IDENTIFICATION OF ACQUIRED MOLECULAR DEPENDENCIES IN ADVANCED BREAST AND OVARIAN CANCERS

by

Nolan Priedigkeit

B.S., Biology, University of Oregon, 2008

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Molecular Pharmacology

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Nolan Priedigkeit

It was defended on

June 19, 2017

and approved by

Robert Edwards, M.D., Chair, Obstetrics and Gynecology, Magee-Women's Hospital

Steffi Oesterreich, Ph.D., Professor, Pharmacology and Chemical Biology

Richard Steinman, M.D., Ph.D., Associate Professor, Department of Medicine

George Tseng, Ph.D., Professor, Department of Biostatistics

Dissertation Advisor: Adrian Lee, Ph.D., Professor, Pharmacology and Chemical Biology

IDENTIFICATION OF ACQUIRED MOLECULAR DEPENDENCIES IN ADVANCED BREAST AND OVARIAN CANCERS

Nolan Priedigkeit, Ph.D.

University of Pittsburgh, 2017

Although individual cancers are driven by heterogeneous processes, cancer mortality has a near universal cause—therapy resistance, recurrence and eventual metastasis to vital organs. Despite great advancements in cancer therapies this past decade, outcomes in patients with advanced disease remain static. In these translational studies, using multiple cohorts of longitudinally collected tumor specimens, we test the hypothesis that relapsed cancers are molecularly distinct from primary disease and acquire druggable vulnerabilities throughout their life histories. As a preliminary study, a targeted gene expression analysis was performed to (1) determine differences in breast cancer (BrCa) intrinsic subtypes between primary tumors and matched brain metastases (BrM) and (2) explore if druggable targets are acquired in metastases. While BrM generally retain their intrinsic molecular subtypes, even after years of dormancy, nearly all gain expression of clinically actionable genes—most notably HER2 (35% of cases). To further assess molecular features acquired in metastases, exome-capture RNA-sequencing on decade-old and degraded tumor specimens was evaluated. Applying this technology, transcriptome-wide acquisitions in BrM were discovered, including highly recurrent expression gains in *RET* (38% of cases). Targeting RET or HER2 using *in vitro*, *ex vivo*, and *in vivo* models produced marked

responses, suggesting RET and HER2-driven signaling as prime targets for patients with BrM. The same approach was applied to estrogen receptor [ER]-positive BrCa bone metastases, which discerned further site-specific acquisitions—such as shifts to Her2 and LumB phenotypes, temporally influenced expression evolution and druggable gains in CDK-Rb-E2F and FGFR-signaling pathways. To determine if these changes are consistent in non-metastatic samples, both RNA expression and DNA changes were assessed in a cohort of ER-positive local recurrences. Limited DNA-level changes, yet highly recurrent transcriptional remodeling events were observed—in particular, losses of *ESR1,* gains of *NTRKs* and upregulation of the cancer stem cell marker *PROM1*. Lastly, these findings were corroborated in ovarian cancer recurrences, where we show fusion RNA transcripts and recurrent outlier expression gains (*NTRK2, INHBA* and *IGF1)* are acquired in relapsed disease. Taken together, these studies establish that cancer recurrences commonly acquire multimodal and readily druggable molecular dependencies, unique from primary tumors, which may have profound clinical implications.

TABLE OF CONTENTS

PREFACE ..... **XVI**

1.0 INTRODUCTION **1**

1.1 BREAST CANCER **1**

1.1.1 Breast cancer subtypes and genomic features **2**

1.1.2 Breast cancer treatment modalities and resistance mechanisms **4**

1.1.2.1 Endocrine therapy **5**

1.1.2.2 HER2 targeted therapy **7**

1.1.2.3 Chemotherapy **9**

1.1.2.4 PI3K/AKT/mTOR and CDK4/6 Targeted Therapy **9**

1.1.3 Breast cancer recurrence and metastasis **11**

1.2 OVARIAN CANCER **17**

1.2.1 Ovarian cancer and genomic features **18**

1.2.2 Ovarian cancer recurrence and chemoresistance **19**

1.3 HYPOTHESIS **21**

2.0 INTRINSIC SUBTYPE SWITCHING AND HER2 GAINS IN BREAST CANCER BRAIN METASTASES **23**

2.1 ABSTRACT **23**

2.2 INTRODUCTION **25**

2.3 MATERIALS AND METHODS 26

2.3.1 Patient samples 26

2.3.2 Tissue processing 28

2.3.3 Clustering and molecular subtyping 28

2.3.4 Recurrent expression alterations 29

2.3.5 Immunohistochemistry 29

2.3.6 Copy number alteration and single nucleotide variant analysis 30

2.3.7 FoundationOne *ERBB2* alterations 31

2.4 RESULTS 31

2.4.1 Expression concordance between primary tumors and matched brain metastases 31

2.4.2 Distinct expression gains of clinically actionable genes 33

2.4.3 DNA-level HER2 acquisitions in breast cancer brain metastases 36

2.4.4 ERBB2 amplifications and SNVs enrichment in brain metastases 36

2.5 DISCUSSION 39

3.0 EXOME-CAPTURE RNA-SEQUENCING OF DECADE-OLD BREAST CANCERS AND MATCHED DECALCIFIED BONE METASTASES IDENTIFIES CLINICALLY ACTIONABLE TARGETS 43

3.1 ABSTRACT 43

3.2 INTRODUCTION 45

3.3 MATERIALS AND METHODS 46

3.3.1 Sample acquisition 46

3.3.2 Tissue processing and RNA extraction 47

3.3.3 Exome-capture RNA-sequencing 47
3.3.4 RNA-sequencing expression quantification and normalization 48
3.3.5 Expression correlations and RNA-seq quality assessment..... 48
3.3.6 *tumorMatch* patient-matched sample identifier 49
3.3.7 Unsupervised hierarchical clustering and intrinsic subtyping 49
3.3.8 Differential gene expression 50
3.3.9 ssGSEA signatures and METABRIC survival analyses 50
3.3.10 Ranked Gene Set Enrichment Analysis (GSEA)..... 51
3.3.11 RBBP8 survival analysis..... 51
3.3.12 Gains and losses in clinically actionable genes 52
3.3.13 Statistical considerations 52
3.4 RESULTS 53
3.4.1 ecRNA-sequencing of aged and decalcified breast cancers..... 53
3.4.2 ecRNA-seq of breast cancer bone metastases..... 54
3.4.3 Clustering and temporal expression shifts 57
3.4.4 Differentially expressed genes..... 59
3.4.5 Dysregulated gene sets and RBBP8 expression loss 61
3.4.6 Expression gains and losses in clinically actionable genes. 62
3.5 DISCUSSION..... 65
4.0 TRANSCRIPTOME-WIDE IDENTIFICATION OF RET AND HER2 SIGNALING AS RECURRENTLY ENRICHED DEPENDENCIES IN BREAST CANCER BRAIN METASTASES..... 72
4.1 ABSTRACT..... 72

4.2 INTRODUCTION 74

4.3 MATERIALS AND METHODS 74

4.3.1 Patient and tumor samples 74

4.3.2 Exome-capture RNA-sequencing 75

4.3.3 RNA-sequencing expression quantification and normalization 75

4.3.4 RNA-seq quality assessment 76

4.3.5 tumorMatch patient-matched sample identifier 76

4.3.6 Differential gene expression 76

4.3.7 Correspondence Analysis 77

4.3.8 Unsupervised hierarchical clustering, intrinsic subtyping and HER2 signature 77

4.3.9 Merging of publicly available microarray breast cancer datasets. 78

4.3.10 Contamination model construction 79

4.3.11 Microenvironmental gene deconvolution 79

4.3.12 Brain metastasis-free survival analysis 80

4.3.13 Gene Set Variation Analysis (GSVA) 81

4.3.14 DNA methylation profiling 81

4.3.15 DNA methylation analysis 82

4.3.16 Gains and losses in clinically actionable kinases 82

4.3.17 Immunohistochemistry (IHC) Staining 83

4.3.18 *In vitro* studies 84

4.3.19 Patient-derived brain metastases *ex vivo* culture 85

4.3.20 *In vivo* experiments 85

4.3.21 Test and Control Agents 86
4.3.22 Statistical Considerations 87
4.4 RESULTS 88
4.4.1 Transcriptome evolution in breast cancer brain metastasis 90
4.4.2 Recurrent expression gains of clinically actionable kinase pathways in breast cancer brain metastases 94
4.4.3 Inhibition of RET and HER2 demonstrates significant anti-tumor activity in breast cancer brain metastases *ex vivo* and *in vivo* 97
4.5 DISCUSSION 101
5.0 RECURRENT TRANSCRIPTIONAL REMODELING EVENTS IN LONG-TERM ESTROGEN-DEPRIVED BREAST CANCER RECURRENCES 104
5.1 ABSTRACT 104
5.2 INTRODUCTION 106
5.3 METHODS AND MATERIALS 108
5.3.1 Patient samples, tissue processing and nucleic acid extractions 108
5.3.2 RNA and DNA-sequencing 109
5.3.3 RNA-sequencing quantification and DNA-sequencing alignment 110
5.3.4 DNA-seq recurrence enriched variant determination 110
5.3.5 Copy number alterations 111
5.3.6 Differential gene expression, clustering and outlier gains and losses 112
5.4 RESULTS 114
5.4.1 Expression and copy number alteration profile conservation 114
5.4.2 SNV enrichments and differentially expressed genes 116

5.4.3 Outlier expression gains and losses 119
5.4.4 *ESR1* depleted recurrences 121
5.5 DISCUSSION 124
6.0 ACQUIRED MOLECULAR FEATURES IN RECURRENT CHEMORESISTANT OVARIAN CANCERS 130
6.1 ABSTRACT 130
6.2 INTRODUCTION 132
6.3 METHODS AND MATERIALS 133
6.3.1 Patient Samples 133
6.3.2 Total RNA-sequencing 135
6.3.3 Expression analyses 135
6.3.4 RNA Fusion Detection and RT-PCR validations 136
6.4 RESULTS 137
6.4.1 Acquired expression gains in ovarian cancer recurrences 137
6.4.2 Preserved and acquired fusion RNAs in ovarian cancer recurrences . 141
6.5 DISCUSSION 144
7.0 CONCLUSIONS 148
APPENDIX A 153
APPENDIX B 173
APPENDIX C 175
REFERENCES 176

LIST OF TABLES

Table 1: Abridged clinicopathological features of brain metastasis cases 27

Table 2: Clinicopathological features of breast cancer bone metastasis cohort 55

Table 3: Comprehensive clinical data of brain metastasis RNA-seq cohort 89

Table 4: Clinical features of local recurrence cohort..... 109

Table 5: Differentially expressed genes in long-term estrogen-deprived local recurrences..... 118

Table 6: Differentially expressed genes in *ESR1* depleted recurrences..... 123

Table 7: Early and late disease, patient-matched ovarian cancer cases..... 134

Table 8: Multi-gene test classifications in patient-matched pairs..... 173

Table 9: Fusion validation primers 174

LIST OF FIGURES

Figure 1: Breast cancer subtype, treatments and resistance mechanisms 4

Figure 2: Metastatic breast cancer site frequencies 12

Figure 3: Metastatic breast cancer overall survival 13

Figure 4: Metastatic breast cancer SNV enrichments 22

Figure 5: Transcriptional similarity between primary breast cancers and matched brain metastases 32

Figure 6: Recurrent expression alterations in breast cancer brain metastases 34

Figure 7: Expression alterations in clinically actionable genes. 35

Figure 8: *ERBB2* gains in breast cancer brain metastases 38

Figure 9: Exome-capture RNA-sequencing of aged, FFPE and decalcified tumors 56

Figure 10: Unsupervised clustering, intrinsic subtype shifts and temporal evolution of ER-positive bone metastases. 58

Figure 11: Differentially expressed genes in patient-matched bone metastases 60

Figure 12: Dysregulated gene sets and *RBBP8* loss in breast cancer bone metastases 62

Figure 13: Recurrent, clinically actionable expression gains and losses in ER-positive bone metastasis 64

Figure 14: Transcriptome evolution in breast cancer brain metastasis. 92

Figure 15: Recurrent expression gains of clinically actionable kinase pathways in breast cancer brain metastases. 95

Figure 16: Inhibition of RET and HER2 demonstrates significant anti-tumor activity in breast cancer brain metastases *ex vivo* and *in vivo*. 99

Figure 17: Transcriptional and CNA conservation in ER-positive local recurrences 115

Figure 18: SNV enrichments in ER-positive local recurrences. 117

Figure 19: Outlier expression gains and losses in ER-positive local recurrences 120

Figure 20: *ESR1* depleted recurrences 122

Figure 21: Unsupervised clustering and *tumorMatch* in ovarian cancer cohort 139

Figure 22: Differentially expressed genes and outlier expression events in ovarian cancer recurrences 140

Figure 23: Fusion RNA landscape in recurrent ovarian cancer 143

Figure 24: Fold change density distribution in patient-matched pairs 153

Figure 25: ER expression loss in breast cancer brain metastases 154

Figure 26: Expression correlation plots of ecRNA-seq sample sets 155

Figure 27: ecRNA-seq QC metrics for patient-matched sample cohort 156

Figure 28: tumorMatch: Proportion of shared variants (POSV) between samples in patient-matched cohort 156

Figure 29: Gene Ontology: Biological Processes (GO:BP) gene overlaps for differentially expressed gene sets 157

Figure 30: Case-specific expression fold-change distributions and expression alteration thresholds 158

Figure 31: All recurrent expression alterations in clinically actionable genes 159

Figure 32: Brain metastasis cohort RNA-seq quality metrics. 160

Figure 33: *tumorMatch* in brain metastasis cohort. 161

Figure 34: Gene-level fold change density distribution in brain metastasis cohort. 161

Figure 35: Cabozantinib and afatinib efficacy in MDA-231-BRM2 and LY2 models 162

Figure 36: Identified brain metastasis genes predict brain-relapse free survival in primary breast tumors. 163

Figure 37: Brain metastasis gene deconvolution 164

Figure 38: Metastasis-free survival with deconvoluted brain metastasis genes 165

Figure 39: HER2 and ER in breast cancer brain metastases. 166

Figure 40: RET protein expression in brain parenchyma and metastases. 167

Figure 41: DNA-seq target interval coverages 168

Figure 42: Copy number call distribution and correlation with expression values 169

Figure 43: *tumorMatch* in local recurrences cohort 170

Figure 44: *KLK7* and *PROM1* basal breast carcinoma expression 170

Figure 45: Overlap of differentially expressed genes between local recurrences and ER+ LTED lines 171

Figure 46: NDRG1 expression in PAM50 subtypes and survival influence in ER-positive breast cancer 172

Figure 47: RT-PCR validation of late-disease acquired fusions. 172

PREFACE

This work is the product of an outstanding group and builds upon a foundation of scientific mentors, role models and friends I've had the privilege to learn from throughout the years. Firstly, Dr. Adrian Lee has been an ideal thesis advisor. He was one of the first people I met in Pittsburgh and it was largely an opportunity to work with him, and the group he and Dr. Steffi Oesterreich put together at the Women's Cancer Research Center, that drew me to this program. I have learned so much in his lab—how to approach scientific questions, how to mentor, how to lead and how to navigate academia and research pragmatically and with integrity. I will forever be in debt to him for always putting my training first, taking the time to teach me so much, challenging me and entrusting me with such compelling projects. I look forward to continuing our work together and seeing what new ideas we can come up with.

I also thank my previous scientific mentors. Dr. Eric Selker inspired a love for basic science at the University of Oregon, Dr. Roderick Capaldi and Dr. Petr Hájek showed me how to be a bench scientist, devise a hypothesis and execute an experiment, and Dr. John Martignetti introduced me to genetics and genomics, serving as an inspiration for becoming a physician scientist and pursuing translational research.

I of course thank the members of my thesis committee. Dr. Richard Steinman, for both providing a critical, scientifically tuned eye and being a fantastic physician scientist role model, Dr. Robert Edwards for his sharp clinical and translational research perspective, Dr. George Tseng for his exceptional bioinformatics and biostatistics expertise and Dr. Steffi Oesterreich—for her amazing scientific intuition and unwavering support of students. I'd also like to thank Dr. Nancy Davidson for serving as my career advisor throughout graduate school, Dr. Bruce Freeman and Dr. Patrick Pagano for leading a stellar, supportive and interdisciplinary graduate program as well as the leadership and my fellow collegues of the University of Pittsburgh MSTP.

I show my deepest gratitude to the members of the Women's Cancer Research Center. Dr. Ryan Hartmaier introduced me to bioinformatics, made it approachable and gave me the courage and confidence to pursue it. Dr. David Boone, one of the best scientific educators I've come across, always brought something valuable to any scientific—or not-so-scientific—discussion and is an inspiring mentor. And to my fellow colleagues in the lab—Ahmed, Ali, Amir, Becky, Beth, Courtney, Daniel, Emily, Gonghong, Julie, Kevin, Nadine, Nick, Rekha, Tian, Tiffany and Zeynep—it's been so much fun to learn from and work with all of you. I know it'll be tough to find such a well-rounded group in the future.

Our collaborators were essential to these projects. Dr. Peter Lucas contributed vital insight—as he served as our core pathologist. William Horne, Dr. Carlos Castro and Christina Kline provided sequencing, tissue processing and tissue acquisition support, respectively. Dr. Rebecca Watters was a key contributor to Chapter 3 and Dr. Ryan Hartmaier was instrumental to Chapters 2 and 4. Particular credit is due to our collaborators from the Royal College of Surgeons (RCSI) in Ireland—Dr. Damir Varešlija, Dr. Ailis Fagan and Dr. Leonie Young. Dr. Varešlija and Dr. Young were co-leaders in the design of Chapter 4 and their group performed

the crucial cancer modeling work and kinase inhibitor studies. As such, they stand as co-first/corresponding authors. Dr. Ailis Fagan and Li Zhu from Dr. George Tseng's lab provided computational support, particularly with the survival, methylation and deconvolution analyses.

Without our clinical collaborators—Drs. Sarah Taylor, Robert Edwards, Kunle Odunsi, Shannon Puhalla, Pablo Leone, Rohit Bhargava and Ronald Hamilton—this work would not be possible. In particular, I thank Dr. Adam Brufsky, not only for contributing to this work, but also for serving as a clinical clerkship mentor during graduate school. Moreover, the unnamed patients and clinicians who contribute to research efforts must be recognized. Without their support, advocacy and sacrifice, this type of research would not be possible.

Finally, I would like to thank my close friends and family. Brady, Derek, Jack, Jared, Matt, Ryan and Sam—I'm so appreciative that we stay in touch, you derelicts keep me grounded and laughing. I'd like to thank my parents, Dan and Cathy, and my siblings, Rob, Taycee, Zach, Danny and Kristine—for their support, providing me with a sense of humor and reminding me to never take anything too seriously. And most importantly, I thank my partner—for her strength, support, patience, humor, smile and love.

Thank you all for helping me find and enabling a pursuit that feels meaningful.

I am endlessly grateful.

1.0 INTRODUCTION

1.1 BREAST CANCER

Breast cancer represents the greatest burden of noncutaneous cancer in women—responsible for approximately 233,000 new diagnoses and over 40,000 deaths annually[1]. Despite these numbers, in the past decades, outcomes in breast cancer have encouragingly grown steadily with current 5-year survival estimates well over 90%[2]. This progress has largely been due to earlier detection and an elevated understanding of the disease's biology, which has led to effective targeted therapies in the form of small molecules and biologics.

Nonetheless, like many epithelial cancers, breast cancer remains a complex and heterogeneous disease. Although groundbreaking success has been made in identifying targetable drivers in breast cancer, there is still a gap in knowledge on how breast cancers progress—particularly what mechanisms drive a breast cancer's ability to resist therapies and metastasize. This lack of understanding has in part resulted in stagnant outcomes in patients with advanced or recurrent disease. For the past decade, 5-year survival rates for patients with distant metastases is unchanged at approximately 25%[3]. Thus, to identify therapeutic targets in patients with advanced disease, we sought to characterize molecular features that make metastatic or therapy resistant tumors distinct from treatment-responsive primary tumors.

1.1.1 Breast cancer subtypes and genomic features

Breast cancer has served as a prototype for targeted therapies in epithelial cancers. One of the first major observations defining a molecular "cancer subtype"—that is, a subset of tumors from the same tissue of origin with distinct molecular features—was the association between breast cancer estrogen receptor [ER] expression and responses to adrenalectomies by Jensen *et al*. in 1971[4]. Further studies into this association solidified the concept that a large proportion of breast cancers are driven by estrogen—ultimately leading to the development and wide clinical adoption of estrogen-depleting therapies. In the late 1980s, the *ERBB2* gene was found to be amplified and its protein product, human epidermal growth factor receptor 2 (HER2), highly expressed in approximately 20% of breast cancers[5]. A humanized monoclonal antibody that targets HER2, trastuzumab, was rationally developed a decade later and is now standard of care. This concept of tailoring a treatment decision to the molecular properties of individual tumors was inaugurated in breast cancer—and is now a driving force for precision medicine in oncology[6]. With the advent and wide adoption of genome-wide interrogation techniques such as microarrays and massively parallel sequencing, even greater insight into breast cancer subtypes and potentially targetable cancer drivers has been elucidated.

Perhaps the most well established genome-wide interrogation of breast cancer was the discovery of the intrinsic molecular subtypes—subclasses of breast cancers with similar expression profiles. Preceding this work, breast cancer subtyping was limited by analyzing protein expression of single molecules such as ER and HER2. When performing unsupervised hierarchical clustering on the entire transcriptome, breast tumors predictably segregated ER-

positive and ER-negative tumors; however, more subtypes emerged[7–9]. These five observed subtypes (LumA, LumB, HER2, Basal and Normal-like) were further refined to an expression-based classifier of 50 genes, now colloquially known as the PAM50, and carried prognostic consequences that was previously unappreciated[10]. Consequently, a version of the PAM50 classifier was recently FDA-approved for identifying risk of distant recurrence in breast cancer patients (Prosigna™) and inspired the development of other multi-gene expression-based tests routinely used in the clinic to guide treatment decisions, such as MammaPrint™ and Oncotype DX™ [11,12].

Further granulation of breast cancer was elucidated by larger scale collaborative molecular characterization efforts such as The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)[13,14]. Despite the additional characterization of thousands of tumors, recurrent driver mutations were quite rare—only *PIK3CA, TP53* and *GATA3* were mutated in greater than 10% of tumors in both TCGA and METABRIC— and the core intrinsic subtypes still serve as the dominant classifiers to place breast cancers into a biological framework[15]. It must be noted; however, that within the major subtypes, further delineations can be made. As an example, basal cancers alone can be broken up into many distinct biological categories based on transcriptome expression—such as basal immunomodulatory, luminal androgen receptor and mesenchymal-like subtypes[16,17].

Importantly, the molecular taxonomy of breast cancer has been exclusively defined in treatment-naïve and relatively indolent primary tumors. How this taxonomy changes or remains consistent following therapies, such as in the context of recurrence or metastases, and the clinical implications of these alterations remain unaddressed and largely unknown.

1.1.2 Breast cancer treatment modalities and resistance mechanisms

Adjuvant breast cancer treatment can be classified into four major categories (1) local therapy, which includes surgery and radiotherapy, (2) hormone therapy, (3) targeted therapy and (4) chemotherapy (Figure 1A). Although the ultimate choice will be made by the physician and patient—treatments are relatively consistent based on the molecular subtype of the tumor. Patients with hormone receptor positive or luminal tumors are almost unambiguously provided an estrogen depleting therapy, patients with HER2-positive tumors are generally offered HER2-targeted therapies and patients with triple-negative or basal breast cancers, which lack expression of any targets, are uniformly offered chemotherapy[18–20]. Patients with ER-positive and/or HER2-positive breast cancers identified as "high risk" for recurrence—either in their clinical presentation or with a companion multi-gene expression-based test—may also be provided with chemotherapy[11,12,21].

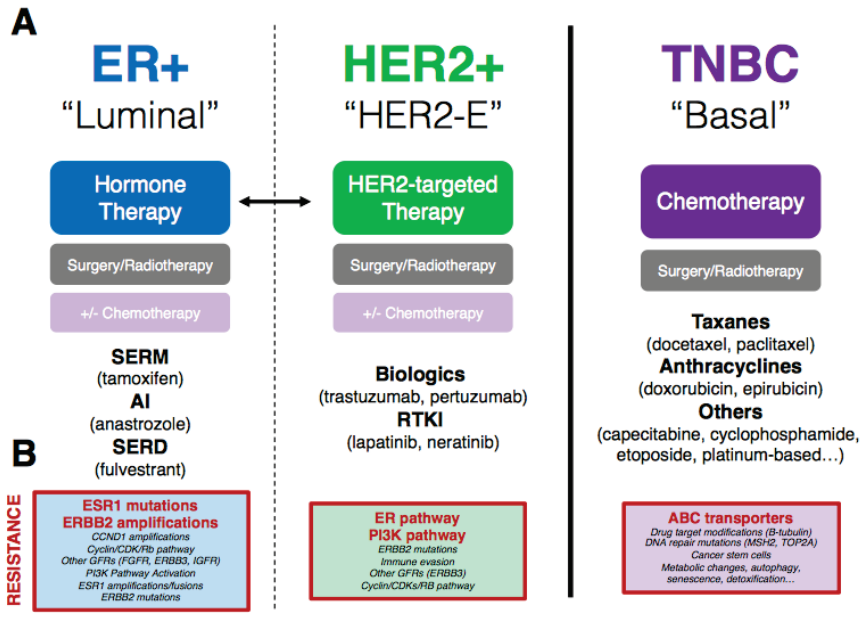

Figure 1: Breast cancer subtype, treatments and resistance mechanisms

1.1.2.1 Endocrine therapy

The mainstay treatment of ER-positive breast cancer is endocrine therapy, which aims to limit 17ß-estradiol (E2) action in breast cancer cells. E2 is the most potent estrogen steroid hormone, diffuses through cell membranes, binds to its receptors (ER-alpha, encoded by *ESR1* and ER-beta) and causes ER dimerization. The dimerized, E2-bound complex then binds to DNA and mediates transcription of a host of E2-regulated genes[22]. This regulation is further influenced by the presence or absence of a series of coregulators, such as *NCOA* family members, and chromatin interactions[23–25]. The ER complex can also bind to other transcription factors (e.g. SP1) and trigger more indirect effects on transcription[26]. Furthermore, non-genomic action of ER has been observed, whereby E2 binding can rapidly induce mitogenic signaling cascades—including IGF, PI3K and MAPK-driven pathways[27–30]. Kinase regulators of these pathways can also act bidirectionally, phosphorylating ER-alpha and regulating its activation[31,32]. In breast cancer cells, this complex and multi-mechanistic ER-driven action ultimately triggers an agglomeration of cancer phenotypes such as cellular proliferation. Modulating this pathway has proven remarkably effective in improving outcomes in patients with breast cancer.

Eliminating E2 effects on breast cancer cells can be accomplished surgically via ovarian ablation or pharmacologically. The first small molecule developed for this purpose was tamoxifen, discovered in the late 1960s, with the first preclinical evidence of its efficacy in breast cancer emerging in the 1970s[33,34]. Tamoxifen is a selective estrogen receptor modulator (SERM) that competes with E2 binding and acts as an antagonist of ER activity in breast cancer cells. Importantly, as their name implies, SERMs activity on estrogen mediated signaling is variable and tissue dependent—as tamoxifen carries agonist properties in the endometrium and bone[35–37]. Another class of pharmacologics, strict competitive antagonists, is selective estrogen

receptor degraders (SERDs) such as fulvestrant—which bind to ER, limit its dimerization and ultimately trigger ER protein degradation[38–41]. Lastly, aromatase inhibitors (AI) are often used in post-menopausal women, which block the conversion of androgens into estrogens by inhibiting the enzyme aromatase in the periphery—mainly within adipose tissue[42]. The use of these three classes of agents has provided vast improvements in recurrence-free and overall survival in patients with breast cancer, yet therapy resistance mechanisms are emerging in the context of long-term recurrences and metastases[43,44].

Many endocrine resistance mechanisms have been elucidated in breast cancer, especially through *in vitro* models—particularly long-term estrogen deprived (LTED) cell lines that have become estrogen independent—and correlating outcome data with molecular or clinical features. The latter method has established that breast cancers can be intrinsically resistant to endocrine therapies, driven collectively by a more proliferative, Luminal B phenotype—which coincides with higher tumor grade, greater Ki67 indices, higher risk scores from multigene signature tests and somatic mutations such as *TP53* and *RUNX1*[45].

Identifying and validating mediators of acquired long-term endocrine therapy resistance in patient samples is challenging due to long periods of disease dormancy in ER-positive patients. Nonetheless, recurrent mechanisms of resistance have emerged. Up to 20% of relapsed tumors tend to lose expression of ER, and some have been shown to activate other growth factor receptors (e.g. *FGFR1, ERBB2, IGF1R)* in its place to preserve mitogenic pathways including MAPK and PI3K[46–50]. Recent sequencing studies have shown that alterations in the *ESR1* gene is a common mechanism of acquired endocrine therapy resistance, as over 20% of ER-positive metastatic breast cancers harbor hotspot *ESR1* mutations—generally occurring in the ligand-binding domain and conferring ligand-independent ER-activity[51–53]. Interestingly, preliminary

evidence suggest *ESR1* mutations are more common in AI-treated therapy, indicating pharmacologic class-specific mechanisms of endocrine therapy resistance may be at play[54]. Alterations in the Rb-CDK-E2F pathway have been proposed as mediators of therapy resistance, such as cyclin D1 (*CCND1)* amplifications, which may explain recent success with cell-cycle inhibitor therapies (Section 1.1.2.4)[13,44]. Other, less widely accepted mechanisms of endocrine resistance have been discovered using cell models, yet still must be validated in more treatment-resistant patient specimens to confirm their clinical importance and actionability—such as *ESR1* fusions and amplifications, stem-cell and mesenchymal-like cell state changes and microenvironment interactions[55,56].

1.1.2.2 HER2 targeted therapy

The second major class of targeted therapy in breast cancer is those targeting HER2. HER2 is a receptor tyrosine kinase, a member of the epidermal growth factor receptor family and is overexpressed and/or mutated in approximately 20% of breast cancers. HER2 signaling is driven by a complex program at the cell surface, whereby it can homo- and heterodimerize with other EGFR members, become phosphorylated and potentiate downstream mediators which encourage cell growth and anti-apoptotic programs—the most well established being the RAS-MAPK and PI3K pathways[57,58]. Two pharmacological options exist for targeting HER2—biologics and small molecule inhibitors.

Trastuzumab is the first rationally designed therapy to target HER2 and although its mechanism of action is complex, the humanized monoclonal antibody has been shown to inhibit downstream signaling by interfering with dimerization and HER2 shedding, increasing HER2 degradation and eliciting antibody-dependent cell-mediated cytotoxicity[59–63]. Pertuzumab, a more recently adopted monoclonal antibody, acts similarly but shows more efficacious and

specific inhibition of HER2 dimerization with ligand-bound HER3. Considering their different epitopes and distinct properties, the combination of the two unsurprisingly show overall survival benefits in patients with metastatic HER2-positve breast cancers[64–67]. The other class of HER2 agents, small molecule tyrosine kinase inhibitors, also show efficacy in HER2-positive disease. This includes lapatinib—a reversible inhibitor that binds to the ATP-binding pocket of the intracellular domain of HER2 preventing phosphorylation—and neratinib—an irreversible inhibitor that binds covalently to cysteine residues and limits the action of the catalytic domain of EGFR family members[68–71]. Importantly, small molecules targeting HER2 may be more efficacious for breast cancer brain metastases, given their ability to cross the blood-brain-barrier more easily than bulky biologics[72].

Resistance to HER2-therapies has been appreciated both intrinsically and in the acquired setting. Like *ESR1,* somatic mutations in *ERBB2* can confer therapy resistance—including a truncated form of HER2 stemming from an alternate transcription initiation site that eliminates the N-terminal extracellular binding region of trastuzumab, p95HER2, and a splice variant, HER2Δ16, that lacks exon 16 and induces stronger dimerization[73,74]. Gatekeeper mutations are emerging as mechanisms of acquired resistance, as was observed in a patient with an activating L869R mutation that was initially responsive to neratinib but became resistant following the acquisition of a T798I mutation[75]. Moreover, L755S mutations can be selected for in HER2-amplified cell lines under the selective pressures of HER2-therapies, suggesting cancers with already amplified *ERBB2* can reestablish therapeutically blunted HER2 signaling by another "hit" in the form of an activating mutation[76]. Other proposed mechanisms of resistance, which still warrant careful evaluation in human specimens and trial settings, are activation of downstream mediators via alternative mechanism—such as *PIK3CA* mutations[77], *PTEN* loss[78,79]

and upregulation of alternate growth factor receptors including EGFR, IGF1R and HER3[80–83]. What has cultivated appreciation recently is trastuzumab's immunologically driven antitumor activity. Trastuzumab's effectiveness can be predicted by a signature of immune genes and the amount of tumor infiltrating lymphocytes[84,85], suggesting resistance mechanisms could be imparted by an altered host immune response. Finally, as discussed above, cross-talk with ER-signaling and activation of cell cycle mediators such as cyclin D1 and CDK4 has been instigated as a mechanism of HER2-therapy resistance[86,87].

1.1.2.3 Chemotherapy

Chemotherapy in breast cancer is initiated for patients with high risk clinical phenotypes, particularly the basal or triple-negative subtype, and metastatic disease. Chemotherapy choice, unlike the aforementioned targeted therapies, is much more variable. Nonetheless, major classes have emerged as more effective than others with combination anthracycline and taxane-based therapies being the standard of care for most patients[88]. Outside of anthracycline and taxane-based treatments, regimens are commonly supplemented with other cytotoxic compounds including cyclophosphamide, etoposide, capecitabine, gemcitabine, vinorelbine and platinum-based therapies such as cisplatin—among others depending on the patient's course, specific responses and tolerances[89]. Given overlapping agents, mechanisms of chemotherapy resistance in the context of breast and ovarian cancers are discussed in Section 1.2.3.

1.1.2.4 PI3K/AKT/mTOR and CDK4/6 Targeted Therapy

Targeting the PI3K/AKT/mTOR axis has been a prime area of investigation in breast cancer given this pathway is a downstream effector of HER2 and ER signaling. Although mTOR inhibitors such as temsirolimus and everolimus have shown promising preclinical evidence,

identifying subsets of patients that benefit most from these compounds has been challenging, especially given additional toxicities. Recent evidence suggests that cancers with hyperactive PI3K pathway activation (e.g. tumors with *PIK3CA* mutations or *PTEN* loss) may show greater response rates[90–93]. Nonetheless, mTOR inhibitors remain an option for patients with advanced, hormone-therapy resistant, metastatic breast cancer with the caveat that large-scale trials have shown minimal gains in overall survival. Promising early phase investigations have recently been completed using PI3K pathway inhibitors, particularly for patients with tumors that harbor *PIK3CA* mutations[94,95]—although it is premature to predict their utility.

