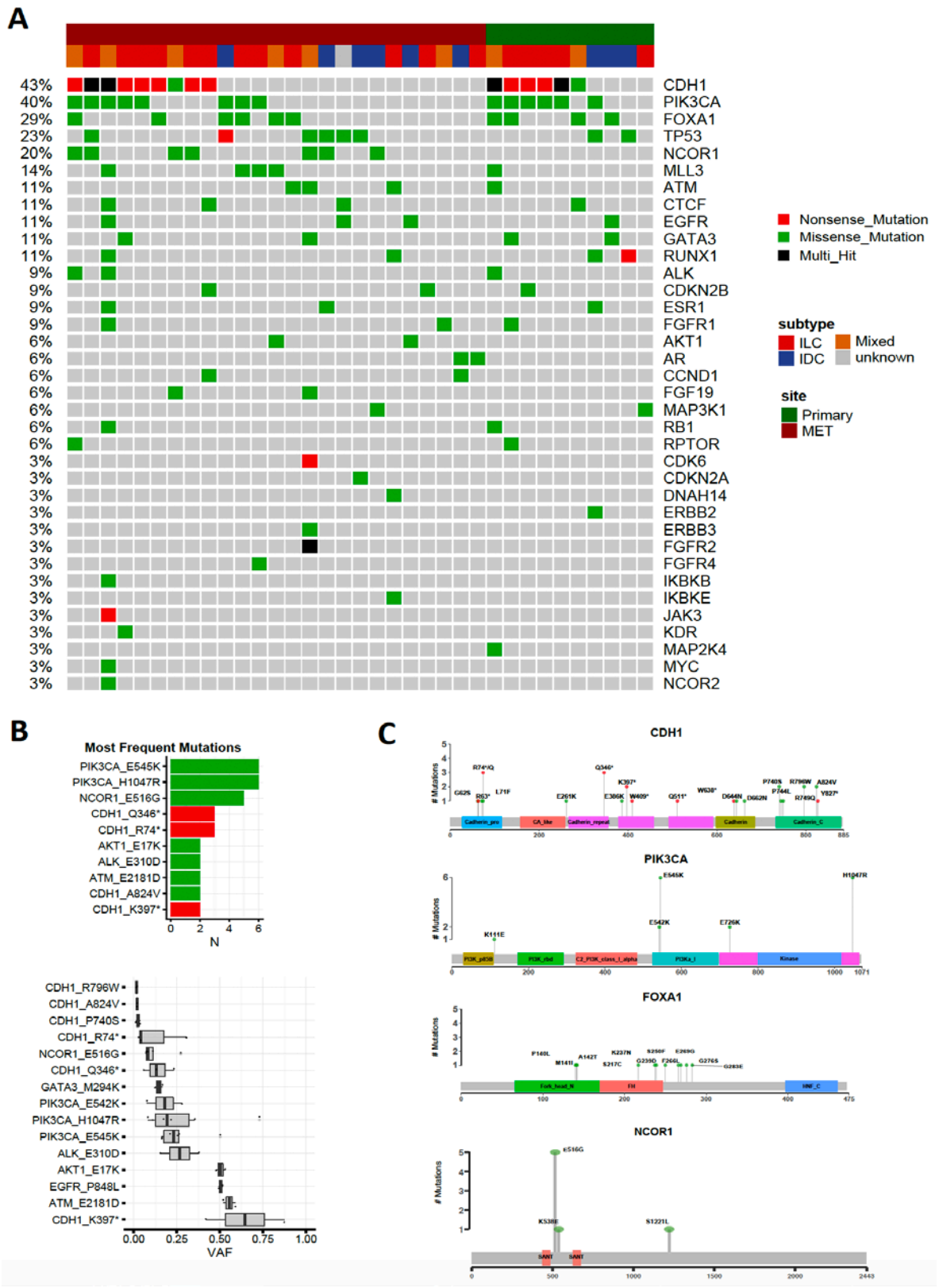


Figure 22: Recurrently mutated genes from MammaSeq analysis

A. Oncoprint visualization of most frequently mutated genes in 35 samples (10 primary and 25 ovarian metastases). Columns are clustered by tumor site (dark green = primary, dark red= metastasis). Type of mutation and tumor subtypes are color coded. **B.** Top: barplot of most recurrent mutations (green= missense, red= non-sense). x-axis represents the number of altered samples. Bottom: multiple boxplots for mutations with highest variant allele frequency (VAF) from the bottom all the way up. **C.** Lollipop plots of the major domains and protein sequence of *CDH1*, *PIK3CA*, and *FOXA1*. Red and green circles represent non-sense and missense mutations, respectively. The length of vertical lines correlates with the frequency of respective mutation as indicated in the y-axis.

Table 7: Comparison of most frequently mutated genes in primary vs metastatic samples (unpaired analysis)

	All primary (n=10)		All Mets (n=25)	
Gene	**samples mutated**	**%**	**samples mutated**	**%**
CDH1	6	60	9	36
PIK3CA	6	60	8	32
FOXA1	4	40	6	24
TP53	2	20	6	24
NCOR1	0	0	7	28

Table 8: Comparison of most frequently mutated genes in primary vs metastatic samples (paired analysis)

	Paired primary (n=9)		Paired Mets (n=10)	
Gene	**samples mutated**	**%**	**samples mutated**	**%**
CDH1	6	66.7	4	40
PIK3CA	5	55.6	3	30
FOXA1	4	44.4	3	30
TP53	1	11.1	1	10
NCOR1	0	0.0	4	40

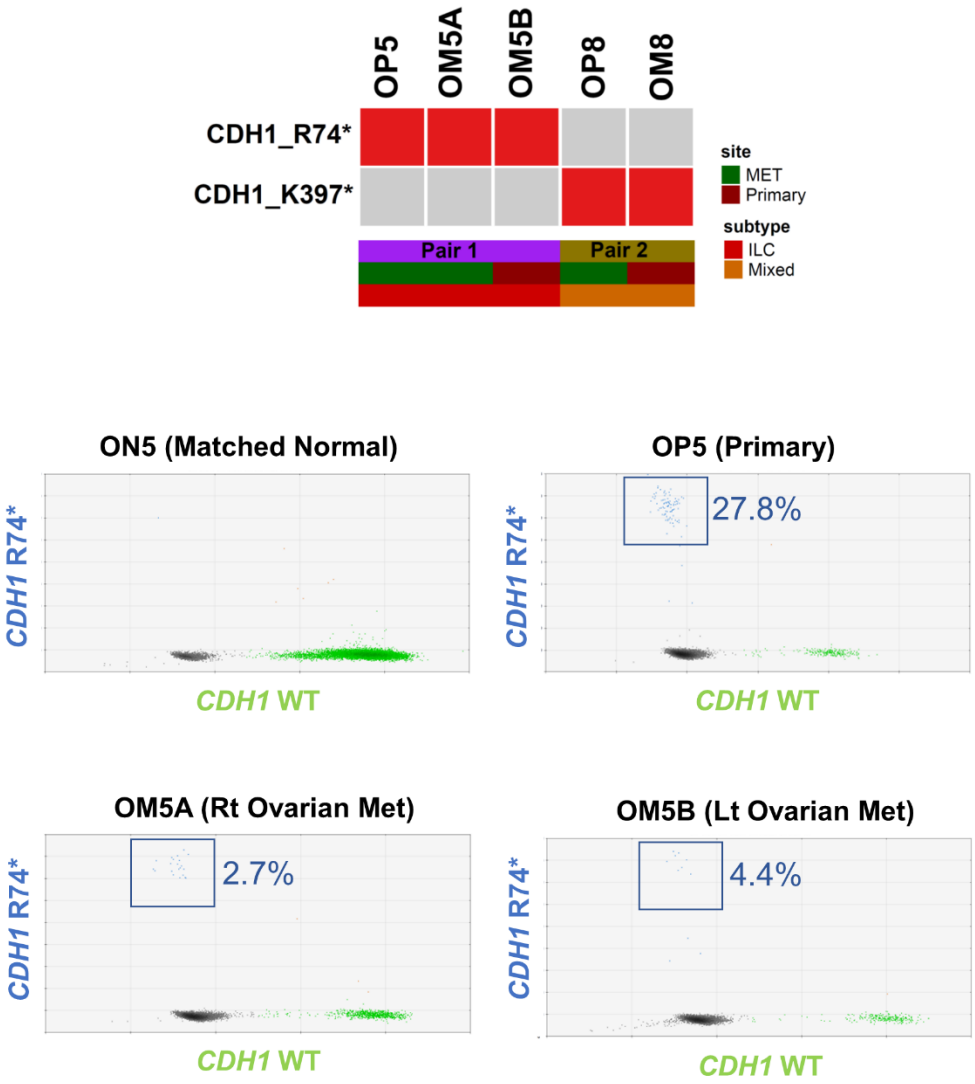


Figure 23: Limited conservation of identical *CDH1* mutation in a subset of the primary-matched ovarian metastases

Top: maintenance of the same *CDH1* mutation in two pairs (Pair1: purple, Pair2: gold). Sample site and histological subtype are color coded. Bottom: scatter plots from ddPCR validation of *CDH1 R74** mutation in Pair1 shows substantial drop of AF in the metastatic samples. Green and blue dots represent droplets with wildtype and mutant genotypes, as indicated in the x- and y- axes, respectively. Black dots indicate droplets without DNA.

4.4 DISCUSSION

Metastatic breast cancer (MBC) has been the main cause of death in breast cancer patients, demonstrating a huge public health burden. With increasing accessibility to metastatic lesions, it has become obvious that MBC is different disease than the primary tumor. In addition, the observation that different breast cancer subtypes can metastasize to unique sites became more observable. Such unique spread could be due to early primary tumor alterations associated with late events, and/or distant site-specific favorable microenvironment for metastasis.

The unique biology of ILC is potentially an important factor to its unique metastatic spread to the ovary and GI system. There is convincing evidence that ILC is the most strongly subtype influenced by menopausal hormones, as indicated by increased risk with exposure to hormone therapy [263,264]. In addition, although it has been shown that ILC is associated with older patient age compared to IDC, age at diagnosis has not been explored within the context of unique metastasis [158,159,265]. Clinicopathological analysis of our ovarian metastasis cohort revealed a median initial age at the diagnosis with primary tumor of 40.5 years which is much younger than reported for ILC patients overall. Furthermore, comparison with two independent metastatic cohorts showed significantly younger median age at diagnosis with primary tumor versus bone (54 years) and brain (53 years) metastases.

There is little evidence that the loss of E-cadherin, which is the hallmark of ILC subtype, is associated with incidence of metastasis to the ovaries [266]. However, although that about 95% of ILC cases show E-cadherin loss, and less than 5% exhibit ovarian metastasis, it is very unlikely that E-cadherin is the main driver for such metastasis [16,161]. More evidence suggests that unless coupled with other alterations (e.g. *PTEN* or *P53* loss), E-cadherin inactivation alone is insufficient for tumor formation [267–269]. Our immunohistochemistry characterization of the ovarian metastases

showed maintained E-cadherin loss and expression for the ILC and IDC subtypes respectively. This supports the notion that E-cadherin loss is involved in identity of the ILC subtype, rather than being involved in the unique metastasis to the ovaries.

Successful molecular interaction between tumor cells and microenvironment is essential to facilitate metastatic spread. As rich endogenous source for hormones (e.g. estrogen), ovaries may create a plausible microenvironment for ILC to metastasize (which seems to be more hormone-driven compared to IDC). Although will need functional validations, our RNA-seq analysis suggests increased calcium signaling in the ovarian metastases, a novel finding that seem to be mainly driven by the calcium-sensing receptor (*CASR*). Additional analysis of clinically actionable changes showed enrichment for N-myc proto-oncogene (*MYCN*) overexpression, another novel finding that has not been explored in breast cancer metastasis.

CASR activation/inactivation has been to linked to multiple genetic disorders outside of cancer. Inactivating mutations cause familial hypocalciuric hypercalcemia (FHH) which can be more severe if mutations were present in homozygous form [270,271]. On the other hand, activating mutations have shown to cause autosomal dominant hypocalcemia and Bartter's syndrome type V [272].

Unlike normal cells, in breast cancer cells, *CASR* seem to increase PTHrP secretion due to a switch to its G-protein preference, which in turn increases calcium availability for proliferation and migration [273,274]. There is increasing evidence that *CASR* signaling pathway may contribute to survival and proliferation of breast cancer cells in high-calcium microenvironment such as bone [273]. High extracellular calcium is one of the most important factors involved in this process. Indeed, calcium extracellular stimulation of *CASR* in the bone-preferring MDA-MB-231 cell lines significantly promoted migration, which was reverted by *CASR* siRNA knockdown [275]. Of note,

the migratory effect was partly through the ERK/MAPK pathway which also upregulated in our ovarian metastasis cohort. Moreover, *CASR* activation in response to calcium has been shown to upregulate *WNT5A*, which in turn stimulate proliferation in an autocrine manner through β-catenin [276]. The high correlation between these three molecules in our study is consistent with these findings. Overall, these discoveries warrant further functional validations, as this pathway might serve as novel therapeutic strategy for breast cancer metastasis to the ovaries.

N-myc (*MYCN*) is a central transcription factor involved in many vital cellular process such as cell growth and proliferation. *MYCN* deregulation is well-established in many tumors such as neuroblastoma, medulloblastoma, and rhabdomyosarcoma [277]. Genomic amplifications of *MYCN* is the most consistent genetic aberration in neuroblastoma, detected in about 20% of cases, and associated with poor outcome [278]. The role of *MYCN* in breast cancer progression and metastasis remain to be explored, although there is limited evidence that it might be associated with clinical outcome of the patients [279]. Our study showed enrichment for *MYCN* upregulation in the ovarian metastasis, in addition to our endocrine-resistant ILC cell line models. Analysis of the METABRIC and TCGA data showed significantly higher expression of *MYCN* in ILC vs IDC patients. Moreover, patients with higher expression of *MYCN* were associated with disease-specific survival (DSS). Collectively, these data suggest a central role for *MYCN* in ILC progression and metastasis.

Our targeted sequencing characterization validated the unique mutational landscape of ILC. In addition to the expected high frequency of *CDH1* mutations, there was high relevance of *PIK3CA* mutations in our cohort, similar to what have been reported previously [16,177]. The *PIK3CA* *E545K* and *H107R* were the most recurrent mutations in this gene. Of note, these mutations have shown the ability to form tumors in mice by activation of PI3K signaling and resistance to

chemotherapy [280–282]. I also identified frequent mutations in *FOXA1* which is involved in genomic action of ER. Clustering in the DNA binding domain, these mutations have potential impact of ER signaling, as *FOXA1* has a key role in the transcription factor complex required for transcription of *ESR1* regulated genes [283]. Indeed, it has been shown that such mutations are associated with higher *FOXA1* and *ESR1* activity, and lower methylation rates [16]. Our analysis also identified novel enrichment of *NCOR1* mutations in metastatic lesions. *NCOR1* is known as a co-repressor that modulate the activation of ER target genes [284]. Previous studies showed that reduced *NCOR1* expression is correlated with acquired tamoxifen resistance in breast cancer mouse models, associated with shorter relapse-free and overall survival in breast cancer [68,284]. These findings support functional consequences in *NCOR1* mutations and deletions (Chapter 3) and their influence in endocrine resistance.

5.0 CONCLUSION

Breast cancer is a major public health burden accounting for 30% of all new cancer diagnoses [153]. Similarly, 30% of early-stage breast cancer patients develop recurrence eventually [122]. Our understanding of breast cancer has evolved in the last decade and taught us that primary tumor and metastasis are not the same disease. Cancer is a smart disease that keep changing its genetic makeup through evolution to gain spread and survival advantages. Metastatic breast cancer remains challenging to treat, and is usually the main cause of death from breast cancer. This is partly due to the way we treat primary tumor and metastasis, considering them similar diseases. Instead, tracking evolution of resistant metastatic subclones can aid in better adaptation of targeted treatment strategy with likely reduced toxicity.

The multiple works presented in this dissertation use state of the art next generation molecular technologies to address distinct molecular changes in endocrine-resistant metastatic breast cancer. In addition, I extend my analysis to unique breast cancer subtypes such as ILC and explore transcriptomic changes involved in its unique metastasis.

Our first work in Chapter 2 represents the most detailed examination of *ESR1* CN analysis in metastatic breast cancer to date. I detected substantial rates of *ESR1* CN amplifications and gains, as well as CN shifts (imbalance between 5' and 3' CN). 11/82 (14%) ER+ tumors showed *ESR1* CN increases with three (4%) CN amplifications, and eight (10%) gains, with a significant enrichment in metastatic samples, and with metastatic site tropism. In contrast, I did not detect any *ESR1* CN amplifications, and only one gain, in ER- tumors. I identified *ESR1* copyshift in five out of the 11 amplified samples. I was able to validate the enrichment of *ESR1* amplifications in metastasis in three independent datasets and show significant association with worse overall

survival. Other frequently amplified genes were *ERBB2* and *GRB7*, while the most frequently deleted gene was *TP53*. I also observed co-amplification of *MDM2* and *ERBB3*, which had not been previously reported. Site-enriched amplifications included *FADD* in bone metastases, and *PTK2* and *PKIA* in brain and GI metastases. Amplifications of the CDK pathway activators *CCND1* and *CDK4/6* were mutually exclusive with deletions of the inhibitory members *CDKN2A*, *CDKN2B*, and *CDKN1B*.

Our findings suggest that, in addition to *ESR1* mutations, *ESR1* amplifications and exon imbalances represent frequent events in endocrine resistant breast cancer that can be of site-tropism nature. Such instances can be optimally suited for diagnostic and/or therapeutic utility in breast cancer. Our analysis also defines a subset of metastatic patients that can be more responsive to selective *CDK4/6* inhibition therapy. Comprehensive profiling of metastatic lesions is warranted, as it has clinical implications on patients with copy number alteration.