Another recent therapy, likely to become widely adopted in hormone receptor positive breast cancers, is targeting CDK4/6—two key kinases responsible for the transition from G0/G1 to S-phase of the cell cycle. Dysregulation of the cell cycle is an established hallmark of cancer; yet, given its ubiquity, targeting this pathway with non-selective inhibitors has been wrought with difficulty and toxicities[96,97]. Some estrogen-receptor positive breast cancers; however, are uniquely dependent on cell cycle mediators, largely mediated by cyclin D1 (*CCND1)* given its amplification, overexpression and association with worse outcomes in a subset of ER-positive breasts cancers[98,99]. Indeed, estrogen signaling is intimately linked to the CDK pathway, as cyclin D1 potentiates E2-regulated genes in the absence of estrogen in breast cancer cells, binds to ER to influence its regulation and can also aid in recruiting transcriptional modulating cofactors[100–103]. Given this link, breast cancer has served as a model to test the efficacy of more selective CDK4/CDK6 inhibitors—particularly ribociclib and palbociclib. These compounds bind reversibly and selectively to the ATP-binding pocket of CDK4 and CDK6, blocking their phosphorylation of the retinoblastoma protein (Rb)—ultimately stalling the cell cycle to G1 and diminishing downstream activation of E2F transcription factors[104,105]. Promising preclinical data

using these inhibitors in luminal breast cancer models eventually paved the way towards clinical trials—where both palbociclib and ribociclib have shown promising trends of longer progression-free survival in patients with advanced disease, even as first-line therapy[106–108]. The compounds have since been FDA-approved for metastatic breast cancers and many trials are ongoing to further refine their utility, with resistance mechanisms already emerging—including upregulation of cyclin D1 via amplification, enhanced CDK2 expression and Rb loss[109].

Taken together, these biologic insights and rationally designed therapies have reduced the burden of breast cancer worldwide. Indeed, tamoxifen, trastuzumab and anastrozole have been added to the World Health Organization's Model List of Essential Medicines. Nevertheless, distinct resistance mechanisms to these therapies have been observed (Figure 1B), many of which are hypothesized to drive breast cancer recurrence, metastasis and ultimately—breast cancer mortality.

1.1.3 Breast cancer recurrence and metastasis

Despite enormous strides in understanding breast cancer biology and an increasing arsenal of targeted therapies, risk of recurrence for patients with even the most indolent subtype, hormone receptor positive breast cancer, can still be over 20% after 5-10 years[110,111]. Breast cancers can return locoregionally, reemerging in breast tissue or the chest wall, or develop into life-threatening metastases. The most common sites of distant metastasis for patients with breast cancer is the bone, liver, lung and brain (Figure 2)[112]. Once metastases are detected, the 5-year survival probability is approximately 20%. Unfortunately, overall survival in patients with distant metastases is unchanged in the past decade—with one study concluding no improvement in metastatic breast cancer outcomes for the past three decades (Figure 3)[3,113]. Clearly, there is an

urgent need to identify mediators of therapy resistance and metastasis in breast cancer; and although progress has been made, comprehensive molecular characterizations of the most common types of breast cancer metastases are limited.

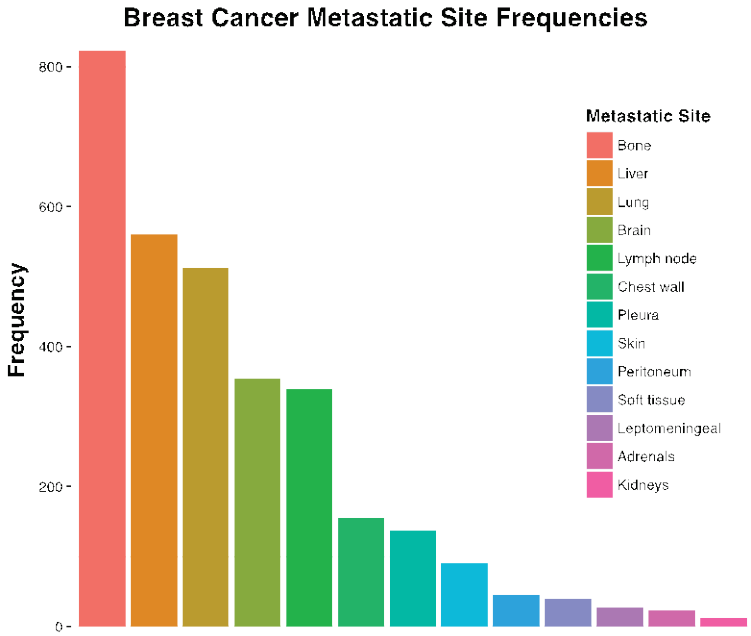

Figure 2: Metastatic breast cancer site frequencies
Metastatic site location(s) from 1202 patients with metastatic breast cancer cared for within the University of Pittsburgh Medical Center (1986-2015, data provided by Dr. Margaret Rosenzeig).

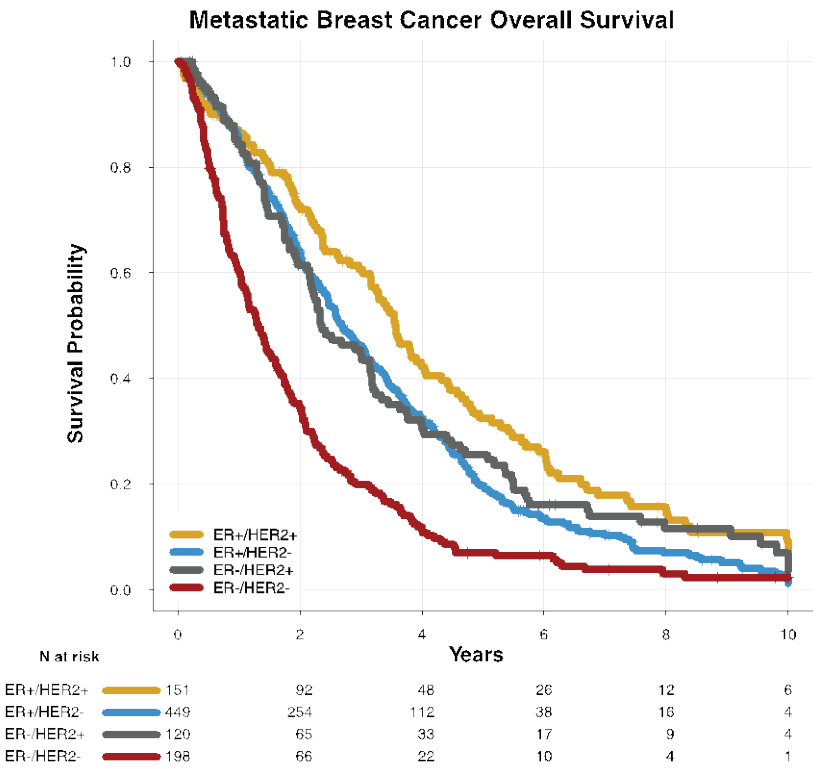

Figure 3: Metastatic breast cancer overall survival
Overall survival of patients diagnosed with metastatic breast cancer (n=918, University of Pittsburgh Medical Center, data provided by Dr. Margaret Rosenzeig), segregated by ER/HER2 statuses. Number at risk table below.

Cancer metastasis is a multi-step process that can take years to develop. Generally, in order to successfully metastasize, cancer cells must co-opt neighboring cells, garner new abilities such as the capacity to invade through adjacent extracellular matrix, intravasate into circulation, shield themselves from the immune system and physical forces encountered in circulation, extravasate into a target tissue and finally adapt themselves to colonize a completely foreign microenvironment—all the while evading the onslaught of therapies[114]. Although complex, each step of this process has been explored using *in vitro* and *in vivo* models and consistent themes and players have emerged.

In order for an epithelial cancer cell to leave its local microenvironment, the cell often undergoes an epithelial-mesenchymal transition (EMT), whereby it takes on mesenchymal-like

features and consequently develops new abilities— such as motility, invasiveness and in breast cancer cells, even more stem-like, tumor-initiating properities[115]. This is coordinated through a suite of transcription factors including Twist, Slug and Snail. Activation of these EMT-transcription factors is precipitated by an impressive number of upstream mediators including TGF-ß, Notch, Wnt, Beta-catenin and growth factor receptor pathways (EGF, FGF)[116]. Additionally, cancer cells have been shown to commandeer surrounding, non-neoplastic cells such as macrophages and stroma, which supply proteases and growth factors to assist a cancer cell's journey through the extracellular matrix towards circulation[117].

Once within circulation, cancer cells interact with other circulating cells to survive. By associating with platelets, malignant cells shield themselves from the survey of immune cells and maintain an EMT-like state through platelet-derived TGFβ[118–120]. Indeed, metastatic breast cancer cells extracted from patient blood are enriched for mesenchymal markers[121]. Prior to extravasating into a target tissue, cancer cells recruit and make tight interactions with neutrophils, hijacking their ability to infiltrate parenchyma by facilitating interactions with endothelial cells[122,123]. Capturing circulating tumor cells is a particularly promising field of translational research in breast cancer, given they have shown prognostic potential and can serve as a source of genomic information in the form of a "liquid biopsy" to identify metastatic evolution and clonal selection of clinically actionable targets[124–126].

The last step, colonization, represents the greatest barrier for cancer cells to successfully metastasize. Metastatic inefficiency—that is, the inability of disseminated, micrometastatic populations to successfully form a macroscopic tumor upon colonization of a foreign organ—has been appreciated in a variety of cancer types. A mere 0.01-0.02% of disseminated cells are able to successfully generate a macometastatic lesion[127–130]. Although most disseminated cancer cells

quickly undergo apoptosis once in a foreign microenvironment, in breast cancer, micrometastases are quite common—approximately 30% of patients have micrometastases in the bone marrow at the time of diagnosis, suggesting cancer cell dissemination can be a very early event, which has been further corroborated in mouse models[131]. These micrometastases can remain dormant for many years and the majority will never reactivate—further supporting the notion of metastatic inefficiency. The mechanisms dictating the survival of dormant cancer cells and how they reactivate to produce macrometastases is an active field of research—with data suggesting particular pathways may play a role. For example, Src signaling is thought to allow dormant cancer cells to survive in bone marrow through activation of Akt[132], DDR1-driven signaling initiates a stem-cell like phenotype that awakens dormant cells in multiple organs[133] and microenvironmental influences—such as the TGFβ-rich perivascular niche[134] and an osteogenic milieu in bone metastases[135]—also promote dormancy and organ-specific colonization.

Because of this intimate relationship between the microenvironment, it is not surprising that certain cancer types have tropisms for colonizing particular organs. Both the intrinsic features of the cancer and the microenvironment contribute to these tropisms—often referred to as the "seed and soil" hypothesis. This is observed clinically, as breast cancer most commonly metastasizes to the lung, brain, bone and liver with subtypes of breast cancer having even more specific tropisms—such as HER2 and basal cancers commonly colonizing the brain and ER-positive tumors homing to the bone[136,137]. *In vivo* models, usually with cells that have been clonally selected after many xenograft passaging from a metastatic organ of interest, have revealed these interactions can be quite complex[138–142]. For example, brain cancer cells can express *PCDH7* highly, which can promote the formation of Cx43 gap junctions between

carcinoma cells and astrocytes—allowing a cGMP driven bidirectional signaling cascade between astrocytes and cancer cells which promotes STAT1 and NFK-B signaling to engender tumor growth and chemoresistance[143]. Despite the elegance of these model systems and groundbreaking insight into these oftentimes complex interactions, the translational applicability of these studies remains questionable, given the difficulty in validating these mechanisms in human disease and the fact that they are generally agnostic to the external selective pressures of therapeutic intervention.

1.2 OVARIAN CANCER

Unlike breast cancer, therapeutic advances in ovarian cancer have been remarkably modest. Although relatively rare, with an estimated 23,000 cases diagnosed each year, over 14,000 patients die annually with greater than 50% of patients succumbing to the disease within 5 years[144,145]. This is largely due to late detection, as ovarian cancers often present as late-stage disease with non-specific gastrointestinal symptoms (e.g. nausea, vomiting, constipation and abdominal discomfort). Counterintuitively, advanced ovarian cancers are exquisitely sensitive to initial cycles of cytotoxic agents—greater than 75% of patients will show a response and over 50% of patients will have a complete response following primary therapy[146]. Unfortunately, most patients will eventually relapse and acquire therapy resistant tumors. Relapsed ovarian cancer is essentially incurable and carries a mere 12 to 24 month median overall survival[147].

The difficulty in treating ovarian cancer can be partially explained by the complexity of its genome. Although pathways such as DNA repair have been identified as central to its pathophysiology, ovarian cancer is uniquely heterogeneous and as a result, no promising drug targets have emerged. Like breast cancer; however, most molecular characterizations have been performed on treatment-naïve primary tumors. The molecular changes that differentiate a treatment-responsive ovarian cancer and an unresponsive, chemoresistant tumor are largely unknown. In these studies, we aimed to build initial insight into this evolution and identify potential mediators of ovarian cancer progression.

1.2.1 Ovarian cancer and genomic features

Ovarian cancer is a heterogeneous disease that can be segregated into distinct histological and molecular subtypes. The focus of this work is on epithelial derived high-grade serous ovarian cancers (HGSOC), which represents the most common and most lethal form of the disease—responsible for over 70-80% of ovarian cancer deaths[148]. Although labeled as ovarian cancer, more recent evidence suggests this neoplasm arises from the epithelial lining of the fallopian tube[149–151]. Deciphering the molecular mechanisms of and identifying potential therapeutic targets for HGSOC has been wrought with difficultly. Nonetheless, as more HGSOCs are characterized by genome-wide interrogations, patterns are emerging.

There are few recurrent single nucleotide mutations in ovarian cancer, with *TP53* mutations being present in nearly all HGSOCs and lower frequency mutations in *BRCA1* and *BRCA2* observed—which have been long known to confer strong genetic susceptibilities[152]. Ovarian cancers, like basal breast cancers, have an unusually high degree of recurrent DNA structural variation. As such, ovarian cancers are categorized as C-class (copy number) driven tumors as opposed to M-class (mutation) driven tumors like renal cell carcinomas[153]. This structurally variant genome often results in gene breakages, which are commonly detected in DNA-repair genes and tumor suppressors[154]. Indeed, with multimodal mechanisms of DNA-repair gene inactivation including mutations, gene disruption via structural variants and silencing via methylation—it is estimated that most ovarian cancers have some degree of DNA repair dysfunction.

The use of PARP inhibitors as a targeted therapy has thus been an active area in ovarian cancer, with recent clinical trials showing increased progression-free survival with their use[155,156]. Other targeted therapies tested include EGFR family member, Src, VEGFA and

estrogen signaling inhibitors—yet all have failed to improve outcomes in overall survival[157–161]. This could be due to improper patient selection, as there are at least four molecular subtypes of HGSOC that may have distinct etiologies and thus may respond differently to targeted agents[152]. Consequently, the foundation of HGSOC treatment remains surgery and chemotherapy—particularly regimens consisting of platinum and taxane-based agents.

1.2.2 Ovarian cancer recurrence and chemoresistance

Ovarian cancer is initially managed with surgery and cytotoxic agents—often referred to as primary surgical cytoreduction or "primary debulking." The goal is to eliminate any evidence of macroscopic disease and can become quite complex depending on the degree of carcinomatous within the peritoneum[162]. The combination of a platinum and taxane-based therapy following surgery has increased outcomes substantially in the past two decades and is first-line therapy for all late-stage patients[163]. Refinements to this core treatment plan, including altered dosing regimens and the use of intraperitoneal chemotherapy, has also contributed to improved outcomes with greater than 70% of patients now showing objective responses[164,165]. Yet, even with these impressive response rates, acquired chemoresistance is common—approximately 80% of ovarian cancer patients will suffer a relapse.

Following primary therapy, patients are clinically categorized with platinum-sensitive or platinum-resistant disease, depending on the timing of disease progression[162]. If progression occurs after 6 months, the cancer is deemed platinum-sensitive and patients will be given additional lines of platinum-based therapies for a future recurrence. If progression occurs before 6 months, the disease is considered platinum-resistant and alternative agents will be used to treat a recurrence such as doxorubicin, topotecan, gemcitabine, etoposide and vinorelbine—very

similar therapeutic choices to advanced, treatment-resistant breast cancers[162]. Other options include bevacizumab, PARP inhibitors and various maintenance regimens—though when ovarian cancers become platinum-resistant and recur, the disease is essentially incurable. Understanding the mechanisms driving resistance to these initially effective therapies is a key step to improve the poor outcomes in patients with relapsed disease.

Chemotherapy resistance is a common problem for advanced breast and ovarian cancers and given overlapping agents, mechanisms mediating this resistance are shared between the two diseases. Perhaps the most well established driver of chemotherapy resistance is through the activation of drug-efflux pumps—membrane bound proteins that eliminate hydrophobic compounds, including cancer agents, from cells. Members of this family, including *ABCB1* and *ABCG2* and their hyperactivation have been implicated in chemotherapeutic resistance in both breast and ovarian cancer[166–168]. Copper efflux transporters, such as *ATP7A*, selectively cause resistance to platinum based therapy in ovarian cancer cells through sequestering agents to intracellular vesicles and limiting their access to DNA, where the drugs crosslink DNA, inhibit DNA repair and drive cancer cells towards apoptosis[169,170]. Alterations in DNA repair, unsurprisingly, are another common pathway dysregulated in both chemoresistant breast and ovarian cancer. Disruption of the mismatch repair system (e.g. *MLH1* and *MSH2*) can cause cancer cells to lose their ability to initiate apoptosis upon DNA damage and although BRCA-deficiency is associated with favorable responses to chemotherapies, BRCA reversion can occur whereby mutant BRCA-cells acquire a wild-type allele after primary treatment, restore their ability for DNA-repair and thus become more resistant to DNA-damaging agents[154,171–173]. Altered cell states is also thought to contribute to disease progression, as malignant cells that switch to more mesenchymal and stem-like phenotypes has been associated with resistance to

cytotoxic agents[168,174–176]. Further mechanisms at play in chemoresistance include alterations in how the agents are metabolized, dysregulation of pro and anti-apoptotic mediators, autophagy and interactions with the microenvironment[177,178]. Again—despite the multi-mechanistic understanding of chemoresistance in breast and ovarian cancer models, molecular characterizations of advanced disease are scant, making it difficult to validate and act on these resistant mechanisms in the clinic.

1.3 HYPOTHESIS

The selection of therapy resistant and metastasis-capable malignant cells is ultimately the process that kills patients with cancer. Despite this fact, as it stands currently, the most well characterized cancers are relatively indolent primary tumors. Recent studies analyzing multiple patient-matched, spatiotemporally collected specimens have suggested that metastatic or recurrent disease acquire clinically meaningful, sometimes patient-specific alterations not appreciated in primary tumors—both at the DNA and transcriptional level[179,180]. Indeed, an analysis of publically available data collated from an academic clinical sequencing center (1272 breast cancers; 795 metastases, 477 primary tumors)[181] shows significantly enriched mutations in metastatic breast cancer versus primary disease (Figure 4).

We hypothesize that advanced breast and ovarian cancers acquire recurrent molecular dependencies, distinct from primary tumors, throughout their life histories. In this collection of studies, we test this hypothesis by defining the altered molecular taxonomy of primary breast cancers following their development of estrogen-independence and

colonization of the brain and bone and ovarian cancers after establishing therapy resistance.

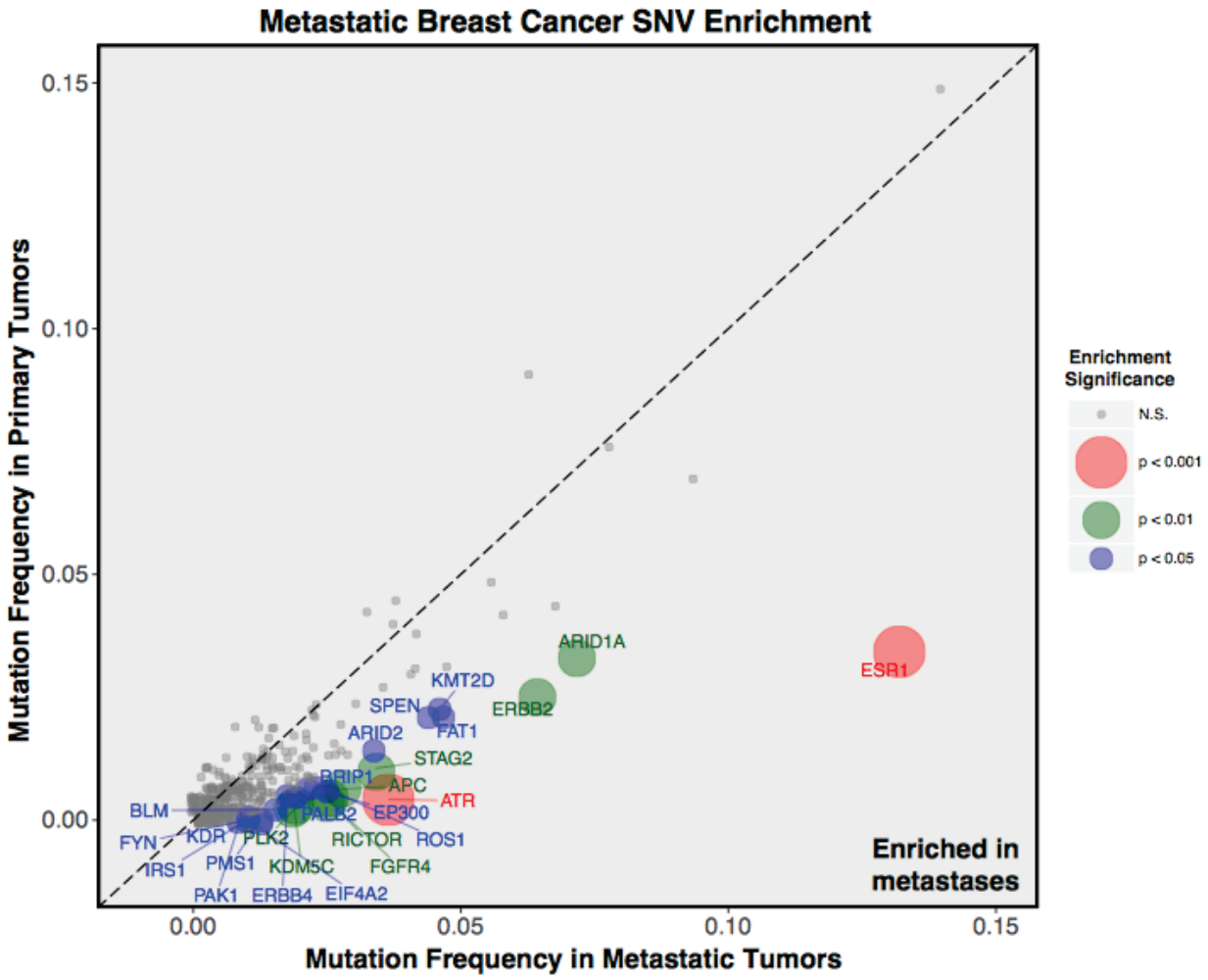

Figure 4: Metastatic breast cancer SNV enrichments
Somatic single nucleotide variation (SNV) frequencies in primary (n=477) and metastatic (n=795) breast cancers. Enrichment p-value in metastases was performed via a Fisher's Exact test and its magnitude is highlighted by both size and color of circles on the plot.

2.0 INTRINSIC SUBTYPE SWITCHING AND HER2 GAINS IN BREAST CANCER BRAIN METASTASES

2.1 ABSTRACT

Breast cancer (BrCa) patients with brain metastases (BrM) have limited therapeutic options. A better understanding of molecular alterations acquired in BrM could identify clinically actionable metastatic dependencies. We aimed to (1) determine whether there are intrinsic subtype differences between primary tumors and matched BrM and (2) uncover BrM-acquired alterations that are clinically actionable. Out of 20 cases, 17/20 BrM retained the PAM50 subtype of the primary BrCa. Despite this concordance, 17/20 BrM harbored expression changes (< or > 2-fold) in clinically actionable genes including gains of *FGFR4* (30%), *FLT1* (20%), *AURKA* (10%) and loss of *ESR1* expression (45%). The most recurrent expression gain was *ERBB2,* which showed a >2-fold expression increase in 7 of 20 BrM (35%). 3 of these 7 cases were HER2-negative, out of 13 HER2-negative in the cohort, in the primary BrCa and became IHC-positive (3+) in the paired BrM with metastasis-specific amplification of the *ERBB2* locus. In an independent dataset, 2 of 9 (22.2%) HER2-negative BrCa switched to HER2-positive with one BrM acquiring *ERBB2* amplification and the other showing metastatic enrichment of the activating V777L *ERBB2* mutation. An expanded cohort revealed that *ERBB2* amplification and/or mutation is enriched in BrM versus local disease (13% local vs 24% BrM, $p<0.001$). Taken together, this

study demonstrates BrCa BrM commonly acquire alterations in clinically actionable genes, with metastasis-acquired *ERBB2* gains in ~20% of HER2-negative cases.

Contributors to study: Nolan Priedigkeit[1], Ryan J. Hartmaier[2], Yijing Chen[1], Damir Vareslija[3], Ahmed Basudan[1], Rebecca J. Watters[1], Roby Thomas[1], Jose P. Leone[4], Peter C. Lucas[1], Rohit Bhargava[1], Ronald L. Hamilton[1], Juliann Chmielecki[2], Shannon L. Puhalla[1], Nancy E. Davidson[1], Steffi Oesterreich[1], Adam M. Brufsky[1], Leonie Young[3], Adrian V. Lee[1]

[1]Departments of Pharmacology and Chemical Biology, Human Genetics, Medicine, and Pathology, Women's Cancer Research Center, Magee-Women's Research Institute, University of Pittsburgh Cancer Institute, PA, USA

[2]Foundation Medicine, Cambridge, MA, USA.

[3]Endocrine Oncology Research Group, Department of Surgery, Royal College of Surgeons in Ireland, Dublin, Ireland

[4]University of Iowa Holden Comprehensive Cancer Center, University of Iowa Hospitals and Clinics, C32 GH. 200 Hawkins Drive, Iowa City, IA, USA

2.2 INTRODUCTION

Brain metastases (BrM) occur in 10-15% of patients with metastatic breast cancer (BrCa) and present a major clinical challenge, overshadowed by a relatively poor 8.5-month median overall survival[182,183]. Limited therapeutic options exist for patients with BrM and current management consists of surgical resection, radiation therapy and chemotherapy. HER2-positive BrM have demonstrated promising responses to HER2-targeted therapies in recent clinical trials, yet more comprehensive studies are needed to confidently define their utility[184,185]. Unfortunately, in patients with HER2-negative BrM, no targeted therapies have shown even modest benefits[186]. Clearly, there is an urgent need to better understand the mechanisms of BrCa metastasis to the brain and to define novel therapeutic targets.

Although metastasis is the major contributor to mortality regardless of cancer type, our understanding of metastatic disease is remarkably limited. The mechanistic drivers of primary BrCa have been well-studied, largely spearheaded by collaborative efforts such as The Cancer Genome Atlas[13]. Metastatic breast cancers are much less well characterized, especially BrM given their relative anatomic inaccessibility. Molecular interrogations of patient-matched primary and metastatic lesions in other cancers have successfully identified mechanisms of tumor evolution and therapy resistance[154,180,187,188]. A recent, pan-cancer study on tumor evolution in BrM, which focused exclusively on single nucleotide variants and copy number changes, revealed metastasis-acquired DNA-level changes that may serve as therapeutic targets[179].

In this study, we performed targeted expression profiling of 127 target genes from five breast cancer prognostic signatures on a molecularly diverse clinical cohort of 20 primary breast tumors and their patient-matched BrM to determine transcriptional differences between primary cancers and BrM and to define metastasis-acquired alterations that may be clinically actionable.

2.3 MATERIALS AND METHODS

2.3.1 Patient samples

Eligible breast cancer cases had paired formalin-fixed paraffin-embedded (FFPE) tissue from primary and resected BrM. Given the rarity of samples, no exclusion criteria were enacted. In total, 20 cases of patient-matched primary breast tumors (10 ER-, 10 ER+) and BrM from two institutions were included—6 pairs from Royal College of Surgeons (RCS), Ireland and 14 pairs from University of Pittsburgh (Pitt), USA (Table 1). This study was reviewed and approved by Institutional Review Boards from both participating institutions (University of Pittsburgh IRB# PRO15050502, Royal College of Surgeons IRB #09-07). An independent, controlled-access dataset of 17 patient-matched samples with brain metastases generated by the Broad Institute was acquired from dbGap (phs000730.v1.p)[179] under the IRB# PRO16030233. A collection of 7,884 breast cancer tumor data (52% metastases, including BrM) was analyzed from Foundation Medicine with study approval by the Western Institutional Review Board (WIRB).

Table 1: Abridged clinicopathological features of brain metastasis cases
Abbreviations: ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma. Hormone receptor status were called from IHC as per ASCO/CAP recommendations[189,190].

Case	Tissue Source	Pathology	ER Status	PR Status	Her2 Status	Endocrine Therapy	HER2 Therapy
RCS_1	RCS	IDC	Neg	Neg	Pos	-	-
RCS_2	RCS	IDC	Neg	Neg	Pos	-	+
RCS_3	RCS	IDC	Pos	Neg	Pos	NA	+
RCS_4	RCS	IDC	Pos	Neg	Neg	+	-
RCS_5	RCS	IDC	Neg	Neg	Neg	-	-
RCS_6	RCS	IDC	Pos	Neg	Neg	+	-
Pitt_6	Pitt	IDC	Neg	Neg	Neg	-	-
Pitt_7	Pitt	IDC	Pos	Neg	Pos	+	+
Pitt_12	Pitt	IDC	Neg	Neg	Neg	-	-
Pitt_17	Pitt	IDC	Pos	Neg	Pos	-	+
Pitt_25	Pitt	IDC	Neg	Neg	Neg	-	-
Pitt_29	Pitt	IDC	Pos	Neg	Neg	-	-
Pitt_47	Pitt	IDC/ILC	Pos	Pos	Pos	+	+
Pitt_51	Pitt	IDC	Pos	Neg	Neg	+	-
Pitt_52	Pitt	IDC	Neg	Pos	Pos	-	+
Pitt_62	Pitt	IDC	Pos	Pos	Neg	+	NA
Pitt_64	Pitt	IDC	Neg	Neg	Neg	-	-
Pitt_68	Pitt	IDC	Neg	Neg	Neg	+	-
Pitt_71	Pitt	IDC	Neg	Neg	Neg	-	-
Pitt_72	Pitt	ILC	Pos	Pos	Neg	+	-

2.3.2 Tissue processing

Formalin-fixed paraffin-embedded (FFPE) tumor blocks were sectioned and H&E staining analyzed by a pathologist for histological and tumor cellularity classifications. All specimens had a tumor cellularity equal to or above 60% except for BM_Pitt_68 (40%) and BM_Pitt_71 (30%). Between four to ten (depending on tumor size) 10-micron FFPE sections immediately adjacent to the H&E-analyzed section were scrolled and pooled for dual DNA/RNA extraction using Qiagen's AllPrep kit according to manufacturer's instructions.

2.3.3 Clustering and molecular subtyping

Hierarchical clustering was performed on normalized expression data (Data Supplement 1: S2, S3). Clustering was performed using the *hclust* function in R, with 1 minus Pearson correlation as distance measures and the "average" agglomeration method. Heatmap was created with *heatmap.3* in R. PAM50 molecular subtyping was performed using *genefu*[191]. To account for PAM50 test set bias, normalized expression data from a cohort of 20 tumor samples with known ER-status were subsampled to create a balanced cohort of ER-positive and ER-negative tumors[192]. A query sample of unknown molecular subtype was added to the balanced cohort. An intrinsic molecular subtype was called for the query sample using the *pam50.robust* model in *genefu*. This method was repeated for all 40 clinical specimens (Data Supplement 1: S4). OncoTypeDX scores were determined using unscaled *genefu* OncoTypeDX scores and a linear model generated from 72 samples with known OncoTypeDX scores as performed previously[193].

2.3.4 Recurrent expression alterations

Fold-change values were calculated between patient-matched primary and metastatic tumors using log2 transformed normalized expression counts for each gene (Data Supplement 1: S5). The mean fold-change between primary and metastatic lesions for all genes across all samples was -0.01 with a standard deviation of 1.04 (Appendix A.1: Figure 24). An 'expression alteration' was defined as a log2 fold-change value greater than or less than one standard deviation from the mean fold-change. Recurrent alterations were plotted using *ComplexHeatmap*[194]. To interrogate clinically significant alterations, the Drug-Gene Interaction (DGIdb 2.0) database was used[195]. All genes were input into the database and only those annotated as 'clinically actionable' (as of March 10th, 2016) were visualized. To plot and statistically assess gene-specific expression differences, the beeswarm R package was used to create ladder plots along with Wilcoxon signed-rank tests on paired (metastasis vs. primary) normalized log2 expression values.

2.3.5 Immunohistochemistry

10 micron FFPE sections were mounted on slides and stained for HER2 and ER as described previously and clinical staining scores were called in accordance with ASCO/CAP recommendations[196].

2.3.6 Copy number alteration and single nucleotide variant analysis

Tumor DNA quality was assessed by an Illumina FFPE QC Kit. DNA with a Delta Cq value below 5 were restored using the Infinium HD FFPE DNA Restore Kit. 200 ng of restored tumor DNA was analyzed on an Illumina iScan System using an Illumina HumanCytoSNP-FFPE v.2.1 BeadChip. *GenomeStudio* was implemented to produce normalized logR intensity values from the two-color readouts using the HumanCytoSNP-12v2.1-FFPE_G.egt cluster file. These values were then analyzed using the *copynumber* package in R[197]. LogR values were preprocessed by excluding outliers via Winsorization and imputing missing measurements as a logR value of 0. Data then underwent multi-sample segmentation and final LogR values and segments were assessed and plotted for chromosome 17 (Data Supplement 1: S6). Raw fastq files from whole-exome sequencing of an independent cohort of 17 patient-matched primary BrCa and BrM were aligned using *bwa* (v0.7.13), sorted with *samtools* (v1.3), duplicates marked and removed with *picardtools* (v1.140) and local realignment performed with *GATK* (v3.4-46)[198–200]. To estimate and plot copy number ratios, *CNVkit* was utilized on processed bam files[201]. A pool of bam files from normal tissue was used as a CNVkit reference. Log2 ratio estimates were then analyzed for metastasis-specific gains in *ERBB2* by performing a student's t-test on primary and metastatic estimated logR values across the 26 *ERBB2* exonic regions (Data Supplement 1: S7). To discover *ERBB2* activating mutations in the HER2-switching PB0049 case, the *ERBB2* region was probed for somatic mutations using *CLC Genomics Workbench* (http://www.clcbio.com, v9.0) and *IGV* (v2.3.60)[202].

2.3.7 FoundationOne *ERBB2* alterations

To test whether *ERBB2* amplification and base pair mutation is metastasis-site specific, changes in this gene were evaluated in an expanded cohort of 7,884 breast tumors enriched for metastatic samples (52%) including liver (16.7%), lung (4.3%), bone (3.6%), and brain (2.0%) that underwent genomic profiling as part of routine clinical care in a CLIA-certified, CAP-accredited, and New York State-accredited laboratory (Data Supplement 1: S8[203]. *ERBB2* alterations were identified as described previously[203,204].

2.4 RESULTS

2.4.1 Expression concordance between primary tumors and matched brain metastases

To determine the transcriptional similarity between primary tumors and patient-matched BrM, unsupervised hierarchical clustering was implemented using normalized gene expression values. This produced three major clades broadly classified as ER-positive, HER2-positive and ER-negative (Figure 5A). The majority of patient-matched pairs (12/20) clustered within a single doublet clade. The 7 pairs that did not cluster within a doublet clustered within the same major clade.

To further interrogate clinically relevant differences between the patient-matched samples, the intrinsic molecular subtype (PAM50) of each tumor was calculated. PAM50 assignments were consistent in 17/20 pairs (Figure 5B) with 3 discordant pairs being Case RCS_2 (LumA to Her2), RCS_4 (LumB to LumA) and Pitt_47 (LumA to Her2). OncotypeDX

scores were also largely unchanged between primary and metastatic tumors, retaining their clinical risk score in 75% of cases (Appendix B: Table 8).

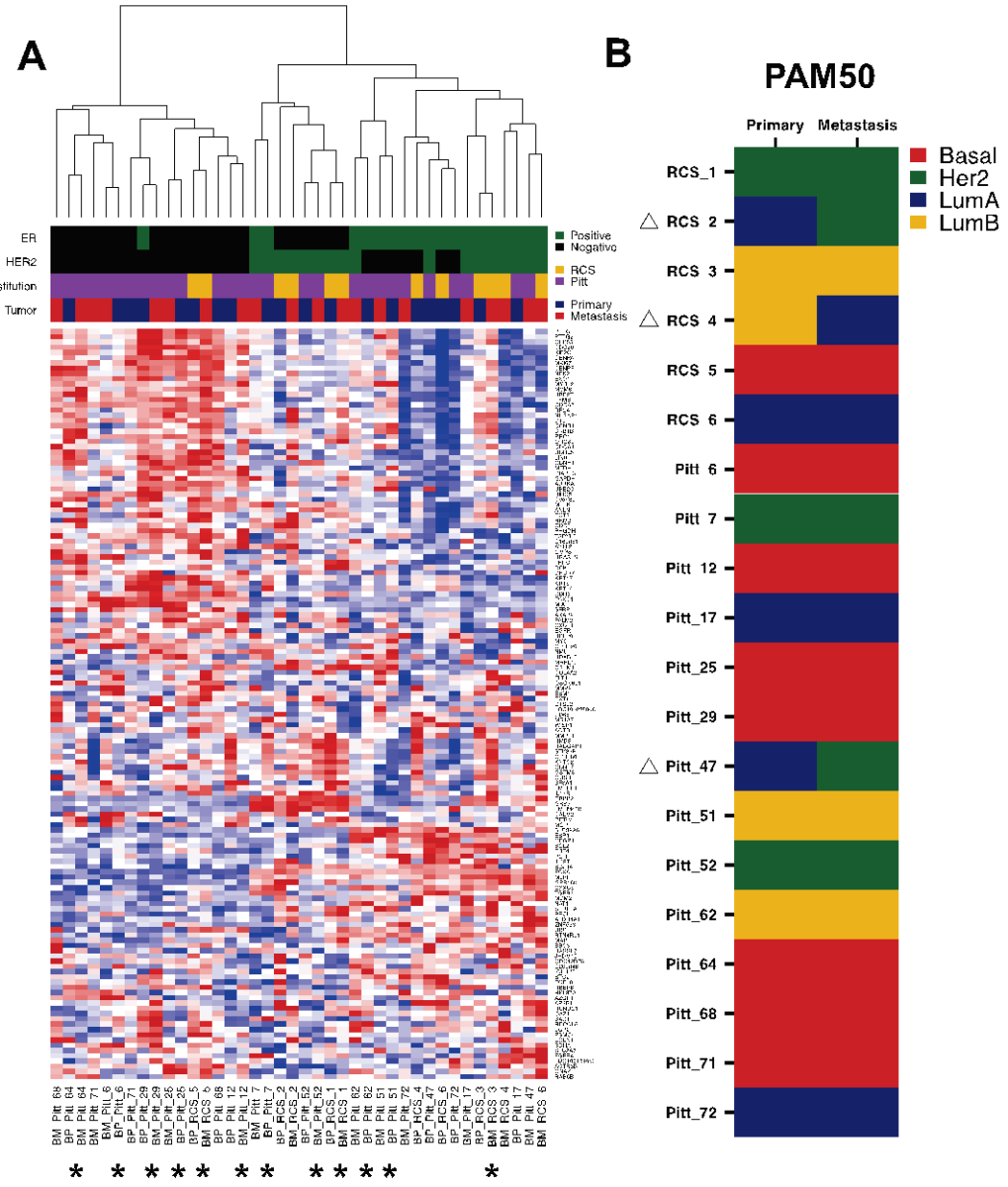

Figure 5: Transcriptional similarity between primary breast cancers and matched brain metastases
(A) Unsupervised hierarchical clustering heatmap of 20 patient-matched cases with hormone status (green = positive, black = negative), tissue site source/institution (yellow = Royal College of Surgeons, Ireland, purple = University of Pittsburgh, USA) and tumor site (blue = primary, red = metastasis) of each sample indicated; BP = Breast Primary, BM = Brain Metastasis. Asterisk below plot indicate patient-matched pairs that clustered in the same doublet of a clade in the dendrogram. **(B)** PAM50 intrinsic molecular subtype calls in patient-matched cases (red = Basal, green = Her2, blue = LumA, yellow = LumB). Discordant pairs are marked with a delta symbol.

2.4.2 Distinct expression gains of clinically actionable genes

Despite a large degree of similarity within patient-matched pairs, 100 genes were recurrently altered (< or > 2-fold expression change) in BrM when compared to patient-matched primaries (Figure 6A). The most recurrently downregulated genes were cytokeratins—*KRT17* being downregulated in 14 of 20 pairs, *KRT5* and *KRT14* in 15 of 20 pairs, $p < 0.001$ (Figure 6B). The most recurrently upregulated genes were *RAB6B* (9 of 20 pairs, $p < 0.01$, Wilcoxon signed-rank test) and *GRB7* (8 of 20 pairs, $p < 0.001$).

Ten genes in the panel are defined as clinically actionable in the DGIdb and many showed BrM-specific changes (Figure 7A). Of these, *ERBB2* was the most recurrent alteration showing at least a 2-fold expression increase in 35% of BrM ($p < 0.05$). 3 of these 7 cases were classified as HER2-negative in the primary tumor. *FGFR4* showed increased expression in 30% of samples, with 3 cases showing >4-fold increase. Other recurrent expression increases included *FLT1* (20%), *AURKA* (10%) and *EGFR* (10%). The most recurrently downregulated gene was *ESR1*, showing a 2-fold decrease of expression in 4 samples and a >4-fold decrease in 5 samples ($p < 0.05$). 2 samples with the greatest fold-change in *ESR1* switched expression from ER-positive to ER-negative levels, while 3 samples with alterations in *ERBB2* went from HER2-negative levels to HER2-positive levels of expression (Figure 7B).

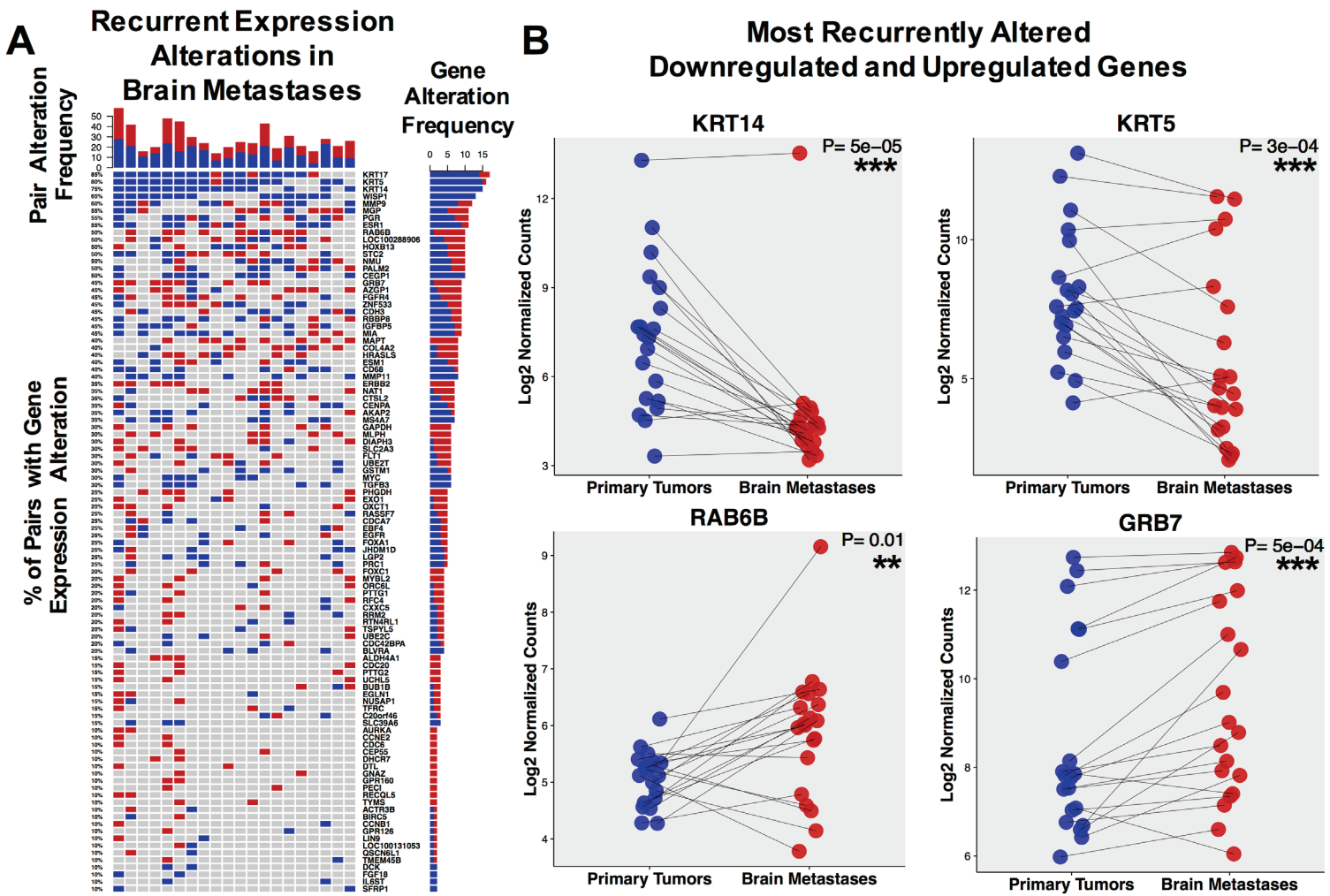

Figure 6: Recurrent expression alterations in breast cancer brain metastases
(A) OncoPrint plot of recurrent expression alterations in 20 cases, ranked by frequency of alteration by gene. Blue tile represents a >2-fold decrease in the patient-matched brain metastasis relative to the primary, while a red tile represents a >2-fold increase. **(B)** Paired ladder plots visualizing case-specific alterations in the most recurrently upregulated and downregulated genes interrogated. Blue dots represent primary tumor expression values (Log2 normalized counts), red dots represent metastatic tumor expression values; p-values (* p <= 0.05, ** p <= 0.01, *** p <= 0.001) shown are from Wilcoxon signed-rank tests.

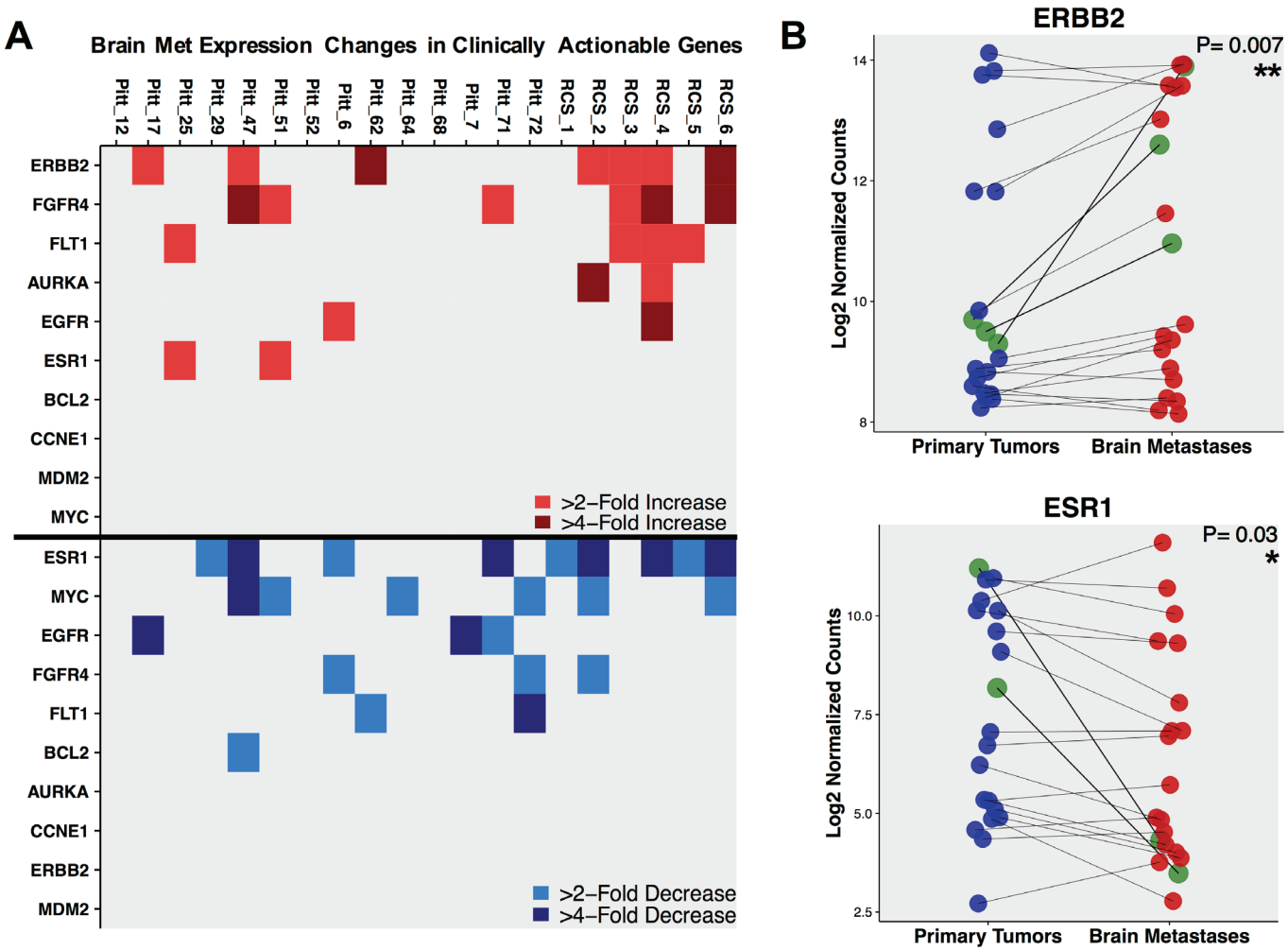

Figure 7: Expression alterations in clinically actionable genes.
(A) Tile plot visualizing expressions alterations in clinically actionable genes. Top panel consists of recurrent increases in expression (light red = >2-fold increase, dark red = >4-fold increase), bottom panel are recurrent decreases in expression (light blue = >2-fold decrease, dark blue = >4-fold decrease) between patient-matched pairs. **(B)** Paired ladder plots of the two most recurrent upregulated and downregulated clinically actionable genes (*ERBB2* and *ESR1*). Green dots represent samples with suspected hormone status switching, p-values (* p <= 0.05, ** p <= 0.01, *** p <= 0.001) shown are from Wilcoxon signed-rank tests.

2.4.3 DNA-level HER2 acquisitions in breast cancer brain metastases

To determine the consequences of *ERBB2* mRNA expression changes, IHC was performed in 3 samples with the greatest fold-changes in *ERBB2*. These samples were HER2-negative in the primary tumors by IHC (out of 13 HER2-negative primaries in the cohort). All three tumors showed significant increases in HER2 IHC scores—RCS_4; 1+ in primary, 3+ in BrM, RCS_6; 1+ in primary, 3+ in BrM, Pitt_62; 0 in primary, 3+ in BrM (Figure 8A). SNP-array CNV analysis revealed HER2-status switching is driven by canonical amplification of the ERBB2 locus (Figure 8B).

We examined *ERBB2* amplification and SNV in an independent cohort (n=17; 9 HER2- and 8 HER2+) of patient-matched breast cancer and BrM analyzed by whole-exome sequencing. One case (Broad_PB0150) showed metastasis-specific copy number gain in *ERBB2*, which was consistent with the case being HER2-negative in the primary and HER2-positive in the metastasis (Figure 8C, top). Another case (Broad_PB0049) switched from HER2-negative to positive, but no significant DNA-level gains in the BrM was found; however, there was an enrichment—from an allele frequency of 39% to 69%—of a somatic V777L activating mutation in the metastasis (Figure 8C, bottom).

2.4.4 ERBB2 amplifications and SNVs enrichment in brain metastases

To generalize these observations and test whether changes in HER2 status is specific to BrM, we analyzed a cohort of 7,884 breast cancers (7,265 with unambiguous tissue site information) representing 3135 cases of local disease and 4,130 cases of metastases for amplifications and/or

SNV in *ERBB2*. Comparing all local and metastatic tumors from all sites (Figure 8D, top) showed no significant difference; however, there was a strong and significant enrichment of *ERBB2* alterations specifically in brain metastases (24%) compared to local disease (13%) (Fisher-exact $p < 0.0005$; Figure 8D, bottom). While SNVs were too infrequent to determine a significant enrichment in BrM, there was a slight enrichment (albeit not statistically significant) for *ERBB2* SNVs with concurrent amplification (Data Supplement 1: S8).

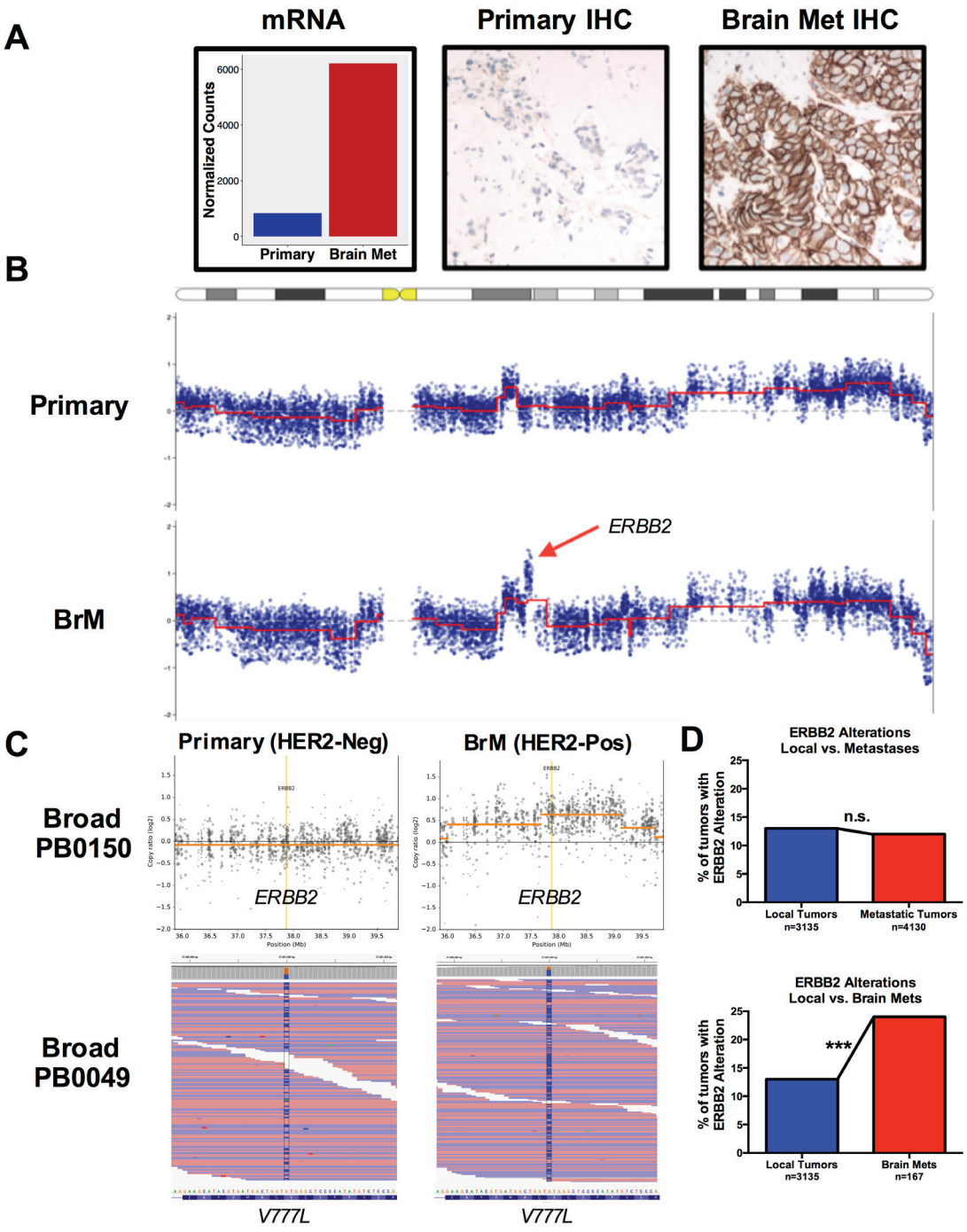

Figure 8: *ERBB2* gains in breast cancer brain metastases
(A) Pitt_62 normalized counts of mRNA expression (blue = primary tumor, red = brain metastasis) with Primary and Metastatic IHC staining of HER2. **(B)** LogR value plot for chromosome 17 in primary and brain metastasis (BrM). *ERBB2* region highlighted with a red arrow. **(C)** Top; Broad PB0150 CNVkit LogR plots from primary and brain metastasis. Segmented LogR ratio means are marked with horizontal orange lines across a 4 MB region surrounding the *ERBB2* locus (marked with a vertical yellow line). Bottom; PB0049 V777L activating *ERBB2* mutation in primary and brain metastasis as visualized in *IGV*. G to C variant highlighted in blue, along with variant frequency barplots above. **(D)** Top; *ERBB2* alterations (amplification or mutation, red) in 3135 local tumors and 4130 metastatic tumors. Bottom, *ERBB2* alterations in local tumors and 167 brain metastases.

2.5 DISCUSSION

The brain is a common and catastrophic site of metastasis for breast cancer patients. Our understanding of metastasis-specific gene expression is limited, as are the options for treatment. In this study, the largest of its kind in BrCa BrM to date, we found that patient-matched primary BrCa and BrM have similar gene signatures; however, when examining on a gene-level, many recurrent changes were observed in clinically actionable genes. Importantly, we found that ~20% of HER2-negative breast cancers recur as HER2-positive with either *ERBB2* amplification and/or gains in *ERBB2*-activating SNVs. Taken together, these observations have immediate clinical implications, as they (1) reveal that BrM acquire expression gains distinct from the primary tumor in targetable genes, many of which have open clinical trials or FDA-approved therapies, (2) establish that approximately 20% of patients with HER2-negative BrCa acquire BrM-specific *ERBB2* amplifications and/or activating SNVs that may be sensitive to existing therapies and (3) suggest that therapies and trial eligibilities on limited, single-platform molecular data from a primary tumor may engender missed opportunities in advanced cancer settings.

A high degree of transcriptional similarity was observed between primary tumors and BrM, both across the entire gene set and when performing clinical gene signature assignments. Transcriptional concordance between patient-matched primaries and metastases has been appreciated in many cancer types[205–208]. Given the observed metastatic inefficiency of tumor cells to colonize a distant site, it is perhaps surprising that many tumors are similar to their patient-matched primary. A limitation to this study and others like it; however, is that

intermediate mechanisms of metastasis—such as those dictating intravasation, EMT, survival in circulation, extravasation, initial colonization and MET—are masked given the binary comparison of a primary tumor and a metastatic endpoint[114]. Further studies focusing on the intermediate steps, such as profiling of circulating tumor cells, would complement matched-sample studies[121,209].

Nonetheless, even with a remarkable transcriptional similarity between paired samples, clinically actionable alterations were identified in all but 3 pairs, including *ERBB2* which showed expression increases in 35% of BrM. The importance of *ERBB2* in BrM, including mechanistic evidence of HER2's contributions to brain metastasis in vivo, has been increasingly appreciated in the past decade with the most recent ASCO recommendation supporting HER2 testing in the metastatic setting[189,210]. Duchnowska et al. and Thomson et al. reported a HER2-negative to HER2-positive switching frequency of 16% and 18% respectively in BrM via IHC, with a portion of these BrM showing no copy number gains[211,212]. In a pan-cancer expression analysis of unmatched BrM, Saunus et al. found that breast cancer BrM have higher *ERBB2* expression than BrM from other sites[213]. Additionally, in a single matched case, the authors identified a case that switched from a HER2-negative primary to HER2-positive BrM. The patient was treated with trastuzumab and lapatinib in the advanced setting and had a clinical response. The results herein further reinforce the notion that HER2 expression gains in BrCa BrM are relatively common (~35%), and even occur in HER2-positive disease—as 4 of 7 patients who were initially HER2-positive showed *ERBB2* expression increases. Analysis of HER2-switching samples showed expression gains are partially driven by classical amplification of the *ERBB2* locus. Interestingly, the HER2-switching Broad_PB0049 case showed no copy number gains in the BrM, yet harbored an enrichment of the activating V777L *ERBB2* mutation[214], suggesting a

heretofore unknown mechanism of *ERBB2* gain in metastatic tumors. These observations, especially in the context of patients classified as HER2-negative in the primary tumor, present immediate personalized treatment options.

A hurdle to molecularly profiling brain metastases to inform therapeutic decisions; however, is their relative anatomic inaccessibility, as biopsying or resecting a brain metastasis can be impractical. Yet, for patients with inaccessible tumors, DNA-level alterations in *ERBB2* can be assessed in tumor DNA from circulating-free DNA and circulating tumor cells. Although there is evidence suggesting tumor DNA from brain lesions is more difficult to detect, recent advances in detection technologies and successful detection of circulating tumor DNA in cerebral spinal fluid are encouraging and should be further investigated for patients with BrM, especially given our observations[215–217].

Novel recurrent targetable alterations beyond *ERBB2* were also discovered, including expression increases in *FGFR4* (30% of pairs), *FLT1* (20%), *AURKA* (10%) and *EGFR* (10%). Each one of these targets have clinical trials ongoing and our results suggest that trial eligibility requiring expression of these markers (NCT02325739) should assess the metastatic tumor, especially given up to 6-fold expression changes (i.e. *FGFR4*) in metastases relative to primaries.

Significant loss of gene expression from the primary to metastatic lesions was also observed. The most recurrent expression losses involved cytokeratins. Cytokeratins have shown a complex role in oncogenesis and breast cancer metastasis, with loss of cytokeratin expression being a hallmark of EMT and metastasis[116,218,219]. Notably, CK5, CK14, and CK17 are expressed in basal and myoepithelial cells, suggesting the loss of keratin gene expression may be due to the departure from the breast environment.

One of the most recurrently downregulated genes was *ESR1*, showing a 2-fold decrease in expression in 45% of tumors, with some cases changing expression from ER-positive to ER-negative levels. For example, case Pitt_47 harbored a >6-fold decrease in ESR1 expression in the brain metastasis and this loss of transcript expression was confirmed at the protein level via IHC (Appendix A.1: Figure 25). This patient received endocrine therapy and such recurrent losses in *ESR1* have important implications, as loss of ER expression and coincidental activation of other mitogenic pathways is an established mediator of estrogen independence and hormone therapy resistance and has been shown to be prognostically significant[44,220]. Notably, this brain metastasis had a greater than two-fold increase of *ERBB2* and greater than four-fold increase in *FGFR4*, perhaps suggesting these two mediators can maintain tumor growth in the absence of ER.

To conclude, this study identifies that breast cancer brain metastases are remarkably similar to patient-matched primary tumors transcriptionally; yet, despite this similarity, recurrent expression changes in clinically actionable genes are common. These results support the notion that that metastatic tumors may be considered distinct from primary tumors and provides rationale to comprehensively profile metastatic lesions to inform clinical decisions, such as targeted therapies and trial eligibilities, in advanced breast cancer. Furthermore, approximately 20% of HER2-negative patients show CNV and SNV gains in HER2 across multiple cohorts, which warrants immediate clinical investigation as many of these patients will not be provided HER2-targeted therapies.

3.0 EXOME-CAPTURE RNA-SEQUENCING OF DECADE-OLD BREAST CANCERS AND MATCHED DECALCIFIED BONE METASTASES IDENTIFIES CLINICALLY ACTIONABLE TARGETS

3.1 ABSTRACT

Bone metastases (BoM) are a significant cause of morbidity in patients with Estrogen-receptor (ER)-positive breast cancer, yet characterizations of human specimens are limited. In this study, exome-capture RNA-sequencing (ecRNA-seq) on aged (8-12 years), formalin-fixed paraffin-embedded (FFPE) and decalcified cancer specimens was first evaluated. Gene expression values and RNA-seq quality metrics from FFPE or decalcified tumor RNA showed minimal differences when compared to matched flash-frozen or non-decalcified tumors. ecRNA-seq was then applied on a longitudinal collection of 11 primary breast cancers and patient-matched *de novo* or recurrent BoM. BoMs harbored shifts to more Her2 and LumB PAM50 intrinsic subtypes, temporally influenced expression evolution, recurrently dysregulated prognostic gene sets and altered expression of clinically actionable genes, particularly in the CDK-Rb-E2F and FGFR-signaling pathways. Taken together, this study demonstrates the use of ecRNA-seq on decade-old and decalcified specimens and defines expression-based tumor evolution in long-term, estrogen-deprived metastases.

Contributors to study: Nolan Priedigkeit[1,2,4§], Rebecca J. Watters[2,3§], Peter C. Lucas[2,6], Ahmed Basudan[2,4,5], Rohit Bhargava[6], William Horne[7], Jay K. Kolls[7], Zhou Fang[8], Margaret Q. Rosenzweig[9], Adam M. Brufsky[2,10], Kurt R. Weiss[3], Steffi Oesterreich[1,2,4*], Adrian V. Lee[1,2,4,5*]

[§]Authors contributed equally to this work

*Shared senior authorship

[1]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, Pennsylvania

[2]Women's Cancer Research Center, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania

[3]Department of Orthopedic Surgery, University of Pittsburgh, Pittsburgh, PA

[4]Magee-Women's Research Institute, Magee-Women's Research Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

[5]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA

[6]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

[7] Richard King Mellon Foundation Institute for Pediatric Research, Children's Hospital of Pittsburgh of University of Pittsburgh Medical Center (UPMC), Pittsburgh, Pennsylvania, USA.

[8]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania

[9]Acute and Tertiary Care Department, University of Pittsburgh School of Nursing, Pittsburgh, Pennsylvania.

[10]Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

3.2 INTRODUCTION

Bone metastases (BoM) occur in approximately 65-75% of breast cancer patients with relapsed disease, resulting in significant comorbidities such as fractures and chronic pain[221]. Following colonization to the bone, breast cancer cells exploit the local microenvironment by activating osteoclasts, which in turn provides proliferative fuel for tumor cells[222]. This process is targeted clinically using anti-osteoclast agents such as bisphosphonates and RANKL inhibitors, yet these therapies do not confer significant survival benefits[223].

Importantly, the majority of breast cancers that metastasize to bone are estrogen receptor (ER)-positive and present clinically in the context of long-term endocrine therapies such as selective estrogen receptor modulators and aromatase inhibitors (Section 1.1.2.1)[224]. *In vivo* models of BoM have unfortunately been somewhat restricted to ER-negative disease due to the more indolent characteristics of ER-positive cell lines[225]. Molecular characterizations of ER-positive specimens that have recurred in an estrogen-deprived system, which represents the major burden of breast cancer BoM, are thus essential to reinforce the significant scientific contributions made using *in vivo* bone metastasis models[135,226–228]. Nonetheless, datasets are currently limited, in part due to the practical difficulties of obtaining and processing human BoM specimens[229].

Large-scale molecular characterizations of patient-matched samples—primary tumors and synchronous or asynchronous matched metastases—show that metastatic lesions acquire features distinct from primary tumors that are either clinically actionable or confer therapy resistance[179,180,230]. Indeed, current treatment guidelines in breast cancer recommend a biopsy to guide therapy in advanced disease if possible and our previous work (Chapter 2) further supported this notion[21]. Unfortunately, BoM often undergo harsh decalcification procedures with

strong acids to eliminate calcium deposits prior to specimen sectioning. Decalcification degrades nucleic acids and can alter results of immunohistochemistry[231–233]. Furthermore, formalin-fixed paraffin embedding (FFPE)—often performed in concert with decalcification—causes severe degradation and hydrolysis of RNA[234]. In light of this, new capture-based methods of nucleic acid sequencing on aged FFPE specimens have shown efficacy in identifying DNA variants and even guiding care in academic centers[235–237]. Exome-capture RNA-sequencing (ecRNA-seq) is less well characterized in aged tumor samples, although recent studies on FFPE specimens have shown promising expression correlations with flash-frozen tissues[238–240].

Because of the untapped potential of archived, decalcified BoM specimens, the burden of BoM in breast cancer patients and the lack of long-term endocrine treated tumor datasets, the performance of ecRNA-seq from decade-old, degraded and decalcified tumor samples was first assessed. Following this evaluation, ecRNA-seq was then applied to a collection of 11 ER-positive patient-matched primary breast cancers and bone metastases to define transcriptional evolution in breast cancer cells following metastatic colonization in the bone and years of endocrine therapy.

3.3 MATERIALS AND METHODS

3.3.1 Sample acquisition

Eleven sets of formalin-fixed paraffin-embedded (FFPE) primary breast tumors and patient-matched bone metastases (total of 22 samples) were obtained from the Health Sciences Tissue Bank, a certified honest broker facility at the University of Pittsburgh that maintains an IRB-

approved protocol for collecting excess tissue and biological materials. A molecular pathologist reviewed hematoxylin and eosin slides from each sample and then subsequently cut 0.6-1 mm cores from the paraffin block exclusively from regions of high tumor cell purity for RNA extraction. De-identified clinical and biological data were collected under the approval of the University of Pittsburgh Institutional Review Board (Protocol numbers: PRO14040193 and PRO10050461).

3.3.2 Tissue processing and RNA extraction

Tissues were digested over-night with shaking at 300 rpm at 56 °C in PKD buffer with the addition of proteinase K (Qiagen). RNA extraction was then performed with Qiagen's FFPE RNeasy kit (Qiagen, Cat#73504) according to the manufacturer's instructions under sterile RNase/DNase free conditions. RNA concentration was determined with the Qubit 3.0 Fluorometer (ThermoFisher Scientific). Quality RNA integrity number (RIN) scores and fragment sizes (DV200 metrics) were obtained utilizing either the Agilent 2100 Bioanalyzer or the Agilent 4200 TapeStation.

3.3.3 Exome-capture RNA-sequencing

Sequencing library preparation was performed using a minimum of 25 ng of RNA according to Illumina's TruSeq RNA Access Library Preparation protocol. Indexed, pooled libraries were then sequenced on the Illumina NextSeq 500 platform with a High Output flow cell producing

stranded, paired-end reads (2 X 75 bp). A target count of 50 million reads per sample was used to plan indexing and sequencing runs.