In the next Chapter, I further characterize distinct CN changes in the understudied breast cancer subtype ILC, and their clinical relevance. I detected *ESR1* gains and amplifications in 17/71 (24%) of primary ILC tumors, with 10 gains (14%) and 7 amplifications (10%). I also observed significant enrichment of *ESR1* amplifications in patients with recurrence (39%) versus those without recurrence (19%). As expected, the patients with CN amplifications conferred strong trend towards worse recurrence-free survival. Analysis of other cancer driver genes revealed frequent amplifications, including *CCND1*(32%), *ERBB2* (18%), *MDM4* (17%), and *MYC* (17%) which mostly correlated with mRNA expression. The high frequency of *ERBB2* amplifications was unexpected given it has not been reported in ILC previously. Interestingly, HER2 IHC staining revealed small pockets of HER+ subpoulations. Our further investigation in TCGA breast cancer dataset showed that 25% of the cases with *ERBB2* amplifications could not be identified as HER2+

by IHC. Although they were HER2 IHC-, these cases exhibited both higher mRNA and protein expression than the cases without amplifications. Furthermore, the same group was also associated with enrichment of the molecular HER+ PAM50 signature, as evidenced by interrogation of RNA-seq data.

In summary, our results demonstrate frequent *ESR1* CN amplifications in primary ILC, especially in samples with recurrence, suggesting a role in endocrine resistance. Our findings also propose a unique HER2 group of negative IHC score but with DNA amplifications. The high correlation with mRNA and protein expression demands further investigation of benefit from HER2 therapy in this group.

In Chapter 3, in addition to clinical characterization, I perform exome-captured RNA-seq to understand the transcriptomic changes involved in the unique ILC metastasis to the ovaries. Our clinicopathological evaluation of 12 primary-matched ovarian pairs interestingly exhibited younger median age at diagnosis of 40.5 years. Comparison with two independent metastatic cohorts of bone (n=11, 54 years) and brain (n=21, 53 years) showed younger age at diagnosis in the ovarian metastasis cohort. Our Ecad/p120 IHC staining confirmed the loss of E-cadherin in the ILC subtype and it was maintained in the metastatic samples, supporting its importance for identity of the ILC subtype. Using RNA-seq data, our bioinformatics approach assessed how frequent the ovarian metastases can switch from one molecular subtype to another compared to their primary tumors. 10 out of 13 (77%) pairs maintained similar identical subtype. Of these, nine (90%) pairs were LumA, while only one pair was with the Her2 subtype. Although three pairs showed shifts, they still maintained a noteworthy shared proportion of an identical subtype with their matched primary. Next, our differential expression and pathway analyses revealed unique transcriptomic alterations in the pathways of WNT, glutamate and calcium receptors, and MAPK/ERK. These

changes seem to be mainly driven by *CASR, WNT5A, GRM7,* and *GRM8*. Expression analysis comparisons of these genes in three independent metastatic cohorts (brain, bone, and GI) suggested that these changes are ovarian-metastasis-specific. The high correlation between *CASR* and *WNT5A* suggests a cross-talk between them. In fact, *CASR* activation in response to calcium has been shown to upregulate *WNT5A,* which in turn stimulate proliferation in an autocrine manner. In the last part of this chapter, our analysis of clinically actionable genes identified distinct expression gains and losses that are enriched in the ovarian metastasis cohort. The most prominent gains were *FGFR4* and *MYCN* in 77% and 46% of the pairs, respectively. On the other hand, the most notable expression losses included *SOX10, IGF2,* and *PDCD1* in 62%, 54%, and 54% of the patients, respectively. *FGFR4* has been a main focus in our lab and there is increasing supporting evidence of its role in ILC progression. *MYCN* upregulation represents a novel target for ILC progression and metastasis. Our analysis of LTED ILC cell lines models showed increased expression of *MYCN* compared to parental cell lines. Validation analyses in TCGA and METABRIC datasets revealed significantly higher *MYCN* expression of *MYCN* in ILC vs IDC patients and association with worse disease-specific survival (DSS). Using targetd Ion Torrent approach (MammaSeq^{TM}), I further validated the enrichment of *PIK3CA* and *FOXA1* mutations reported previously. In addition, I identified novel mutations in the ER co-repressor *NCOR1* that was enriched in metastatic lesions.

Overall, I identify novel findings and targets that potentially play central role in the unique metastasis of ILC to the ovaries, which can be utilized to serve as novel therapeutic options. Our ongoing research is focused on *in-vitro* and *in-vivo* validations to comprehensively evaluate further the impact of these alteration on metastatic breast cancer.

APPENDIX A: SUPPLEMENTARY FIGURES

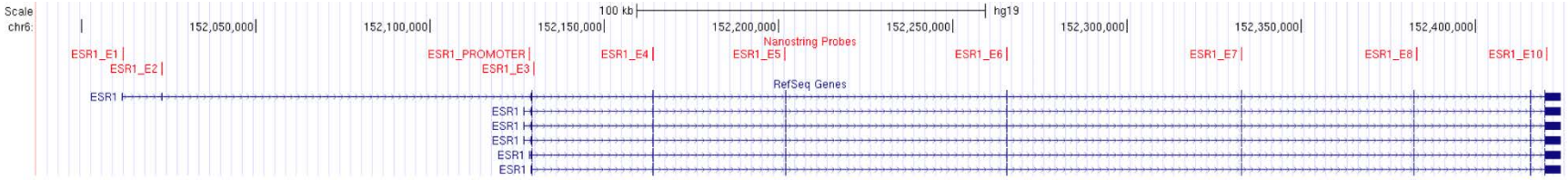
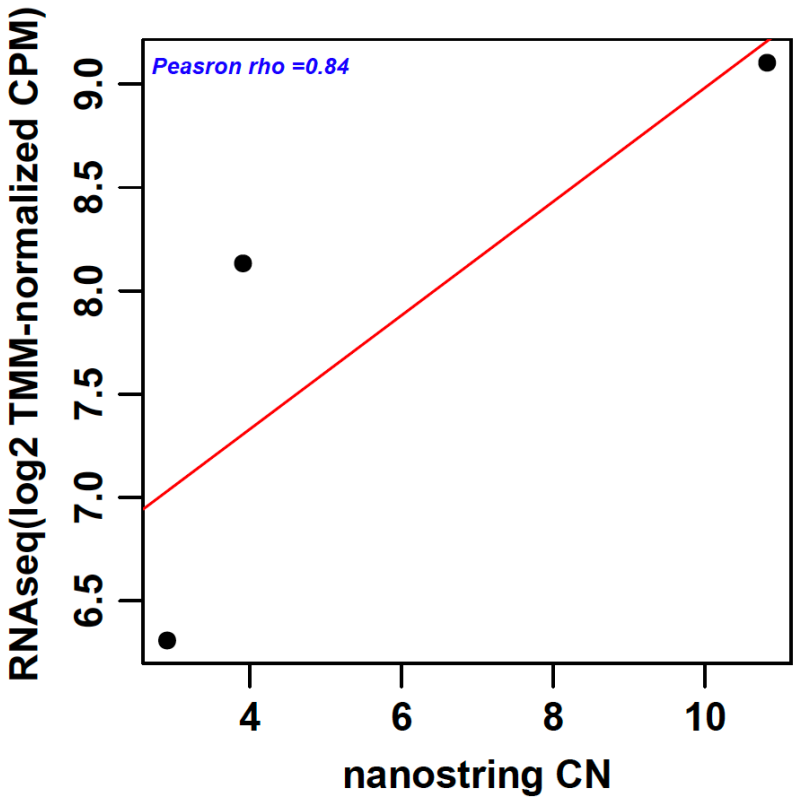


Figure 24: Genomic organization of *ESR1* nanoString probes
Genomic coordinates for nanoString probes comprehensively covering *ESR1* exons are highlighted in red. Different *ESR1* isoforms (RefSeq genes) are listed in blue for comparison. Exons are represented by blocks connected by horizontal lines representing introns. Thinner blocks represent untranslated regions (UTRs).



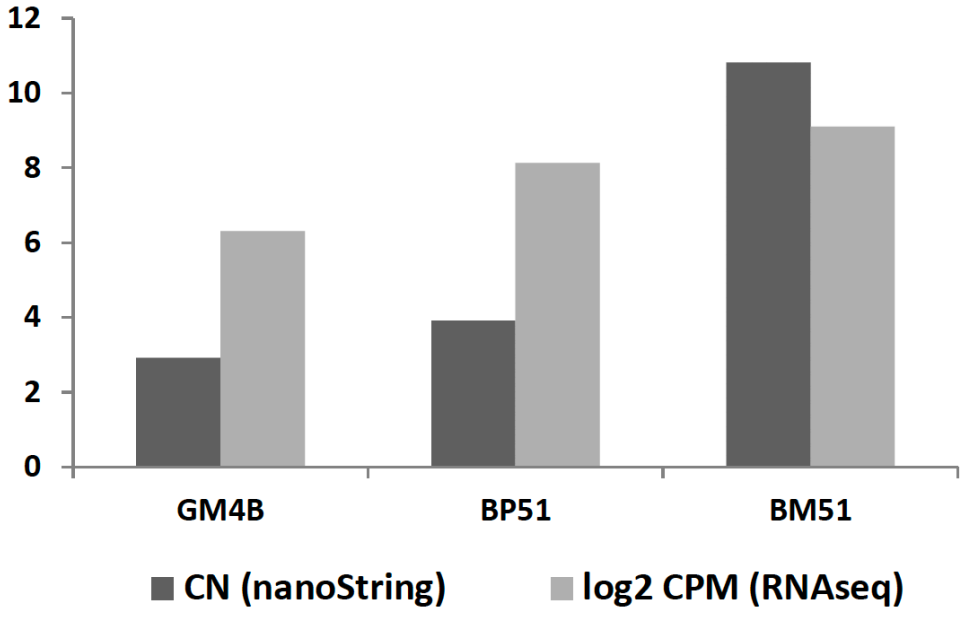


Figure 25: Correlation of *ESR1* amplification with mRNA expression
The data shows dose-dependent correlation between copy number (nanoString) and expression (RNAseq) for 3 amplified samples.

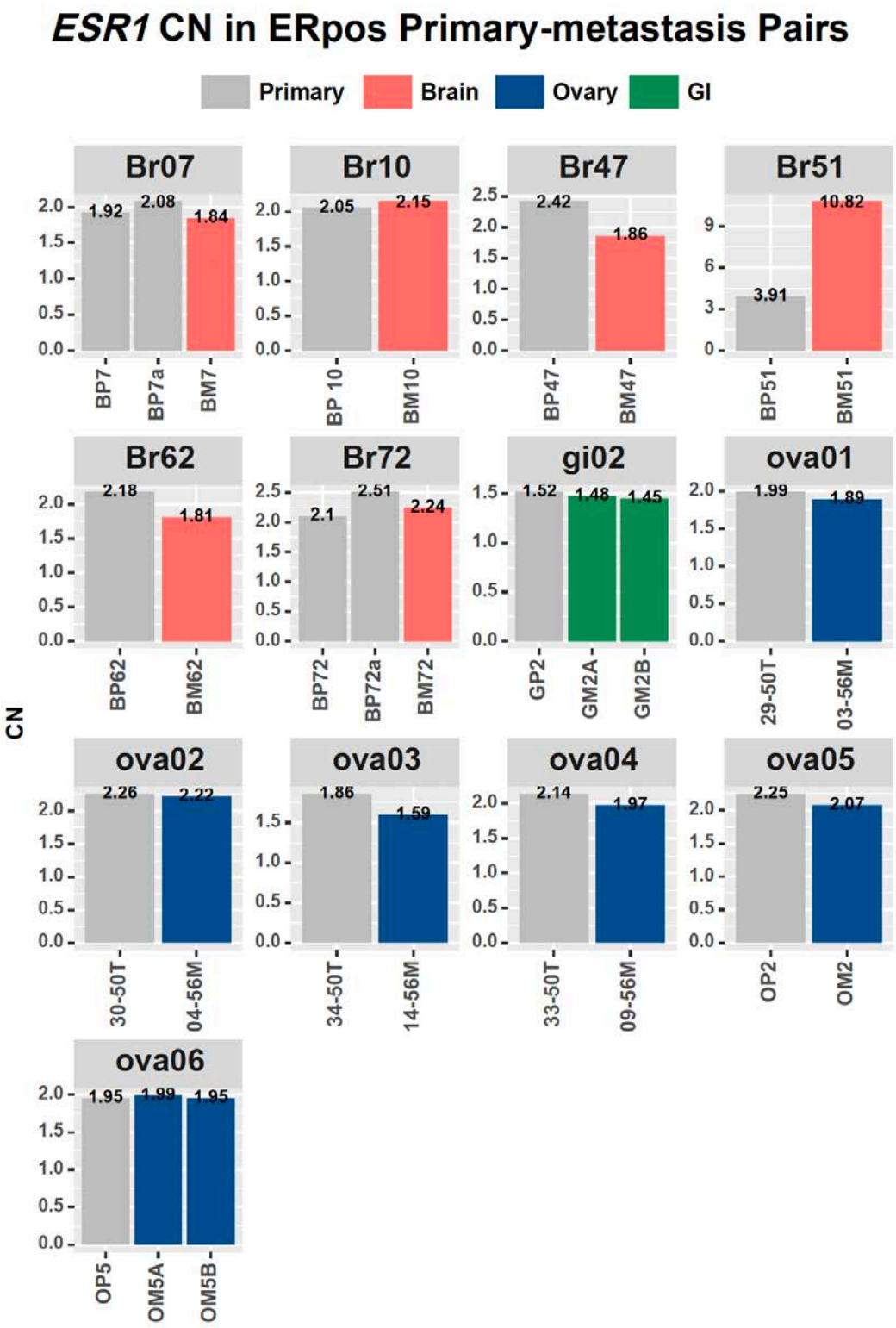


Figure 26: *ESR1* copy number status in ER+ primary-metastasis pairs
Each pair is plotted separately. Y-axis: CN= copy number call. Different colors represent different sites. Br=brain pair; ova= ovarian pair; gi= gi pair. Pair Br51 showed a clear increase in *ESR1* copy number for the brain metastasis.

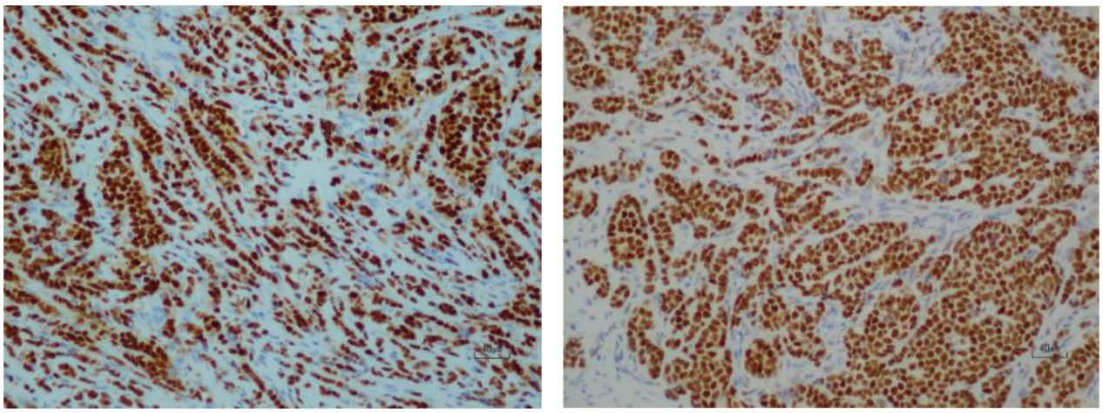


Figure 27: ER IHC for non-amplified samples
IHC staining of *ESR1* non-amplified samples shows strong protein expression. Correspondent *ESR1* DNA copy number calls are indicated below each image.