3.3.4 RNA-sequencing expression quantification and normalization

RNA transcripts from paired-end FASTQ files were mapped and quantified using k-mer based lightweight-alignment with seqBias and gcBias corrections (Salmon v0.7.2, quasi-mapping mode, 31-kmer index built from GRCh38 Ensembl v82 transcript annotations)[241]. Transcript-level abundance estimates were collapsed to gene-level estimates using tximport2[242]. To filter out non- or low expressed genes, only genes harboring a TPM value of more than 0.5 in at least 10% of samples were considered. Gene-level counts or log2 transformed TMM-normalized CPM (log2normCPM) values were implemented for subsequent analyses[243,244].

3.3.5 Expression correlations and RNA-seq quality assessment

Exome-capture RNA-seq was performed on two cohorts: 1) a set of four aged (ranging from 8 – 12 years) primary breast cancer specimens that at the time of surgical resection were split in half and either immediately embedded in optimal cutting temperature (OCT) compound and flash-frozen for storage at -80C, or formalin-fixed paraffin embedded (FFPE) and stored at room temperature. A second cohort consisted of three breast cancer bone metastases that at the time of resection were split in half and either decalcified or nondecalcified and processed to FFPE. These datasets were quantified and normalized as described above. Pearson *r* correlations between all samples were determined using log2normCPM values. Reads and mapping rates were obtained from *Salmon*. More detailed RNA-seq metrics were calculated and plotted using

QoRTs (v1.1.8) following two-pass read alignment with STAR (v2.4.2a) for the 11 patient-matched cases[245,246].

3.3.6 *tumorMatch* patient-matched sample identifier

To confirm samples were patient-matched, variants from RNA-seq were called using *GATK's Best Practices for variant calling on RNA-seq*[247]. Output .vcf files were then provided to *tumorMatch*, a custom *R* script that analyzes a pool of .vcf files and calculates the proportion of shared variants (POSV) between each .vcf. These proportion values were visualized using *corrplot* in *R*[248].

3.3.7 Unsupervised hierarchical clustering and intrinsic subtyping

Hierarchical clustering was performed using the *heatmap.3* function (https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R) in R on log2normCPM values of the top 5% most variable genes (defined by IQR) with 1 minus Pearson correlations as distance measurements and the "average" agglomeration method. PAM50 calls were generated using the *molecular.subtyping* function in *genefu*[191]. A separate cohort of exome-capture RNA-sequencing expression data from primary tumors (n = 12 ER-negative, 9 ER-positive) was merged with the bone metastasis cohort to help account for test-set bias and increase the stability of the PAM50 assignments[192]. To call PAM50 subtypes, for each query sample in the bone metastasis cohort a random subset of primary tumor expression data was added to enforce a balanced distribution of ER-positive and ER-negative tumors. This was repeated 20 times and the discrete PAM50 subtype was designated as the mode of this 20-fold

PAM50 assignment test while the final probability score was an average of all 20 probability scores from *genefu*.

3.3.8 Differential gene expression

Salmon gene-level counts with effective lengths of target transcripts were used to call differentially expressed genes (DEGs) between primary tumors and bone metastases using DESeq[249]. Given samples were patient-matched, a multi-factor design was implemented (~Patient + Tumor [i.e. primary vs. metastasis]). Genes with an FDR adjusted p-value of less than 0.10 were assigned as differentially expressed. An unclustered heatmap using log2normCPM values from the 207 DEGs, first segregated by metastatic log2FoldChange gains and losses and then sorted by DESeq2 adjusted p-values, was created in R using heatmap.3. Differentially expressed genes within the *MsigDB* database that were gained or lost in bone metastases were separately interrogated for gene ontology (GO: Biological Process) enrichment by computing significant (top 10 gene sets) gene overlaps using the MsigDB online tool[250].

3.3.9 ssGSEA signatures and METABRIC survival analyses

Microarray expression along with disease-specific survival (DSS) data was obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) through Synapse (https://www.synapse.org/, Synapse ID: syn1688369), following IRB approval for data access from the University of Pittsburgh[14]. Normalized expression values from IHC-confirmed ER-positive tumors were used to develop a single-sample gene-set enrichment score (ssGSEA) for strongly DEGs (adjusted p-value < 0.05) between primary tumors and bone metastases[251]. 48

genes that carried positive log2FoldChange values and had a corresponding gene expression value in METABRIC were assigned to the "boneMetSigUp" signature; 74 genes with negative log2FoldChange values were assigned to the "boneMetSigDown" signature. A ssGSEA score for each sample from both gene sets was calculated using the ssGSEA method implemented in the *GSVA R* package[252]. Binary dichotomization of samples (low vs. high) based on ssGSEA signature score strata (10th, 25th, 50th, 75th, 90th percentiles) and log-rank testing were used to assess significant differences in DSS[253]. The strata with the most significant log-rank p-values were plotted using *survminer* from CRAN[254].

3.3.10 Ranked Gene Set Enrichment Analysis (GSEA)

To determine pathways significantly enriched or lost in breast cancer bone metastases versus patient-matched primaries, GSEA analyses were performed using gene sets with coordinately expressed genes representing specific biological and cancer-related pathways (MSigDB: H and C6 sets). Input into GSEA was a ranked list (DESeq2 log2FoldChange values) of 21,702 genes. Enrichment scores, significance values and plots were generated using default settings of the Broad Institute's *javaGSEA* Desktop Application (v2.2.3).

3.3.11 RBBP8 survival analysis

RBBP8 expression was further interrogated and plotted using log2normCPM values from patient-matched tumors. *RBBP8* expression influence on DSS in METABRIC ER-positive patients was interrogated as described above. *RBBP8* expression influence on bone-met free

survival (BMFS) was assessed by querying a GCRMA-normalized microarray expression dataset (GSE12276) from 204 primary tumors and associated survival data as described above[139].

3.3.12 Gains and losses in clinically actionable genes

Clinically actionable gene set was obtained using the Drug Gene Interaction Database (DGIdb 2.0)[195]. Considering metastatic fold-change distributions calculated from log2normCPM values for all genes were slightly different for each case, stringent case-specific fold-change thresholds were used to transform continuous fold-change values into discrete "expression alterations." More specifically, if the fold-change value for a clinically actionable *GENE_X* was greater than the 95th percentile of all gene fold-change values in that case, *GENE_X* would be designated as a significant, case-specific expression gain. If the fold-change value for *GENE_Y* was lower than the 5th percentile, *GENE_Y* was designated as a significant, case-specific expression loss (Supplementary Data S13). After assigning discrete expression alteration calls to clinically actionable genes, data was visualized using the *OncoPrint* function in *ComplexHeatmap*[194].

3.3.13 Statistical considerations

To determine differentially expressed genes between patient-matched primary tumors and bone metastases, *DESeq2* was used. *DESeq2* is designed for RNA-seq gene-based count abundance estimates and assigns differential expression *p-values* based on a negative binomial distribution. For Kaplan-Meier curves, the logrank test was used to determine statistically significant differences in event probabilities (i.e. death or time to metastasis) based on binary expression or

signature strata. For single gene queries, paired Wilcoxon-signed ranked tests on log2normCPM values were used.

3.4 RESULTS

3.4.1 ecRNA-sequencing of aged and decalcified breast cancers

To determine the feasibility of sequencing an aged, FFPE and decalcified tumor cohort, ecRNA-seq on two separate sample sets was performed. The first sample set included four cases of primary breast tumors that at the time of resection, were split in two. One part was flash-frozen and stored at -80 C and the other tumor section was formalin-fixed paraffin embedded and stored at room temperature. Storage times ranged from 8.2 to 12.3 years. Post-alignment RNA-sequencing QC analyses showed differences in GC content and insert size, yet gene body coverage and transcript diversity assignments were largely similar (Figure 9A). After quantifying and normalizing gene abundances, expression correlations between frozen and FFPE matched samples were assessed using log2normCPM values. *Pearson r* correlations ranged from 0.929 to 0.963, with an average correlation of 0.953 (Figure 9B). The same analysis was performed using a second sample set of matched FFPE-decalcified and FFPE-non-decalcified samples. Again, no concerning deviations in RNA-seq quality metrics were observed between the two differently processed sample groups (Figure 9C) and Pearson *r* expression correlations ranged from 0.936 to 0.969 (Figure 9D). Furthermore, correlation matrices of the two sample sets showed matched tumor sample expression values were more similar to each other than expression values from tumors with equivalent processing and storage (Appendix A.2: Figure 26). Full RNA-seq metrics

from the QC analysis did reveal differences in some metrics between FFPE and flash-frozen tissue (i.e. splice junction loci number), that may be informative for other applications such as indel mutation calling or isoform detection (Data Supplement 2: S1, S2). In summary, ecRNA-seq shows outstanding quality metrics for analysis of aged FFPE and decalcified bone metastases samples.

3.4.2 ecRNA-seq of breast cancer bone metastases

Following the validation of ecRNA-seq, a cohort of 11 ER-positive patient-matched primary tumors and BoMs was acquired through the University of Pittsburgh Health Science Tissue Bank (Table 2, Data Supplement 2: S3). Abstracted clinical records showed that nearly all patients (10/11) were documented as having received adjuvant endocrine therapy, and bone metastasis free survival ranged from 0 (de novo bone metastasis) to greater than 5 years with the most common site of bone metastasis being the vertebral column.

ecRNA-seq was performed on the 22 samples yielding an average read count of 58,294,593 and an average *Salmon* transcript mapping rate of 92.6% (Data Supplement 2: S4). Consistent with the initial quality control studies above, quality metrics on these samples showed consistent gene body coverage, GC content, insert sizes and transcript diversity regardless of decalcification status (Appendix A.2: Figure 27, Data Supplement 2: S5). Furthermore, since samples within the cohort had been surgically excised and banked many years apart, all paired specimens underwent an analysis of shared variants, which confirmed tumor pairs were patient-matched (Appendix A.2: Figure 28).

Table 2: Clinicopathological features of breast cancer bone metastasis cohort
Abbreviations: Dx, diagnosis; Tx, therapy; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; BoM, bone metastasis; BMFS; bone metastasis free survival; OS, overall survival

Case	Age Dx	Hist Subtype	Stage	ER Prim	PR Prim	HER2 Prim	BoM Location	BoM Decal	Endo Tx	HER2 Tx	Radio Tx	Chemo Tx	BMFS	OS
17	54	IDC	IIIA	Pos	Pos	Neg	Ileum	Yes	Yes	No	Yes	Yes	24	46
19	50	IDC w/ lobular features	IV	Pos	Pos	Neg	Vertebra	No	Yes	No	Yes	No	0	75
22	60	IDC	IIA	Pos	Pos	Neg	Femur	No	Yes	No	Yes	Yes	18	37
31	59	IDC w/ lobular features	IIB	Pos	Pos	Neg	Vertebra	Yes	Yes	No	Yes	Yes	43	55
34	38	IDC	IIIA	Pos	Pos	Neg	Vertebra	Yes	Yes	No	Yes	Yes	65	130
43	65	IDC	IV	Pos	Pos	Neg	Vertebra	Yes	Yes	No	Yes	No	0	54
44	56	IDC	IA	Pos	Pos	Pos	Femur	No	NA	Yes	Yes	Yes	23	42
48	49	ILC	IIIC	Pos	Pos	Neg	Vertebra	No	Yes	No	Yes	Yes	28	68
55	56	IDC	IV	Pos	Pos	Neg	Femur	No	Yes	No	NA	No	0	137
60	44	IDC	IIB	Pos	Pos	Neg	Sacrum	Yes	Yes	No	Yes	Yes	46	53
A25	39	IDC	IIIA	Pos	Pos	Neg	Femur	Yes	Yes	Yes	Yes	Yes	38	57

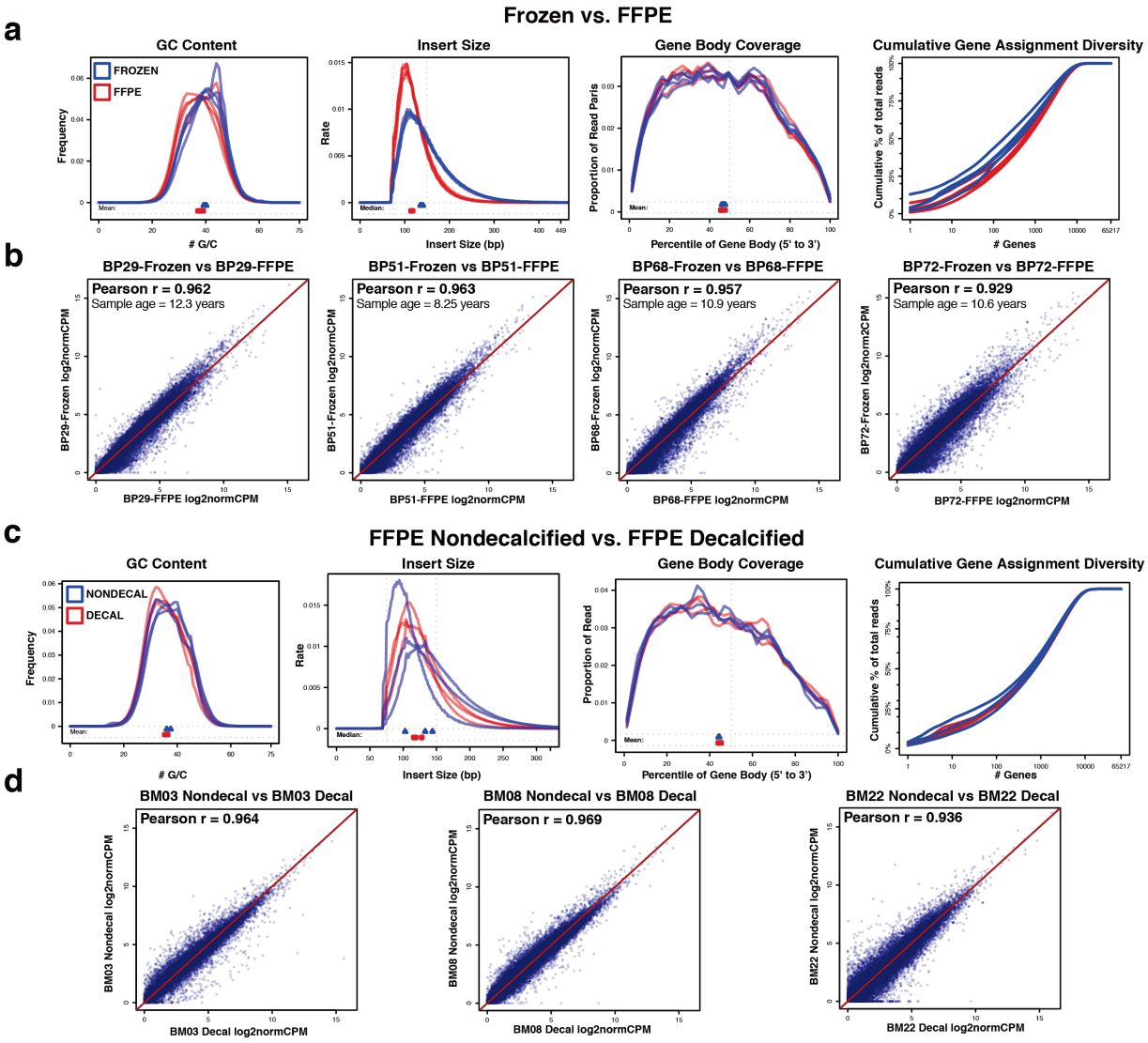

Figure 9: Exome-capture RNA-sequencing of aged, FFPE and decalcified tumors
(A) RNA-seq quality metrics (GC content, insert size, gene body coverage and cumulative gene assignment diversity) of aged and tumor-matched FFPE and flash-frozen (FF) sample; FF samples in blue, FFPE samples in red. **(B)** Expression value correlations between four sets of matched tumor samples (FF vs. FFPE) along with Pearson *r* correlations and sample ages. **(C)** RNA-seq quality metrics of matched non-decalcified and decalcified samples; non-decalcified samples in blue, decalcified samples in red. **(D)** Expression correlations between three sets of matched tumor samples (non-decalcified vs. decalcified) along with Pearson *r* correlations.

3.4.3 Clustering and temporal expression shifts

Unsupervised hierarchical clustering of patient-matched pairs revealed that decalcification of BoMs did not produce independent clades, with 5 of 11 BoM clustering in the same doublet clade as their matched primary (denoted with * in Figure 10A). Notably, 3 of the 5 doublet clustering cases were de novo metastases. Discrete PAM50 intrinsic subtype assignments were identical in 6 of 11 pairs. 2 pairs switched from LumA to LumB in the metastasis, 1 pair from LumB to LumA, 1 pair from LumB to Her2 and another was classified as Normal subtype in the primary tumor and LumB in the BoM (Figure 10B). To obtain more granularity than discrete PAM50 calls, probability scores for each PAM50 subtype were assigned (Figure 10B and Data Supplement 2: S6). Her2 and LumB profile gains (defined as a probability gain of >10% in a matched BoM) were the most common—being observed in 4 of 11 cases (Figure 10B). Given observed shifts in expression profiles of bone metastases and doublet clustering of de novo bone metastases, temporal influence on transcriptional evolution was analyzed. Pearson *r* correlations between each patient-matched pair using log2normCPM expression values were utilized as a metric for transcriptional similarity. Expression pair similarity was significantly correlated (Pearson $r = -0.864$, $p\text{-value} < 0.001$) with time from primary tumor diagnosis to bone metastasis (Figure 10C).

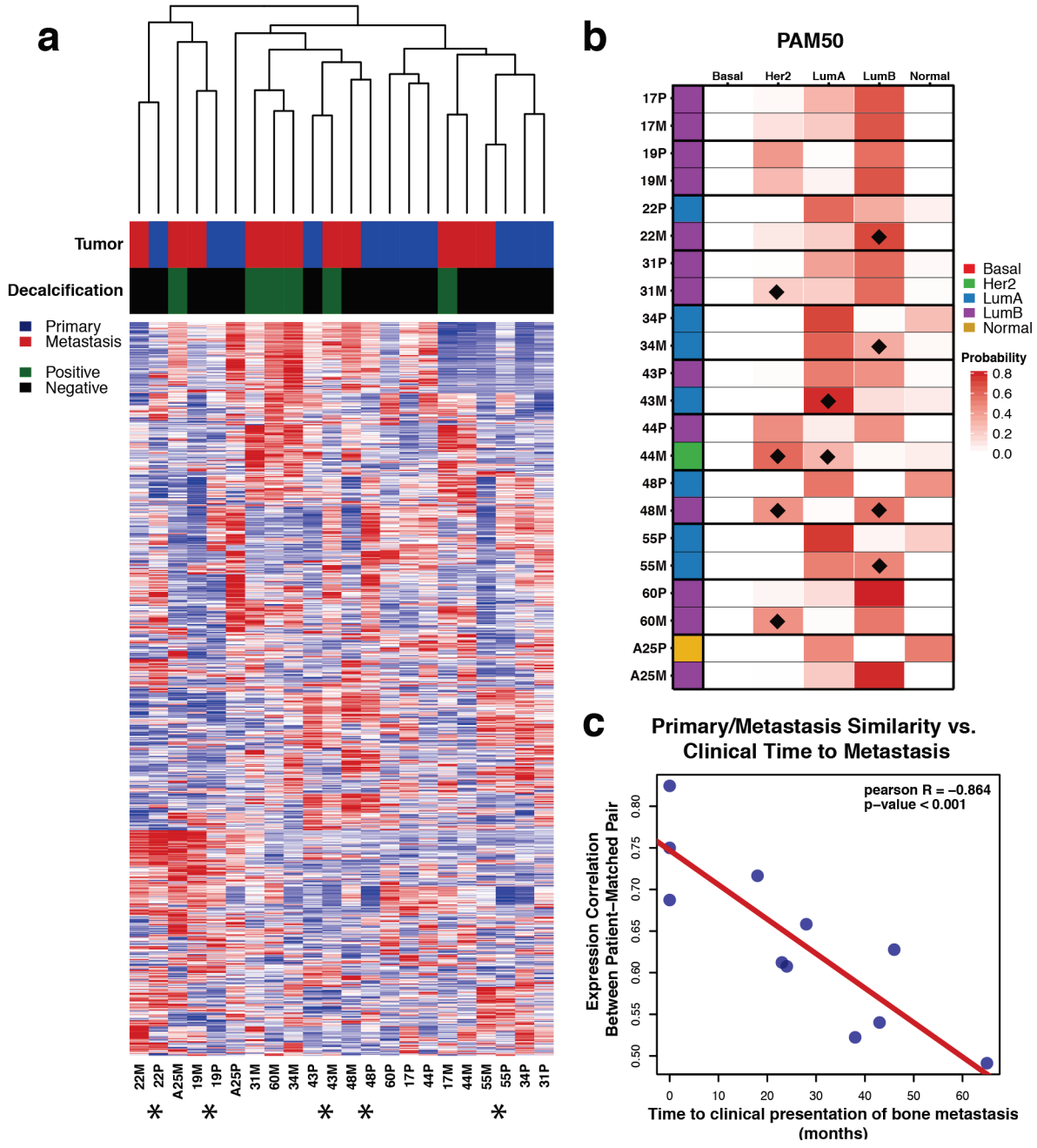

Figure 10: Unsupervised clustering, intrinsic subtype shifts and temporal evolution of ER-positive bone metastases.
(A) Unsupervised hierarchical clustering heatmap (red = high relative expression, blue = low relative expression) of patient-matched pairs using the top 5% most variable genes (n = 1096) across the cohort. Tumor (primary in blue, metastasis in red) and decalcification status (positive in green, negative in black) indicated. Asterisks below heatmap designate patient-matched pairs that cluster in a single doublet clade. **(B)** Discrete PAM50 assignments (red = basal, green = HER2, blue = LumA, purple = LumB, yellow = Normal) and PAM50 probabilities for patient-matched pairs. PAM50 probability shifts in metastases (if greater than 10%) are marked with a black diamond. **(C)** Correlation of patient-matched tumor expression similarity versus clinical time to metastasis with Pearson *r* value and correlation p-value.

3.4.4 Differentially expressed genes

To determine genes consistently up- or downregulated in bone metastases, a paired DESeq2 differential gene expression analysis was performed. 207 genes were differentially expressed (FDR adjusted *p-value* < 0.10)—80 genes with increased and 127 genes with decreased expression in bone metastases (Figure 11A, Data Supplement 2: S7). Gene ontology analysis was performed to determine biological processes represented in the up- and downregulated gene sets. Generally, genes within osteogenic programs showed the most significant increases in expression while muscle-related, adhesion and motility gene sets were found to be significantly lost in bone metastases (Figure 11A, Data Supplement 2: S8, Appendix A.2: Figure 29). Given that a subset of these genes may be mediating therapy resistance and/or distant metastases, single sample gene set enrichment analysis (ssGSEA) scores[251] were calculated using tumor expression data from patients with long-term outcomes in METABRIC[14]. Two separate gene lists were created to build the signatures—representing the most significantly upregulated (boneMetSigUp) and downregulated (boneMetSigDown) genes in bone metastases (Data Supplement 2: S9). Tumors intrinsically expressing higher boneMetSigUp and lower boneMetSigDown ssGSEA scores conferred worse (log-rank *p-value* < 0.001) disease-specific survival outcomes (Figure 11B). To increase the power of discerning gene expression effects due to long-term estrogen deprivation, a differential gene expression analysis was performed excluding the treatment-naïve, de novo bone metastases. This yielded a list of 612 differentially expressed genes (Data Supplement 2: S10), some of which were not detected as differentially expressed with treatment-naïve de novo bone metastasis cases included.

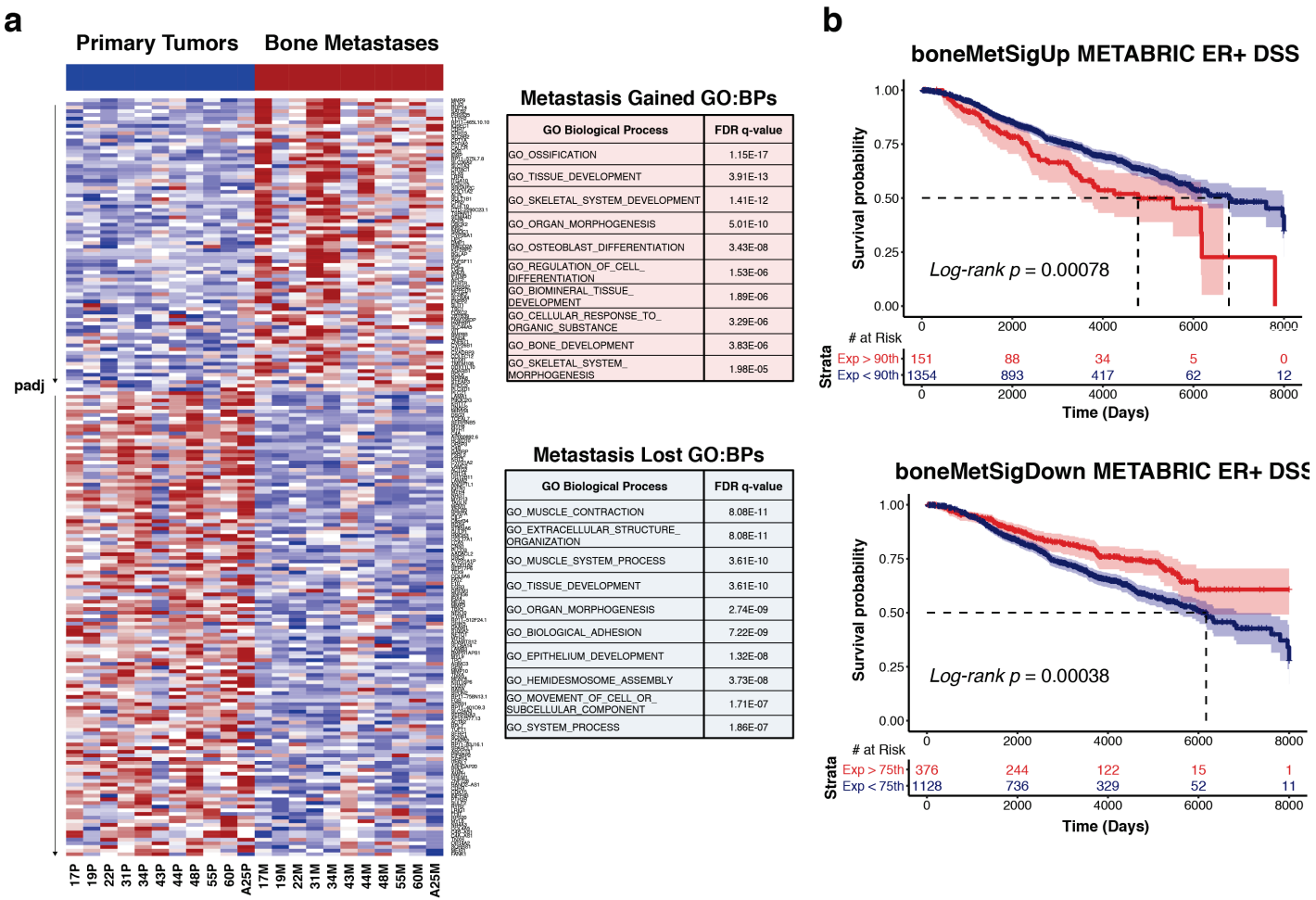

Figure 11: Differentially expressed genes in patient-matched bone metastases
(A) Left, heatmap (red = high relative expression, blue = low relative expression) of log2normCPM values from 207 differentially expressed genes (FDR adjusted p-value < 0.10) between primary tumors and patient-matched bone metastases. Heatmap is segregated into two sections; genes with log2FoldChange > 0 on top and genes with log2FoldChange < 0 on bottom. Each section is gene-sorted by adjusted p-values. Right, Gene Ontology: Biological Process gene overlap analysis for genes with significant expression gains (top, red) and losses (bottom, blue) in bone metastases. Top 10 pathways are shown alongside FDR adjusted q-values. **(B)** Disease-specific survival outcome differences in ER-positive METABRIC tumors using boneMetSigUp (top) and boneMetSigDown (bottom) expression scores as strata. 95% confidence intervals are highlighted along with log-rank p-values and associated risk table.

3.4.5 Dysregulated gene sets and RBBP8 expression loss

To determine pathway level changes in breast cancer bone metastases, a pre-ranked GSEA was performed. All genes were ranked by DESeq2 calculated log2 fold-changes (metastasis vs. primary, Data Supplement 2: S11) and then analyzed for enrichments using Molecular Signature Database (MsigDB) gene sets (http://software.broadinstitute.org/gsea/msigdb, H: Hallmark gene sets, C6: Oncogenic signatures)[250]. This yielded several significantly metastasis-enriched and metastasis-diminished gene sets (FDR *q-val* < 0.10, Data Supplement 2: S12). The three most significantly enriched gene sets in metastases involved E2F transcription factor targets, genes mediating the G2M checkpoint and an experimental perturbation gene set consisting of genes up-regulated with knockdown of *RBBP8* in a breast cell line (Figure 12A)**.** Other upregulated gene sets included hedgehog signaling and gene sets associated with Rb loss and KRAS gains. The three most significantly negatively correlated gene sets consisted of an NFKb/TNF gene set, genes involved in epithelial mesenchymal transition (EMT) and an embryonic development gene set. We further interrogated *RBBP8* due to it being the most significant gene set enriched in bone metastasis. As predicted by the enrichment, bone metastases carried significant *RBBP8* expression loss (Wilcoxon-signed rank *p-value* = 0.02), with 5 of 11 metastases [45%] having at least a 2-fold decrease in expression versus patient-matched primaries (Figure 12B). Tumors intrinsically expressing lower levels of *RBBP8* showed worse disease-specific and bone metastasis-free survival outcomes (Figure 12C).

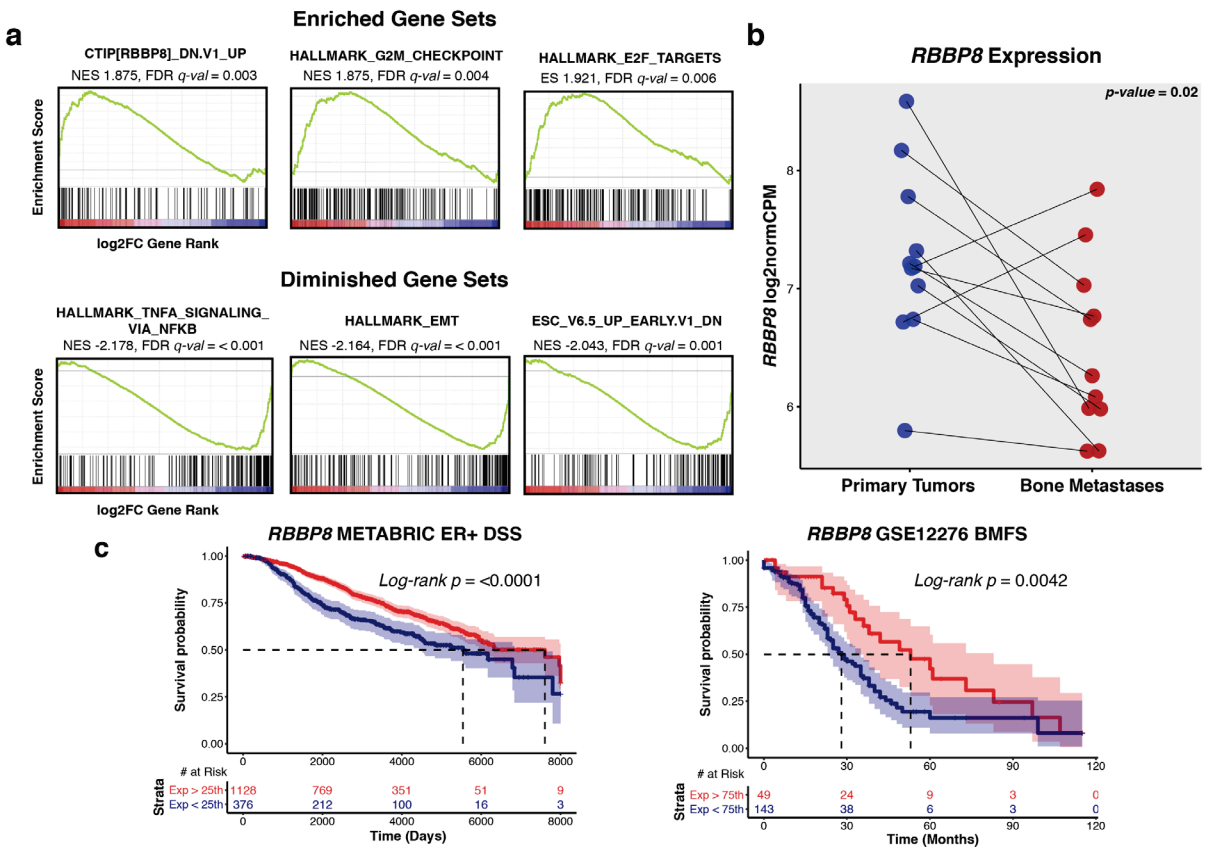

Figure 12: Dysregulated gene sets and *RBBP8* loss in breast cancer bone metastases
(A) Top three enriched and depleted gene sets (by FDR q-value) in bone metastases from ranked GSEA analysis. Gene list ranking was performed using log2FoldChange values from DESeq2 differential expression output, where a positive log2FoldChange represents increased expression in metastasis (red) and a negative log2FoldChange represents decreased expression in metastasis (blue). Green line shows running enrichment score as algorithm walks down the ranked gene list. Black vertical lines below curve show where genes within the query gene set are represented in the ranked list. Normalized enrichment score (NES) and FDR q-values are noted below gene set names. **(B)** *RBBP8* expression values (log2normCPMs) in primary tumors (blue) and bone metastasis (red). Pairs are connected with a line and Wilcoxon signed-rank *p-value* is shown. **(C)** Disease-specific survival outcome differences in ER-positive tumors (METABRIC) and bone metastasis free survival differences (GSE12276) using normalized *RBBP8* expression values as strata. 95% confidence intervals are highlighted along with log-rank p-values and risk tables.

3.4.6 Expression gains and losses in clinically actionable genes.

Because of the observed acquisition of clinically actionable targets reported in other studies of paired primary and recurrent tumors (Chapter 2)[179,230], a paired expression analysis to define clinically actionable expression changes in ER-positive bone metastases was performed (Data Supplement 2: S13). Using stringent, case-informed cutoffs for expression alterations (Appendix

A.2: Figure 30), the most common expression losses in bone metastases were *PIK3C2G* [8 of 11, 73%], *ESR1 [7 of 11, 64%] and TUBB3 [6 of 11, 55%]* (Figure 13A and Appendix A.2: Figure 31). Other notable losses included *GREM1*, *PTPRT*, *CDKN2A, KIT* and *GATA3*. The most recurrent expression gains were *FGFR3* [7 of 11, 64%], *EPHA3 and PTPRD* [6 of 11, 55%]. *PDGFRA*, *PTCH1*, *ALK, HGF, FGFR1* and *FGFR4* also showed highly recurrent gains (Figure 5B). Interestingly, some expression gains were absent in *de novo* bone metastasis cases (Cases 19, 53 and 55) yet highly recurrent in long-term endocrine-deprived cases (*EPHA3, PTPRD, PDGFRA, PTCH1),* suggesting clinically actionable, treatment-driven gains in endocrine-resistant breast cancer recurrences.