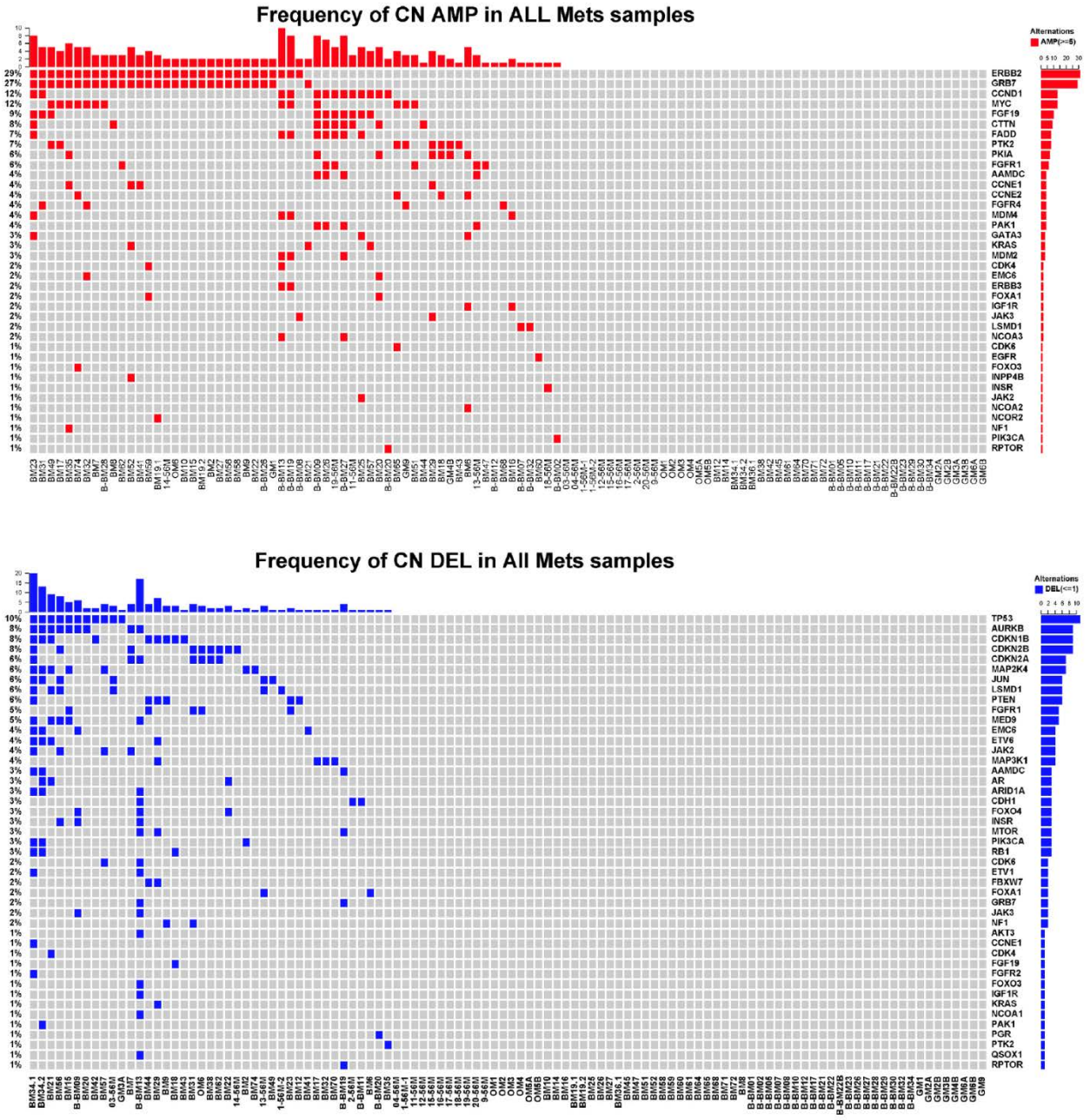


Figure 28: Oncoprint visualization of copy number alterations in clinical samples
Copy number alterations by genes (rows) and samples (columns) show amplifications (top in red) and deletions (bottom in blue) frequency. Top and right-side bars show counts for alteration events within a sample and by gene, respectively.

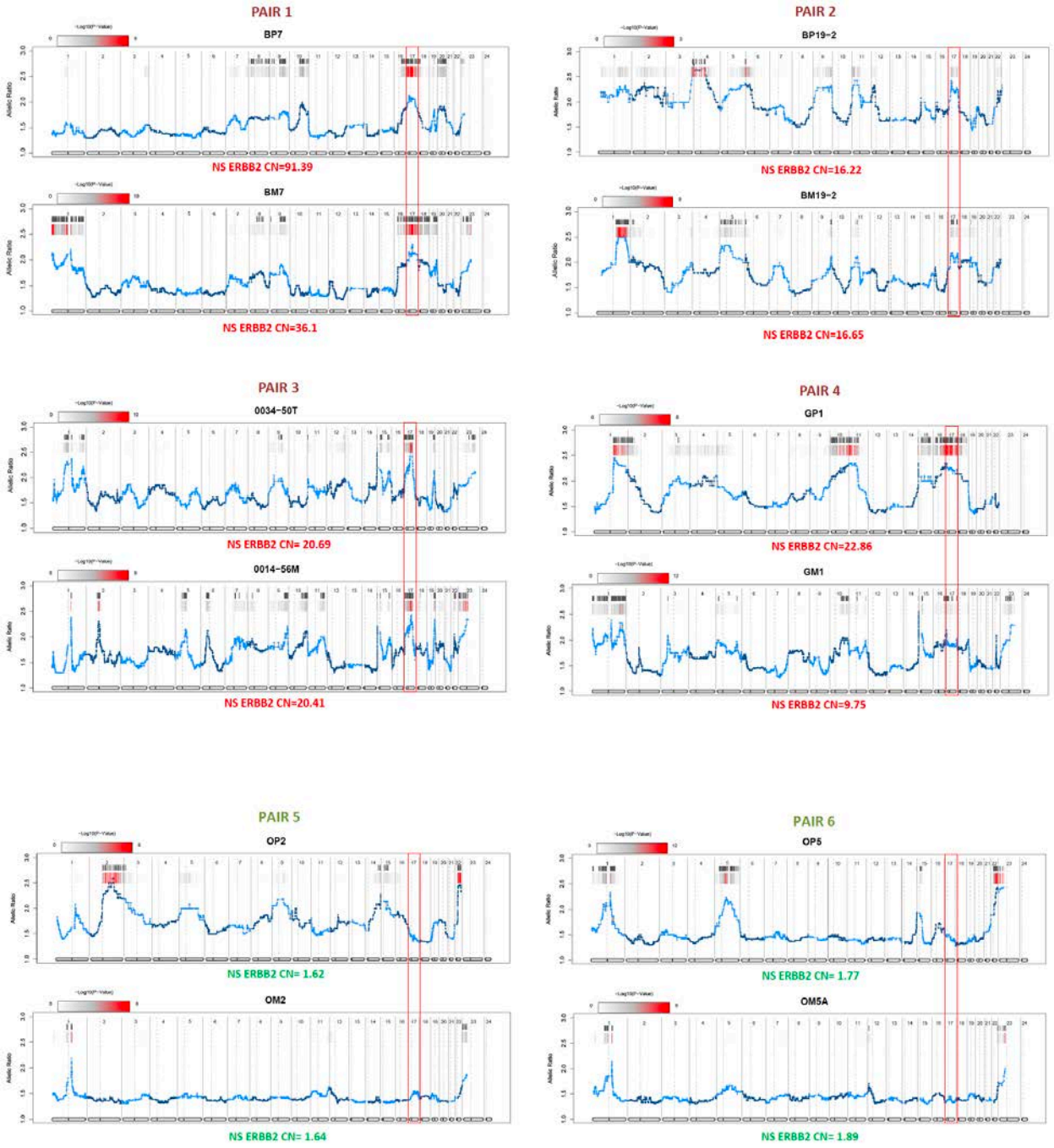


Figure 29: eSNP-Karyotyping analysis of ERBB2 amplified samples
Shown are moving median plots for 6 paired tumors (4 *ERBB2* amplified + two unamplified). Color bars represent FDR-corrected p-value. Positions with $p < 0.01$ are marked by black line. Highlighted in red is the *ERBB2* region in chromosome 17. The left side of the figure shows the copy number calls by nanoString analysis.

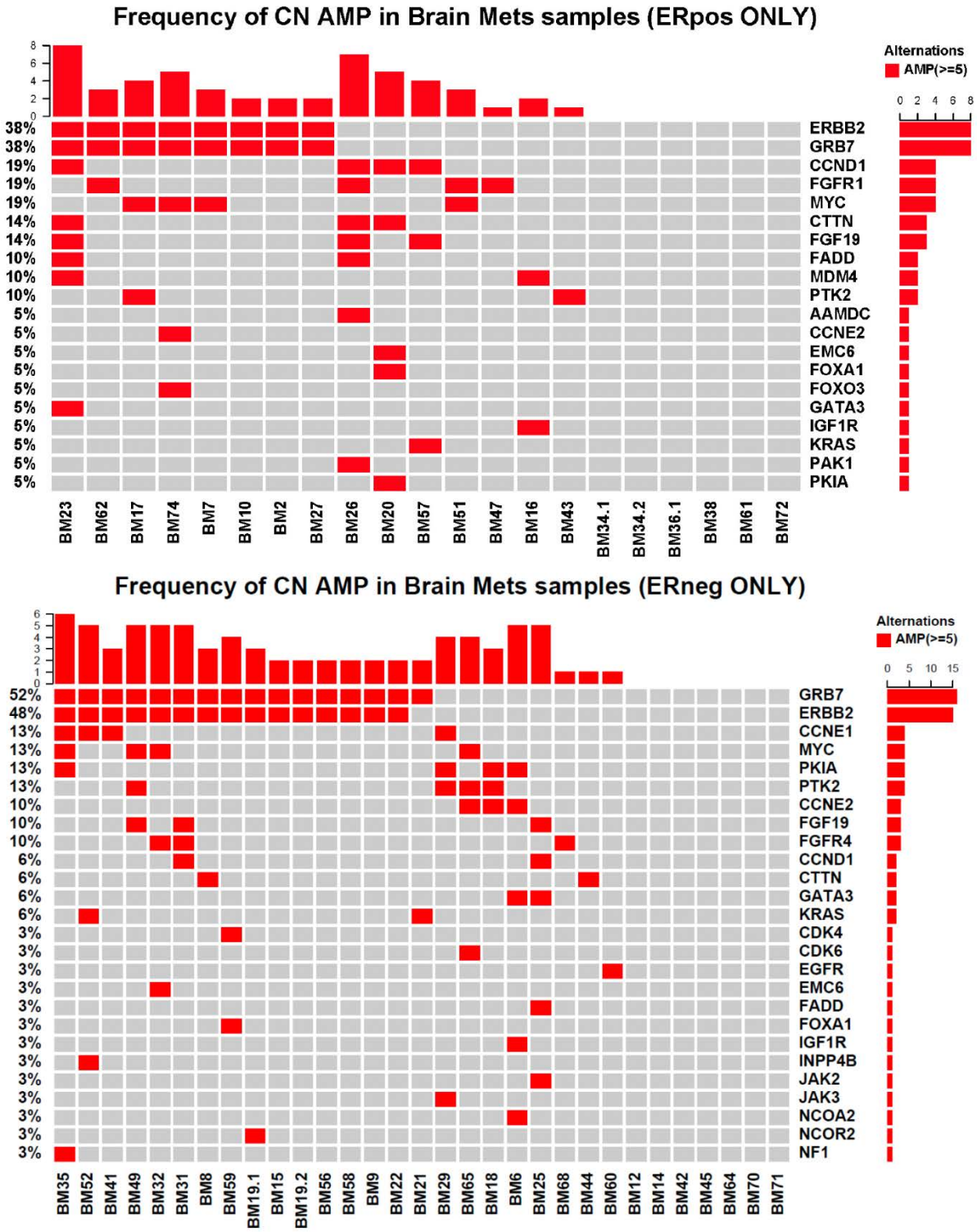
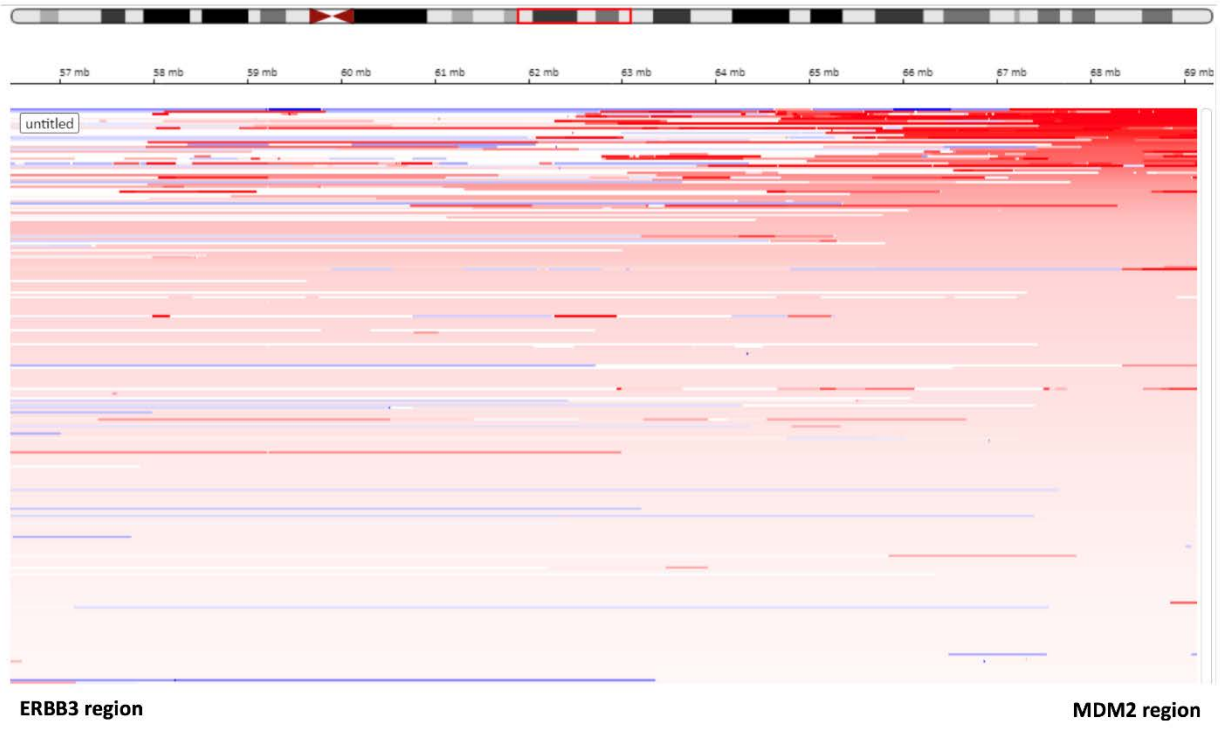


Figure 30: Oncoprint visualization of copy number alterations in ER+ vs ER+ brain metastases
Copy number alterations by genes (rows) and samples (columns) show amplifications frequencies in ER+ (top) and ER- (bottom) samples. Top and right side bars show counts for alteration events within a sample and by gene, respectively. ER+ tumors showed enrichment for FGFR1 amplifications (Fisher's exact test p= 0.0221).

TCGA data

	start	end	distance between the two genes(bp)
MDM2	69,201,956	69,239,214	12,704,665
ERBB3	56,473,641	56,497,291	



	Amplified	non-amplified	%amp
MDM2	36	941	3.7
ERBB3	3	974	0.3

Figure 31: Investigation of MDM2 and ERBB3 amplifications in the TCGA breast dataset.
IGV genomic visualization for MDM2 (far right) and ERBB3 (far left) and the region in between. Each line represents one sample. Intense red or blue color indicate high amplification or deletion, respectively. The clustering of amplifications in MDM2 but not in ERBB3 region suggests separate amplification amplicons.

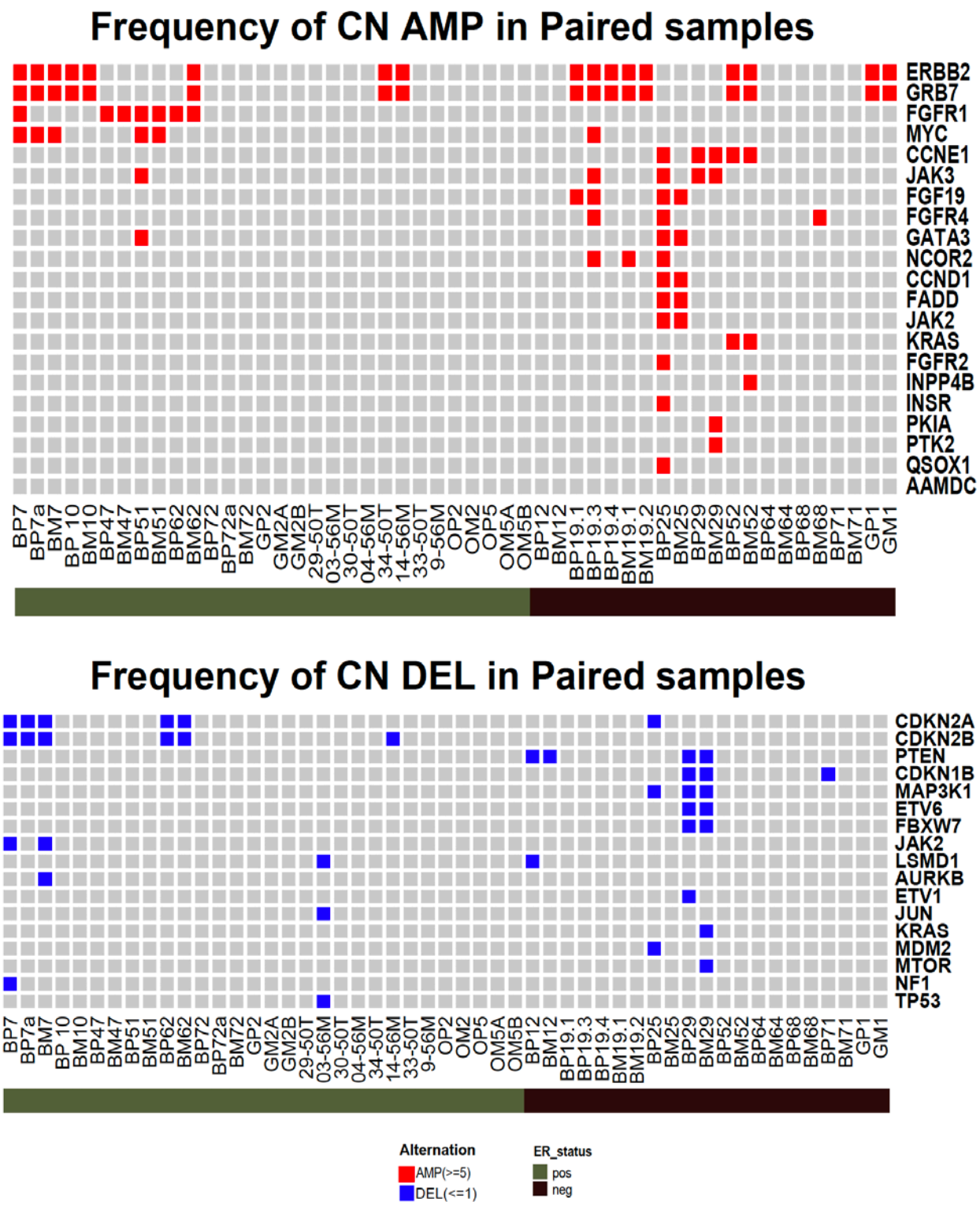


Figure 32: Oncoprint visualization of copy number alterations in paired primary-met samples
Copy number alterations by genes (rows) and samples (columns) show amplifications (amplifications) and deletions frequencies (bottom) in paired-samples (columns organized by pairs). Top and right-side bars in each plot show counts for alteration events within a sample and by gene, respectively. Pair BP-BM29 showed increase in *PKIA* and *PTK* CN, while pair BP-BM68 showed increase in *FGFR4* CN.

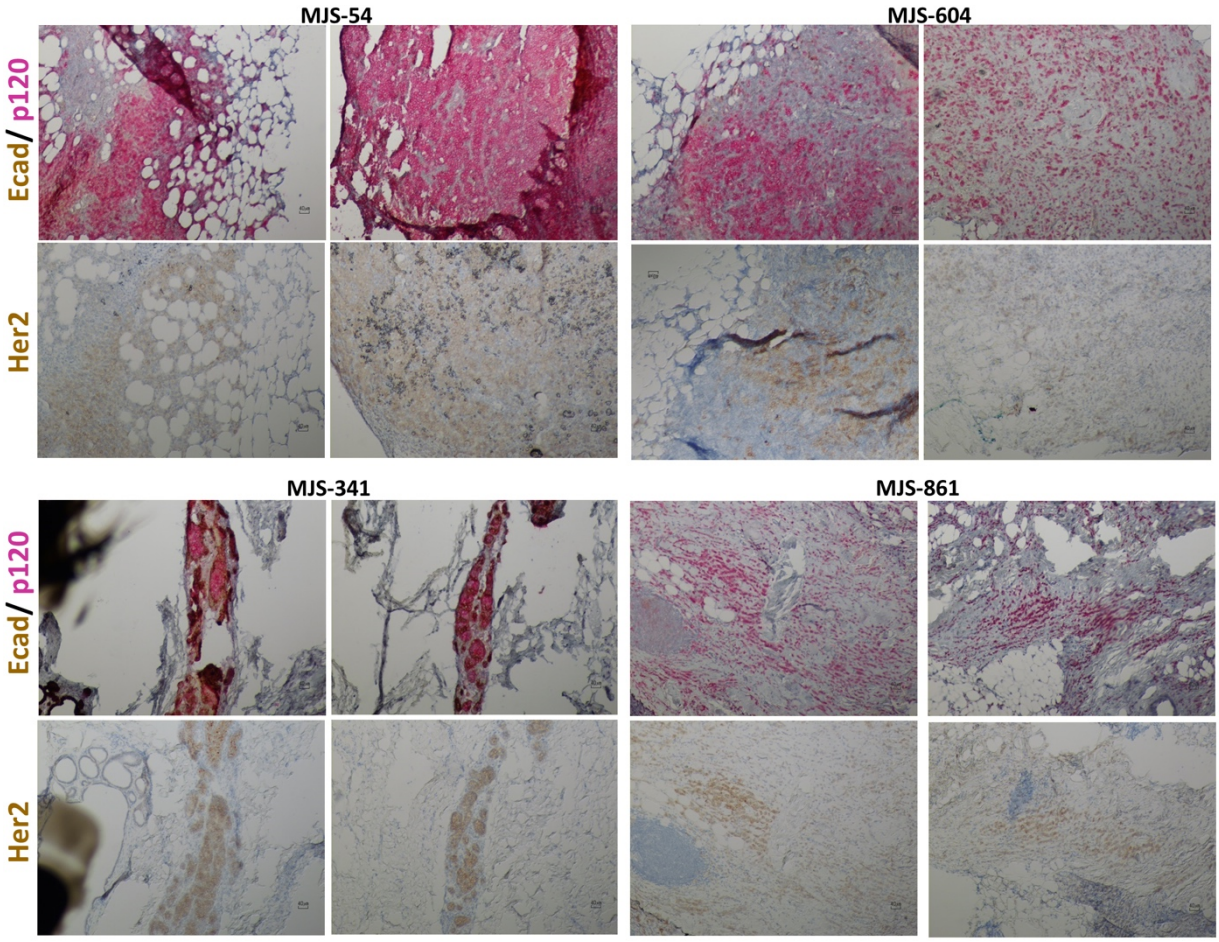


Figure 33: IHC validation of HER2 amplified primary ILC samples
Characterization of four additional ILC primary tumors n Ecad/p120 (top panel) and HER2 (bottom panel) immunohistochemistry. Two fields per sample are shown.

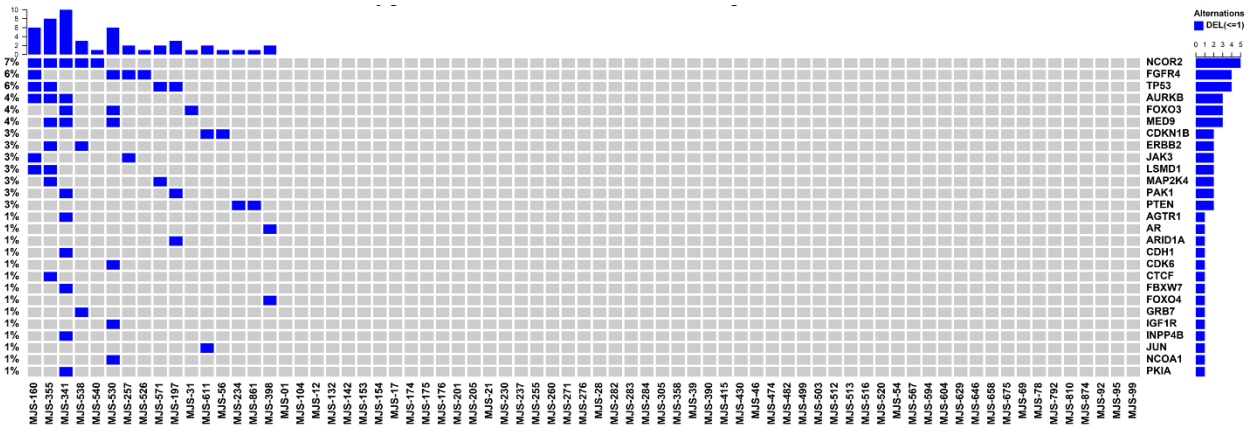


Figure 34: Most frequently deleted genes in primary ILC tumors
Copy number deletions by genes (rows) and samples (columns). Top and right side bars show counts for alteration events within a sample and by gene, respectively.

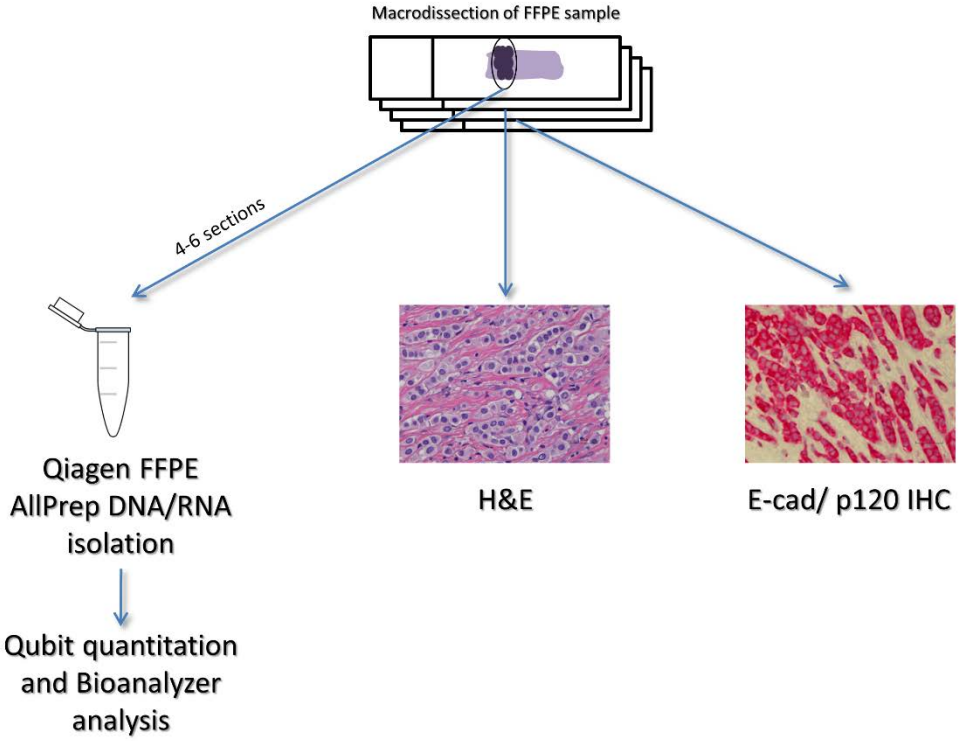
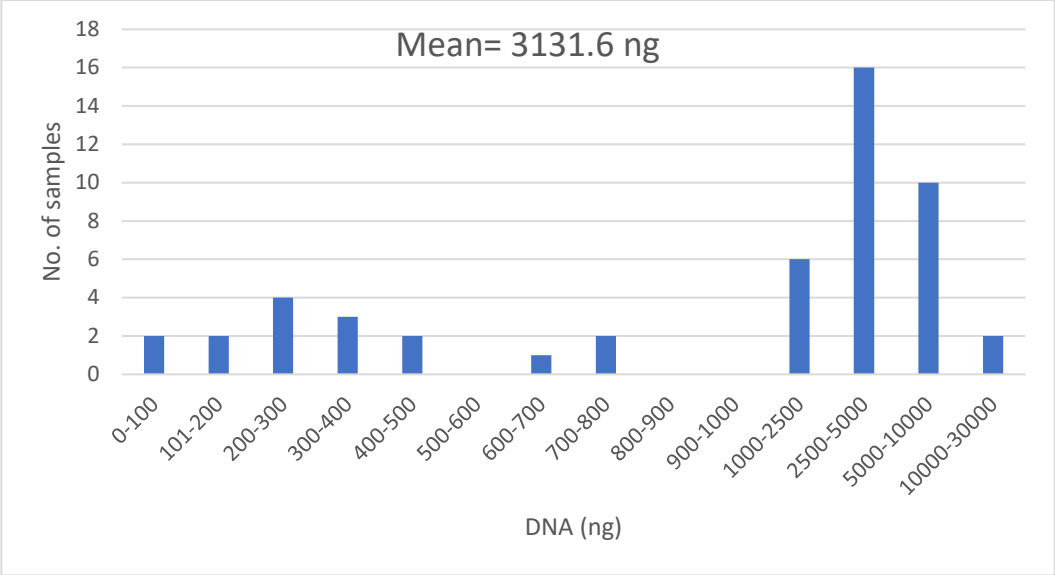


Figure 35: Workflow of formalin-fixed paraffin-embedded (FFPE) processing
Samples were subject to nucleic acid isolation, hematoxylin and eosin (H&E) and Ecad/p-120 staining to confirm the characteristic growth patterns of the ILC subtype.



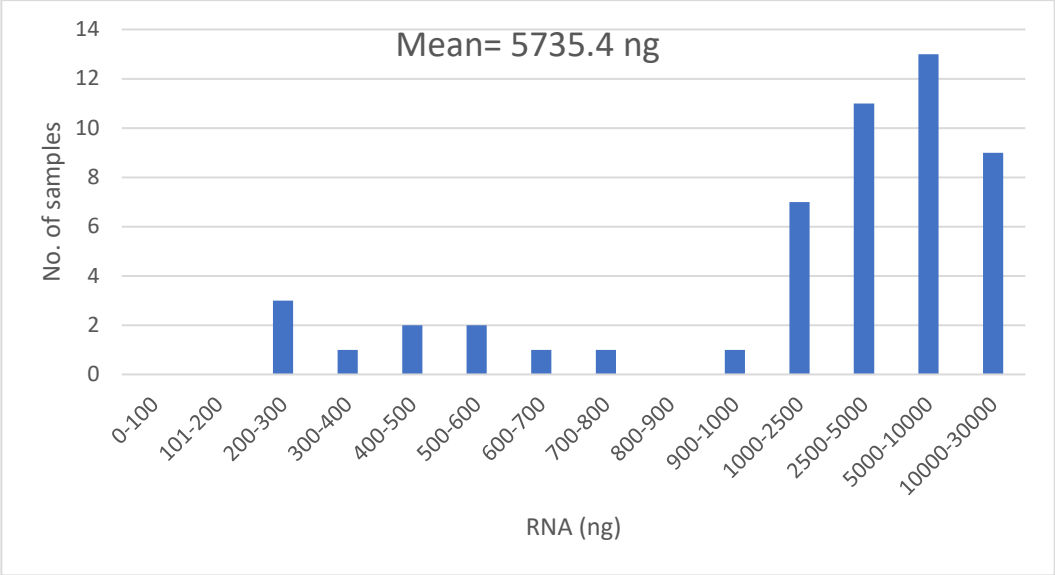


Figure 36: Distribution of the samples with different yields per nucleic acid type

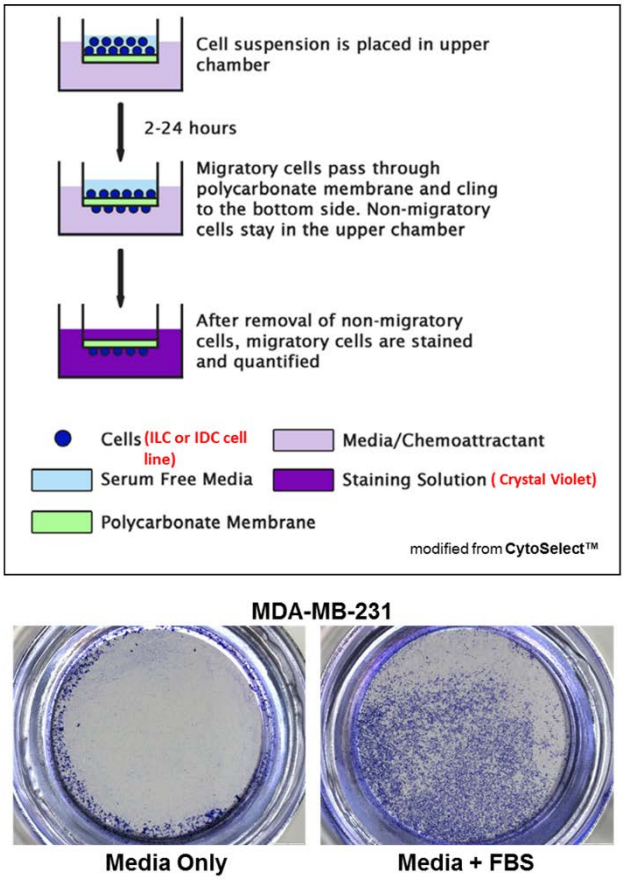


Figure 37: workflow of migration assay

Cell lines are overnight serum starved prior to running the assay. Desired attractant (e.g. ovaries or FBS) are added to the lower well of the migration plate. Cell suspension is added to the inside of each insert and incubated at 37Cº with 5% CO2 for 72 hours. Inserts are then washed with PBS, air dried, and stained with 0.1% crystal violet, then destained with 10% acetic acid. In the final step stained cells are read in spectrophotometer at 590nm. Bottom part of the figure shows results of MDA-MB-231 cells where more cells are migrated toward the FBS attractant.