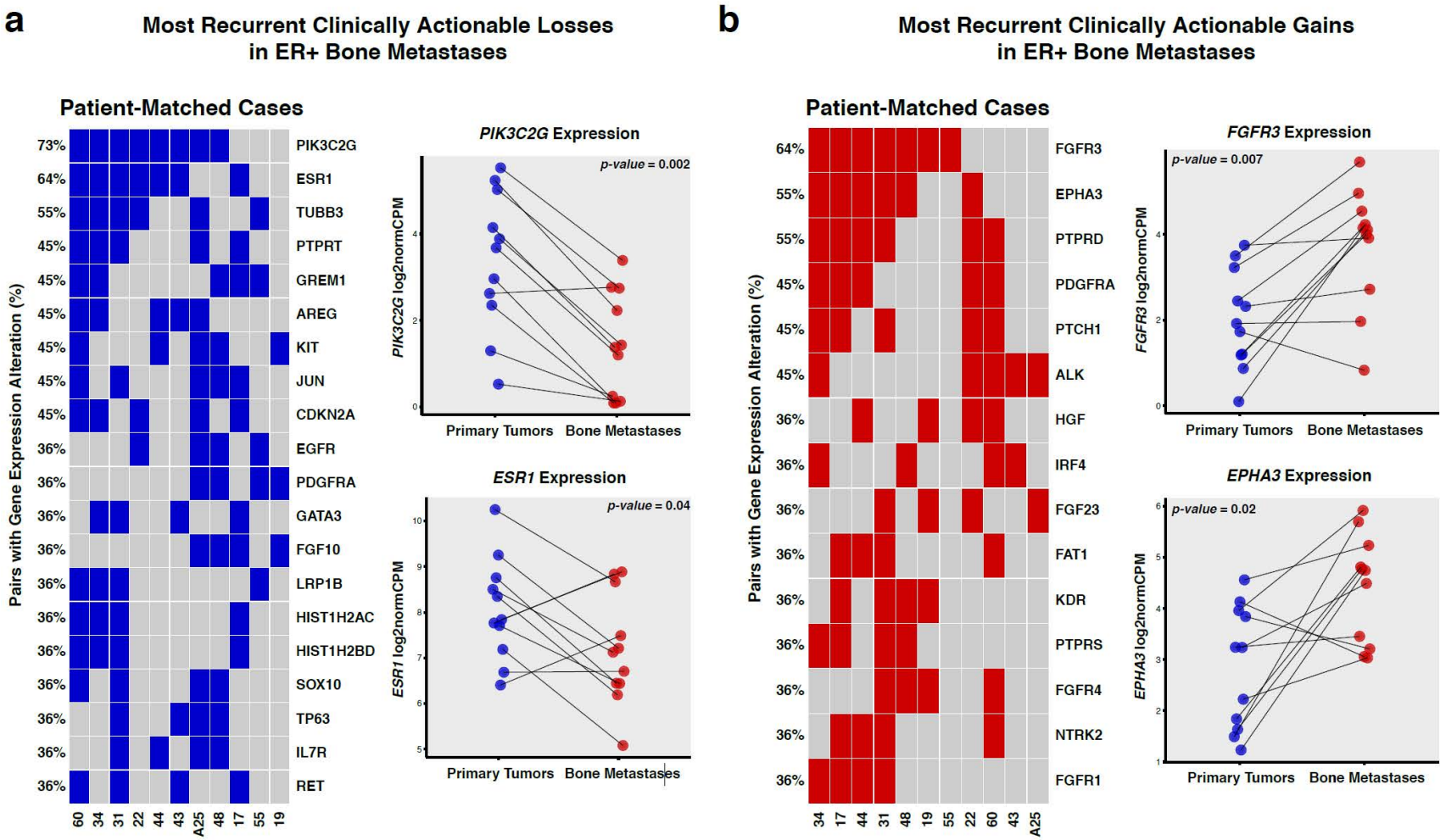

Figure 13: Recurrent, clinically actionable expression gains and losses in ER-positive bone metastasis
(A) Recurrent expression alteration losses, ranked by frequency, for each patient-matched case (columns). Each blue tile represents a bone metastasis with a lower log2FoldChange vs. its matched primary than the case-specific expression loss threshold. Expression values (log2normCPMs) for most recurrent losses (*PIK3C2G, ESR1*) are pair plotted with corresponding Wilcoxon signed-rank test *p-values* noted. **(B)** Recurrent expression alteration gains, ranked by frequency. Red tiles represent bone metastases with higher log2FoldChange than the case-specific expression gain thresholds. The two most recurrent expression gains (*FGFR3, EPHA3)* are also plotted.

3.5 DISCUSSION

Bone is the most common site of distant recurrence for patients with ER-positive breast cancer, yet comprehensive sequencing datasets of endocrine therapy treated, metastatic samples are currently limited. This is in part due the challenge of obtaining tissue, and degradation of nucleic acids caused by decalcification. In this study, we found that aged FFPE and FFPE-decalcified tumors showed highly similar transcript quantification values as matched flash-frozen and FFPE-non-decalcified tumors. As a proof-of-concept, we then applied ecRNA-seq to a cohort of patient-matched primary and bone metastases collected over a period of five years. We identified subtle shifts in intrinsic subtypes and found a strong temporal influence on transcriptional evolution in breast cancer recurrences. Furthermore, we created several differentially expressed gene sets/signatures that are prognostic and point towards acquired *RBBP8* loss, CDK-Rb-E2F and FGFR pathway gains as mediators of ER-positive breast cancer progression. Lastly, we found bone metastases commonly gain or lose expression in clinically actionable genes, which may be distinct from primary tumors.

ecRNA-seq is an effective method for quantifying expression on aged, FFPE and decalcified tumor specimens. Previous work has assessed nucleic acid amplification success, DNA-sequencing and RNA integrity metrics using decalcified samples[233,255,256]; however, a comprehensive analysis of RNA-sequencing, to our knowledge, has not yet been performed. Consistent with only very minor differences between GC content, insert sizes and other QC metrics, gene expression values between aged matched FFPE/flash-frozen and FFPE-decalcified/FFPE-non-decalcified tumors are highly correlated (Pearson r range 0.929 – 0.969).

This study reinforces and should encourage the use of capture-hybridization approaches to sequence RNA from retrospectively collected, low yield, highly degraded and decalcified archival specimens (Data Supplement 2: S14)[238–240]. Expanding sample sets and modalities for genome-wide characterization, especially for rare specimen cohorts that may be impractical to obtain prospectively in large numbers, will accelerate translational discoveries.

Given promising results from our evaluation, we applied ecRNA-seq in a proof-of-concept effort to characterize the transcriptome of 11 archival patient-matched ER-positive primary and recurrent metastases— 3 cases having treatment-naïve, *de novo* bone metastases and 8 recurrent cases harboring long-term endocrine-therapy treated metastases. In the recurrent cases, bone metastasis-free survival ranged from 18 to 65 months. Despite a large portion of the bone metastases being decalcified, global transcriptome QC metrics showed similar features (i.e. GC content, insert sizes, gene body coverage and transcript assignment diversity) and no outliers. Consistent with this, unsupervised hierarchical clustering showed no distinct clusters of decalcified samples, with 5 bone metastases clustering in the same doublet clade as their patient-matched primary breast cancer. Interestingly, 3 of these doublet clustering pairs were clinically *de novo*, treatment naïve bone metastases, implying limited transcriptional evolution from the primary tumor in synchronous metastases. This was further corroborated with a striking negative correlation between patient-matched expression similarity and time to bone metastasis, suggesting metachronous metastases that present later in their treatment course are more dissimilar from their derived primary lesions. Intrinsic subtyping revealed 5 of the 11 cases changed PAM50 subtypes, with 3 cases switching to LumB in the metastasis and another switching to Her2. Subtle Her2 and LumB profile shifts were also the most common when observing continuous PAM50 probability scores, even in samples that remained concordant in

their discrete PAM50 assignments. A recent, targeted expression study analyzed PAM50 assignments in 123 matched breast cancer metastases and the authors found similar frequencies of LumB and Her2 acquisitions in ER-positive metastatic tumors[257]. Given this transcriptional evolution to more LumB and Her2 profiles, a thoughtful reevaluation of therapy selection in the advanced and perhaps the adjuvant setting may be necessary.

We found 207 genes to be differentially expressed between primary tumors and patient-matched bone metastases. The top upregulated genes belonged to osteogenic gene sets—*BGLAP, RANKL, PTH1R* all showing significant expression gains—and supports *in vivo* modelling observations of breast cancer osteomimicry and hijacking of the bone microenvironment[258]. Downregulated gene sets included genes involved in broad categories such as cellular adhesion, hemidesmosome assembly and epithelium development, pointing towards specific biological programs lost following metastatic colonization. Moreover, when either the upregulated or downregulated genes are expressed coordinately in primary tumors, we found that they confer worse and better outcomes respectively in ER-positive tumors, suggesting some tumors may develop these transcriptional programs early in their evolution. Lastly, a differential expression analysis between endocrine naïve primary tumors and long-term endocrine treated bone metastases identified a larger list of differentially expressed genes. Importantly, known mediators of endocrine resistance are represented in the list, including dysregulated expression of *Wnt* family members[259], expression gains in *FGFR1*[50], *FOXC1*[260] and loss of *ESR1* expression[44]. Notably, many of these genes do not overlap with the differential expression analysis that included the de novo metastases, suggesting expression alterations specific to late recurrent therapy-treated tumors. This non-overlapping gene set included a greater than 2-fold average expression gain of *ABCG2* in therapy-exposed metastases—a multidrug resistance protein shown

to be active in breast cancer[261,262]—and loss of *CDKN2A*. *CDKN2A* encodes *p16,* a negative regulator of CDK4/CDK6 and is located on a common somatically deleted region (9p21) in cancer[263]. Given recent success of CDK4/CDK6-inhibiting compounds (palbociclib and ribociclib) in treating ER-positive breast cancers, this recurrent, acquired, metastatic-specific loss of *CDK2NA* is a clinically important observation[106–108].

Following significant gene-level changes, a gene set enrichment analysis defined enriched and diminished pathways in breast cancer bone metastases. Enriched genes included those involved in G2M checkpoint and E2F targets. Consistent with the observed LumB enrichments, breast cancer cells appear to develop a more proliferative phenotype following bone colonization and the strong enrichment of E2F signature in metastatic disease again highlights the CDK-Rb-E2F pathway as a potential actionable target. Interestingly, another study that utilized a targeted gene expression platform found proliferative gene signatures in ER-positive metastases may be more accurate at predicting overall survival than signatures in the primary tumor[257]. A survival analysis for this work was impractical given the small set of patient-matched pairs, but future meta-analyses are warranted to determine if gene expression signatures in metastases are better predictors of overall survival in the advanced setting, especially given the significant transcriptomic shifts observed in this study.

The most significant gene set enriched in bone metastasis was an experimental perturbation gene set involving the knockdown of the tumor suppressor *RBBP8*[264]. *RBBP8* (also known as CtIP) binds directly to Rb, mediates cell cycle regulation, helps maintain genomic stability and loss of *RBBP8* incurs tamoxifen resistance and sensitizes breast cancer cells to PARP inhibition *in* vitro[265–268]. Concordant with the GSEA analysis, bone metastases have significant expression loss of *RBBP8,* with 45% of cases showing a greater than 2-fold decrease

in expression. We found low *RBBP8* expression in ER-positive tumors confers poorer disease-specific survival and bone metastasis-free survival outcomes. These observations point to *RBBP8* loss in metastatic breast cancers as being a prime, perhaps therapeutically relevant candidate for further preclinical investigations.

Lastly, considering we have previously shown that brain metastases acquire highly recurrent gains in clinically actionable genes (Chapter 2)[230], particularly in HER2, we analyzed the same set of genes in bone metastases. All tumors harbored significant gains and losses, some of which were highly recurrent. *PIK3C2G*, a relatively uncharacterized gene in the PI3K pathway, was the most recurrent gene expression loss. Other notable losses included *ESR1*, *CDKN2A* and *GATA3*—genes that have already been implicated in endocrine therapy resistance in experimental models. Intriguingly, *GATA3* is one of the most recurrently mutated genes in breast cancer, being particularly enriched in ER-positive disease[13]. Moreover, *GATA3* inhibits breast cancer metastasis in various model systems and given losses of *GATA3* in ER-positive bone metastases are common, further evaluation of *GATA3* as a potentially targetable breast cancer metastasis suppressor gene should be encouraged[260,269,270]. Metastatic gains included FGFR family members (*FGFR3*, *FGFR4*, *FGFR1*), *ALK* and *KDR*—all protein products having small molecules currently in clinical trials. Interestingly, some highly recurrent expression gains (i.e. *EPHA3*, *PTPRD*, *PDGFRA*, *PTCH1*) were exclusive to long-term endocrine treated bone metastases suggesting them as prime, clinically actionable candidate mediators of therapy resistance. Collectively, these observations provide yet further evidence of acquired transcriptional programs in metastatic lesions and suggests that precision care in breast cancer should be informed by molecular features of advanced tumors in order to not miss targetable metastatic dependencies acquired in advanced disease.

Although this study points towards ecRNA-seq as being a viable option to characterize the transcriptome of archived, decalcified specimens, there are limitations. Firstly, multiple methods are used for decalcification with varying effects on nucleic acids and we were unaware of this information for the profiled specimens, as it is rarely recorded in clinical notes[233]. Secondly, in primary versus metastatic expression studies, it is difficult to deconvolute expression contributions from tumors versus the altered microenvironment of the distant organ site. To limit these artifacts in this study, regions of high tumor cellularity in the bone metastasis were cored by a trained molecular pathologist for RNA extraction, which is corroborated by RNA-seq derived tumor purity estimates—as no significant tumor purity differences between primary and metastatic tumors (Data Supplement 2: S15) were observed[271]. Nonetheless, single-cell sequencing approaches of metastatic tumors will be essential to bring cell-level resolution to transcriptional studies of metastatic tumors. Novel computational methods that deconvolute heterogeneous sample sets, until single-cell sequencing becomes more widely adopted, will also be essential[272–274]. All of this withstanding, features of the data are encouraging such as patient-matched tumors clustering together, intuitive PAM50 assignments, corroboration of other groups' findings and treatment-specific gains and losses. Finally, a limitation of this study is the small sample size. Hopefully, these results will encourage the use of ecRNA-seq to transcriptionally profile other highly degraded samples and begin a collection of genomic data from metastatic or rare tissues for integration. Importantly, de-identified clinical data should be provided alongside the sequencing, as in this study, to allow more fluid merging of datasets and inspire clinical phenotype-driven analyses.

Taken together, this study both validates the use of ecRNA-seq to transcriptionally profile highly degraded RNA from decade-old and decalcified tumor specimens and defines

multiple acquired and lost transcriptional programs in ER-positive bone metastases. We highlight acquired changes in the CDK-Rb-E2F and FGFR pathways, particularly relevant given the recent clinical use of CDK4/6 inhibitors, and point towards *RBBP8* as a particularly compelling candidate in breast cancer progression. We also find significant gains in clinically actionable genes that may have not been appreciated in primary tumors, reinforcing the need for longitudinal characterizations of cancer specimens to guide clinical care.

4.0 TRANSCRIPTOME-WIDE IDENTIFICATION OF RET AND HER2 SIGNALING AS RECURRENTLY ENRICHED DEPENDENCIES IN BREAST CANCER BRAIN METASTASES

4.1 ABSTRACT

Breast cancer brain metastases (BrM) are defined by complex adaptations to both adjuvant treatment regimens and the brain microenvironment. Consequences of these alterations remain poorly understood, as does their potential for clinical targeting. In this study, the most comprehensive of its kind to date, we extensively characterized the BrM-altered transcriptome across 21 patient-matched primary breast tumors and their associated brain metastases. We observe that breast cancer cells shift their expression profile following colonization in the brain parenchyma and demonstrate recurrent gains in RET expression and HER2 signaling. In line with these observations, inhibition of aberrant RET and HER2 results in significant anti-tumor activity in BrM patient-derived xenograft models and patient resected BrMs cultured *ex-vivo*. Altogether, our study identifies recurrent, acquired vulnerabilities in BrM that warrant immediate clinical investigation.

Contributors to study: Damir Varešlija[1*], Nolan Priedigkeit[5,A*], Ailís Fagan[1], Siobhan Purcell[1], Nicola Cosgrove[1], Philip J O'Halloran[2], Elspeth Ward[1], Sinéad Cocchiglia[1], Ryan Hartmaier[5], Carlos A. Castro[A], Li Zhu[B], George C. Tseng[B], Peter C. Lucas[3], Shannon L Puhalla[4], Adam M. Brufsky[4], Ronald L. Hamilton[3], Aju Mathew[4], Jose P. Leone[4], Lance Hudson[9], Róisín Dwyer[10], Sudipto Das[8], Darran P. O'Connor[8], Patrick G. Buckley[7], Michael Farrell[7], Arnold DK Hill[9], Steffi Oesterreich[5,A], Adrian V. Lee[5, 6,A*] and Leonie S. Young[1*]

*These authors contributed equally to this work

[1]Endocrine Oncology Research Group, Department of Surgery, Royal College of Surgeons in Ireland, Dublin, Ireland

[2]Department of Neurosurgery, National Neurosurgical Center, Beaumont Hospital, Dublin, Ireland.

[3]Department of Pathology, [4]Medicine, [5]Pharmacology and Chemical Biology, [6]Human Genetics, [A]Women's Cancer Research Center, Magee-Women's Research Institute, [B]Biostatistics, University of Pittsburgh Cancer Institute, University of Pittsburgh, PA, USA

[7]Departments of Neuropathology, [8]Molecular and Cellular Therapeutics and [9]Surgical Research, Royal College of Surgeons in Ireland, Dublin, Ireland

[10]Discipline of Surgery, School of Medicine, Lambe Institute for Translational Research, National University of Ireland, Galway, Ireland

4.2 INTRODUCTION

Given acquired transcriptional changes demonstrated in Chapters 2 and 3—especially involving genes that are readily druggable—and the technical success of exome-capture RNA-sequencing on highly degraded, FFPE biospecimens, we revisited the brain metastases cohort to define transcriptome-wide changes that occur in breast cancer cells following colonization. The goal of this study was to (1) further challenge the notion that brain metastases (BrM) are molecularly distinct from primary tumors with a more unbiased, comprehensive, transcriptome-wide methodology and (2) demonstrate preclinical evidence that targeting these acquired or enriched features may be viable therapeutic options for patients with BrM.

4.3 MATERIALS AND METHODS

4.3.1 Patient and tumor samples

Informed consent from all eligible patients was received and the study was approved by Institutional Review Boards from both participating institutions (University of Pittsburgh IRB# PRO15050502, Royal College of Surgeons IRB #13/09; ICORG 09/07). Eligible cases had patient-matched formalin-fixed paraffin-embedded (FFPE) tissue from primary and resected BrM (Table 3; Data Supplement 3: S1) processed for analysis. Tumor tissues were subjected to

neuropathological review to confirm histology and tumor cell content. Between four to ten (depending on tumor size) 10-micron FFPE sections immediately adjacent to the H&E-analyzed section underwent dual DNA/RNA extraction using Qiagen's AllPrep kit according to manufacturer's instructions.

4.3.2 Exome-capture RNA-sequencing

Library preparation was performed using 100 ng of RNA and Illumina's TruSeq RNA Access Library Preparation protocol. Indexed, pooled libraries were then sequenced on a High Output flow cell with an Illumina NextSeq 500 (paired-end reads, 2 X 75 bp). A target of 25-50 million reads per sample was used to plan indexing and sequencing runs.

4.3.3 RNA-sequencing expression quantification and normalization

FASTQ files were quantified using k-mer based lightweight-alignment (*Salmon* v0.7.2, quasi-mapping mode, 31-kmer index established from GRCh38 Ensembl v82 transcript annotations, seqBias and gcBias corrections)[241]. Read counts and percentage alignment were calculated (Data Supplement 3: S2). Transcript abundance estimates were collapsed to gene-level values using tximport[275]. To exclude non- or lowly expressed genes, only genes with a TPM value greater than 0.5 in at least 10% of samples were considered for clustering, gene set enrichment and clinically actionable kinase evaluation. Log2 transformed TMM-normalized CPM (log2normCPM) values were implemented for subsequent analyses[243,244].

4.3.4 RNA-seq quality assessment

Reads and mapping rates were obtained from *meta_info* files output by *Salmon*. More detailed RNA-seq metrics (Data Supplement 3: S3; Appendix A.3: Fig. 32) were calculated and plotted using QoRTs (v1.1.8) following two-pass read alignment with STAR (v2.4.2a) for the 21 patient-matched cases[245,246].

4.3.5 tumorMatch patient-matched sample identifier

To validate samples collected over many years and across institutions were patient-matched, RNA-seq variants were generated using *GATK's Best Practices for variant calling on RNA-seq*[247]. Output .vcf files were then provided to *tumorMatch*, a custom *R* script that analyzes pools of .vcf files and provides a proportion of shared variants (POSV) value for each sample pairing (Appendix A.3: Fig. 33) These proportion values were visualized using *corrplot* in *R*.

4.3.6 Differential gene expression

Salmon gene-level counts with effective gene lengths were provided to *DESeq2* to identify differentially expressed genes[249]. Given patient-matched samples, a multi-factor design was used (model = ~Patient + Tumor [i.e. primary vs. metastasis]). Genes were considered to be up or down-regulated if they exhibited a log2 fold change of greater than ± 1.5 and an adjusted p-value of <0.05 (Data Supplement 3: S4).

4.3.7 Correspondence Analysis

Correspondence analysis was carried out using the made4 package in R on the regularized log data from the 20,219 Ensembl genes of the 21 patient matched breast cancer primary and BrM samples[276]. The samples were then visualized in three dimensions with the first 3 components of the correspondence analysis, representing the clinical variables that describe the most variance, using the *rgl* package in R.

4.3.8 Unsupervised hierarchical clustering, intrinsic subtyping and HER2 signature

Hierarchical clustering was performed with the *heatmap.3* function (https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R) in R on log2normCPM values from the top 2000 most variable genes (defined by IQR) with 1 minus Pearson correlations as distance measurements and the "average" agglomeration method. PAM50 assignments were made using the *molecular.subtyping()* function in *genefu*[191]. To account for test-set bias, a random subset of primary tumor expression data was added to each query sample's PAM50 expression set to ensure an even distribution of ER-positive and ER-negative tumors[192]. This process was repeated 20 times. The discrete PAM50 assignment was finalized as the mode of the 20-fold PAM50 assignment tests while the probability score was an average of 20 probability scores (Data Supplement 3: S5). HER2 signature scores for each sample was calculated using the *genefu sig.score()* function along with the HER2 gene module[277] (Data Supplement 3: S6). Correlations between HER2 signature change (HER2 signature in met – HER2 signature in primary) and *ESR1* expression fold-changes were assessed using Spearman

r correlations. In correlation plot, but not Spearman r calculation, case 47_Pitt was excluded for better visualization given extreme *ESR1* log2FoldChange of approximately -6.

4.3.9 Merging of publicly available microarray breast cancer datasets.

The raw .CEL files from primary breast cancer microarray gene expression data from GSE12276 (Affymetrix HGU133plus2) and GSE2034 (Affymetrix HGU133a) were downloaded and subsequently merged using *inSilicoMerging* package[278] in R. There were 22,277 probes in common between the two datasets. The probe expression values were normalized using GCRMA and were then corrected for batch effects using the *ComBat* tool[279]. These merged primary breast cancer datasets had clinical information regarding time to distant metastasis and the site of metastasis: brain (n=23), lung (n=65), bone (n=171) and liver (n=66). Some of the primary samples metastasized to multiple sites. The merged dataset was used for site of metastasis survival analysis described below. In a separate analysis, public breast cancer metastasis microarray data (.CEL files) were downloaded and analyzed: GSE14017 (Affymetrix HGU133plus2) and GSE14018 (Affymetrix HGU133a). These were processed in the same manner as the public primary breast cancer datasets above (GSE12276 and GSE2034). This resulted in 22,277 probes in common between the two datasets that were normalized and batch corrected. The merged dataset contained breast cancer metastasis samples in the brain (n=19), lung (n=18), liver (n=5) and bone (n=16).

4.3.10 Contamination model construction

A pseudo brain transcript contamination model was created to estimate likely gene expression differences due strictly to an altered microenvironment. We first randomly drew RNA-sequencing data from 100 TCGA breast cancer primaries consisting of 60% ER+ / 40% ER- to match the proportion in the cohort, and randomly drew another 100 RNA-seq datasets generated from brain tissues (cerebellum and cortex) from GTEx. The two data sets were normalized by correcting for library size in *DESeq2*. Each brain-contaminated TCGA sample was then created by mixing the counts from one TCGA primary and one brain tissue. In this way, we artificially introduced brain transcript contamination into TCGA primaries, and a paired *DESeq2* analysis comparing contaminated TCGA and pure TCGA primaries was performed to determine genes likely to be differentially expressed solely due to brain contamination. To better select the mixing weight of brain tissues so that the brain contamination model was similar to our cohort (i.e. brain metastasis vs. primary), we selected 200 brain genes known to be highly expressed in brain but not in breast. We assume those brain genes should have similar fold changes in experiment and contamination model comparisons, if the brain contaminations are similar. We tried various mixing weights ranging from 0.01-0.20, and selected 0.07 as this carried the least sum of squared difference.

4.3.11 Microenvironmental gene deconvolution

The *(1,314)* genes that were found to be up-regulated in the BrM samples relative to the primary (using the paired DESeq2 analysis described above) were filtered to remove potential brain contaminating genes. We applied a stringent filter to define brain metastasis genes. This filter

used a log2 fold change cut off of greater than 1.5 in the experimental model comparison (brain metastasis vs. primary) and a log2 fold change of less than 1 in the contamination model comparison (TCGA brain contaminated vs. pure TCGA). This aimed to catch genes which are highly expressed in the brain metastasis but which are not typically highly expressed in the brain contaminated breast cancer model. Ensembl gene IDs were converted HGNC IDs using *biomaRt* (Data Supplement 3: S7)[280]. Deconvoluted genes were assessed using an independent dataset, GSE52604. Data from GSE52604 was downloaded and analyzed for differentially expressed genes between non-neoplastic brain (n=10) and breast cancer to brain metastases samples (n=35) using the *limma* package in R[281]. Differentially expressed genes were defined as having a log2 fold change > 1.5 and Benjamini-Hochberg adjusted p-value < 0.05.

4.3.12 Brain metastasis-free survival analysis

The *(249)* deconvoluted brain metastasis genes were assessed for higher expression in brain metastasis relative to other metastatic sites using the merged publicly available datasets described above (GSE14017 and GSE14018). Not all genes had probes present on the array. Probes that had a mean 1.5-fold change higher in the brain metastasis relative to their mean expression in the three other sites of metastasis (lung, liver and bone) were deemed to be brain metastasis specific and were subjected to site of metastasis survival analysis described below. The merged public primary breast cancer datasets described above (GSE12276 and GSE2034) were used for site of metastasis survival analysis. Ensembl IDs/HGNC IDs were converted to probe IDs. Not all genes were represented as probes on the array. The expression values for any matching probes were converted to a z-score (as calculated per probe). The mean of the z-score per sample was then used to create a signature for the selected probes. This was then used to

carry out Cox proportional hazards test to the different sites of metastasis. All analysis was carried out using the *survival* package in R. Kaplan-Meier plots were created using the *survplot* package in R.

4.3.13 Gene Set Variation Analysis (GSVA)

To determine oncogenic pathways enriched in breast cancer brain metastases, a *GSVA* analysis was performed with cancer pathway gene perturbation sets (n = 190) that harbor coordinately expressed gene members (MSigDB: C6 set)[250,252]. Input into the *GSVA* algorithm was log2normCPM values from genes within each gene set. A GSVA score representing potential pathway enrichments, using the maximum difference GSVA enrichment score, was assigned to each sample for each gene set. A paired Wilcoxon-signed rank test (metastasis vs. primary) was then implemented on GSVA scores to determine pathway enrichments in brain metastases. Significantly enriched gene set candidates were defined as those that carried a Benjamini and Hochberg adjusted p-value < 0.05 and a greater than 0 difference between mean metastasis GSVA scores and mean primary GSVA scores. GSVA scores for these values were then plotted using heatmap.3 and sorted by adjusted p-value.

4.3.14 DNA methylation profiling

DNA was bisulfite converted (BS) and the efficiency of BS conversion was assessed using 2 X 50 base pair (bp) sequencing of the libraries on the Illumina MiSeq platform. Once the libraries demonstrated >98% BS conversion efficiency, methyl-capture was carried out for each of the DNA libraries generated. The captured sample was assessed for overall quality using a

bioanalyzer, followed by sequencing in a 2 X 125 bp fashion on the HiSeq2000 v4.0 Illumina platform.

4.3.15 DNA methylation analysis

Paired end raw reads were analyzed with *fastqc* (http: //www.bioinformatics. babraham.ac.uk/projects/fastqc/) for initial quality control. *Trim Galore* (v.0.4.2) (http: //www.bioinformatics. babraham.ac.uk/ projects/trim_galore/) was used to remove sequencing adapters and remove base calls with a Phred quality score < 20. Trimmed reads were aligned to the human genome (hg19) reference using *bwa-meth* (v.0.10) with default parameters. *Picardtools* was used to mark and remove likely duplicates. *Bis-SNP BisulfiteGenotyper*[282] was used to identify single-nucleotide polymorphisms (SNPs) and insertion/deletion events (indels) and *BisulfiteIndelRealign* was used to realign reads. Analysis of processed BS-seq data was conducted in R using the *methylKit* package[283]. CpG level methylation calls with a minimum coverage of 10 reads for each were read into R and % methylation levels calculated by counting the ratio of C/(C+T) at each base. Differential methylation analysis was performed using methylKit which uses the Fisher's Exact Test and p-value adjustment using the SLIM method. A CpG was considered to be differentially methylated if methylation difference (Δ) = (μ normal – μtumor) was greater than 0.3 and a q-value < 0.01.

4.3.16 Gains and losses in clinically actionable kinases

Clinically actionable and kinase gene sets were obtained from the Drug Gene Interaction Database (DGIdB 2.0) and the overlap between the two sets were used to define clinically

actionable kinases (n = 105)[195]. Continuous expression fold-changes calculated from log2normCPM values (Data Supplement 3: S8) were transformed to discrete, stringent expression gains by defining an "expression gain" as a log2FoldChange greater than the 95th percentile (log2FoldChange = 1.198) of all gene and case fold-changes (Appendix A.3: Figure 34). After assigning discrete expression gains, data for recurrent gains (n > 1 pair) was visualized using the *OncoPrint* function in *ComplexHeatmap*[194]. *NTRK2* and *NTRK3* were excluded from the OncoPrint due to the kinases being highly expressed in normal brain, making it difficult to discern if unusually high expression gains (10 of 21 cases for *NTRK2,* 16 of 21 cases for *NTRK3*) were due to the altered microenvironment or tumor.

4.3.17 Immunohistochemistry (IHC) Staining

4-micron-thick paraffin sections were mounted on slides and stained for protein of choice using Dako EnVisionTM Kit, as described previously[284]. Briefly, heat antigen retrieval was carried out with 10 mM sodium citrate buffer (pH 6.0) for 20 min. Sections were treated with peroxidase block (Dako), and then incubated for 1 hour at 25˚C using recommended dilutions of the following antibodies: Ki67 (mouse monoclonal Dako clone MIB-1, M2740, Lot #A97064), ER (mouse monoclonal, novacastra leica, NCL-L-ER-6F11, Lot #6043537), PanCK (Mouse monoclonal, novacastra leica, NCL-L-AE1/AE3, Lot #6038590), HER2 (Mouse monoclonal, novacastra leica, NCL-L-CB11, Lot #6046036), RET (Rabbit polyclonal, Sigma Prestige Antibodies, Lot #A97064). All images were captured at 10x or 20x magnifications, and quantifications of Ki67 were performed following recommended guidelines[285]. ki67 staining was confirmed and analyzed on 3 non-consecutive slides at least 10 sections apart. Images were

scored as the percentage of Ki67 positive tumor cell nuclei per total tumor cell nuclei in each captured field.

4.3.18 *In vitro* studies

Estrogen receptor positive, endocrine therapy-resistant LY2 cells were a kind gift from R. Clarke, Georgetown University, Washington DC. LY2 cells were maintained in Phenol red-free Modified Eagle Medium (MEM) with 1x L-glutamine (L-Glut) and supplemented with 10% fetal bovine serum (FCS) (Invitrogen). They have demonstrated ability to readily metastasize to distant organs[286], including the brain (Appendix A.3: Figure 35). MDA-231-BrM2 cell line was obtained from the Massague Lab, MSKCC, New York. This metastatic variant of the MDA-231 has metastatic selectivity for the brain[139]. MDA-231-BrM2 cells were maintained in Dulbecco's MEM with 1x L-glutamine (L-Glut) and supplemented with 10% fetal bovine serum (FCS) (Invitrogen). All cells were maintained at 37˚C, 5% CO^2 in a humidified incubator. All cell lines were authenticated according to ATCC guidelines and mycoplasma tested (Mycoalert plus, Lonza) prior to undertaking functional studies. No cell lines used in this paper are listed in the database of commonly misidentified cell lines maintained by ICLAC. Cells were treated with Cabozantinb (10nM), Afatinib (25nM) or vehicle (%DMSO). Cellomics Cell Motility Kit (Thermo Scientific, K0800011) was used to assess individual cell movement after 24 hours as per manufacturer's instructions using cells seeded at $1x10^4$cells/mL. Mean track areas (minimum of 100 cell tracks per condition) were analyzed with Olympus cell imaging software. For growth assays, cells were treated at experimentally determined time-points and concentrations. MTS reagent (Sigma Aldrich) was added after 3 days and the resultant colorimetric outputs analyzed by measuring the absorbance at 490nm using a spectrophotometer (Perkin Elmer, USA).

4.3.19 Patient-derived brain metastases *ex vivo* culture

Written informed consent was obtained from patients, and fresh brain metastases were acquired from patients undergoing neurosurgery under the Institutional Review Board (IRB) approved protocol (Royal College of Surgeons IRB #13/09; ICORG 09/07). To establish patient-derived brain metastatic *ex vivo* models, fresh intact tumor tissue was collected, de-identified and placed in DMEM/F12 on ice immediately after surgical resection from the brain. Within 30-40 minutes, under sterile conditions, the brain metastatic tissue was cut and dissected into 2-4mm^3 fragments. These tumor fragments were placed on pre-soaked 1 cm^3 hemostatic gelatin dental sponges (Vetspon, Novartis) as described previously in brain/breast supporting media consisting of human mammary epithelial media (HMEC), supplemented with B27 (Life Technologies), N2 (Life Technologies), at 37 °C and 5% CO2[287]. The *ex vivo* brain tumors were cultured and treated with Cabozantinb (10nM), Afatinib (25nM) or vehicle (%DMSO) for 72hrs after which they were paraffin embedded and IHC stained. The viability of the tumors was evaluated by screening for necrosis of the tissue and using proliferation markers to confirm viable, proliferating cells. Schematic of the explant procedure utilized in this study is shown in Figure 16A.