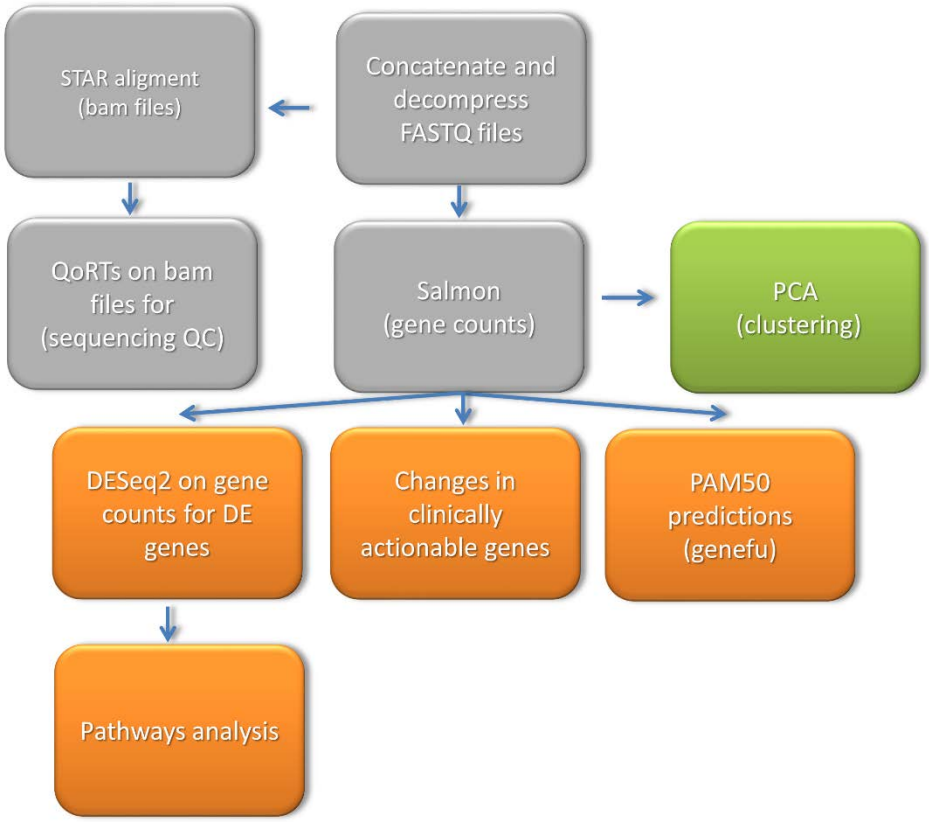


Figure 38: General overview of RNA-seq analyses workflow

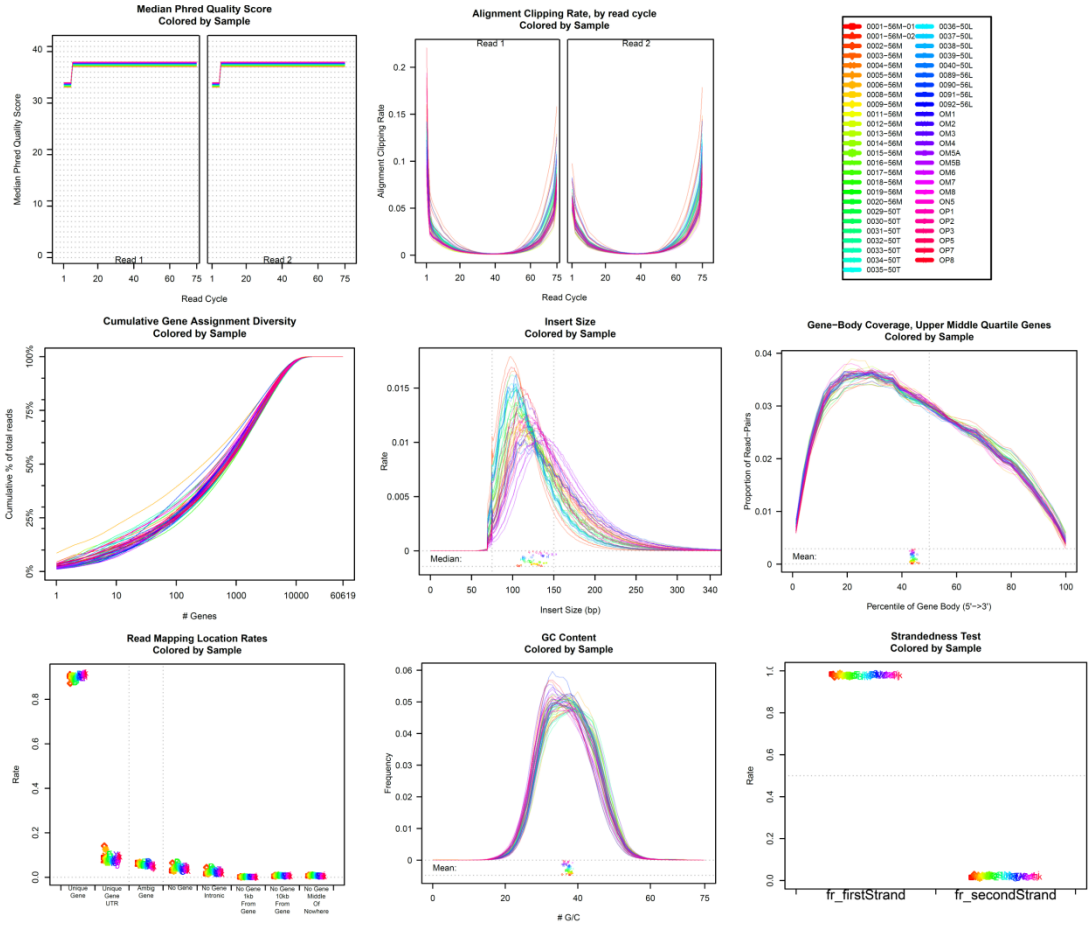


Figure 39: QoRTs QC metrics
Multiple RA-seq quality metrics for primary and ovarian metastases including base phred scores, clipping rates, gene assignment diversity, insert size, gene-body coverage, mapping rates, GC content, and strandedness. Each sample is color coded differently.

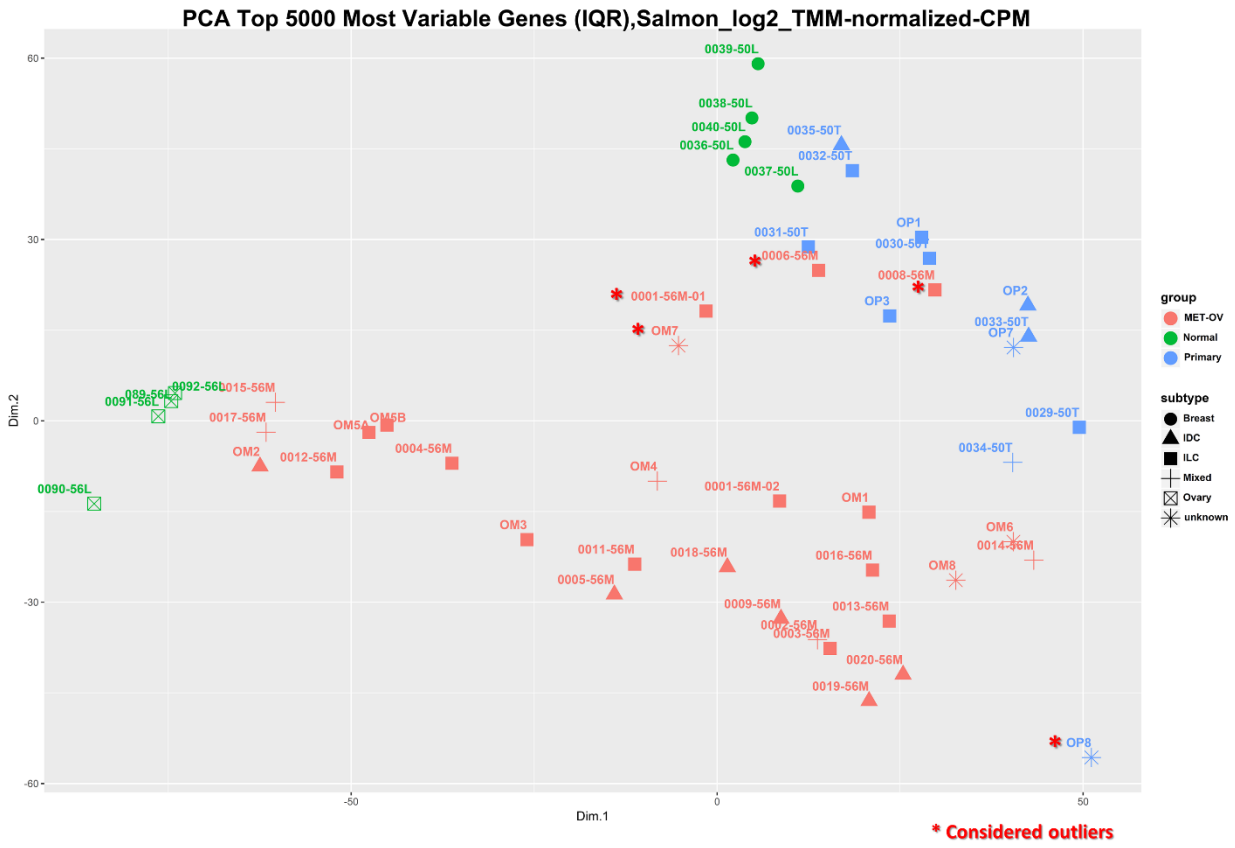


Figure 40: principal component analysis of the primary and ovarian metastasis cohort
Clustering of primary and ovarian metastases along with normal breast and ovary samples by principal component analysis (PCA) using top 5000 most variable genes (interquartile ranges method). Site source for samples are color coded while histological subtypes have different shapes.

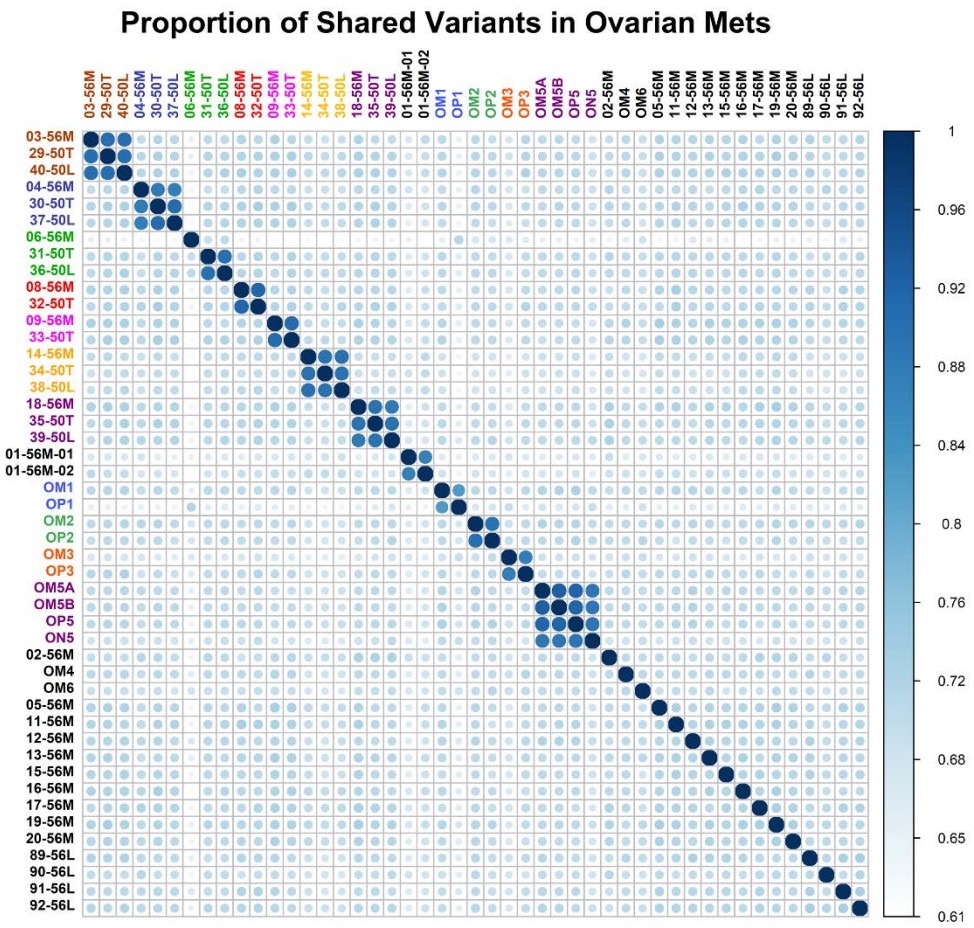


Figure 41: Molecular validation of the matched primary-met pairing status
Correlation plot for proportion of shared variants values in paired samples. Larger and darker circles indicate higher correlation. Paired samples share similar labelling color.

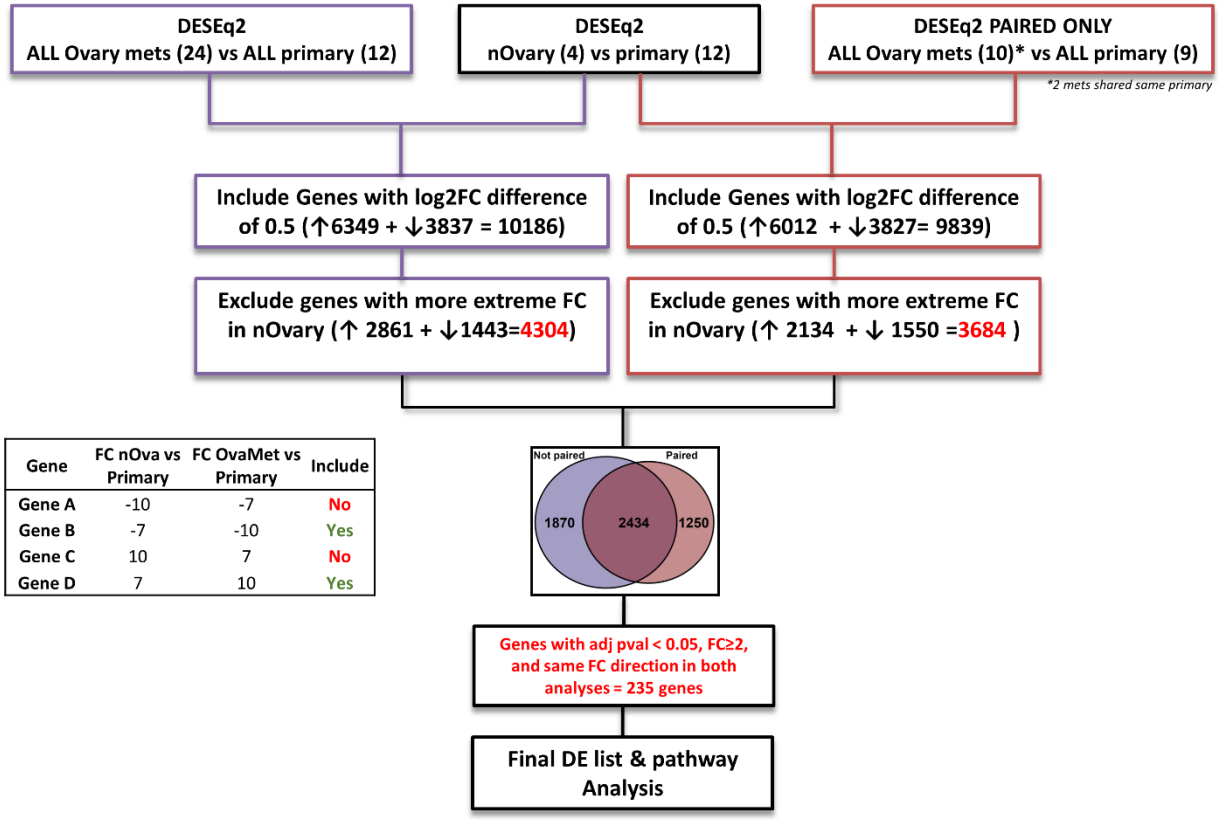


Figure 42: Novel workflow to identify metastasis candidate genes

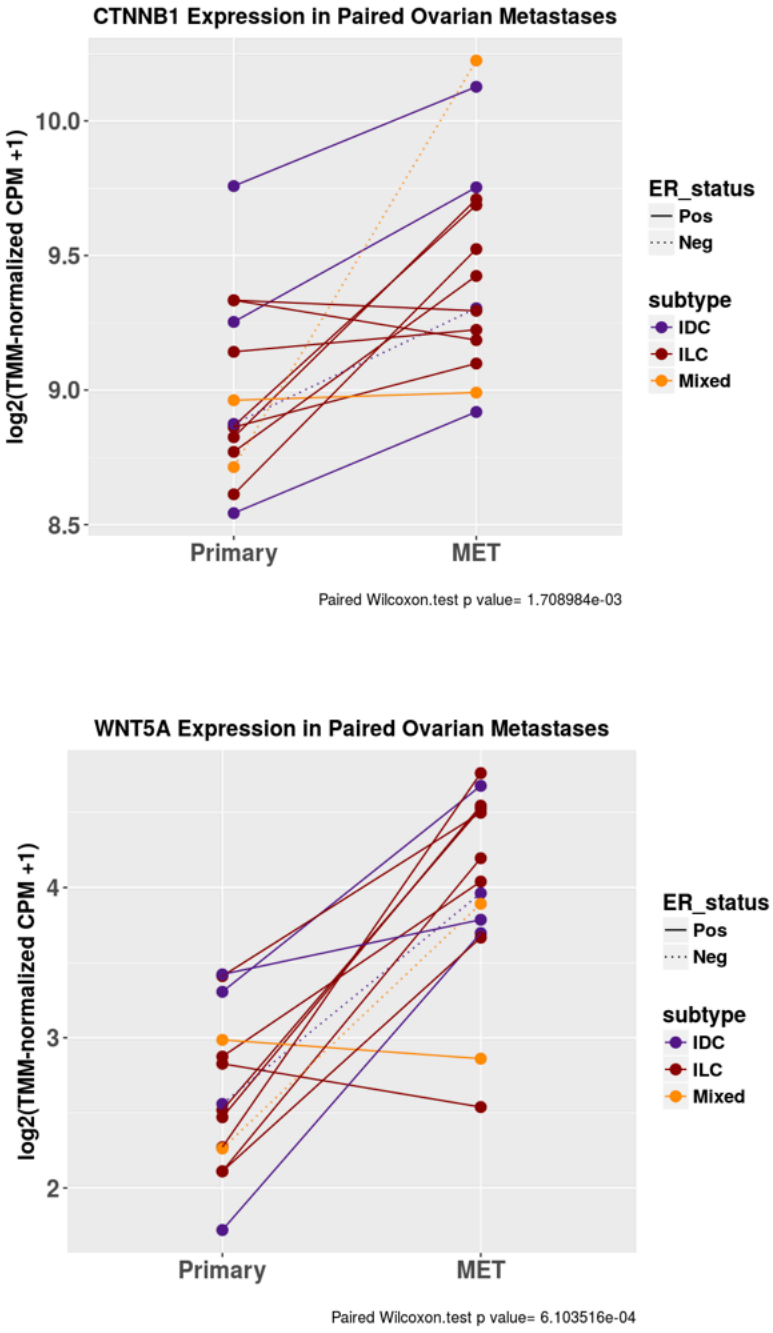


Figure 43: Enrichment of CTNNB1 and WNT5A expression in the ovarian metastatic cohorts
Paired spaghetti plots to visualize the change of gene expression between primary (left) and its matched ovarian metastasis (right). Sample histological subtypes are color-coded and ER status denoted by line type.

APPENDIX B: SUPPLEMENTARY TABLES

Table 9: Primers/probes design for ddPCR (*ESR1*, *EIF2C*, *AP3P1*)

robe/primer name	Sequence	company
ER_CNV_Nterm_F	TAA TAG GAC ATA ACG ACT ATA TGT GTC CAG	IDT
ER_CNV_Nterm_R	GGC AGC TCT TCC TCC TGT TT	IDT
ER_CNV_Nterm_P	- /56-FAM/CCA CCA ACC /ZEN/AGT GCA CCA TTG AT/3IABkFQ/	IDT
EIF2C	http://www.bio-rad.com/en-us/prime-pcr-assays/assay/dhsacp1000002-primepcr-ddpcr-copy-number-assay-eif2c1-human	Bio-Rad
AP3B1	http://www.bio-rad.com/en-us/prime-pcr-assays/assay/dhsacp1000001-primepcr-ddpcr-copy-number-assay-ap3b1-human	Bio-Rad

Table 10: *ESR1* copy number alterations by ER status

Site	N	n (%)		
		AMP	Gain	Total
Primary	11	0.00	1 (9.1)	1 (9.1)
Brain	31	0.00	0.00	0.00
Bone	6	0.00	0.00	0.00
Ovaries	1	0.00	0.00	0.00
GI	3	0.00	0.00	0.00
Total	52	0.00	1.92	1.92

N= number of samples included in the cohort; n and (%)= count and percentage of samples with alterations for the indicated site, respectively.

Table 11: Multiple Fisher exact comparisons for *ESR1* amplifications

comparison	P-value (ER+ only)
primary vs brain	1.0000
primary vs bone	**0.0463**
primary vs ovaries	0.4054
primary vs GI	0.5000
brain vs bone	0.0548
brain vs ovaries	0.2326
brain vs GI	0.5453
bone vs ovaries	**0.0017**
bone vs GI	0.6214
ovaries vs GI	0.2143

Table 12: *ESR1* 5`-3` copy number imbalance

ID	ave_Ex3toEx6	ave_Ex7toEx10	5/3_ratio	shift
BP51	4.440	3.407	1.303	yes
BM51	11.888	10.173	1.168	no
BM57	8.490	8.343	1.018	no
B-BM10	4.530	2.813	1.610	yes
B-BM12	12.250	4.540	2.698	yes
B-BM13	57.740	16.393	3.522	yes
B-BM19	105.380	34.003	3.099	yes
B-BM20	4.238	3.660	1.158	no
B-BM28	3.658	3.273	1.117	no
B-BM32	2.470	2.637	0.937	no
GM4B	2.948	2.797	1.054	no

ratio≥1.3

Wilcoxon matched-pairs signed rank test (p-value)= 0.0024

Table 13: Detailed clinical characteristics of the ovarian metastatic cohort.

Case (Met;Primary)	Study.Met. Location	Primary Histology	Menopausal. Status.At.Dx	ER.Prim	PR.Prim	HER2.Prim	Adj.Endocrine.Tx	Adj.Radiotherapy	Adj.Chemotherapy	Adj.HER2.Tx	ER.MET	PR.MET	HER2.MET	Vital. Status	MFS	Note
OP1; OM1	ovary	ILC	Post	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Pos	NA	Dead	126	Paired samples
OP2; OM2	ovary	IDC	Pre	Neg	Neg	Neg	No	Yes	Yes	No	wkPos	NA	NA	Dead	19	
OP3; OM3	ovary	ILC	Pre	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	NA	NA	Alive	60	
OP5; OM5A; OM5B	ovary	ILC	Pre	Pos	Pos	Equi	Yes	Yes	Yes	No	Pos	NA	NA	Alive	20	
OP7; OM7	ovary	IDC	Pre	Neg	Neg	NA	NA	NA	NA	No	NA	NA	NA	Dead	NA	
OP8; OM8	ovary	Mixed	Post	Pos	Pos	NA	Yes	NA	NA	No	NA	NA	NA	Dead	62	
WCRC-0003-56M; WCRC-0029-50T	ovary	ILC	Pre	Pos	Pos	Equi	Yes	Yes	Yes	No	Pos	Pos	NA	Dead	22	
WCRC-0004-56M; WCRC-0030-50T	ovary	ILC	Post	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Pos	NA	Alive	19	
WCRC-0006-56M; WCRC-0031-50T	ovary	ILC	Pre	Pos	Pos	Equi	Yes	Yes	Yes	No	Pos	Pos	NA	Dead	169	
WCRC-0009-56M; WCRC-0033-50T	ovary	IDC	Pre	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Neg	NA	Dead	93	
WCRC-0014-56M; WCRC-0034-50T	ovary	Mixed	Pre	Pos	Neg	Pos	Yes	No	Yes	No	Pos	Neg	NA	Dead	87	
WCRC-0018-56M; WCRC-0035-50T	ovary	IDC	Pre	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Neg	NA	Dead	39	
WCRC-0013-56M	ovary	ILC	Pre	Pos	Pos	Neg	NA	Yes	Yes	No	Pos	Pos	NA	Dead	207	Orphan samples
WCRC-0011-56M	ovary	ILC	Pre	Pos	Pos	Neg	Yes	No	Yes	No	Pos	Pos	NA	Alive	196	
WCRC-0016-56M	ovary	ILC	Post	Pos	Pos	Neg	Yes	No	Yes	No	Pos	Pos	NA	Dead	NA	
WCRC-0002-56M	ovary	IDC	Pre	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Pos	NA	Dead	241	
WCRC-0005-56M	ovary	IDC	NA	Pos	Pos	Pos	Yes	No	Yes	Yes	Pos	Pos	NA	Dead	74	
WCRC-0019-56M	ovary	IDC	Pre	Pos	Neg	Neg	Yes	No	Yes	No	Pos	NA	NA	Dead	NA	
WCRC-0020-56M	ovary	IDC	Pre	Pos	NA	NA	Yes	No	Yes	No	Pos	NA	NA	Dead	225	
WCRC-0012-56M	ovary	ILC	Pre	Pos	Neg	Neg	Yes	No	Yes	No	Pos	NA	NA	Dead	NA	
WCRC-0001-56M	ovary	ILC	Pre	Pos	Pos	Neg	No	No	Yes	No	Pos	Neg	NA	Alive	336	
WCRC-0017-56M	ovary	IDC	Pre	Pos	Pos	Pos	No	Yes	No	No	Pos	NA	NA	Dead	28	
WCRC-0015-56M	ovary	IDC	Pre	Pos	Pos	Neg	Yes	Yes	Yes	No	Pos	Pos	NA	Dead	94	
OM4	ovary	IDC	Post	Pos	NA	Neg	Yes	No	No	No	Pos	Pos	NA	Alive	NA	
OM6	ovary	IDC	Pre	Neg	Neg	Pos	No	Yes	Yes	Yes	Neg	Neg	NA	Dead	71	

Abbreviations: ILC, invasive lobular carcinoma; IDC, invasive ductal carcinoma; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Dx, diagnosis; Adj Tx, adjuvant therapy; MFS, metastasis-free survival

Table 14: Tumor cellularity for each sample stratified by type, site, and subtype.

ID	type	site	subtype	Tumor cellularity (%)
WCRC-0001-56M-01	met	Ovary	ILC	40-50
WCRC-0001-56M-02	met	Ovary	ILC	90
WCRC-0003-56M	met	Ovary	ILC	90
WCRC-0002-56M	met	Ovary	mixed	90
WCRC-0004-56M	met	Ovary	ILC	80
WCRC-0008-56M	met	Omentum	ILC	50
WCRC-0032-50T	primary	Breast	ILC	50
WCRC-0014-56M	met	Ovary	mixed	70
WCRC-0018-56M	met	Ovary	IDC	70
WCRC-0019-56M	met	Ovary	IDC	80
WCRC-0029-50T	primary	Breast	ILC	70
WCRC-0030-50T	primary	Breast	ILC	30
WCRC-0031-50T	primary	Breast	ILC	50
WCRC-0033-50T	primary	Breast	IDC	70
WCRC-0034-50T	primary	Breast	IDC with lob features	70
WCRC-0035-50T	primary	Breast	IDC	30
WCRC-0011-56M	met	Ovary	ILC	80
WCRC-0012-56M	met	Ovary	ILC	90
WCRC-0013-56M	met	Ovary	ILC	90
WCRC-0016-56M	met	Ovary	ILC	90
WCRC-0009-56M	met	Ovary	IDC	90
WCRC-0040-50L	normal breast	Breast	Normal Br	normal
WCRC-0037-50L	normal breast	Breast	Normal Br	normal
WCRC-0036-50L	normal breast	Breast	Normal Br	normal
WCRC-0039-50L	normal breast	Breast	Normal Br	normal
WCRC-0006-56M	met	Ovary	ILC	20
WCRC-0005-56M	met	Ovary	IDC	90
WCRC-0015-56M	met	Ovary	mixed	90
WCRC-0038-50L	normal breast	Breast	Normal Br	normal
WCRC-0089-56L	normal ova	Ovary	Normal ova	normal
WCRC-0090-56L	normal ova	Ovary	Normal ova	normal
WCRC-0091-56L	normal ova	Ovary	Normal ova	normal
WCRC-0092-56L	normal ova	Ovary	Normal ova	normal
WCRC-0017-56M	met	Ovary	mixed	80
WCRC-0020-56M	met	Ovary	IDC	90
OP1	primary	Breast	ILC	20
OM1	met	Ovary	ILC	80
OP2	primary	Breast	IDC	50
OM2	met	Ovary	IDC	30
OP3	primary	Breast	ILC	30
OM3	met	Ovary	ILC	90
OM4	met	Ovary	IDC with lob features	80
OP5	primary	Breast	ILC	50
ON5	normal breast	Breast	Normal Br	normal
OM5A	met	Ovary	ILC	60
OM5B	met	Ovary	ILC	70
OM6	met	Ovary	NA	90
OP7	primary	Breast	NA	core
OM7	met	Ovary	NA	core
OP8	primary	Breast	NA	core
OM8	met	Ovary	NA	core

Table 15: Overview of clinical samples stratified by site and subtype

Subtype	Primary	Ova Met	Normal Br	Total
ILC	7	14	4	25
IDC	4	7	1	12
Mixed	2	6	1	9
unknown	0	1	0	1
Total	13	28	6	47

4 normal ovaries were also sequenced to subtract site-specific genes contamination

APPENDIX C: DATA SUPPLEMENTS

Data Supplement 1: FREQUENT ESR1 AND CDK PATHWAY COPY NUMBER ALTERATIONS IN METASTATIC BREAST CANCERS

(see supplemental file Data Supplement 1)

Data Supplement 2: ENRICHMENT OF ESR1 COPY NUMBER AMPLIFICATIONS IN ENDOCRINE RESISTANT INVASIVE LOBULAR CARCINOMA

(see supplemental file Data Supplement 2)

Data Supplement 3: COMPREHENSIVE MOLECULAR AND CLINICOPATHLOGICAL LANDSCAPE OF UNIQUE INVASIVE LOBULAR CARCINOMA METASTSIS TO THE OVARIES

(see supplemental file Data Supplement 3)

BIBLIOGRAPHY

1. Elston, C. W. & Ellis, I. O. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19,** 403–410 (1991).

2. Weigelt, B., Geyer, F. C. & Reis-Filho, J. S. Histological types of breast cancer: How special are they? *Mol. Oncol.* **4,** 192–208 (2010).

3. Fisher, E. *et al.* The pathology of invasive breast cancerA Syllabus Derived from Findings of the National Surgical Adjuvant Breast Project (Protocol No. 4). *Cancer* **36,** 1–85 (1975).

4. Li, C. I. *et al.* Changing incidence rate of invasive lobular breast carcinoma among older women. *Cancer* **88,** 2561–9 (2000).

5. Biglia, N. *et al.* Increased incidence of lobular breast cancer in women treated with hormone replacement therapy: implications for diagnosis, surgical and medical treatment. *Endocr. Relat. Cancer* **14,** 549–67 (2007).

6. Malhotra, G. *et al.* Histological, molecular and functional subtypes of breast cancers. *Cancer Biol. Ther.* **10,** 955–60 (2010).

7. Makki, J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clin. Med. Insights. Pathol.* **8,** 23–31 (2015).

8. Arthur, L. M. *et al.* Molecular changes in lobular breast cancers in response to endocrine therapy. *Cancer Res.* (2014). doi:10.1158/0008-5472.CAN-14-0620

9. Dixon, J. *et al.* Infiltrating lobular carcinoma of the breast. *Histopathology* **6,** 149–61 (1982).

10. Korkola, J. E. *et al.* Differentiation of Lobular versus Ductal Breast Carcinomas by Expression Microarray Analysis Differentiation of Lobular versus Ductal Breast Carcinomas by Expression. 7167–7175 (2003).

11. Sikora, M. *et al.* Invasive lobular carcinoma of the breast: patient response to systemic endocrine therapy and hormone response in model systems. *Steroids* **78,** 568–75 (2013).

12. Mamtani, A. & King, T. A. Lobular Breast Cancer: Different Disease, Different Algorithms? *Surg. Oncol. Clin. N. Am.* **27,** 81–94 (2018).

13. Morrogh, M. *et al.* Cadherin-catenin complex dissociation in lobular neoplasia of the breast. *Breast Cancer Res. Treat.* **132,** 641–52 (2012).

14. Dabbs, D. J. *et al.* Lobular Neoplasia of the Breast Revisited With Emphasis on the Role of E-Cadherin Immunohistochemistry. *Am. J. Surg. Pathol.* **37,** e1–e11 (2013).

15. McCart Reed, A. *et al.* Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. *Breast Cancer Res.* **17,** 12 (2015).

16. Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163,** 506–19 (2015).

17. Wang, Y. *et al.* The Role of Snail in EMT and Tumorigenesis. *Curr. Cancer Drug Targets* **13,** 963–972 (2013).

18. Gall, T. M. H. & Frampton, A. E. Gene of the month: E-cadherin (CDH1). *J. Clin. Pathol.* **66,** 928–32 (2013).

19. Davis, M. A., Ireton, R. C. & Reynolds, A. B. A core function for p120-catenin in cadherin turnover. *J. Cell Biol.* **163,** 525–534 (2003).

20. Peifer, M. & Yap, A. S. Comment Traffic control: p120-catenin acts as a gatekeeper to control the fate of classical cadherins in mammalian cells. *J. Cell Biol. JCB ⁄ J. Cell Biol.* **114374,** 21–9525 (2003).

21. Chen, X. *et al.* p120 catenin associates with kinesin and facilitates the transport of cadherin-catenin complexes to intercellular junctions. *J. Cell Biol.* **163,** 547–57 (2003).

22. Reynolds, A. B. & Roczniak-Ferguson, A. Emerging roles for p120-catenin in cell adhesion and cancer. *Oncogene* **23,** 7947–7956 (2004).

23. Sarrió, D. *et al.* Cytoplasmic localization of p120ctn and E-cadherin loss characterize lobular breast carcinoma from preinvasive to metastatic lesions. *Oncogene* **23,** 3272–3283 (2004).

24. Shibata, T. *et al.* Cytoplasmic p120ctn regulates the invasive phenotypes of E-cadherin-deficient breast cancer. *Am. J. Pathol.* **164,** 2269–78 (2004).

25. Jensen, E. *et al.* Estrogen receptors and breast cancer response to adrenalectomy. *Natl. Cancer Inst. Monogr.* **34,** 55–70 (1971).

26. Jensen, E. V. On the mechanism of estrogen action. *Perspect. Biol. Med.* **6,** 47–59 (1962).

27. Kennedy, B. J. Hormone therapy for advanced breast cancer. *Cancer* **18,** 1551–7 (1965).

28. Jordan, V. C. & Brodie, A. M. H. Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids* **72,** 7–25 (2007).

29. Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235,** 177–82 (1987).

30. Slamon, D. J. *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244,** 707–12 (1989).

31. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).

32. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10869–74 (2001).

33. Parker, J. S. *et al.* Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).

34. Paik, S. *et al.* A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* **351,** 2817–2826 (2004).

35. Paik, S. *et al.* Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor–Positive Breast Cancer. *J. Clin. Oncol.* **24,** 3726–3734 (2006).

36. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).

37. van't Veer, L. J., Paik, S. & Hayes, D. F. Gene Expression Profiling of Breast Cancer: A New Tumor Marker. *J. Clin. Oncol.* **23,** 1631–1635 (2005).

38. Dowsett, M. *et al.* Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J. Clin. Oncol.* **31,** 2783–90 (2013).

39. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

40. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–52 (2012).

41. Nik-Zainal, S. *et al.* Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **149,** 979–993 (2012).

42. Goldhirsch, A. *et al.* Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. *Ann. Oncol.* **18,** 1133–1144 (2007).

43. Maughan, K. L., Lutterbie, M. A. & Ham, P. S. Treatment of breast cancer. *Am. Fam. Physician* **81,** 1339–46 (2010).

44. Ferguson, T. *et al.* Taxanes for adjuvant treatment of early breast cancer. *Cochrane Database Syst. Rev.* CD004421 (2007). doi:10.1002/14651858.CD004421.pub2

45. von Minckwitz, G. *et al.* S3-2: Neoadjuvant Chemotherapy Adapted by Interim Response Improves Overall Survival of Primary Breast Cancer Patients – Results of the GeparTrio Trial. *Cancer Res.* **71,** S3-2-S3-2 (2011).

46. Stockwell, S. Classics in oncology. George Thomas Beatson, M.D. (1848-1933). *CA. Cancer J. Clin.* **33,** 105–21

47. Clark, G. M., Osborne, C. K. & McGuire, W. L. Correlations between estrogen receptor, progesterone receptor, and patient characteristics in human breast cancer. *J. Clin. Oncol.* **2,** 1102–1109 (1984).

48. Jensen, E. V & Jordan, V. C. The estrogen receptor: a model for molecular medicine. *Clin. Cancer Res.* **9,** 1980–9 (2003).

49. Jordan, V. C. Effect of tamoxifen (ICI 46,474) on initiation and growth of DMBA-induced rat mammary carcinomata. *Eur. J. Cancer* **12,** 419–24 (1976).