4.3.20 *In vivo* experiments

All animal experimental procedures were conducted under IACUC approval. The following work was conducted in collaboration with Champions Oncology, using Champions Oncology breast cancer brain metastases patient-derived xenograft (PDX) model CTG-1520. Immunocompromised female nu/nu nude mice (Harlan Laboratories, USA) between 5-8 weeks of age were housed on irradiated, Alpha-twist-enriched 1/8" corncob bedding (Sheperd) in

individual HEPA ventilated cages (Innocage® IVC, Innovive USA) on a 12-hour light-dark cycle at 68-74°F (20-23°C) and 30-70% humidity. Animals were fed water (reverse osmosis, 2 ppm Cl2) and an irradiated test rodent diet (Teklad 2919; 19% protein, 9% fat, and 4% fiber) ad libitum. Mice were implanted subcutaneously into the left flank with the tumor fragments. Tumor growth was monitored twice a week using digital calipers and the tumor volume (TV) was calculated using the formula ($0.52 \times [\text{length} \times \text{width}^2]$). When the TV reached approximately 150-300 mm^3, animals were matched by tumor size and assigned into control or treatment groups (n = 4/group). Mice were treated for 4 cycles of once daily for 5 days followed by 2 days off (QDx5 on, 2 off) via oral gavage (PO) of vehicle, 20 mg/kg Afatinib or 30 mg/kg Cabozantinib. Researchers were not blinded to the treatment groups. Effects on tumor growth were evaluated by measuring percent tumor growth inhibition (TGI). Tumor size and body weight were measured twice weekly. The study was terminated when the mean tumor volume in the control group reached approximately 1500 mm^3. At study completion, tumors were collected from all animals in each group. The tumors were bisected: half was flash frozen and stored at -80 °C; the other half was processed for FFPE. Tumors that were < 250 mm^3 were processed as a single flash frozen sample and no FFPE material was available. Tolerability was assessed by body weight loss, lethality, and clinical signs of adverse treatment-related side effects of which there were none.

4.3.21 Test and Control Agents

Afatinib (#CT-BW2992) and Cabozantinib (CT-XL184) were obtained from Chemietek, USA. Agents were stored at -20 °C in the dark. The vehicle was 0.5 % methylcellulose, 0.4% Tween-80, 10% dimethyl sulfoxide (DMSO) in deionized water. Afatinib dosing solution (2 mg/mL)

was prepared by adding 0.72 mL of DMSO to 14.4 mg Afatinib in an amber dosing vial with a few glass beads, followed by vortex/sonication to yield a clear solution. A volume of 6.48 mL 0.5 % methylcellulose, 0.4% Tween-80 in deionized water was added and then followed by vortexing to form a solution. The dosing solution was stored at 2-4˚C for up to 7 days. Cabozantinib dosing solution (3 mg/mL) was prepared by adding 0.72 mL of DMSO to 21.6 mg Cabozantinib in an amber dosing vial with a few glass beads, followed by vortex/sonication to yield a clear solution. A volume of 6.48 mL 0.5 % methylcellulose, 0.4% Tween-80 in deionized water was added and then followed by vortexing to form a solution. The dosing solution was stored at 2-4˚C for up to 7 days.

4.3.22 Statistical Considerations

Statistical parameters including the exact value of n in terms of number of samples and models for each figure are reported in the Figures and the Figure Legends. All results are shown as mean +/- s.e.m., unless otherwise indicated. $P < 0.05$ was considered to indicate statistical significance throughout the study. Differentially expressed genes between patient-matched primary tumors and brain metastases were determined with *DESeq2*—a computational platform that utilizes a negative binomial distribution to assign differential expression p-values for gene-based count abundance estimates derived from RNA-seq. For survival analyses, logrank tests were used to illustrate statistically significant differences in event probabilities (i.e. death or time to metastasis)[253]. For single gene queries, paired Wilcoxon-signed ranked tests (primary vs. metastasis) on log2normCPM values were used. All cell-based *in vitro* experiments were independently repeated three times in triplicate. Two-way Student's t-test was used to compare two groups of independent samples. For *ex vivo* Ki67 analyses p values were obtained using one-

way analysis of variance (ANOVA), followed by Dunnett's test (GraphPad Prism). For *in vivo* study statistical comparisons of tumor volumes were conducted using one-way ANOVA followed by Newman-Keuls multiple comparison test (GraphPad Prism). $p < 0.05$ was considered to be statistically significant. The investigators were blinded to allocation for *ex vivo* and immunohistochemical analyses. With respect to randomization, for animal experiments, tumor-bearing mice of similar tumor burden were equally divided into the control and experimental groups for subsequent drug treatment which was not blinded. No statistical method was used to predetermine sample size. For animal experiments, efforts were made to attain the scientific aims of this study with the minimum number of animals taking into consideration PDX tumor growth deviations and results of previously tested agents. For *in vitro*, *ex vivo* and *in vivo* studies, all completed experiments are reported.

4.4 RESULTS

In order to identify recurrent alterations that can guide improvement in BrM treatment, we undertook an analysis of a clinical cohort of patient-matched primary breast and paired BrM (n=21) (Table 3; Data Supplement 3: S1). We performed genome-wide exome-capture RNA-seq. This method yields highly concordant expression values when compared to matched frozen samples (Chapter 3).

Table 3: Comprehensive clinical data of brain metastasis RNA-seq cohort

Abbreviations: Dx.Age, age at primary breast diagnosis; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; pos, positive; neg, negative; -,not determined; NA, not available; BrM, brain metastasis; DFS, disease free survival, time from primary diagnosis to first recurrence; BMFS, brain metastases free survival, time from primary diagnosis to death or last follow-up; SPBM, survival post brain metastasis, time from brain metastasis to death or last followup; OS, overall survival, time from primary diagnosis to death or last follow up.

	Primary Breast Tumor					Pre Brain Metastases Treatment					Brain Metastases			Post Brain Metastases Treatment					Progression (months)			
Case	Histology	Dx.Age	ER	PR	HER2	Endocrine	Radiotherapy	Chemotherapy	HER2	Recurrence.Prior.To.BrM	ER	PR	HER2	Endocrine	HER2	Radiotherapy	Chemotherapy	Status	DFS	BMFS	SPBM	OS
1_RCS	IDC	49	Neg	Neg	Pos	No	No	Yes	Yes	No	Neg	Neg	Pos	No	Yes	Yes	Yes	Dead	20	20	11	32
2_RCS	IDC	58	Neg	Neg	Pos	No	Yes	Yes	Yes	Yes	Neg	Neg	Pos	No	Yes	Yes	No	Dead	61	67	48	108
3_RCS	IDC	61	Pos	Neg	Pos	No	No	Yes	Yes	No	Pos	Neg	Pos	Yes	No	Yes	No	Alive	37	37	67	76
4_RCS	IDC	53	Pos	Neg	Neg	Yes	Yes	Yes	No	No	Pos	Neg	Pos	No	No	Yes	No	Dead	66	66	24	90
5_RCS	IDC	38	Neg	Neg	Neg	No	No	No	No	No	Neg	Neg	Neg	No	No	Yes	Yes	Dead	23	23	17	40
6_RCS	IDC	45	Pos	Neg	Neg	Yes	No	No	No	Yes	Pos	Neg	Pos	No	No	Yes	No	Dead	53	53	21	74
6_Pitt	IDC	66	Neg	Neg	Neg	No	Yes	Yes	No	Yes	Neg	Neg	Neg	No	No	Yes	Yes	Dead	23	25	12	37
7_Pitt	IDC	40	Pos	Neg	Pos	Yes	Yes	Yes	Yes	-	Pos	NA	Pos	NA	Yes	Yes	Yes	Dead	0	5	13	18
12_Pitt	IDC	38	Neg	Neg	Neg	No	Yes	Yes	No	-	Neg	Neg	Neg	No	No	Yes	Yes	Dead	0	31	14	46
17_Pitt	MDC	36	Pos	Neg	Pos	No	Yes	Yes	Yes	No	Pos	Pos	Pos	Yes	Yes	Yes	No	Dead	12	12	30	42
19-2_Pitt	IDC	57	Neg	Neg	Pos	No	Yes	Yes	No	No	Neg	NA	Pos	NA	NA	Yes	Yes	Dead	17	17	16	33
25_Pitt	IDC	66	Neg	Neg	Neg	No	Yes	Yes	No	Yes	Neg	NA	Neg	No	No	No	Yes	Dead	24	40	3	43
29_Pitt	IDC	52	Neg	Neg	Neg	No	Yes	Yes	No	No	Neg	NA	Neg	NA	NA	NA	NA	Dead	22	22	5	28
47_Pitt	IDC/ILC	53	Pos	Pos	Pos	Yes	Yes	Yes	No	Yes	Pos	Neg	Pos	Yes	Yes	Yes	Yes	Dead	83	151	74	225
51_Pitt	IDC	60	Pos	Neg	Neg	Yes	Yes	Yes	No	Yes	Pos	Pos	Neg	No	No	Yes	Yes	Dead	18	57	9	66
52_Pitt	IDC	62	Neg	Pos	Pos	No	Yes	Yes	Yes	Yes	Neg	Pos	Pos	No	Yes	Yes	Yes	Dead	36	55	7	63
62_Pitt	IDC	63	Pos	Pos	Neg	Yes	No	Yes	Yes	Yes	Pos	NA	Pos	No	Yes	Yes	Yes	Dead	39	53	6	60
64_Pitt	IDC	39	Neg	Neg	Neg	Yes	Yes	Yes	No	Yes	Neg	Neg	Neg	No	No	Yes	Yes	Dead	75	89	5	94
68_Pitt	IDC	51	Neg	Neg	Neg	Yes	Yes	Yes	No	No	Neg	NA	Neg	Yes	No	Yes	No	Dead	20	20	114	135
71_Pitt	IDC	26	Neg	Neg	Neg	No	Yes	Yes	No	No	Neg	Neg	Neg	No	No	Yes	Yes	Alive	25	25	147	173
72_Pitt	ILC	55	Pos	Pos	Neg	Yes	Yes	Yes	No	-	Pos	Pos	Neg	NA	No	Yes	Yes	Dead	0	31	5	37

4.4.1 Transcriptome evolution in breast cancer brain metastasis

Differential gene expression analyses revealed a catalogue of recurrently altered genes in BrM (Up in BrM n=1314; Down in BrM 1702; DESeq; fc>1.5, p-adj< 0.05) (Data Supplement 3: S4). Correspondence analysis showed that despite a marked gene expression divergence from primary tumors to BrM, transcriptome variability was largely due to intrinsic molecular subtypes (Figure 14A). Indeed, unsupervised hierarchical clustering revealed three major clusters—estrogen receptor (ER)-positive, HER2-positive, and ER-negative disease. 38.1% (8/21) of the patient-matched primary and metastatic tumor samples clustered as related pairs in the dendrogram (Figure 14B). Furthermore, PAM50 subtyping revealed 19/21 brain metastases retained the intrinsic subtype of the matched primary tumor (Figure 14C), consistent with our previous observations using targeted nanoString analysis (Chapter 2)[230]. Despite this broad conservation, 10/21 brain metastases showed deviations (>10%) of PAM50 subtype probabilities from their patient-matched primaries with the most common shifts being gains in Her2 and LumB profiles (Figure 14C; Data Supplement 3: S5), in line with our observations in Chapter 3 and recent PAM50 analyses in metastatic tumors[257].

To identify determinants of brain metastasis proficiency, we interrogated the overexpressed BrM genes in an expression dataset with multiple metastatic sites[132]. Of the 1314 Up in BrM genes, we focused on those expressed in BrM cohorts at a higher level (>1.5-fold) than in metastases from other sites. 7.9% of the genes satisfied this criteria (Figure 14D; Appendix A.3: Figure 36). Notably, in established cohorts of primary breast cancer tumors with extended follow-up expression of this BrM-related gene set significantly associated with brain (p=0.016) and lung relapse (p=3.2e-^{05}) but not relapse to either bone or liver (Figure 14E,

Appendix A.3: Figure 36)[139,288]. To further define brain tumor-associated genes, we developed a brain deconvolution approach to remove potentially contaminating non-neoplastic brain genes (Appendix A.3: Figure 37). Deconvoluted BrM gene set had a highly significant association with brain relapse (Figure 14F/G; Appendix A.3: Figure 38).

Beyond identifying alterations in genes important in the brain metastatic process including enrichment in genes implicated in vascular co-option (*L1CAM*)[143] and metastatic outgrowth (*SOX2*)[289], using gene set enrichment analysis[252]; we further delineated expression changes in BrM from matched primaries by identifying several oncogenic pathway gains in BrM[290]. These included several gene sets associated with cell cycle dysregulation (E2F3, RB), proto-oncogenes (KRAS, ALK) and kinase-driven pathways (SRC, mTOR, HER2) (Figure 14H).

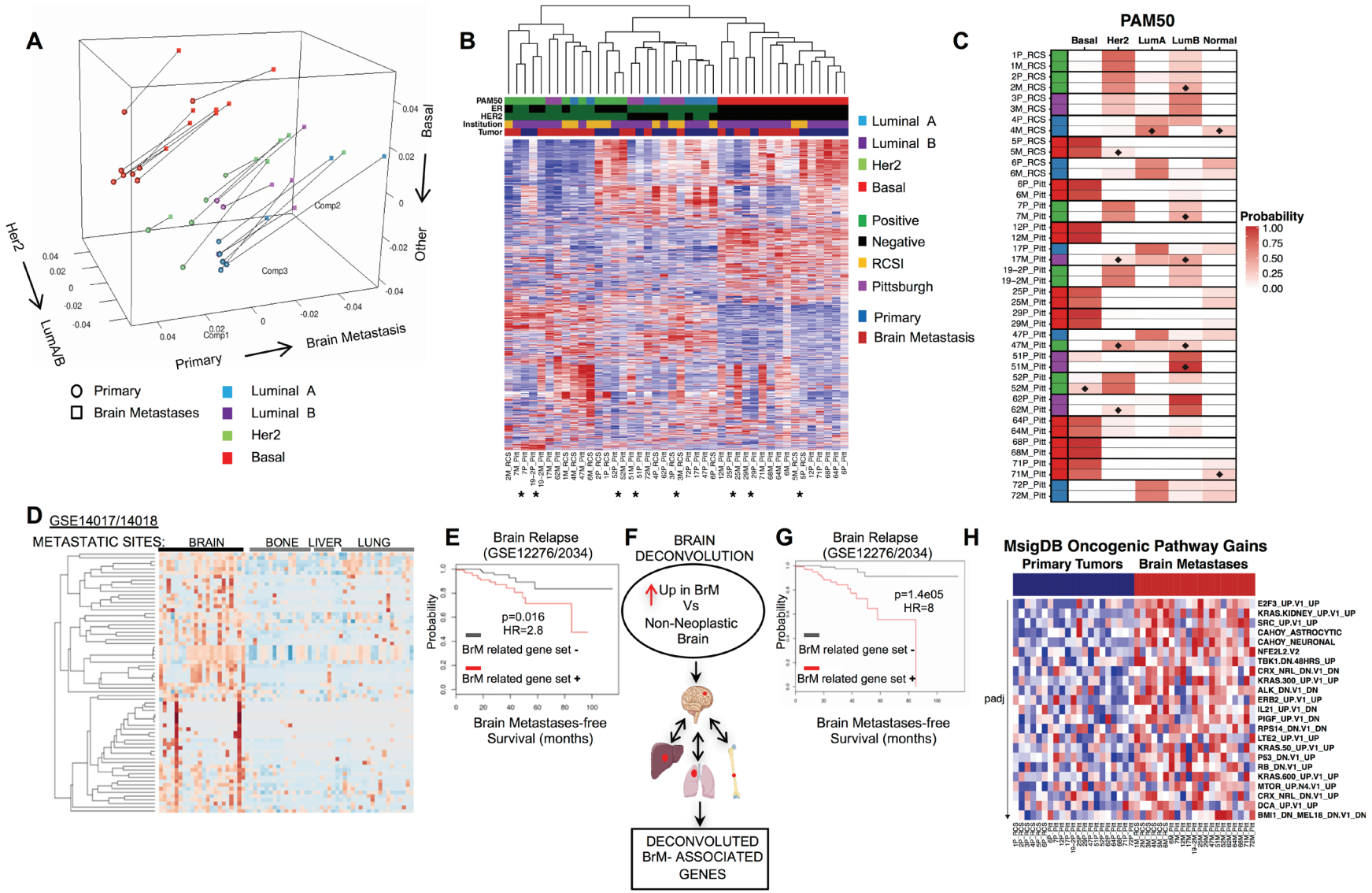

Figure 14: Transcriptome evolution in breast cancer brain metastasis.

(A) Correspondence analysis showing overall trends in paired samples using the gene expression of all protein coding genes. The matched primary (circles) and the metastasis samples (squares) are paired via a connecting line. The first component ("Comp1") represents the strongest trend and splits the samples from the primary to the metastasis, the other two components split the samples by intrinsic subtype. **(B)** Unsupervised hierarchical clustering heatmap. Patient-matched primary and metastatic tumor samples that clustered as related pairs in the dendrogram are indicated with an asterisk. **(C)** PAM50 intrinsic molecular subtype calls in patient-matched samples. Probability for each subtype is a mean of all 20-fold test probabilities; tile plot denotes this probability for each subtype. Diamonds indicate brain metastases with >10% probability gain in PAM50 subtypes. Legend denotes PAM50 subtype (blue=luminal A, purple=luminal B, green=Her2, red=basal), hormone status (green=positive, black=negative), tissue source (yellow = Royal College of Surgeons, Ireland, purple = University of Pittsburgh, US) and tumor site (blue=primary, red=metastasis). **(D)** Recurrent differentially up-regulated genes (n=1314) were screened in two merged public metastatic cohorts (GSE14017/18). Heatmap displays 62 genes whose expression was upregulated in brain metastases but not in metastases to lung, liver, or bone (BrM-related gene set). **(E)** Kaplan–Meier curves for brain metastasis-free survival of BrM-related gene set status in two cohorts (n=268) (GSE12276/2034). p value based on log rank test. **(F)** Schematic of the workflow for uncovering decontaminated brain metastases related genes. **(G)** Kaplan–Meier curves for brain metastasis-free survival on the basis of decontaminated BrM-related gene set (n=11) status in two cohorts (n=268) (GSE12276/2034). p value based on log rank test. **(H)** GSVA analysis utilizing MsigDB Oncogenic Pathway (MsigDB). Heatmap illustrates brain metastasis enriched pathways (FDR adjusted Wilcoxon signed-ranked P-value < 0.05) in brain metastases vs. primaries.

4.4.2 Recurrent expression gains of clinically actionable kinase pathways in breast cancer brain metastases

Using an established HER2 signature[277], we explored HER2 pathway activation in BrM given demonstrated HER2 expression increases in up to 35% of BrMs relative to matched primaries (Chapter 2) and a significant HER2 pathway enrichment from GSVA. Here, we show 15/21 pairs harbored elevated HER2 signature scores in the BrM relative to the matched primary (Figure 15A; Appendix A.3: Figure 39). Significant HER2 signature gains were not restricted to cases that switched from HER2-negative to HER2-positive in the BrM, implying that BrM outgrowth may be dependent on subclones with the highest levels of HER2 activation in the primary tumor (Appendix A.3: Figure 39). Indeed, tumors that switched from HER2-negative to HER2-positive in the brain metastasis had intermediate HER2 signature scores in the primary tumor (Figure 15B). Loss of *ESR1* gene expression, a known byproduct of hormone therapy resistance[46,284], correlated with increases in HER2 signature (Figure 15C, Appendix A.3: Figure 39). In the case 4_RCS, loss of *ESR1* was accompanied by enhanced *ESR1* hypermethylation acquired in BrM compared to the primary tumor (Figure 15C).

Given BrM-acquired gains in multiple kinase-driven signaling pathways, we examined clinically actionable kinases for recurrent expression gains (DGIdb 2.0). The most recurrent expression gains in BrM were *RET* and *ERBB2* (both gained in 38% of brain metastases) (Figure 15F). Alterations observed in *RET* mRNA in BrM were also confirmed at the protein level by IHC (Figure 15G, Appendix A.3: Figure 40).

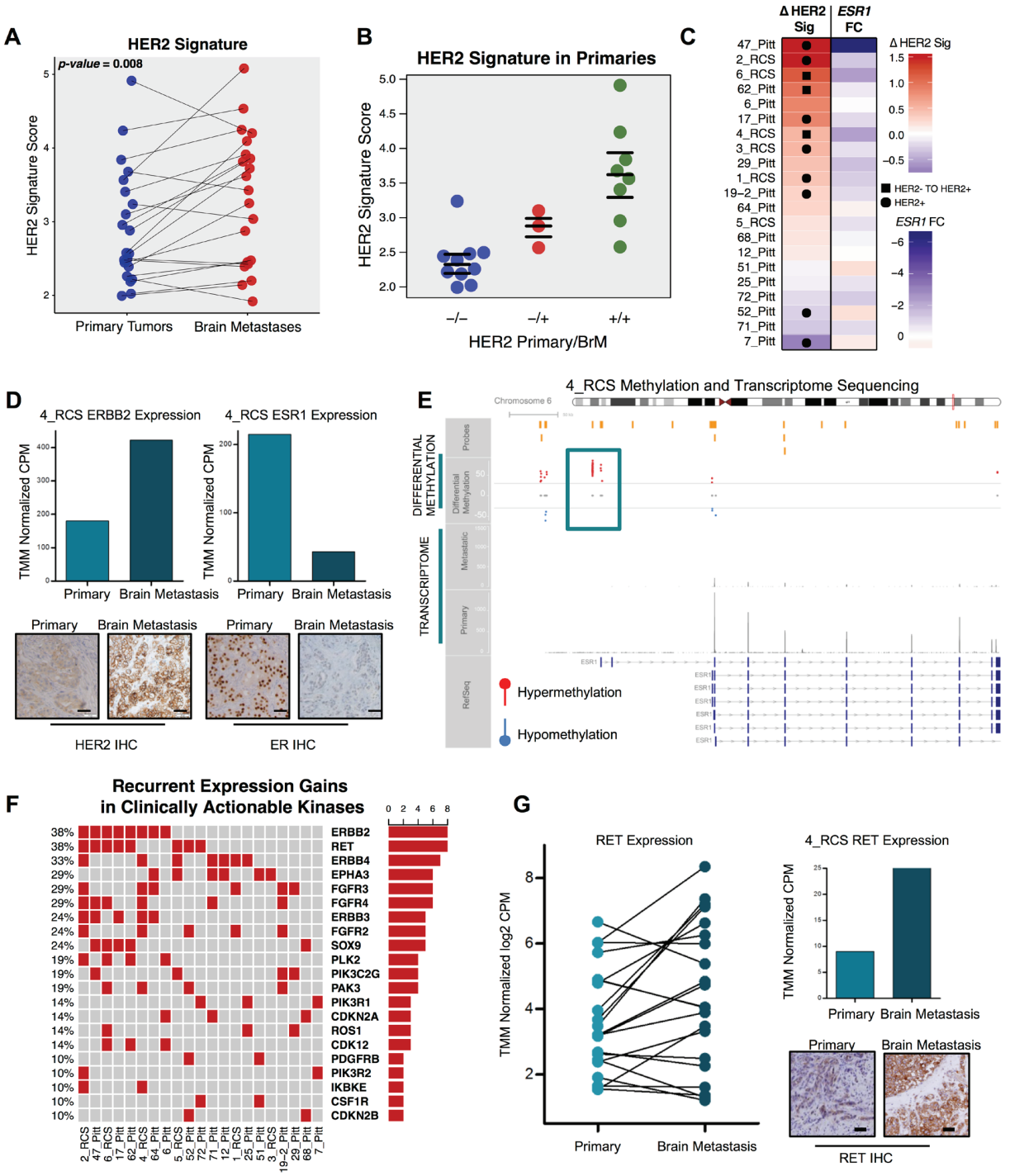

Figure 15: Recurrent expression gains of clinically actionable kinase pathways in breast cancer brain metastases.

(A) Paired ladder plot of established genes represented in the HER2 signature depicts the expression change in patient matched cases (p=0.008; Wilcoxon signed-rank test; primaries vs. brain metastases). Blue dots represent primary tumor signature scores and red dots represent metastatic tumor signature scores. (B) Scatter plot of HER2 signature score in primary tumors. Blue dots (-/-) represent patient-matched are HER2 negative in both the primary

and metastatic tumors, red dots (-/+) represent patient-matched cases that switched from HER2 negative to positive whereas green dots (+/+) represent HER2 positive tumors that have further activation in HER2 pathway. **(C)** Tile plot indicates gain of HER2 signature or loss of *ESR1* expression. Squares represent patient-matched cases that switched from HER2 negative to positive whereas circles represent HER2 positive tumors that have further activation in HER2 pathways. **(D)** Primary and metastatic log2normCPM values of *ESR1*/*ERBB2* from case 4_RCS, along with immunohistochemistry protein analysis. Images shown are 20x; scale bars correspond to 50μm. **(E)** *ESR1* gene differentially methylated regions (DMR) identified with methyl capture sequencing are illustrated and were identified by comparing 4_RCS case primary and brain metastasis. Plot shows regions of hypermethylation and hypomethylation found in ESR1 gene. **(F)** OncoPrint of clinically actionable kinases (DGIdb) with discrete expression gains in brain metastases. **(G)** Paired ladder plot of *RET* expression in patient-matched cases. Light green dots represent primary tumor expression values and dark green dots represent metastatic tumor expression values (log2norm CPM). Primary and metastatic IHC staining of RET from case 5_RCS; along with TMM normalized CPM counts. Images shown are 20x; scale bars correspond to 50μm.

4.4.3 Inhibition of RET and HER2 demonstrates significant anti-tumor activity in breast cancer brain metastases *ex vivo* and *in vivo*

Given RET and HER2 have FDA-approved agents and were the most recurrent expression gains, we next evaluated the effect of RET and HER2 inhibition in BrM models using RET inhibitor cabozantinib, and pan-EGFR-pathway inhibitor afatinib. A small molecule inhibitor of HER2 was selected over the large biologic trastuzumab due to the reported observations that the blood brain barrier may prevent its uptake to therapeutically efficacious levels[291]. *In vitro* we observed that inhibition with either cabozantinib or afatinib had a significant effect on the cellular viability and migratory capacity of TNBC[139] and ER positive brain-colonizing cell lines (Appendix A.3: Figure 35).

For preclinical assessment of the efficacy of cabozantinib and afatinib on BrM, we developed an *ex vivo* culture of BrM samples obtained from patients undergoing BrM resection. These models fully capture the histological, cellular and molecular components of the epithelial tumor interacting with the cells of glial origin, thus recapitulating components of the brain microenvironment (Figure 16A). *Ex vivo* Patient 1 (x-BrM T606) had endocrine-resistant disease, with loss of ER expression resulting in a triple negative brain metastatic tumor, whereas Patient 2 (x-BrM T347) gained *ERBB2* amplification. *Ex vivo* Patient 3 (x-BrM 681) was treatment naïve. The pathology of these metastatic tumors mirrored the key receptor subtype alterations observed in our sequencing study. We observed strong tumor specific RET expression in all *ex vivo* models used in this study (Figure 16B). HER2 was highly expressed in x-BrM T347 and T681, whereas x-BrM T606 had a weak expression, and was clinically graded as +1. Cabozantinib demonstrated substantial anti-tumor efficacy in x-BrM T606, T347 and T681

demonstrated by a significant decrease in proliferating cells (ki67+) compared to vehicle treated tumors (Figure 16B). Likewise, we see a significant tumor response to afatinib inhibition in x-BrM T606, T347 and T681 (Figure 16B). Of note, afatinib had an anti-proliferative effect independently of HER2 amplified status as evidenced in x-BrM T606, which may be due to activation of additional members of the EGFR pathway.

We next evaluated the effect of cabozantinib and afatinib in BrM patient-derived xenograft (PDX/ CTG-1520) derived from a triple negative tumor (Data Supplement 3: S1). The metastatic tumor expressed high levels of RET and was clinically HER2 negative (+1) (Figure 16C). PDX tumors were transplanted subcutaneously as grafts into immunocompromised mice and were allowed to grow to a volume of 150–300 mm^3. The tumor-bearing mice were then treated with cabozantinib (30 mg/kg), afatinib (20 mg/kg) or vehicle control via oral gavage for 20 days (5 days on/ 2 days off). At the conclusion of the study, both agents showed similar and significant anti-tumor activity compared to vehicle treatment in the BrM PDX model (cabozantinib 86%TGI; afatinib 91%TGI) (Figure 16C).

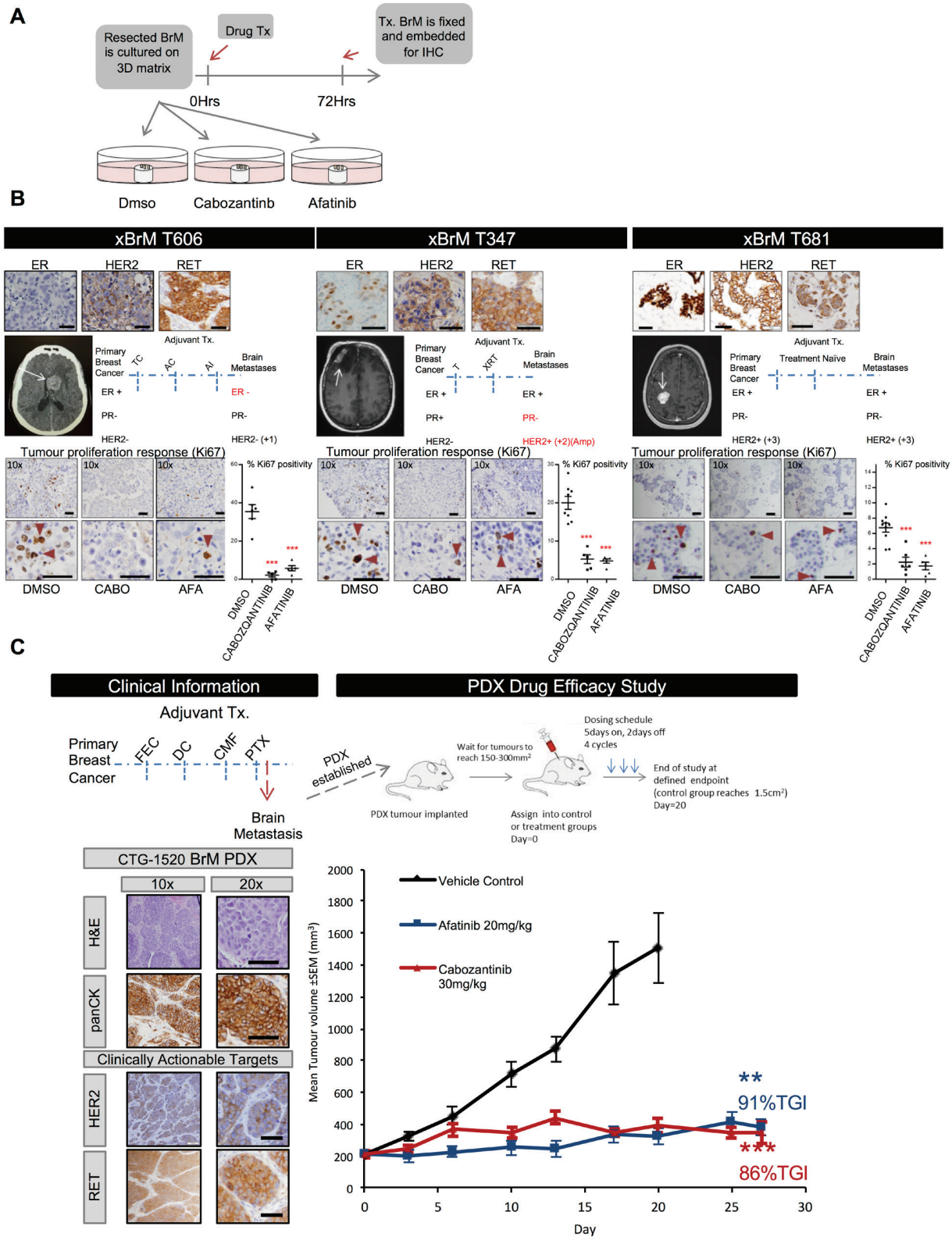

Figure 16: Inhibition of RET and HER2 demonstrates significant anti-tumor activity in breast cancer brain metastases *ex vivo* and *in vivo*.

(A) Schematic of the *ex vivo* experimental set up. **(B)** Brain metastatic tissue (x-BrMT606, T347 and T681) was treated with vehicle (0.1%DMSO), 10nM cabozantinib and 25nM afatinib and processed as described. IHC was carried out to profile ER, HER2 and RET of the *ex vivo* sample. MRI/CTI images of the brain metastases resected are shown. ER, PR and HER2 status in primary and brain metastases are indicated alongside adjuvant treatment received prior to resection. Representative images of IHC analyses of Ki67 tumors treated for 72hrs with indicated treatments (positive cells indicated with red triangles). All scale bars, 50 μm. Error bars represent mean ± s.e.m. (n = 5–10 images per group). ***P < 0.001, one-way analysis of variance (ANOVA), followed by Dunnett's test. **(C)** Schematic indicates clinical information pertaining to brain metastases (BrM) PDX CTG-1520 and the experimental design of the *in vivo* experiment. Treatment schedule was 4 cycles (QDx5 on/ 2 off) via oral gavage of vehicle (black line), 30mg/kg cabozantinib (red line) and 20mg/kg afatinib (blue line). Representative IHC images of H&E, pan cytokeratin, HER2 and RET are shown. Scale bars, 50 μm. Effects on tumor growth were evaluated with % tumor growth inhibition (%TGI). The tumor growth curve shows mean tumor volume +/- S.E.M. (n = 4 per treatment group). **P < 0.01, ***P < 0.001, one-way ANOVA test followed by Newman-Keuls multiple comparison test. Tx, treatment; xBrM, brain metastases explant; TC, taxol/carboplatin; AC, cyclophosphamide/doxorubicin; AI, aromatase inhibitor; T, taxol; XRT, radiotherapy; FEC, fluorouracil (5FU), epirubicin, cyclophosphamide; DC, docetaxel/carboplatin; CMF, cyclophosphamide, methotrexate, 5-fluorouracil; PTX, paclitaxel (Performed by Dr. Damir Vareslija).

4.5 DISCUSSION

Brain relapse can occur rapidly or in numerous cases many years after a primary diagnosis, a facet of BrM latency reflected in our clinical cohort. Genomically, analyses of BrM suggest that cancer cells continue to evolve upon colonization of the brain parenchyma, with mutations that are both common and distinct to originating tumors. The observations presented here expand upon these findings and establish recurrent transcriptional reprogramming events in breast cancer cells following brain colonization, shedding new light on the biology of BrM and potential therapeutic targets.

Our studies revealed a comprehensive list of genes enriched in BrM, including genes previously implicated in experimental models in the early events of vascular co-option[292], and those found to be essential for early survival and brain metastatic outgrowth[289]. Our work also points to many novel candidate BrM genes, whose exact role in BrM is open to further analyses but that appear specific to the brain parenchyma. This BrM-related gene set significantly associated with brain-relapse in primary tumors. Given the overlap with lung relapse and the limited available datasets, these observations are not interpreted as a gene signature capable of predicting brain relapse with high selectivity. More complete analyses can be undertaken as further relevant cohorts become available. Nevertheless, these collective shifts in gene expression signify a molecularly dynamic tumor adapting to its new microenvironment which have a large degree of metastatic selectivity and clinical relevance.