50. Richardson, D. N. The history of Nolvadex. *Drug Des. Deliv.* **3,** 1–14 (1988).

51. Turner, R. T. *et al.* Tamoxifen prevents the skeletal effects of ovarian hormone deficiency in rats. *J. Bone Miner. Res.* **2,** 449–56 (1987).

52. Gottardis, M. *et al.* Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. *Cancer Res.* **48,** 812–5 (1988).

53. Wakeling, A. E. Similarities and distinctions in the mode of action of different classes of antioestrogens. *Endocr. Relat. Cancer* **7,** 17–28 (2000).

54. Agrawal, A. *et al.* Biological effects of fulvestrant on estrogen receptor positive human breast cancer: short, medium and long-term effects based on sequential biopsies. *Int. J. cancer* **138,** 146–59 (2016).

55. Boér, K. Fulvestrant in advanced breast cancer: evidence to date and place in therapy. *Ther. Adv. Med. Oncol.* **9,** 465–479 (2017).

56. Miller, W. R. Aromatase inhibitors: mechanism of action and role in the treatment of breast cancer. *Semin. Oncol.* **30,** 3–11 (2003).

57. Miller, W. R. & Larionov, A. A. Understanding the mechanisms of aromatase inhibitor resistance. *Breast Cancer Res.* **14,** 201 (2012).

58. Reinbolt, R. E. *et al.* Endocrine Therapy in Breast Cancer: The Neoadjuvant, Adjuvant, and Metastatic Approach. *Semin. Oncol. Nurs.* **31,** 146–155 (2015).

59. Nourmoussavi, M. *et al.* Ovarian ablation for premenopausal breast cancer: A review of treatment considerations and the impact of premature menopause. *Cancer Treat. Rev.* **55,** 26–35 (2017).

60. Conte, B., Poggio, F. & Del Mastro, L. Luteininzing hormone releasing hormones analogs in combination with tamoxifen for the adjuvant treatment of premenopausal women with hormone receptor positive breast cancer. *Expert Opin. Pharmacother.* **18,** 1357–1362 (2017).

61. Fan, W., Chang, J. & Fu, P. Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies. *Future Med. Chem.* **7,** 1511–9 (2015).

62. Gutierrez, M. C. *et al.* Molecular Changes in Tamoxifen-Resistant Breast Cancer: Relationship Between Estrogen Receptor, HER-2, and p38 Mitogen-Activated Protein Kinase. *J. Clin. Oncol.* **23,** 2469–2476 (2005).

63. Encarnación, C. A. *et al.* Measurement of steroid hormone receptors in breast cancer patients on tamoxifen. *Breast Cancer Res. Treat.* **26,** 237–46 (1993).

64. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45,** 1446–1451 (2013).

65. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* **45,** 1439–1445 (2013).

66. Oesterreich, S. & Davidson, N. E. The search for ESR1 mutations in breast cancer. *Nat. Genet.* **45,** 1415–6 (2013).

67. Osborne, C. K. *et al.* Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J. Natl. Cancer Inst.* **95,** 353–61 (2003).

68. Lavinsky, R. M. *et al.* Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 2920–5 (1998).

69. Span, P. *et al.* Cyclin-E is a strong predictor of endocrine therapy failure in human breast cancer. *Oncogene* **22,** 4898–4904 (2003).

70. Butt, A. *et al.* Downstream targets of growth factor and oestrogen signalling and endocrine resistance: the potential roles of c-Myc, cyclin D1 and cyclin E. *Endocr. Relat. Cancer* **12,** S47–S59 (2005).

71. Chu, I. M., Hengst, L. & Slingerland, J. M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat. Rev. Cancer* **8,** 253–267 (2008).

72. Pérez-Tenorio, G. *et al.* Cytoplasmic p21WAF1/CIP1 correlates with Akt activation and poor response to tamoxifen in breast cancer. *Int. J. Oncol.* **28,** 1031–42 (2006).

73. Arpino, G. *et al.* Crosstalk between the Estrogen Receptor and the HER Tyrosine Kinase Receptor Family: Molecular Mechanism and Clinical Implications for Endocrine Therapy Resistance. *Endocr. Rev.* **29,** 217–233 (2008).

74. Chakraborty, A. K., Welsh, A. & DiGiovanna, M. P. Co-targeting the insulin-like growth factor I receptor enhances growth-inhibitory and pro-apoptotic effects of anti-estrogens in human breast cancer cell lines. *Breast Cancer Res. Treat.* **120,** 327–335 (2010).

75. Kern, F. G. *et al.* Transfected MCF-7 cells as a model for breast-cancer progression. *Breast Cancer Res. Treat.* **31,** 153–65 (1994).

76. Morgan, L. *et al.* Elevated Src kinase activity attenuates Tamoxifen response in vitro and is associated with poor prognosis clinically. *Cancer Biol. Ther.* **8,** 1550–8 (2009).

77. Shoman, N. *et al.* Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Mod. Pathol.* **18,** 250–259 (2005).

78. Ferguson, K. M. Structure-Based View of Epidermal Growth Factor Receptor Regulation. *Annu. Rev. Biophys.* **37,** 353–373 (2008).

79. Moasser, M. M. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* **26,** 6469–6487 (2007).

80. Carter, P. *et al.* Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 4285–9 (1992).

81. Molina, M. A. *et al.* Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res.* **61,** 4744–9 (2001).

82. Clynes, R. *et al.* Inhibitory Fc receptors modulate in vivo cytoxicity against tumortargets. *Nat. Med.* **6,** 443–446 (2000).

83. Arnould, L. *et al.* Trastuzumab-based treatment of HER2-positive breast cancer: an antibody-dependent cellular cytotoxicity mechanism? *Br. J. Cancer* **94,** 259–267 (2006).

84. Klapper, L. *et al.* Tumor-inhibitory antibodies to HER-2/ErbB-2 may act by recruiting c-Cbl and enhancing ubiquitination of HER-2. *Cancer Res.* **60,** 3384–8 (2000).

85. Nagata, Y. *et al.* PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6,** 117–127 (2004).

86. Junttila, T. T. *et al.* Ligand-Independent HER2/HER3/PI3K Complex Is Disrupted by Trastuzumab and Is Effectively Inhibited by the PI3K Inhibitor GDC-0941. *Cancer Cell* **15,** 429–440 (2009).

87. Zhang, S. *et al.* Combating trastuzumab resistance by targeting SRC, a common node downstream of multiple resistance pathways. *Nat. Med.* **17,** 461–469 (2011).

88. von Minckwitz, G. *et al.* Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N. Engl. J. Med.* **377,** 122–131 (2017).

89. Scheuer, W. *et al.* Strongly enhanced antitumor activity of trastuzumab and pertuzumab combination treatment on HER2-positive human xenograft tumor models. *Cancer Res.* **69,** 9330–6 (2009).

90. Gianni, L. *et al.* Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol.* **13,** 25–32 (2012).

91. Swain, S. M. *et al.* Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. *N. Engl. J. Med.* **372,** 724–734 (2015).

92. Arkin, M. & Moasser, M. M. HER-2-directed, small-molecule antagonists. *Curr. Opin. Investig. Drugs* **9,** 1264–76 (2008).

93. Rusnak, D. W. *et al.* The effects of the novel, reversible epidermal growth factor receptor/ErbB-2 tyrosine kinase inhibitor, GW2016, on the growth of human normal and tumor-derived cell lines in vitro and in vivo. *Mol. Cancer Ther.* **1,** 85–94 (2001).

94. Li, D. *et al.* BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* **27,** 4702–4711 (2008).

95. Geyer, C. E. *et al.* Lapatinib plus Capecitabine for HER2-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **355,** 2733–2743 (2006).

96. Guarneri, V. *et al.* Preoperative Chemotherapy Plus Trastuzumab, Lapatinib, or Both in Human Epidermal Growth Factor Receptor 2–Positive Operable Breast Cancer: Results of the Randomized Phase II CHER-LOB Study. *J. Clin. Oncol.* **30,** 1989–1995 (2012).

97. Gajria, D. & Chandarlapaty, S. HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev. Anticancer Ther.* **11,** 263–275 (2011).

98. Scott, G. K. *et al.* A truncated intracellular HER2/neu receptor produced by alternative RNA processing affects growth of human carcinoma cells. *Mol. Cell. Biol.* **13,** 2247–57 (1993).

99. Christianson, T. A. *et al.* NH2-terminally truncated HER-2/neu protein: relationship with shedding of the extracellular domain and with prognostic factors in breast cancer. *Cancer Res.* **58,** 5123–9 (1998).

100. Scaltriti, M. *et al.* Expression of p95HER2, a Truncated Form of the HER2 Receptor, and Response to Anti-HER2 Therapies in Breast Cancer. *JNCI J. Natl. Cancer Inst.* **99,** 628–638 (2007).

101. Hanker, A. B. *et al.* An AcquiredHER2T798IGatekeeper Mutation Induces Resistance to Neratinib in a Patient with HER2 Mutant-Driven Breast Cancer. *Cancer Discov.* **7,** 575–585 (2017).

102. Xu, X. *et al.* HER2 Reactivation through Acquisition of the HER2 L755S Mutation as a Mechanism of Acquired Resistance to HER2-targeted Therapy in HER2+Breast Cancer. *Clin. Cancer Res.* **23,** 5123–5134 (2017).

103. Alimandi, M. *et al.* Cooperative signaling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. *Oncogene* **10,** 1813–21 (1995).

104. Holbro, T. *et al.* The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc. Natl. Acad. Sci.* **100,** 8933–8938 (2003).

105. Lu, Y. *et al.* Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). *J. Natl. Cancer Inst.* **93,** 1852–7 (2001).

106. Shattuck, D. *et al.* Met Receptor Contributes to Trastuzumab Resistance of Her2-Overexpressing Breast Cancer Cells. *Cancer Res.* **68,** 1471–1477 (2008).

107. Berns, K. *et al.* A Functional Genetic Approach Identifies the PI3K Pathway as a Major Determinant of Trastuzumab Resistance in Breast Cancer. *Cancer Cell* **12,** 395–402 (2007).

108. Kataoka, Y. *et al.* Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. *Ann. Oncol.* **21,** 255–262 (2010).

109. Foster, J. *et al.* Estrogens and cell-cycle regulation in breast cancer. *Trends Endocrinol. Metab.* **12,** 320–7 (2001).

110. Nardone, A. *et al.* The changing role of ER in endocrine resistance. *Breast* **24 Suppl 2,** S60-6 (2015).

111. O'Leary, B., Finn, R. S. & Turner, N. C. Treating cancer with selective CDK4/6 inhibitors. *Nat. Rev. Clin. Oncol.* **13,** 417–30 (2016).

112. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

113. Zwijsen, R. M. *et al.* CDK-independent activation of estrogen receptor by cyclin D1. *Cell* **88,** 405–15 (1997).

114. Finn, R. S. *et al.* PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res.* **11,** R77 (2009).

115. Brufsky, A. M. & Dickler, M. N. Estrogen Receptor-Positive Breast Cancer: Exploiting Signaling Pathways Implicated in Endocrine Resistance. *Oncologist* theoncologist.2017-0423 (2018). doi:10.1634/theoncologist.2017-0423

116. Campbell, I. G. *et al.* Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res.* **64,** 7678–81 (2004).

117. Levine, D. A. *et al.* Frequent mutation of the PIK3CA gene in ovarian and breast cancers. *Clin. Cancer Res.* **11,** 2875–8 (2005).

118. Saal, L. H. *et al.* PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res.* **65,** 2554–9 (2005).

119. Miller, T. *et al.* Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. *Breast Cancer Res.* **13,** 224 (2011).

120. Yardley, D. A. *et al.* Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis. *Adv. Ther.* **30,** 870–84 (2013).

121. Dawood, S. & Rugo, H. S. Targeting the host immune system. *Curr. Opin. Support. Palliat. Care* **10,** 336–342 (2016).

122. O'Shaughnessy, J. Extending Survival with Chemotherapy in Metastatic Breast Cancer. *Oncologist* **10,** 20–29 (2005).

123. Rabbani, S. A. & Mazar, A. P. Evaluating distant metastases in breast cancer: from biology to outcomes. *Cancer Metastasis Rev.* **26,** 663–74 (2007).

124. Valastyan, S. & Weinberg, R. A. Tumor Metastasis: Molecular Insights and Evolving Paradigms. *Cell* **147,** 275–292 (2011).

125. Paget, S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* **8,** 98–101 (1989).

126. Kozłowski, J., Kozłowska, A. & Kocki, J. Breast cancer metastasis - insight into selected molecular mechanisms of the phenomenon. *Postepy Hig. Med. Dosw. (Online)* **69,** 447–51 (2015).

127. Brooks, S. *et al.* Molecular interactions in cancer cell metastasis. *Acta Histochem.* **112,** 3–25 (2010).

128. Tse, J. C. & Kalluri, R. Mechanisms of metastasis: epithelial-to-mesenchymal transition and contribution of tumor microenvironment. *J. Cell. Biochem.* **101,** 816–29 (2007).

129. Lorusso, G. & Rüegg, C. New insights into the mechanisms of organ-specific breast cancer metastasis. *Semin. Cancer Biol.* **22,** 226–33 (2012).

130. Läubli, H. & Borsig, L. Selectins promote tumor metastasis. *Semin. Cancer Biol.* **20,** 169–77 (2010).

131. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

132. Kaplan, R. N., Psaila, B. & Lyden, D. Bone marrow cells in the 'pre-metastatic niche': within bone and beyond. *Cancer Metastasis Rev.* **25,** 521–9 (2006).

133. Weil, R. *et al.* Breast cancer metastasis to the central nervous system. *Am. J. Pathol.* **167,** 913–20 (2005).

134. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410,** 50–6 (2001).

135. Goss, P. E. & Chambers, A. F. Does tumour dormancy offer a therapeutic target? *Nat. Rev. Cancer* **10,** 871–7 (2010).

136. Toft, D. & Gorski, J. A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. *Proc. Natl. Acad. Sci. U. S. A.* **55,** 1574–81 (1966).

137. Kuiper, G. *et al.* Cloning of a novel receptor expressed in rat prostate and ovary. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 5925–30 (1996).

138. Thomas, C. & Gustafsson, J.-Å. Estrogen receptor mutations and functional consequences for breast cancer. *Trends Endocrinol. Metab.* **26,** 467–476 (2015).

139. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).

140. Joseph, R. *et al.* Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. *Mol. Syst. Biol.* **6,** 456 (2010).

141. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481,** 389–93 (2012).

142. Björnström, L. & Sjöberg, M. Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes. *Mol. Endocrinol.* **19,** 833–842 (2005).

143. Fox, E. M., Andrade, J. & Shupnik, M. A. Novel actions of estrogen to promote proliferation: integration of cytoplasmic and nuclear pathways. *Steroids* **74,** 622–7 (2009).

144. Lee, A. V, Cui, X. & Oesterreich, S. Cross-talk among estrogen receptor, epidermal growth factor, and insulin-like growth factor signaling in breast cancer. *Clin. Cancer Res.* **7,** 4429s–4435s; discussion 4411s–4412s (2001).

145. Jeselsohn, R. *et al.* Emergence of Constitutively Active Estrogen Receptor- Mutations in Pretreated Advanced Estrogen Receptor-Positive Breast Cancer. *Clin. Cancer Res.* **20,** 1757–1767 (2014).

146. Garcia, T. *et al.* Identification of a variant form of the human estrogen receptor with an amino acid replacement. *Nucleic Acids Res.* **17,** 8364 (1989).

147. McGuire, W. L., Chamness, G. C. & Fuqua, S. A. Abnormal estrogen receptor in clinical breast cancer. *J. Steroid Biochem. Mol. Biol.* **43,** 243–7 (1992).

148. Karnik, P. *et al.* Estrogen receptor mutations in tamoxifen-resistant breast cancer. *Cancer Res.* **54,** 349–53 (1994).

149. Roodi, N. *et al.* Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *J. Natl. Cancer Inst.* **87,** 446–51 (1995).

150. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep.* **4,** 1116–30 (2013).

151. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-α: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res.* **73,** 6856–64 (2013).

152. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45,** 1446–51 (2013).

153. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA. Cancer J. Clin.* **68,** 7–30 (2018).

154. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA. Cancer J. Clin.* **66,** 7–30 (2016).

155. Aguirre-Ghiso, J. A. Models, mechanisms and clinical evidence for cancer dormancy. *Nat. Rev. Cancer* **7,** 834–46 (2007).

156. Steeg, P. S. Targeting metastasis. *Nat. Rev. Cancer* **16,** 201–218 (2016).

157. Mariotto, A. *et al.* Estimation of the Number of Women Living with Metastatic Breast Cancer in the United States. *Cancer Epidemiol. Biomarkers Prev.* **26,** 809–815 (2017).

158. Arpino, G. *et al.* Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res.* **6,** R149 (2004).

159. Pestalozzi, B. C. *et al.* Distinct Clinical and Prognostic Features of Infiltrating Lobular Carcinoma of the Breast: Combined Results of 15 International Breast Cancer Study Group Clinical Trials. *J. Clin. Oncol.* **26,** 3006–3014 (2008).

160. Zhang, X. *et al.* Metastasis dormancy in estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **19,** 6389–97 (2013).

161. Arpino, G., Bardou, V., Clark, G. & Elledge, R. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res* **6,** R149–R156 (2004).

162. Metzger Filho, O. *et al.* Relative Effectiveness of Letrozole Compared With Tamoxifen for Patients With Lobular Carcinoma in the BIG 1-98 Trial. *J. Clin. Oncol.* **33,** 2772–9 (2015).

163. Sikora, M. J. *et al.* Invasive lobular carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and altered tamoxifen response. *Cancer Res.* **74,** 1463–74 (2014).

164. Harris, M. *et al.* A comparison of the metastatic pattern of infiltrating lobular carcinoma and infiltrating duct carcinoma of the breast. *Br. J. Cancer* **50,** 23–30 (1984).

165. Ferlicot, S. *et al.* Wide metastatic spreading in infiltrating lobular carcinoma of the breast. *Eur. J. Cancer* **40,** 336–41 (2004).

166. He, H., Gonzalez, A., Robinson, E. & Yang, W. T. Distant Metastatic Disease Manifestations in Infiltrating Lobular Carcinoma of the Breast. *Am. J. Roentgenol.* **202,** 1140–1148 (2014).

167. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45,** 1127–33 (2013).

168. Holst, F. *et al.* Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nat. Genet.* **39,** 655–60 (2007).

169. Brown, L. A. *et al.* ESR1 gene amplification in breast cancer: a common phenomenon? *Nat. Genet.* **40,** 806–807 (2008).

170. Horlings, H. M. *et al.* ESR1 gene amplification in breast cancer: a common phenomenon? *Nat. Genet.* **40,** 807-8–2 (2008).

171. Reis-Filho, J. S. *et al.* ESR1 gene amplification in breast cancer: a common phenomenon? *Nat. Genet.* **40,** 809–810 (2008).

172. Vincent-Salomon, A. *et al.* ESR1 gene amplification in breast cancer: a common phenomenon? *Nat. Genet.* **40,** 809–809 (2008).

173. Albertson, D. G. Conflicting evidence on the frequency of ESR1 amplification in breast cancer. *Nat. Genet.* **40,** 821–822 (2008).

174. Holst, F. & Singer, C. F. ESR1-Amplification-Associated Estrogen Receptor α Activity in Breast Cancer. *Trends Endocrinol. Metab.* **27,** 751–752 (2016).

175. cBioPoratal for cancer genomics. Available at: http://www.cbioportal.org. (Accessed: 3rd January 2017)

176. Cancer, T. & Atlas, G. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

177. Desmedt, C. *et al.* Genomic Characterization of Primary Invasive Lobular Breast Cancer. *J. Clin. Oncol.* (2016). doi:10.1200/JCO.2015.64.0334

178. Fumagalli, D. *et al.* Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **27,** 1860–6 (2016).

179. Fu, X. *et al.* FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E6600–E6609 (2016).

180. Matissek, K. J. *et al.* Expressed Gene Fusions as Frequent Drivers of Poor Outcomes in Hormone Receptor Positive Breast Cancer. *Cancer Discov.* CD-17-0535 (2017). doi:10.1158/2159-8290.CD-17-0535

181. Priedigkeit, N. *et al.* Intrinsic Subtype Switching and Acquired *ERBB2* / *HER2* Amplifications and Mutations in Breast Cancer Brain Metastases. *JAMA Oncol.* (2016). doi:10.1001/jamaoncol.2016.5630

182. Hartmaier, R. J. *et al.* Recurrent hyperactive ESR1 fusion proteins in endocrine therapy resistant breast cancer. *Ann. Oncol.* (2018). doi:10.1093/annonc/mdy025

183. Edgren, H. *et al.* Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.* **12,** R6 (2011).

184. Iwakawa, R. *et al.* Genome-wide identification of genes with amplification and/or fusion in small cell lung cancer. *Genes. Chromosomes Cancer* **52,** 802–16 (2013).

185. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–307 (2012).

186. Soule H & McGrath CM. Immortal human mammary epithelial cell lines. *US Pat. 5,026,637* (1991).

187. Hindson, B. J. *et al.* High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Anal. Chem.* **83,** 8604–8610 (2011).

188. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* **45,** 1439–45 (2013).

189. Craig, D. W. *et al.* Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. *Mol. Cancer Ther.* **12,** 104–16 (2013).

190. Chang, K. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45,** 1113–1120 (2013).

191. Griffith, M. *et al.* DGIdb: mining the druggable genome. *Nat. Methods* **10,** 1209–10 (2013).

192. nSolver™ Analysis Software | NanoString Technologies. Available at: http://www.nanostring.com/products/nSolver/.

193. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

194. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).

195. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–4 (2012).

196. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6,** pl1-pl1 (2013).

197. R Core Team (2013). R: The R Project for Statistical Computing. Available at: https://www.r-project.org/.

198. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* btw313 (2016). doi:10.1093/bioinformatics/btw313

199. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2009).

200. Griffin, O. (2016, Dec 19). *Heatmap 3*. Retrieved from https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R.

201. American Assocation for Cancer Research (AACR) Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* **7,** 818–831 (2017).

202. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23,** 703–713 (2017).

203. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–10 (2009).

204. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144,** 27–40 (2011).

205. Saito, M. *et al.* Expression screening of 17q12-21 amplicon reveals GRB7 as an ERBB2-dependent oncogene. *FEBS Lett.* **586,** 1708–1714 (2012).

206. Sahlberg, K. K. *et al.* The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells. *Mol. Oncol.* **7,** 392–401 (2013).

207. Jacot, W. *et al.* The HER2 amplicon in breast cancer: Topoisomerase IIA and beyond. *Biochim. Biophys. Acta - Rev. Cancer* **1836,** 146–157 (2013).

208. Hui, R. *et al.* EMS1 amplification can occur independently of CCND1 or INT-2 amplification at 11q13 and may identify different phenotypes in primary breast cancer. *Oncogene* **15,** 1617–1623 (1997).

209. Shibata, T. *et al.* Genetic Classification of Lung Adenocarcinoma Based on Array-Based Comparative Genomic Hybridization Analysis: Its Association with Clinicopathologic Features. *Clin. Cancer Res.* **11,** 6177–6185 (2005).

210. Zaharieva, B. M. *et al.* High-throughput tissue microarray analysis of 11q13 gene amplification (CCND1, FGF3, FGF4, EMS1) in urinary bladder cancer. *J. Pathol.* **201,** 603–608 (2003).

211. Ormandy, *et al.* Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast Cancer Res. Treat.* **78,** 323–35 (2003).

212. Hui, R. *et al.* EMS1 gene expression in primary breast cancer: relationship to cyclin D1 and oestrogen receptor expression and patient survival. *Oncogene* **17,** 1053–1059 (1998).

213. Borg, A. *et al.* Association of INT2/HST1 coamplification in primary breast cancer with hormone-dependent phenotype and poor prognosis. *Br. J. Cancer* **63,** 136–42 (1991).

214. Borg, A. *et al.* c-myc amplification is an independent prognostic factor in postmenopausal breast cancer. *Int. J. cancer* **51,** 687–91 (1992).

215. Mitsui, F. *et al.* Non-incidental coamplification of Myc and ERBB2 and Myc and EGFR, in gastric adenocarcinomas. *Mod. Pathol.* **20,** 622–631 (2007).

216. Ooi, A. *et al.* Gene amplification of Myc and its coamplification with ERBB2 and EGFR in gallbladder adenocarcinoma. *Anticancer Res.* **29,** 19–26 (2009).

217. Courjal, F. *et al.* DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours. *Br. J. Cancer* **74,** 1984–9 (1996).

218. Holst, F. Estrogen receptor alpha gene amplification in breast cancer: 25 years of debate World Journal of Clinical Oncology. *World J Clin Oncol World J Clin Oncol April* **7,** 160–173 (2016).

219. Tomita, S. *et al.* Estrogen receptor alpha gene ESR1 amplification may predict endocrine therapy responsiveness in breast cancer patients. *Cancer Sci.* **100,** 1012–7 (2009).

220. Ooi, A. *et al.* Gene amplification of ESR1 in breast cancers-fact or fiction? A fluorescence in situ hybridization and multiplex ligation-dependent probe amplification study. *J. Pathol.* **227,** 8–16 (2012).

221. Markiewicz, A. *et al.* Prognostic significance of ESR1 amplification and ESR1 PvuII, CYP2C19*2, UGT2B15*2 polymorphisms in breast cancer patients. *PLoS One* **8,** e72219 (2013).

222. Nielsen, K. V. *et al.* Amplification of ESR1 may predict resistance to adjuvant tamoxifen in postmenopausal patients with hormone receptor positive breast cancer. *Breast Cancer Res. Treat.* **127,** 345–355 (2011).

223. Song, R. X. *et al.* Effect of long-term estrogen deprivation on apoptotic responses of breast cancer cells to 17beta-estradiol. *J. Natl. Cancer Inst.* **93,** 1714–23 (2001).

224. Kota, K. *et al.* Estradiol as a Targeted, Late-Line Therapy in Metastatic Breast Cancer with Estrogen Receptor Amplification. (2017). doi:10.7759/cureus.1434

225. Palmieri, D. *et al.* Her-2 overexpression increases the metastatic outgrowth of breast cancer cells in the brain. *Cancer Res.* **67,** 4190–8 (2007).

226. Wolff, A. C. *et al.* Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *J. Clin. Oncol.* **31,** 3997–4013 (2013).

227. Gillett, C. *et al.* Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. *Cancer Res.* **54,** 1812–7 (1994).

228. Weinstat-Saslow, D. *et al.* Overexpression of cyclin D mRNA distinguishes invasive and in situ breast carcinomas from non-malignant lesions. *Nat. Med.* **1,** 1257–60 (1995).

229. Kenny, F. S. *et al.* Overexpression of cyclin D1 messenger RNA predicts for poor prognosis in estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **5,** 2069–76 (1999).

230. McIntosh, G. G. *et al.* Determination of the prognostic value of cyclin D1 overexpression in breast cancer. *Oncogene* **11,** 885–91 (1995).

231. Yu, Q. *et al.* Requirement for CDK4 kinase function in breast cancer. *Cancer Cell* **9,** 23–32 (2006).

232. Cairns, P. *et al.* Frequency of homozygous deletion at p16/CDKN2 in primary human tumours. *Nat. Genet.* **11,** 210–212 (1995).

233. Cen, L. *et al.* p16-Cdk4-Rb axis controls sensitivity to a cyclin-dependent kinase inhibitor PD0332991 in glioblastoma xenograft cells. *Neuro. Oncol.* **14,** 870–81 (2012).

234. Konecny, G. E. *et al.* Expression of p16 and Retinoblastoma Determines Response to CDK4/6 Inhibition in Ovarian Cancer. *Clin. Cancer Res.* **17,** 1591–1602 (2011).

235. Logan, J. E. *et al.* PD-0332991, a potent and selective inhibitor of cyclin-dependent kinase 4/6, demonstrates inhibition of proliferation in renal cell carcinoma at nanomolar concentrations and molecular markers predict for sensitivity. *Anticancer Res.* **33,** 2997–3004 (2013).

236. Pal, S. *et al.* Exceptional Response to Palbociclib in Metastatic Collecting Duct Carcinoma Bearing a *CDKN2A* Homozygous Deletion. *JCO Precis. Oncol.* 1–5 (2017). doi:10.1200/PO.17.00017

237. Gao, J., Adams, R. P. & Swain, S. M. Does CDKN2A loss predict palbociclib benefit? *Curr. Oncol.* **22,** e498-501 (2015).

238. Finn, R. S. *et al.* The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *Lancet Oncol.* **16,** 25–35 (2015).

239. Goncalves, R. *et al.* New concepts in breast cancer genomics and genetics. *Breast Cancer Res.* **16,** 460 (2014).

240. Gendoo, D. M. A. *et al.* Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32,** 1097–1099 (2016).

241. Green, A. *et al.* Prognostic significance of steroid receptor co-regulators in breast cancer: co-repressor NCOR2/SMRT is an independent indicator of poor outcome. *Breast Cancer Res.* **10,** P63 (2008).

242. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep.* **4,** 1116–30 (2013).

243. Tajiri, R. *et al.* Clonal profiling of mixed lobular and ductal carcinoma revealed by multiplex ligation-dependent probe amplification and fluorescence in situ hybridization. *Pathol. Int.* **64,** 231–6 (2014).

244. Romond, E. H. *et al.* Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer. *N. Engl. J. Med.* **353,** 1673–1684 (2005).

245. Chemotherapy With or Without Trastuzumab After Surgery in Treating Women With Invasive Breast Cancer - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT01275677. (Accessed: 10th March 2018)

246. Smith, I. *et al.* 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *Lancet* **369,** 29–36 (2007).

247. Varga, Z., Tubbs, R. R. & Moch, H. Concomitant detection of HER2 protein and gene alterations by immunohistochemistry (IHC) and silver enhanced in situ hybridization

(SISH) identifies HER2 positive breast cancer with and without gene amplification. *PLoS One* **9,** e105961 (2014).

248. Sauter, G. *et al.* Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. *J. Clin. Oncol.* **27,** 1323–33 (2009).

249. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA. Cancer J. Clin.* **62,** 10–29 (2012).

250. Patro, R. *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14,** 417–419 (2017).

251. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

252. Hartley, S. W. *et al.* QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* **16,** 224 (2015).

253. Priedigkeit, N. *et al.* Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI Insight* **2,** (2017).

254. Simko, T. W. and V. R package 'corrplot': Visualization of a correlation matrix. (2016).

255. Kamburov, A. *et al.* ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res.* **37,** D623–D628 (2009).

256. Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1143

257. Douville, C. *et al.* CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* **29,** 647–648 (2013).

258. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–91 (2016).

259. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–11 (2001).

260. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

261. Mayakonda, A. & Koeffler, H. P. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv* 52662 (2016). doi:10.1101/052662

262. Sikora, M. J. *et al.* WNT4 mediates estrogen receptor signaling and endocrine resistance in invasive lobular carcinoma cell lines. *Breast Cancer Res.* **18,** 92 (2016).

263. Li, C. I. *et al.* Use of menopausal hormone therapy and risk of ductal and lobular breast cancer among women 55-74 years of age. *Breast Cancer Res. Treat.* **145,** 481–9 (2014).

264. Fournier, A. *et al.* Use of Different Postmenopausal Hormone Therapies and Risk of Histology- and Hormone Receptor–Defined Invasive Breast Cancer. *J. Clin. Oncol.* **26,** 1260–1268 (2008).

265. Li, C. I., Uribe, D. J. & Daling, J. R. Clinical characteristics of different histologic types of breast cancer. *Br. J. Cancer* **93,** 1046–52 (2005).

266. Kuwabara, Y. *et al.* Establishment of an ovarian metastasis model and possible involvement of E-cadherin down-regulation in the metastasis. *Cancer Sci.* **99,** 1933–9 (2008).

267. Boussadia, O. *et al.* E-cadherin is a survival factor for the lactating mouse mammary gland. *Mech. Dev.* **115,** 53–62 (2002).

268. Derksen, P. W. B. *et al.* Mammary-specific inactivation of E-cadherin and p53 impairs functional gland development and leads to pleomorphic invasive lobular carcinoma in mice. *Dis. Model. Mech.* **4,** 347–358 (2011).

269. Boelens, M. C. *et al.* PTEN Loss in E-Cadherin-Deficient Mouse Mammary Epithelial Cells Rescues Apoptosis and Results in Development of Classical Invasive Lobular Carcinoma. *Cell Rep.* **16,** 2087–2101 (2016).

270. Pidasheva, S. *et al.* Impaired cotranslational processing of the calcium-sensing receptor due to signal peptide missense mutations in familial hypocalciuric hypercalcemia. *Hum. Mol. Genet.* **14,** 1679–1690 (2005).

271. Egbuna, O. I. & Brown, E. M. Hypercalcaemic and hypocalcaemic conditions due to calcium-sensing receptor mutations. *Best Pract. Res. Clin. Rheumatol.* **22,** 129–148 (2008).

272. Mancilla, E. E., De Luca, F. & Baron, J. Activating Mutations of the Ca2+-Sensing Receptor. *Mol. Genet. Metab.* **64,** 198–204 (1998).

273. Kim, W. & Wysolmerski, J. J. Calcium-Sensing Receptor in Breast Physiology and Cancer. *Front. Physiol.* **7,** 440 (2016).

274. Mamillapalli, R *et al.* Switching of G-protein Usage by the Calcium-sensing Receptor Reverses Its Effect on Parathyroid Hormone-related Protein Secretion in Normal *Versus* Malignant Breast Cells. *J. Biol. Chem.* **283,** 24435–24447 (2008).

275. Saidak, Z. *et al.* Extracellular calcium promotes the migration of breast cancer cells through the activation of the calcium sensing receptor. *Exp. Cell Res.* **315,** 2072–2080 (2009).

276. Popp, T. *et al.* Wnt5a/β-Catenin Signaling Drives Calcium-Induced Differentiation of Human Primary Keratinocytes. *J. Invest. Dermatol.* **134,** 2183–2191 (2014).

277. Ruiz-Pérez, M. V., Henley, A. B. & Arsenian-Henriksson, M. The MYCN Protein in Health and Disease. *Genes (Basel).* **8,** (2017).

278. Dzieran, J. *et al.* MYCN-amplified neuroblastoma maintains an aggressive and undifferentiated phenotype by deregulation of estrogen and NGF signaling. *Proc. Natl. Acad. Sci. U. S. A.* **115,** E1229–E1238 (2018).

279. Mizukami, Y. *et al.* N-myc protein expression in human breast carcinoma: prognostic implications. *Anticancer Res.* **15,** 2899–905

280. Isakoff, S. J. *et al.* Breast Cancer–Associated *PIK3CA* Mutations Are Oncogenic in Mammary Epithelial Cells. *Cancer Res.* **65,** 10992–11000 (2005).

281. Bader, A. G., Kang, S. & Vogt, P. K. Cancer-specific mutations in PIK3CA are oncogenic in vivo. *Proc. Natl. Acad. Sci.* **103,** 1475–1479 (2006).

282. Zhao, J. J. *et al.* The oncogenic properties of mutant p110 and p110 phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc. Natl. Acad. Sci.* **102,** 18443–18448 (2005).

283. Liu, Z. *et al.* Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* **159,** 358–73 (2014).

284. Zhang, Z. *et al.* NCOR1 mRNA is an independent prognostic factor for breast cancer. *Cancer Lett.* **237,** 123–9 (2006).