Metastatic colonization and BrM outgrowth depends on key adaptive pathway and alterations, and we demonstrate recurrent enrichments in druggable kinase-driven signaling. We

show conclusive activation of the HER2 pathway in BrM, especially important given the acquired HER2 mutational burden verified in BrM[179,230]. The observed and reinforced HER2 pathway gains have a number of immediate implications; (1) the observation of HER2 switching underlies the importance of dynamic tracking of tumor evolution; (2) intra-tumor heterogeneity of HER2 should be incorporated into routine breast pathology given data indicating the ability of subclones to evolve and adapt in BrM; and (3) HER2 inhibitors may be effective in patients with non-HER2 amplified (+2) metastatic BrM.

Notably, our transcriptional approach revealed no loss in *PTEN* expression, which has been proposed as a potential driver of PI3K/AKT activation in BrM[293,294]. This concordance in *PTEN* expression in patient-matched samples has previously been reported[295], and does not rule out its potential significance in BrM, particularly in *PTEN*-mutated BrM. Perhaps more importantly, *ESR1*, key clinically actionable gene, demonstrated consistent depletion in BrM compared to primary tumors. This loss of *ESR1* gene expression, a known feature of hormone therapy resistant disease, correlated with increases in HER2 signature. We further show ER loss in brain metastases can be epigenetically driven, suggesting further mechanistic studies into this process—especially as it relates to coincident HER2 activation—may be informative. Additionally, the exact point at which these *ESR1/ERBB2* alterations are acquired in the multi-step metastatic process is unclear and could be addressed through analyzing advanced but non-metastatic lesions or through profiling circulating tumor cells. Overall, these observations reinforce the dynamic regulatory interactions between ER and HER2 and expand its importance to the clinical setting of brain metastases[296].

Lastly, we define recurrent RET enrichments as a novel target for breast cancer BrM. Expression and activation of RET contributes to disease progression in multiple tumor types and

has been implicated in therapy resistance in breast cancer models[297–299]. We demonstrate significant anti-tumor efficacy of cabozantinib in *ex vivo* models of BrM that highly express RET. A recent clinical trial of cabozantinib monotherapy in heavily pretreated metastatic breast cancer demonstrated clinical activity including objective response and disease control[300]. However, the response to cabozantinib could perhaps be augmented through inhibiting other receptor tyrosine kinases (RTK) including MET and VEGFR[301]. In future studies, the impact of the tumor cell-brain parenchyma interaction should be assessed in the context of intracranial, orthotopic PDX models. Interestingly, we note that 24% of BrM cases demonstrated dual activation of both RET and HER2, and in those instances it is plausible to assume that combination therapy could amplify an anti-tumor response. As such, combining RET-specific inhibitors[302] with drugs targeting its downstream effectors (i.e. mTOR) may increase efficacy and improve overall benefit of blocking RET in BrM[293,297,303].

Collectively, though limited overall intrinsic clinical subtype switching is observed, our study demonstrates that BrMs undergo significant transcriptome shifts upon colonization. Enhanced cancer cell dependency on aberrant kinase pathways facilitates survival and outgrowth advantages—presenting therapeutic opportunities for BrM that are distinct from their matched primary tumors. These translational pre-clinical results deliver a compelling proof-of-principle for exploiting acquired vulnerabilities in advanced cancers.

5.0 RECURRENT TRANSCRIPTIONAL REMODELING EVENTS IN LONG-TERM ESTROGEN-DEPRIVED BREAST CANCER RECURRENCES

5.1 ABSTRACT

Resistance to endocrine therapies is a hallmark of advanced estrogen receptor (ER) positive breast cancers. In this study, we undertook a longitudinal analysis of 12 local recurrences that grew long-term (median time to recurrence 3.7 years) in an estrogen-deprived environment and compared them to features of their matched primary tumor using hybrid-capture DNA and RNA sequencing of approximately 1400 genes. Despite being up to 7 years removed from the primary lesion, the majority of recurrences harbored similar transcriptional and copy number profiles. Only two genes, *AKAP9* and *KMT2C,* were found to have enrichment in mutation allele frequencies in more than one local recurrence. Other enriched mutations, which were found only in a single case, included SNVs within transcriptional regulators such as *ARID1A, TP53, FOXO1*, *NCOA1* and *NCOR2*. One local recurrence showed enrichment of three distinct activating *PIK3CA* mutations, suggesting a strong, polyclonal selection in that particular tumor. In contrast to DNA-level changes, recurrent mRNA expression alterations were much more common. This included shared outlier gains in *TP63* [n = 4 [42%]), *NTRK2* [n = 5 [42%]), *NTRK3* (n = 4 [33%]), *PAX3* (n = 4 [33%]), *FGFR4* (n = 3 [25%]) and *TERT* (n = 3 [25%]). Recurrent losses involved *ESR1* (n = 5 [42%]), *RELN* (n = 5 [42%]), *SFRP4* (n = 4 [33%]) and

FOSB (n = 4 [33%]). Analysis of a subset of local recurrences that harbored major losses of *ESR1* mRNA expression (42% of recurrences) uncovered shared and distinct transcriptional remodeling events in these tumors—most notably gains in *PROM1* (CD133), a cancer stem cell marker usually expressed in basal cancers. Taken together, this study defines specific, targetable and recurrently acquired transcriptional remodeling events in long-term, hormone therapy treated disease and identifies a relatively common hormone-therapy resistant, *ESR1*-depleted breast cancer subtype that gains basal-like transcriptional traits.

Contributors to study: Nolan Priedigkeit[1,2,3], William Horne[4], Jay K. Kolls[4], Tian Du[2], Peter C. Lucas[2,5], Claire Atwooll[6], Jens-Uwe Blohmer[7], Carsten Denkert[7], Anna Machleidt[7], Barbara Ingold Heppner[7], Steffi Oesterreich[1,2,3], Adrian V. Lee[1,2,3,8]

[1]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, Pennsylvania

[2]Women's Cancer Research Center, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania

[3]Magee-Women's Research Institute, Magee-Women's Research Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

[4]Richard King Mellon Foundation Institute for Pediatric Research, Children's Hospital of Pittsburgh of University of Pittsburgh Medical Center (UPMC), Pittsburgh, Pennsylvania, USA.

[5]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

[6]Illumina, San Diego, CA

[7]Institute of Pathology and Department of Gynecology, Charité University Hospital, Berlin, Germany

[8]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA

5.2 INTRODUCTION

Hormone receptor positive breast cancer has served as a prototype for targeted therapy due to the well-established efficacy of estrogen-depleting small molecules in managing the disease. Largely because of these compounds, breast cancers are somewhat unique in that recurrences can occur years, sometimes decades following the primary diagnosis[111,306–308]. Unsurprisingly, outcomes of patients with locoregional relapse are generally worse than patients with an initial diagnosis, as 10-year median overall survival rates are between 40–70%[309,310]. Given the majority of patients in the past few decades receive long-term maintenance regimens of either a selective estrogen receptor modulator (SERM) or aromatase inhibitor (AI), recurrent breast cancers are occasionally classified as estrogen-independent given their ability to thrive in a continuously estrogen-deprived environment. Identifying the biological mediators that allow breast cancer cells to bypass their dependence on estrogen is a crucial step in understanding advanced breast cancer biology and defining novel therapeutic targets.

Defining these molecular processes in patient samples; however, has been challenging because of the logistics in obtaining well-characterized, longitudinally collected biospecimens. Nevertheless, shared features of more advanced breast cancers have emerged, such as relapsed tumors losing expression of ER and up to 20% of metastatic ER-positive breast cancers acquiring mutations in *ESR1* that confer ligand-independent signaling[51–53]. Other largely accepted mechanisms of estrogen-independence are bypass activations of mitogenic pathways such as MAPK and PI3K through initiating FGFR, EGFR and IGF signaling and exploitation of the Rb-CDK-E2F axis[46–50]. Less well validated mechanisms include *ESR1* fusions and amplifications, stem and mesenchymal cell state transformations and altered states of the microenvironment[55,56].

Recent studies analyzing multiple, longitudinally collected, pre and post-treatment samples have shown clonal evolution and selection in the context of targeted therapies[75,180,188,311–313]. Similar work analyzing hormone-receptor positive breast cancers have almost exclusively been restricted to short-term pre/post neoadjuvant therapy analyses[284,314–316]. The most comprehensive study of this type was a multi-platform effort that characterized the clonal architecture of tumors after four months of AI therapy[317]. Although drastic clonal remodeling was observed at the DNA-level, few recurrent resistance mechanisms were appreciated. The molecular changes that occur in long-term endocrine-deprived tumors, which represent the greatest burden of advanced breast cancer, are still completely unknown.

Thus, to better define both DNA and transcriptional changes that occur in long-term estrogen-independent tumors, we undertook a targeted analysis of 12 paired primary and local recurrences from patients with ER-positive breast cancers that were documented as being treated with some form of estrogen-depleting therapy. The median time to recurrence was 3.7 years, with the longest time to recurrence being over 7 years.

5.3 METHODS AND MATERIALS

5.3.1 Patient samples, tissue processing and nucleic acid extractions

Institutional Review Board approval from both participating institutions (University of Pittsburgh IRB# PRO15050502, The Charité IRB Office) was obtained prior to initiating the study. Inclusion criteria for this study were (1) patients harbored patient-matched formalin-fixed paraffin-embedded (FFPE) tissue from primary breast cancers and local recurrences (Table 4), (2) biospecimens contained macrodissectable regions with sufficient tumor cellularity and (3) disease was treated continuously with a form of estrogen-depleting therapy prior to the recurrence. Biospecimens were reviewed by a trained molecular pathologist to confirm pathology, quantify tumor cellularity and to highlight regions of relatively high tumor cellularity for macrodissection. If a slide region harbored sufficient, microscopically verifiable adjacent normal cells, this region was also dissected and processed for downstream analyses. Between four to ten (depending on tumor size) 10-micron FFPE sections immediately adjacent to the H&E-analyzed section were pooled and underwent dual DNA/RNA extraction using Qiagen's AllPrep kit. Nucleic acids were quantified fluorometrically with a Qubit 2.0 Fluorometer and quality assessed with an Agilent 4200 TapeStation Instrument prior to sequencing.

Table 4: Clinical features of local recurrence cohort
Abbreviations: Dx: Diagnosis; Hist: Histology; ER: estrogen receptor; PR: Progesterone receptor; HER2: human epidermal growth factor 2; Endo: endocrine; Tx: therapy; DFS: disease free survival; SPLR: survival post local recurrence; OS: overall survival; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; IMC: invasive mucinous carcinoma

Case	Age Dx	Hist	Stage	ER Prim	PR Prim	HER2 Prim	Endo Tx	HER2 Tx	Radio Tx	Chemo Tx	DFS	SPLR	Vital Status	OS
ERLR_01	36	IDC/ILC Mixed	I	Pos	Pos	Neg	Yes	No	Yes	Yes	86	132	Alive	218
ERLR_02	54	IDC	IIA	Pos	Neg	Pos	Yes	No	Yes	Yes	61	141	Alive	203
ERLR_03	74	IDC	I	Pos	Pos	NA	Yes	No	Yes	No	76	128	Dead	204
ERLR_05	54	IDC	IIA	Pos	Pos	Neg	Yes	No	Yes	Yes	69	85	Dead	155
ERLR_07	58	IDC	I	Pos	Pos	Pos	Yes	No	Yes	No	19	179	Alive	199
ERLR_08	52	IDC	IA	Pos	Pos	Pos	Yes	Yes	Yes	Yes	37	38	Alive	75
ERLR_09	51	IDC	IA	Pos	Pos	Neg	Yes	No	Yes	No	25	46	Alive	71
ERLR_12	47	IMC	IIA	Pos	Pos	Neg	Yes	No	No	No	26	34	Alive	61
ERLR_14	50	IDC	IA	Pos	Pos	Neg	Yes	No	NA	No	3	26	Alive	29
ERLR_15	65	IDC	IIIC	Pos	Pos	Neg	Yes	No	Yes	No	10	27	Alive	38
ERLR_19	49	IDC w/ lobular features	IIA	Pos	Pos	Neg	Yes	No	No	No	52	8	Alive	61
ERLR_20	42	IDC	IIIA	Pos	Pos	Pos	Yes	Yes	Yes	Yes	59	44	Dead	104

5.3.2 RNA and DNA-sequencing

RNA-seq library preparation was performed for all 12 cases using approximately 100 ng of RNA and Illumina's *TruSight RNA Pan-Cancer* (1385 targets) protocol. DNA-seq library preparation was performed for 10 (6 with associated normal tissue) cases using no less than 30 ng of DNA and Illumina's *TruSeq Exome* protocol with *TruSight RNA Pan-Cancer* probes for hybridization-based capture. Indexed, pooled libraries were then sequenced on Medium Output flow cells using an Illumina NextSeq 500 system (paired-end reads, 2 X 75 bp). A target of 5-10 million reads per sample was used to plan indexing and sequencing runs for RNA-sequencing and a target of 10-15 million reads was used for DNA-sequencing. RNA-sequencing FASTQ files

were quantified with k-mer based lightweight-alignment (*Salmon* v0.7.2, quasi-mapping mode, 31-kmer index using GRCh38 Ensembl v82 transcript annotations, seqBias and gcBias corrections)[241]. *tumorMatch* (Chapter 3, Chapter 4) was used to validate sequencing pairs were patient-matched.

5.3.3 RNA-sequencing quantification and DNA-sequencing alignment

RNA-seq read counts and mapping percentages were calculated (Data Supplement 4: S1) and transcript abundance estimates were collapsed to gene-level with tximport[275]. Log2 transformed TMM-normalized CPM (log2normCPM) values were implemented for subsequent analyses[243,244]. DNA-seq reads were aligned with *bwa –mem* (v.0.7.13) to an hg19 reference, sorted with *samtools* (v1.3), duplicates marked and removed with *picardtools* (v1.140) and local realignment performed with *GATK* (v3.4-46)[198–200]. Average coverage depth for the processed bam file was calculated using *GATK's DiagnoseTargets* and the Illumina *Pan-Can* bed file (Appendix A.4: Figure 41, Data Supplement 4: S2). Metrics for average coverage values across all target intervals were plotted with *ggplot2.*

5.3.4 DNA-seq recurrence enriched variant determination

To determine enriched variants in recurrences versus patient-matched primary tumors, *VarScan2* was implemented[318]. More specifically, primary and recurrent *samtools* mpileup files derived from processed bam files were input into *VarScan2* using *somatic* mode, with somatic p-values representing the significance of a particular variant being acquired or enriched in the recurrence [SS = 1 or SS = 2]. Tumor purity estimates, as assessed by a molecular pathologist, were

included in *VarScan2* to correct contaminating normal cell influence on allele frequencies. The minimum coverage for a variant to be considered was 40X, with a minimum allele frequency (AF) of 0.05 in either the primary or recurrence and a minimum of 5 reads supporting the variant. Germline variants were determined for cases containing a matched normal (ERLR_01, ERLR_02, ERLR_07, ERLR_08, ERLR_12 and ERLR_15) using *VarScan2*'s *germline* mode with the same parameters. VCF output files were then imported into R using the *VariantAnnotation* package[319]. If an adjacent normal sample was available for the case, all germline variants (AF > 0.30) were excluded from subsequent analyses. Additionally, to limit technical artifacts especially considering specimens were formalin-fixed paraffin embedded[320], a "blacklist" of variants was created including those called in at least 3 of the normal samples. Germline and blacklist-removed variants were then annotated with *Annovar*[321]. Lastly, to call recurrence-enriched, potentially pathogenic variants the following inclusion criteria were enacted: (1) VarScan2 somatic p-value < 0.05, (2) > 2-fold gain in allele frequency in the recurrence versus the primary, (3) minimum AF of 0.10 in the recurrence, (4) non-silent and (5) an ExAC AF < 0.01 considering some samples were without a paired normal (Data Supplement 4: S3)[322]. These non-silent, enriched, potentially pathogenic variants were then plotted using the *OncoPrint* function in *ComplexHeatmaps*[194]. A pearson R correlation was calculated between the frequency of enriched variants and disease-free-survival. *PIK3CA* mutations were visualized with *IGV* (2.3.60)[202] and variant allele frequencies were derived from *VarScan2*.

5.3.5 Copy number alterations

To estimate copy number ratios, *CNVkit* was implemented on processed bam files using default settings and the *-drop-low-coverage* option[201]. A pool of bam files from adjacent normal tissue,

sequenced in the same manner, was used as a reference. Probe and segment level copy number estimates were finalized with *CNVkit*'s *call* function, which utilizes circular binary segmentation[323]. To adjust for tumor purity and normal contamination, the *–m clonal* option was used with tumor purities from pathologic evaluations. Copy number ratios were then plotted with the *heatmap* function and copy number values were assessed and plotted with *ggplot2*. Gene-level copy number estimates represent the mean copy number call across all probe targets. *CNVkit* copy number ratios showed a near normal distribution and *ERBB2* copy number values demonstrated a strong correlation (pearson R = 0.924, p-value < 0.001) with expression (Appendix A.4: Figure 42).

5.3.6 Differential gene expression, clustering and outlier gains and losses

Hierarchical clustering was performed using the heatmap.3 function (https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R) in R on log2normCPM values of the top 10% most variable genes (defined by IQR) with 1 minus Pearson correlations as distance measurements and the "average" agglomeration method. Differential expression between primary and recurrent tumors was analyzed with *limma*. Raw counts were input into the *voom* function and quantile normalized prior to fitting the linear model and performing the empirical Bayes method for differential expression[281,324]. The linear model was fitted with a design that accounts for the paired nature of the cohort (model = ~Patient+Tissue [primary or recurrence]). Outlier expression gains and losses were determined for each patient by discretely categorizing genes into one of 5 categories. If log2FC values (i.e. recurrence log2normCPM – primary log2normCPM) for a given gene were less than Q1 – (1.5 X IQR) or Q1 – (3 X IQR), using case-specific log2FC values for all genes as the distribution, that

gene was deemed an “Outlier Loss” or “Extreme Loss” respectively. If log2FC values calculated were greater than Q3 + (1.5 X IQR) or Q3 + (3 X IQR), it was deemed an “Outlier Gain” or “Extreme Gain” respectively. All other genes with intermediate fold changes were classified as “Stable.” To determine subtype expression of *KLK7, PROM1* and *NDRG1,* normalized microarray expression data along with PAM50 calls was obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) through Synapse (https://www.synapse.org/, Synapse ID: syn1688369), following IRB approval for data access from the University of Pittsburgh[14]. Overlap with genes in long-term estrogen deprived, ER-positive breast cancer lines (HCC1428, MCF7, T47D, ZR75.1) was performed by running a separate differential expression analysis (LTED vs. parental lines) on microarray data with *limma*[281,325]. Dysregulated gene overlap was designated if the nominal p-value and FDR-adjusted p-value were both < 0.05 in the local recurrence and LTED differential expression analysis, respectively. Binary dichotomization of METABRIC samples using *NDRG1* expression (>50th percentile, <50th percentile) and log-rank testing were used to assess significant differences in disease-specific survival (DSS) and then Kaplan-Meier curves were plotted with *survminer*[253,254].

5.4 RESULTS

5.4.1 Expression and copy number alteration profile conservation

Unsupervised hierarchical clustering showed most patient matched pairs cluster with their matched primary—regardless of the length of disease free survival (Figure 17A). Unlike a previous transcriptome-wide analysis of primary breast cancers and matched bone metastases (Chapter 3), there was no significant correlation in pair transcriptional similarity and time to recurrence—although a negative correlation was observed (pearson R = -0.37, p-value = 0.236). Only a single recurrence showed marked transcriptional deviation from its matched primary (ERLR_03_R1); whereby it lost ER-positivity and gained HER2-positivity clinically (marked with a Δ). Copy number alterations (CNAs) between primary and recurrences were then analyzed in the targeted capture regions for 10 cases. Similar to expression, copy number alterations were largely consistent among the recurrences when compared to their matched primary (Figure 17B). Two exceptions were recurrences from cases ERLR_01 and ERLR_03, which showed distinct copy number profiles from the matched primary tumors. Notably, unlike case ERLR_03, ERLR_01 retained a similar expression profile. An analysis of shared variants validated both DNA and RNA extracts originated from the same patient (Appendix A.4: Figure 43), excluding the possibility of sample mixup.

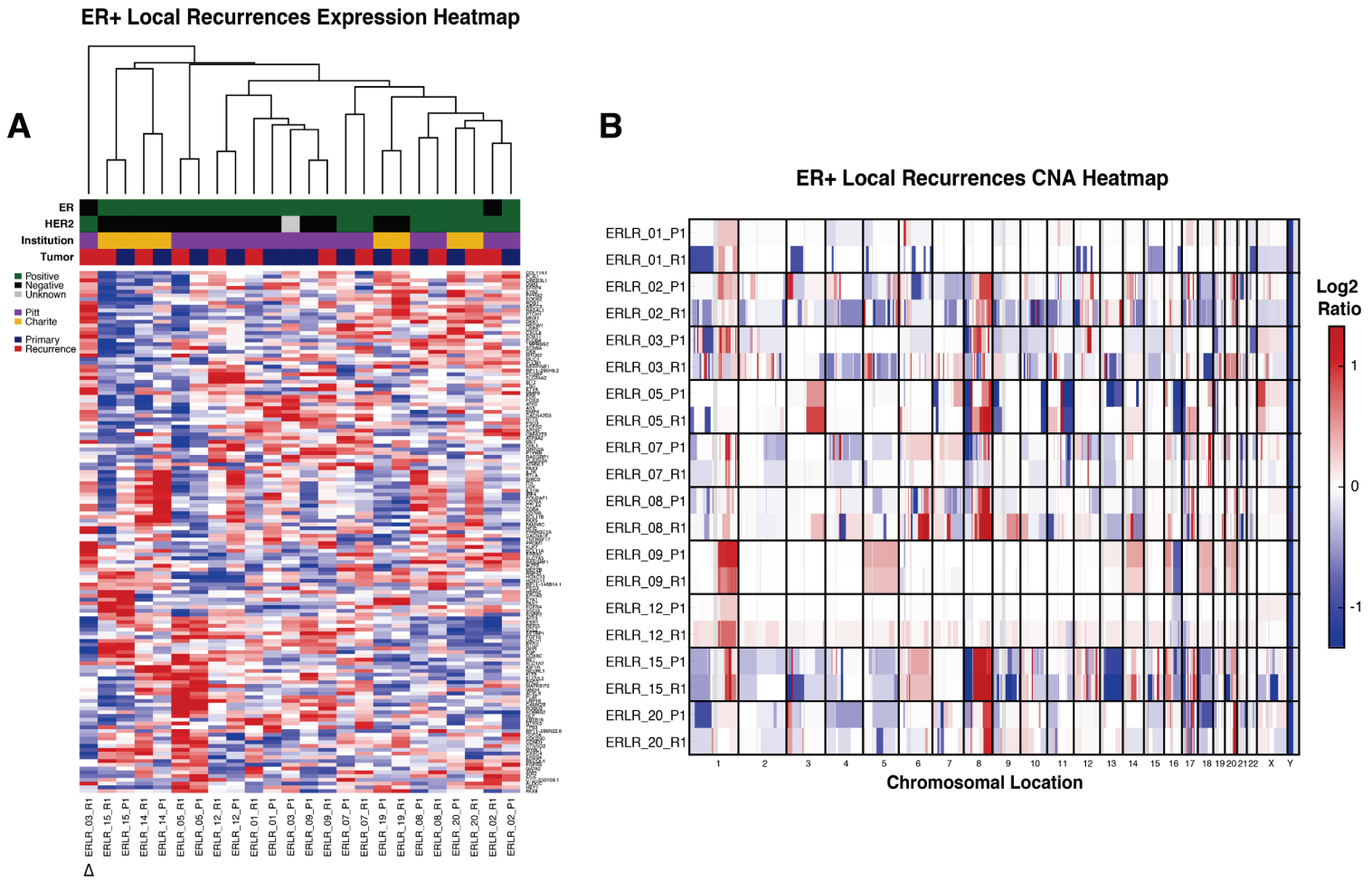

Figure 17: Transcriptional and CNA conservation in ER-positive local recurrences
(A) Unsupervised hierarchical clustering and heatmap on normalized gene expression values from patient-matched pairs (P1 = Primary, R1 = Recurrence). Clinical ER and HER2 status (black = negative, green = positive, grey = unknown), tissue source site (purple = Pitt, yellow = Charite), and tumor type (blue = primary, red = recurrence) are indicated. Delta symbol shows distinct clustering of ERLR_03_R1 away from its matched primary, ERLR_03_P1. **(B)** Heatmap of copy number ratios from patient-matched pairs. Redder regions indicate regions of copy number gain and bluer regions indicate regions of loss.

5.4.2 SNV enrichments and differentially expressed genes

To assess if there are shared features acquired in recurrences, an analysis of enriched single nucleotide variants (SNVs) was performed for the 10 cases that were DNA-sequenced. Two genes were found to be enriched in more than one case (n = 2 [20%]) in local recurrences versus primary tumors, *AKAP9* and *KMT2C* (Figure 18A, Data Supplement 4: S3). The recurrent mutations did not show any features suggesting functional selection, such as being within a conserved functional domain or within a COSMIC[326] hotspot region, making it difficult to assess if these are pathogenic. Other case-specific, n-of-one enriched mutations included a nonsense mutation in *ARID1A* (Case ERLR_20, Primary AF 0.5%, Recurrent AF 16.5%), an acquired *TP53* mutations (Case ERLR_03, Primary AF 0.0%, Recurrence AF 53.4%) and an enriched *NCOR2* mutation (Case ERLR_08, Primary AF 4.4%, Recurrence AF 19.4%). In case ERLR_01, an enrichment of a suite of three somatic mutations in *PIK3CA* was observed (E542K, Q546K, E726K) in the recurrence (Figure 18B). Notably, the number of enriched, non-silent SNVs ranged from 0 to 13 and was positively correlated with clinical time to recurrence (Figure 18C). No acquired *ESR1* mutations were observed, and this was orthogonally confirmed by droplet digital PCR (data not shown, performed by Zheqi Li). A differential expression analysis, comparing all primary tumors versus all local recurrences, yielded no genes passing an FDR corrected p-value of less than 0.05—which is perhaps expected given the heterogeneity of clinical specimens (Data Supplement 4: S4). Nonetheless, 71 genes with an average, *voom* normalized expression value of 2 or greater, a nominal p-value of less than 0.05 and a log2 fold-change greater than +/- 0.5 were identified (Table 5). Some of these genes, including the upregulation of *EPOR, NDRG1, IDH2, CEBPA* and *PTPA* and downregulation of *ESR1, IGF1R,*

NFKB1 and *RUNX2*, are also differentially expressed in long-term estrogen deprived ER-positive cell lines (Appendix A.4: Figure 45).

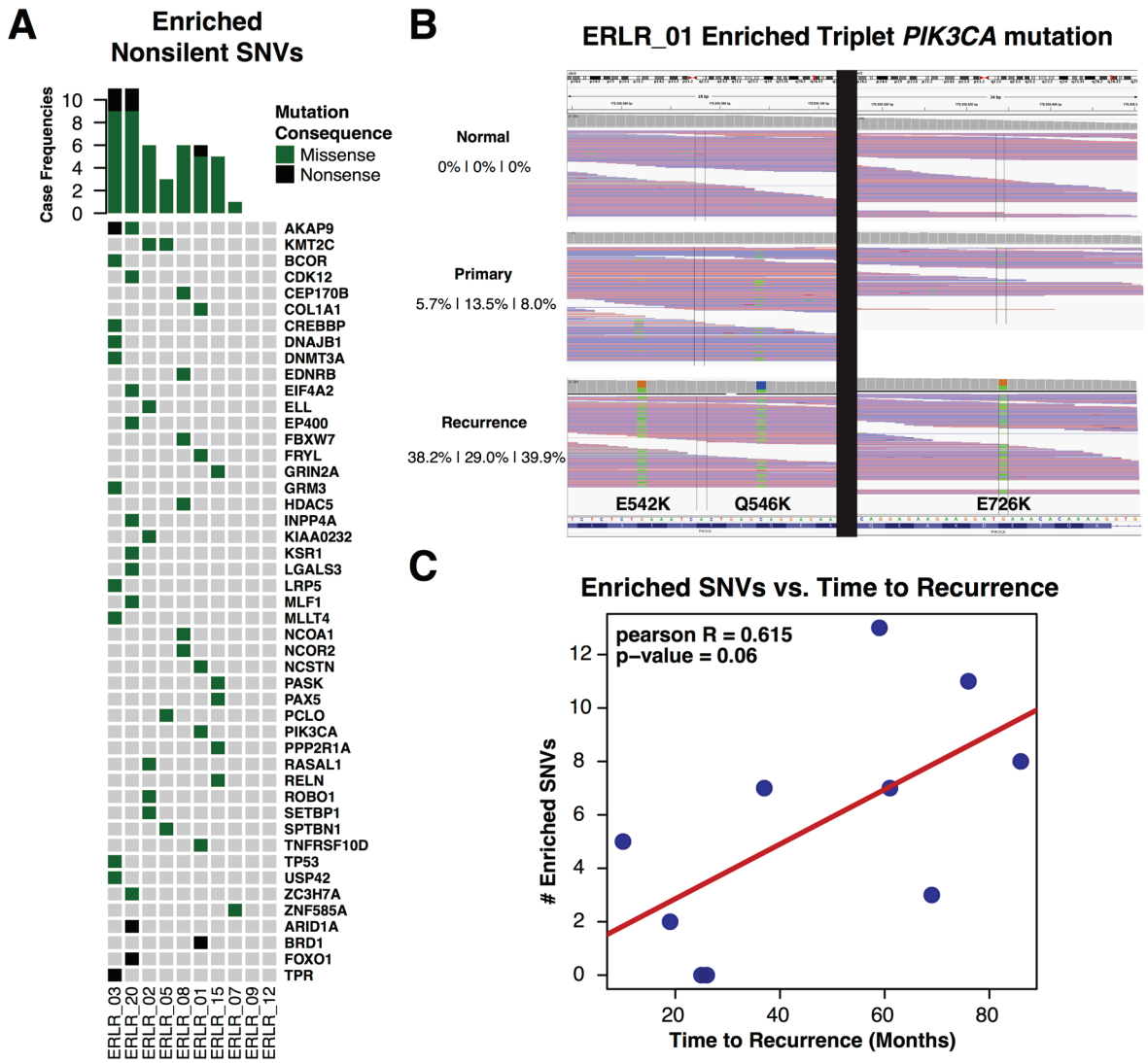

Figure 18: SNV enrichments in ER-positive local recurrences.
(A) OncoPrint of non-silent, enriched single nucleotide variants in patient-matched cases. Missense variants are indicated with a green box and nonsense variants with black. **(B)** Polyclonal, triplet mutation enrichment of *PIK3CA* mutations in case ERLR_01. Collapsed *IGV* alignments are shown, along with allele frequencies, for the normal, primary and recurrence. **(C)** Frequency of enriched, non-silent single nucleotide variants versus time to recurrence along with pearson R and calculated p-value.

Table 5: Differentially expressed genes in long-term estrogen-deprived local recurrences

Gene	Log2FC	*voom* Average Expression	Nominal P-value	FDR Adjusted P-value
DDIT3	0.570	6.241	0.004	0.322
PC	0.536	5.847	0.004	0.322
CEBPA	0.750	4.889	0.005	0.322
E2F1	0.788	5.223	0.011	0.422
FANCA	0.550	7.102	0.012	0.441
RET	0.714	7.015	0.015	0.441
HIST1H3B	0.803	5.165	0.015	0.441
RASGRF1	**1.139**	4.066	0.015	0.441
EPHB6	**1.014**	3.418	0.016	0.441
POLD1	0.505	6.353	0.017	0.441
RECQL4	0.770	6.370	0.020	0.441
SLC7A5	0.579	6.292	0.021	0.441
CEBPB	0.627	7.198	0.023	0.441
CENPU	0.745	7.597	0.023	0.441
ALDOC	0.796	5.265	0.026	0.441
LAMA5	0.542	10.076	0.027	0.441
RPA3	0.626	5.020	0.027	0.441
NDRG1	0.747	9.013	0.029	0.441
PPP1R13L	0.503	5.638	0.029	0.441
NTRK3	**1.334**	3.993	0.029	0.441
EPHA2	0.642	4.947	0.030	0.441
SLC34A2	**2.031**	2.218	0.031	0.441
AURKA	0.582	5.966	0.032	0.443
H2AFX	0.593	4.357	0.037	0.459
EPOR	0.701	3.987	0.039	0.461
VEGFA	0.692	9.212	0.041	0.471
ASPH	0.565	11.252	0.046	0.490
CIT	0.578	7.406	0.048	0.492
SFRP2	**-1.329**	8.934	0.000	0.162
ETV1	**-1.102**	5.522	0.000	0.162
CYP1B1	-0.765	6.342	0.001	0.162
AK5	**-1.399**	4.639	0.001	0.162
AKT3	-0.718	7.441	0.001	0.279
LEF1	-0.520	7.241	0.002	0.289
PDGFRA	-0.893	7.161	0.002	0.313
RUNX2	-0.815	6.515	0.002	0.322
CDC14A	-0.611	6.782	0.004	0.322
IGF1R	-0.807	9.598	0.004	0.322
LHFP	-0.625	6.217	0.005	0.322
HTRA1	-0.816	9.288	0.005	0.322
POSTN	-0.905	12.477	0.005	0.322
ZNF521	-0.747	7.905	0.006	0.322
SFRP4	**-1.306**	7.855	0.006	0.343
ADD3	-0.825	8.199	0.007	0.355
CDH11	-0.618	9.666	0.007	0.355
ARHGAP20	-0.834	4.204	0.009	0.386
DCN	-0.831	11.074	0.009	0.386
ZFPM2	-0.733	6.488	0.009	0.386
GRIN2A	-0.970	2.159	0.010	0.414
RELN	**-1.171**	4.220	0.011	0.422
GRID1	-0.607	3.894	0.014	0.441
EGR2	-0.914	6.350	0.015	0.441
EGR1	-1.000	9.137	0.016	0.441
PQLC3	-0.525	6.472	0.016	0.441
HAS2	-0.747	5.090	0.018	0.441
ESR1	**-1.668**	9.480	0.019	0.441
ATP8A2	-1.279	3.966	0.021	0.441
PRRX1	-0.641	7.975	0.021	0.441
STAT4	-0.886	4.178	0.022	0.441
PRDM16	-0.917	2.311	0.023	0.441
LAMA1	-0.549	5.186	0.027	0.441
IL7R	**-1.025**	5.094	0.028	0.441
COL6A3	-0.591	13.383	0.030	0.441
ALDH1A1	-0.636	6.524	0.030	0.441
RASGRF2	-0.623	6.833	0.030	0.441
NAV3	-0.535	6.228	0.036	0.459
GAS7	-0.561	7.495	0.037	0.459
COL1A2	-0.500	14.574	0.037	0.459
DGKI	-0.722	5.383	0.039	0.461
IL1R1	-0.560	8.198	0.044	0.487

5.4.3 Outlier expression gains and losses

To further explore major expression changes that may be driving recurrence but not shared among all recurrences, an outlier expression analysis was performed using gene-level fold-change values of each patient-matched case. Unlike non-silent SNVs, recurrent transcriptional gains and losses were common (Figure 19A). These included gains and losses in shared pathway members, notably *NTRKs* and *SFRPs* respectively, targetable upregulation of growth factor pathway mediators such as *FGFR4* and *EGF* and outlier gains in the CDK regulator *CCNE1*. 3 of 12 cases also shared outlier expression gains in *TERT*, with case ERLR_14 harboring a particularly extreme enrichment from near undetectable levels in the primary tumor (Figure 19B). Case ERLR_03's recurrence, which was most dissimilar to its patient-matched pair transcriptionally, showed extreme loss and gain of *ESR1* and *ERBB2* respectively. CNA analysis confirmed recurrence-specific *ERBB2* amplification and is consistent with previous studies of endocrine therapy-treated breast cancers selecting for HER2-signaling in more advanced tumors (Chapters 2, 3). The most recurrent outlier loss involved *ESR1*.

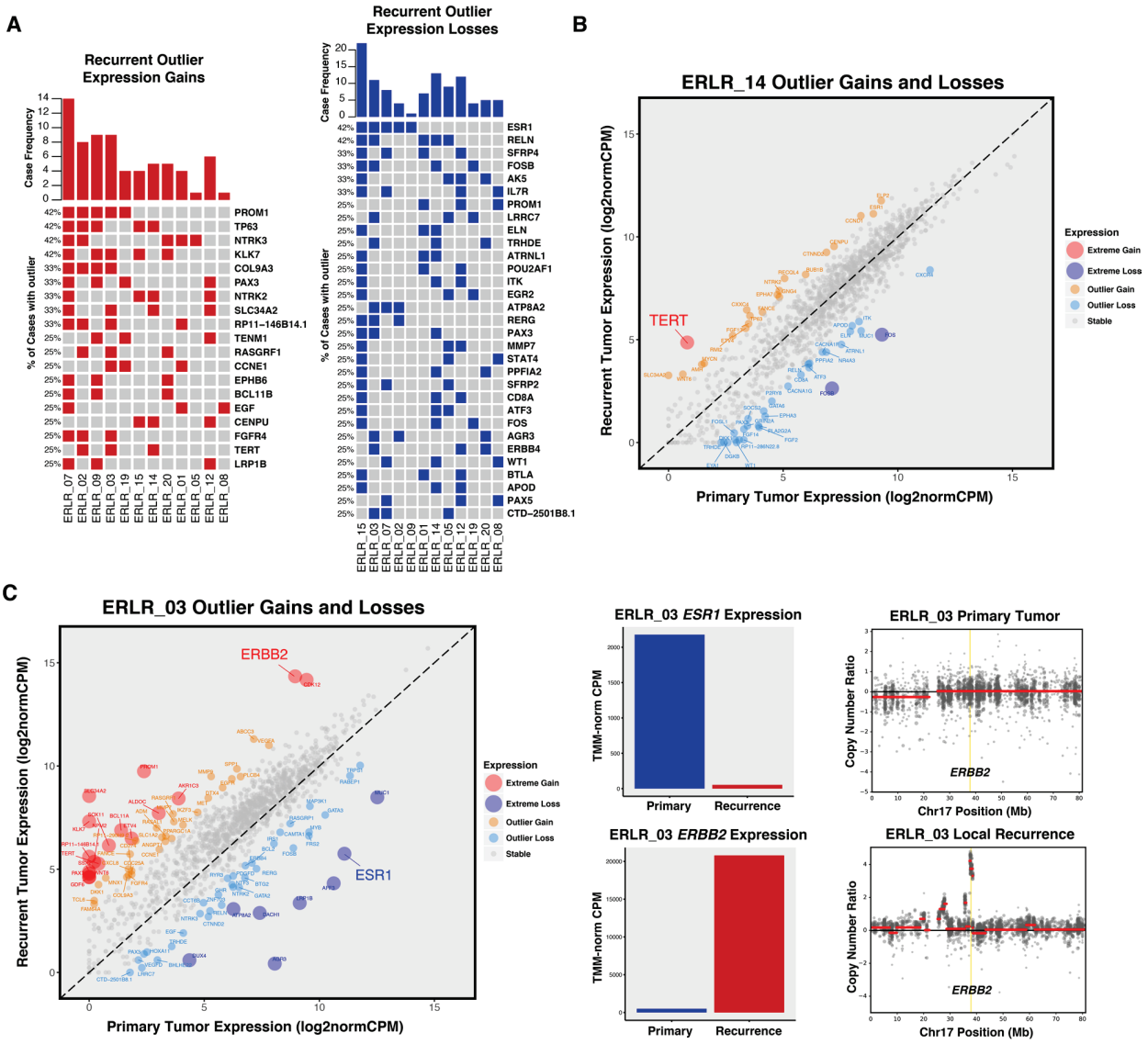

Figure 19: Outlier expression gains and losses in ER-positive local recurrences
(A) OncoPrint of outlier expression gains (red) and outlier expression losses (blue) in ER-positive local recurrences. Genes are sorted by frequency of outlier changes across pairs. **(B)** Extreme expression gain of *TERT* in case ERLR_14; 2 other cases showed similar *TERT* enrichments in recurrent tumors. **(C)** Extreme expression gain and loss of *ERBB2* and *ESR1* respectively. TMM-normalized CPM values of primary (blue) and recurrent (red) tumor. *ERBB2* expression gain is driven by recurrence-specific DNA-level amplification of *ERBB2* locus.

5.4.4 *ESR1* depleted recurrences

Five cases showed outlier expression losses of *ESR1* (Figure 20A). Despite estrogen receptor being the driver of ER-positive breast cancer and a major regulator of transcription; counterintuitively, 4 of 5 of the recurrences which lost *ESR1* expression generally retained the expression profile of their patient-matched primary (Figure 17A). Importantly, many of these cases also harbored very similar CNA profiles (Figure 17B), implying the recurrences were derived from a continuous cancer linage as opposed to being completely distinct breast cancers. Thus, to explore the transcriptional consequences of acquired *ESR1* loss in ER-positive disease and identify potential bypass mechanisms driving *ESR1* independence, a differential expression analysis was performed on the subset of pairs with outlier *ESR1* expression losses. This analysis revealed several recurrently dysregulated genes in *ESR1* depleted recurrences (Figure 20B, Data Supplement 4: S6). Two standout genes, *KLK7* and *PROM1,* showed the highest degree of fold change with a log2 fold-change increase of 5.4 and 3.9 respectively—with some tumors exhibiting changes from near undetectable levels to high expression (Figure 20C). These two genes are more commonly expressed in basal cancers, with *PROM1* being a cancer stem cell marker and luminal lineage factor (Appendix A.4: Figure 44)[327]. Other genes with significant log2 fold-changes > 1 included drug targets such as *FGFR4, KIT, IGF1R* and *BCL-2* (Table 6). *NDRG1,* a particularly compelling candidate since it also showed upregulation in LTED breast cancer models, was further interrogated using METABRIC data. Like *PROM1* and *KLK7*, *NDRG1* is most highly expressed in basal breast cancers; yet, when expressed in ER-positive primary tumors, *NDGR1* confers significantly worse disease-specific survival outcomes (Appendix A.4: Figure 46).

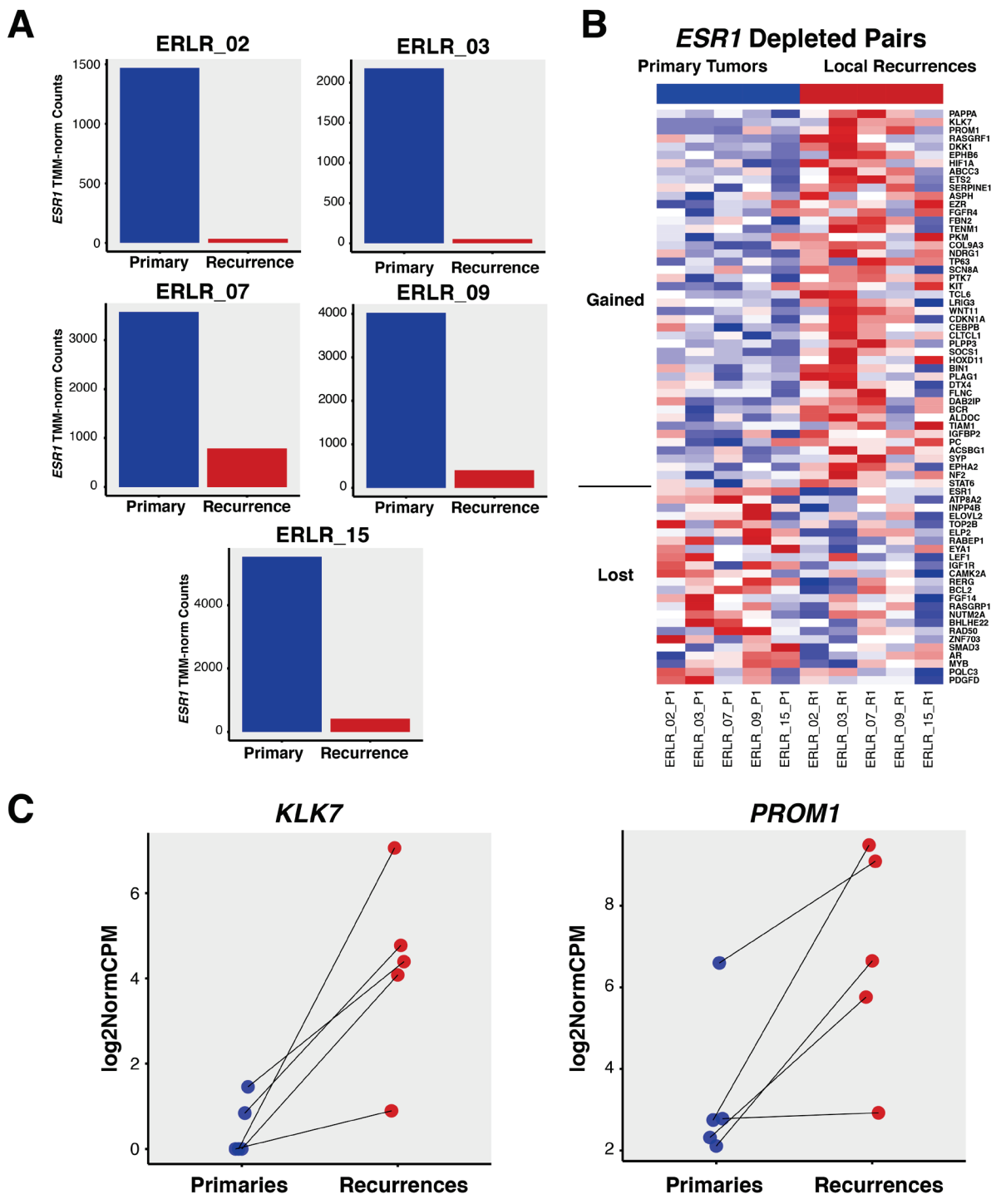

Figure 20: *ESR1* depleted recurrences
(A) TMM-normalized expression of patient-matched local recurrences; primary tumor expression in blue, recurrent tumor expression in red. **(B)** Heatmap of differentially expressed genes (nominal p-value < 0.05) in *ESR1* depleted recurrences versus matched primary tumors. Genes are sorted by p-value and segregated by log2 fold-change values; log2 fold-change > 0 on top, log2 fold-change < 0 on bottom. **(C)** Ladder plots showing log2normCPM expression values for both *KLK7* and *PROM1*, two of the most significantly upregulated genes in local recurrences with the largest average log2 fold-changes.

Table 6: Differentially expressed genes in *ESR1* depleted recurrences

Gene	Log2FC	*voom* Average Expression	Nominal P-value	FDR Adjusted P-value
PAPPA	1.395	6.416	0.001	0.293
KLK7	5.422	1.158	0.001	0.293
PROM1	3.931	5.005	0.002	0.588
RASGRF1	2.307	4.106	0.002	0.588
DKK1	2.732	0.473	0.004	0.614
EPHB6	1.641	3.819	0.005	0.614
ABCC3	1.637	8.010	0.006	0.614
FGFR4	1.515	5.267	0.010	0.695
FBN2	1.010	5.326	0.010	0.695
TENM1	1.326	4.709	0.012	0.705
COL9A3	2.034	2.249	0.014	0.705
NDRG1	1.218	8.945	0.014	0.705
TP63	2.135	4.441	0.018	0.768
SCN8A	1.290	5.881	0.019	0.768
KIT	1.289	6.020	0.020	0.768
TCL6	2.228	-0.254	0.022	0.790
WNT11	1.585	1.256	0.024	0.823
SOCS1	1.534	0.387	0.033	0.911
HOXD11	2.755	-1.369	0.034	0.911
PLAG1	1.275	4.576	0.036	0.911
DTX4	1.185	5.711	0.036	0.911
FLNC	1.588	6.787	0.037	0.911
ALDOC	1.494	5.224	0.039	0.911
ACSBG1	1.843	0.601	0.042	0.915
SYP	1.348	0.862	0.045	0.915
ESR1	-3.952	9.492	0.000	0.146
ATP8A2	-2.599	4.510	0.003	0.588
ELOVL2	-2.090	2.413	0.006	0.614
RABEP1	-1.009	10.352	0.012	0.705
EYA1	-1.494	2.203	0.013	0.705
IGF1R	-1.149	9.083	0.016	0.747
CAMK2A	-1.391	2.742	0.016	0.747
RERG	-1.413	6.562	0.018	0.768
BCL2	-1.055	6.619	0.020	0.768
FGF14	-1.393	2.430	0.023	0.790
RASGRP1	-1.044	6.799	0.027	0.857
BHLHE22	-1.822	0.822	0.035	0.911
ZNF703	-1.811	4.865	0.038	0.911
MYB	-1.179	8.857	0.045	0.915

5.5 DISCUSSION

In this study, a targeted RNA/DNA analysis of approximately 1,400 cancer genes in ER-positive primary breast cancers and matched long-term, endocrine therapy treated local recurrences was performed. We found a profound conservation of transcriptional and copy number profiles—suggesting that even after 7 years of dormancy and the onslaught of therapies, recurrent breast cancers retain their intrinsic molecular features. An analysis of recurrence-enriched single nucleotide variants revealed limited recurrent mutation events, including no acquired *ESR1* mutations, yet notable "n-of-one" mutation evolution was observed—such as case ERLR_01 which showed three distinct, recurrence-enriched *PIK3CA* mutations. The most striking changes in long-term estrogen-deprived tumors; however, were highly recurrent (up to 42%), outlier expression changes. An analysis of tumors with the most recurrent outlier loss, *ESR1,* revealed concurrent upregulation of genes typically expressed in basal breast cancers, such as *PROM1, KLK7* and *NDGR1*.

Nearly all recurrences are more similar transcriptionally to their matched primaries than to other, long-term estrogen deprived tumors—reinforcing the notion that advanced cancers generally retain their core transcriptional programming, even after nearly a decade of dormancy[26-29]. This transcriptional conservation appears to be even more pronounced than metastatic lesions (Chapter 2,3,4)— perhaps due to an unaltered microenvironment and the greater, multistep selective pressures required for cells to seed a foreign organ[328]. Furthermore, amplifications and deletions of recurrences are markedly similar to primaries, supporting recent evidence from breast cancer single-cell sequencing that structural variation is likely an early

event and many CNAs, even in metachronous therapy-resistant tumors, may be shared by the majority of subclones[329]. An important exception to this conservation was ERLR_03_R1, a recurrence with a completely unique transcriptional and copy number profile than its matched primary. Evidence has emerged of so-called 'collision tumors', whereby two synchronous, distinct cancers can merge anatomically and only under the selective pressures of therapy or through deep sequencing , their individuality can be unmasked[317,330]. Indeed, this "recurrence" switched to ER-negative/HER2-positive from ER-positive/HER2-negative clinically, and thus could represent a completely different cancer than the primary tumor. Countering this notion; however, were shared variants with similar allele frequencies between the primary and recurrent tumors (data not shown)—although it is difficult to make this assessment conclusively given a matched normal from this patient was not available.

Limited shared, non-silent SNVs were discovered in these specimens, with *AKAP9* (R3320W, S319*) and *KMT2C* (T1969I, Y366N, R894Q) being the only two genes that harbored recurrence-enriched mutations in greater than one case. These mutations are not in a conserved functional domain nor in a hotspot location, making it difficult to assess their pathogenic roles. *AKAP9* and *KMT2C* also encode relatively large gene products (3911 and 4911 amino acids, respectively) which may increase the likelihood of obtaining a so called passenger mutation by chance. Nevertheless, *KMT2C* and other lysine methyltransferases have been implicated in breast cancer pathology, argued as potential drivers in large-scale sequencing studies of primary tumors and *KMT2C* mutations specifically may confer hormone therapy resistance in breast cancer models[15,331,332]. Case ERLR_20 harbored an enriched nonsense mutation in *ARID1A* (Q1424*, primary AF 0.5%, recurrence AF 16.54%). Notably, somatic mutations in this chromatin remodeling gene are frequent in gynecologic cancers with compelling data supporting *ARID1A*

as a tumor suppressor. *ARID1A* is also associated with more unfavorable tumor features in breast cancer and is enriched in metastatic breast cancers versus primary tumors (Figure 4), further suggesting a role in disease progression[333–335]. A single recurrent cancer (ERLR_01_R1) showed enrichment of three somatic hotspot *PIK3CA* mutations (E542K, Q546K, E726K), suggesting an extreme, polyclonal selection within that particular tumor. Given the likely dependency this tumor carries on *PI3K* signaling and recent early phase trials for *PIK3CA* mutant cancers[94,95], enriched mutations found early in local recurrences may represent a particularly compelling method of rational drug selection or planned trial enrollment if this patient were to progress to a more advanced disease. SNVs within genes that act as corepressors and coactivators, some with direct influences on estrogen receptor mediated transcription, were found to be enriched in recurrences—such as *NCOA1*, *NCOR2, FRYL* and *CREBBP*—along with transcription factors including *PAX5*, *FOXO1* and *TP53*. Finally, we observed a positive correlation between the frequency of acquired, non-silent SNVs and disease-free survival—validating the concept that surviving cancer cells after initial therapy acquire potentially pathogenic mutations as they lay dormant and undetectable over time. As more long-term estrogen deprived breast cancers are characterized, the selection of advanced disease driver mutations—which may be distinct from primary disease such as *ESR1* mutations—will become clearer.

Given the heterogeneity of clinical specimens makes it difficult to rely on typically used differential expression analyses—since resistant mechanisms of individual tumors may be distinct—we undertook an analysis of outlier expression gains and losses to identify more extreme transcriptional reprogramming events within individual cases that may be driving estrogen independence. Surprisingly, unlike SNVs, recurrent outlier transcriptional gains and losses were quite common. Particularly compelling outlier events included recurrent gains within

shared pathway members, such as near mutually exclusive upregulations of *NTRK2* [n = 5 [42%]) and *NTRK3* (n = 4 [33%]). Tropomyosin-related kinases have been historically associated with psychiatric disorders and neural development; however, their role in cancer has been increasingly appreciated given their involvement in recurrent, oncogenic fusions[336–338]. Notably, activation of *NTRK's* mediates downstream signaling pathways typically associated with breast carcinomas, including PI3K and MAPK, and small molecule inhibitors of this family are currently being tested in solid tumor trials (NCT02568267)[339]. Other notable pathway member changes included loss of Wnt antagonists *SFRP2* (n = 3 [25%]) and *SFRP4* (n = 4 [33%]). *SFRP2* is hypermethylated and silenced in a subset of breast cancers[340,341] and experiments in model systems have shown cross-talk between ER and Wnt signaling that may mediate endocrine therapy resistance[259,342]. Other recurrent gains included *FGFR4* (n = 4 [33%]), *TERT* (n = 3 [25%]) and *CCNE1* (n = 3 [25%])—particularly relevant given the recent success of CDK inhibitors in hormone-positive disease and the burgeoning use of *FGFR* inhibitors against solid malignancies[343].

The most recurrent outlier expression loss was *ESR1,* which was diminished in 42% of long-term estrogen-deprived local recurrences. Interestingly, the loss of *ESR1* for the majority of cases was not associated with a dramatic change in the tumors' transcriptional profile. To further explore this counterintuitive result, given *ESR1* is a master regulator of transcription and a driver of luminal breast cancers, we identified genes that were consistently altered in *ESR1* depleted recurrences. The most substantial gains in *ESR1* depleted tumors are genes generally expressed in basal breast cancers—such as *NDRG1, DKK1, KIT, KLK7, PROM1* and *COL9A3*—and genes significantly lost in the *ESR1* depleted subset are generally downregulated in basal cancers—*EVLOVL2, BCL2, IGF1R, MYB, ESR1, RABEP* and *ATP8A2* (MsigDB:

SMID_BREAST_CANCER_BASAL_DN/UP gene lists)[250]. These results reveal a distinct *ESR1*-depleted subtype of advanced breast cancers that obtain more basal-like characteristics.

The greatest fold-change difference in *ESR1* depleted recurrences was the upregulation of *PROM1*. *PROM1* is a marker for tumor-initiating cancer stem cells and plays a key role in determining ER-positive luminal cell fate during differentiation from multipotent stem cells[327], suggesting long-term endocrine deprived breast cancer cells may enrich themselves with stem-like progenitors to achieve estrogen-independence. Indeed, *PROM1* has been shown to mediate endocrine therapy resistance in breast cancer models through IL6/Notch3 signaling [344,345]. Here, we show that a large portion of long-term endocrine resistant breast cancers may be exploiting this transcriptional reprogramming. Importantly, this gene has been shown to be immunogenic in melanoma and glioma, suggesting it may be a prime target for immunotherapy—benefits of which has not yet been realized for breast cancer[346,347]. Finally, *NDRG1*, also significantly upregulated in *ESR1* depleted recurrences and generally expressed in basal cancers, showed differential expression in three distinct, ER-positive LTED cell lines. *NDRG1* is a suspected metastasis suppressor gene. Counterintuitively, we see upregulation of this gene in resistant disease and show increased expression confers worse survival outcomes in ER-positive primary tumors[348]. Further functional studies assessing the mechanistic and biological consequences of these transcriptional reprogramming events will be essential.

A pertinent point these results raise is the potential benefit of integrating longitudinal, targeted RNA-sequencing to inform resistance mechanisms and therapeutic targets in breast cancers. In this study, we find limited DNA-level enrichments yet highly recurrent, acquired transcriptional remodeling events from primary to advanced cancers, including a few of which that are immediately targetable such as *NTRKs, FGFR4* and *CCNE1.* Overall, this work may

challenge our lack of emphasis on RNA-level changes, particularly those that can be elucidated from longitudinal biopsies, in clinical profiling of tumors—especially in breast cancers considering they are driven by transcriptional regulators rather than recurrent DNA-level changes.

Collectively, these results begin to unravel the complex adaptations that breast cancer populations undergo when under the selection of estrogen depleting therapies long-term. We identify acquired DNA-level mechanisms of resistance, such as mutations in *ARID1A* and polyclonal selection of *PIK3CA* mutations—but more importantly, uncover the most recurrent genomic adaptations taking place appear to be at the transcriptional level. These include targetable outlier gains and modifications in *NTRKs* as well as a distinct population of *ESR1* depleted recurrences that enrich themselves with genes generally expressed in basal breast cancers—such as *PROM1* and *NDRG3*. Preclinical, mechanistic investigations into these temporally altered genes are immediately warranted given they may uncover novel and targetable mechanisms of endocrine therapy resistance in advanced breast cancers.

6.0 ACQUIRED MOLECULAR FEATURES IN RECURRENT CHEMORESISTANT OVARIAN CANCERS

6.1 ABSTRACT

80% of patients with late-stage serous ovarian cancer (OvCa) recur after an initial treatment response, with the majority of relapsed tumors developing deadly resistance to subsequent therapies. Identifying molecular mediators accountable for this increased malignancy is essential to improve the tragic 12-18 month prognosis for OvCa patients with recurrent disease. To decipher the molecular features driving relapsed ovarian cancers towards therapy resistance, we undertook a transcriptome analysis of 19 longitudinally collected patient-matched pairs representing early and late disease (median disease-free interval of 37 months). We identify a suite of genes consistently upregulated in ovarian cancer recurrences, the most significant being *NTRK2* (adjusted p-value < 0.001) —a targetable tyrosine kinase. Given the lack of targeted therapies available for ovarian cancer, we pragmatically screened for additional clinically actionable targets by defining outlier expression gains and losses in recurrences. The most shared outlier gains were *INHBA* (n = 8 [44%]) and *IGF1* (n = 7 [39%]). Because of the structurally unstable ovarian cancer genome, we then analyzed the cohort for cancer-specific (i.e. absent in a comprehensive panel of normal tissues) fusion RNA transcripts. Globally, 18 of 19 recurrent cancers acquire cancer-specific fusion RNAs that are undetectable in the early lesion. We

subsequently validate an in-frame, late-disease specific fusion between *TOP2A*, a target of doxorubicin and known chemoresistance mediator, and *STAU1*. Lastly, we identify a recurrent, in-frame fusion (*CCDC6-ANK3)* with distinct breakpoints that is maintained in both primary and recurrent lesions and also expressed in the *OVCAR3* cell line. Collectively, these results define multimodal transcriptomic mechanisms of ovarian cancer evolution in late disease—and point towards highly shared acquisitions in recurrences (*NTRK2, INHBA, IGF1,* as well as acquired in-frame fusions) as compelling candidates for disease progression and further preclinical investigation.

Contributors to this study: Nolan Priedigkeit[1,2,7,8], Sarah Taylor[5,6,7,8], Shannon Grabosch[4,7,8], Jahnik Kurukulasuriya[7,8,9], Silvia Liu[4,7,8], Peter C. Lucas,[3,7,8], Ester Elishaev[3,5,7,8], George C. Tseng[4,7,8], Kunle Odunsi[10], Robert P. Edwards[5,6,7,8], Adrian V. Lee[2,7,8]

[1]Medical Scientist Training Program, University of Pittsburgh, Pittsburgh, PA

[2]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, PA

[3]Department of Pathology, University of Pittsburgh, Pittsburgh, PA

[4]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA

[5]Division of Gynecologic Oncology, University of Pittsburgh, Pittsburgh, PA

[6]Department of Obstetrics, Gynecology & Reproductive Sciences, University of Pittburgh, PA

[7]University of Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, PA

[8]Magee-Women's Research Institute, University of Pittsburgh, Pittsburgh, PA

[9]Allderdice High School, Pittsburgh, PA

[10]Department of Gynecologic Oncology, Roswell Park Cancer Institute, Buffalo, NY

6.2 INTRODUCTION

High-grade serous ovarian cancer is a deadly disease with limited options—not one targeted therapy tested clinically has shown overall survival benefits. As such, cytotoxic agents remain the mainstay therapy, particularly platinum and taxane-based chemotherapies[349]. Despite the high mortality of the disease; counterintuitively, over 50% of patients have a complete clinical remission following primary therapies[146]. This short-term success is diminished by the fact that greater than 80% of tumors will recur, after which, the disease is generally incurable with a median overall survival of only 1 to 2 years[147]. Understanding the molecular mediators that permit ovarian cancer to be exquisitely sensitive to chemotherapies at the onset, yet its recurrence to reemerge as viciously chemoresistant, is essential to improve outcomes.

Like almost all other cancers, our understanding of ovarian cancer stems from large-scale genomic characterizations of treatment-naïve primary tumors. From these studies, ovarian cancer is thought to be driven by genomic instability rather than single nucleotide mutations considering the disease is dominated by recurrent structural variation—such as copy number alterations and rearrangements[153]. The only genomic hallmarks thus far are nearly ubiquitous *TP53* mutations and defects in DNA-repair genes, which are disrupted by nucleotide level mutations or an intervening structural variant that causes gene breakage[154]. The adaptations that occur following therapies; however, are largely unexplored.

The only comprehensive, longitudinal study of treatment-resistance ovarian cancers was a whole-genome characterization of 13 primary and relapsed pairs[154]. The authors found little shared potential mechanisms of resistance, but identified *BRCA* mutation reversion events, histologic switching to a more stromal phenotype and recurrent promoter-driven fusions involving *ABCB1*, the *MDR1* drug efflux pump for various cytotoxic agents including paclitaxel,

as likely mediators[154]. Importantly, this study focused on DNA-level changes and most of the recurrent specimens were not solid tumors, but rather malignant ascites cells—as recurrent tumors are rarely biopsied or banked.

In this study, we aimed to expand the characterizations of late ovarian cancer by performing, to our knowledge, the largest transcriptome-wide molecular characterization of paired primary and solid tumor recurrent disease to date.

6.3 METHODS AND MATERIALS

6.3.1 Patient Samples

Institutional Review Board approval from both participating institutions (University of Pittsburgh IRB #PRO15050502, IRB0406147, Roswell Park Cancer Institute IRB #215512) was obtained. Inclusion criteria for this study were (1) patients harbored patient-matched frozen tissue from primary ovarian cancer and a later recurrence (referred to as "early" and "late" disease respectively, Table 7), (2) biospecimens contained regions with sufficient tumor cellularity (> 30%, median in cohort 80%) and (3) RNA integrity scores (RIN) was sufficient for total RNA-sequencing (RIN > 5, median in cohort 7.7). Both a top and bottom slide of the whole tumor, with RNA extraction slides in between, were reviewed by a trained molecular pathologist to confirm pathology and to quantify tumor cellularity. Six, 25-micron frozen OCT-embedded sections were pooled and underwent RNA extraction using Qiagen's RNeasy protocol. Nucleic acids were quantified fluorometrically with a Qubit 2.0 Fluorometer and quality assessed with an Agilent 4200 TapeStation Instrument to determine RIN scores prior to sequencing.

Table 7: Early and late disease, patient-matched ovarian cancer cases
Abbreviations: Met NOS: Metastasis not otherwise specified; NA: Data not currently available.
Note: Case OVCA_19 was excluded from expression analyses given this was a case sequenced by the TCGA. A different sequencing platform was used and harbored profound batch effects vs. all other pairs.

Case	Disease Interval (months)	Early Disease Site	Late Disease Site
OVCA_01	32	Ovary	Small Bowel
OVCA_02	22	Omentum	Met NOS
OVCA_03	72	Met NOS	Colon
OVCA_04	24	NA	Lymph Node
OVCA_05	37	NA	Met NOS
OVCA_06	48	Ovary	Ovary
OVCA_07	88	Met NOS	Abdominal Wall
OVCA_08	18	Omentum	Lymph Node
OVCA_09	73	Spleen	Ovary
OVCA_10	62	Ovary	Met NOS
OVCA_11	24	NA	NA
OVCA_12	37	Ovary	Lymph Node
OVCA_13	25	Ovary	Colon
OVCA_14	55	Ovary	Abdominal Wall
OVCA_15	54	Ovary	Lymph Node
OVCA_16	6	Omentum	Spleen
OVCA_17	7	Omentum	Spleen
OVCA_18	6	Endometrium	Abdominal Wall
OVCA_19	64	Ovary	Pelvic Mass

6.3.2 Total RNA-sequencing

RNA-seq library preparation was performed for 18 early and late disease ovarian cancer pairs using approximately 500 ng of RNA and Illumina's *TruSeq Stranded Total RNA-seq* with Ribo-depletion protocol. Indexed, pooled libraries were then sequenced on High Output flow cells using an Illumina NextSeq 500 system (paired-end reads, 2 X 150 bp). A target of 50 million reads per sample was used to plan indexing and sequencing runs. OVCA_19 was previously sequenced by The Cancer Genome Atlas using different sequencing parameters and was included in the fusion transcript analysis, and not the differential expression analyses, given a large batch effect observed with expression values.

6.3.3 Expression analyses

Adapter-trimmed RNA-sequencing FASTQ files were quantified with k-mer based lightweight-alignment (*Salmon* v0.8.2, quasi-mapping mode, 31-kmer index using GRCh38 Ensembl v88 transcript annotations, seqBias and gcBias corrections)[241]. *tumorMatch* (Chapter 3, Chapter 4, Chapter 5) was used to validate sequencing pairs were patient-matched. RNA-seq read counts and mapping percentages were calculated from *salmon* (Data Supplement 5: S1) and transcript abundance estimates were collapsed to the gene-level with *tximport*[242]. Lowly expressed genes were excluded by defining an expressed gene as having a transcripts per million (TPM) value greater than 1.0 in at least 3 samples. Expressed gene counts were then converted to Log2 transformed TMM-normalized CPM (log2normCPM) values. Log2NormCPM values were used for subsequent analyses, such as hierarchical clustering and outlier expression analyses[243,244]. Adapter-trimmed FASTQs were also aligned using 2-pass mode in STAR (v.2.5.3.a) for

visualization in *Integrated Genomics Viewer* (v2.3.60)[202,246]. Hierarchical clustering was performed using the *heatmap.3* function (https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R) in R on the top 20% most variable genes (defined by IQR) with 1 minus Pearson correlations as distance measurements and the "average" agglomeration method. Differential expression between primary and recurrent tumors was analyzed with *DESeq2*[249]. A paired model was used in the differential expression design (model = ~Patient+Tissue [primary or recurrence]) to account for patient-matched samples. Genes were designated as differentially expressed if they carried an FDR-adjusted p-value of less than 0.10. Outlier expression gains and losses were determined for each patient by discretely categorizing genes into one of 5 categories. If log2FC values (i.e. late disease log2normCPM – early disease log2normCPM) for a given gene were less than Q1 – (1.5 X IQR) or Q1 – (3 X IQR), using case-specific log2FC values for all genes as the distribution, that gene was deemed an "Outlier Loss" or "Extreme Loss" respectively. If log2FC values calculated were greater than Q3 + (1.5 X IQR) or Q3 + (3 X IQR), it was deemed an "Outlier Gain" or "Extreme Gain" respectively. All other genes with intermediate fold changes were classified as "Stable."

6.3.4 RNA Fusion Detection and RT-PCR validations

Fusion RNAs were called with *FusionCatcher v0.99.7b*. Default parameters were used. Final-candidate fusion genes were subsequently filtered for cancer-specific fusions by discarding any fusion also detected in the Human Protein Atlas[350] or BodyMap (*EMBL-BMI, E-MTAB-513*) RNA-sequencing datasets. The same fusion analysis was performed on ovarian cancer cell line data from the Cancer Cell Line Encyclopedia (CCLE)[351]. To validate fusions via RT-PCR, cDNA was generated from 250-500 ng of RNA template using *Bio-rad's iScript Reverse*

Transcription Supermix and the manufacturer's protocol. Approximately 0.5 - 1 ul of resulting cDNA was used to perform a 40-cycle PCR with forward and reverse primers flanking the fusion RNA breakpoints. The PCR product was then visualized with *SYBR Safe* following agarose gel electrophoresis. PCR product was then cleaned using *Exiqon's DNA Clean and Concentrator* kit and subjected to Sanger sequencing. Primer sequences used for validations can be found in Appendix B: Table 9.

6.4 RESULTS

6.4.1 Acquired expression gains in ovarian cancer recurrences

To determine global transcriptome differences between matched pairs, unsupervised hierarchical clustering was performed using normalized expression values. Nine pairs clustered in the same doublet clade of their patient-matched primary, suggesting a profound transcriptional conservation between the recurrence and the early lesion (Figure 21A). To confirm samples were patient-matched, given up to 88 months between early and late disease surgeries, an analysis of shared variants was performed. All pairs harbored a higher proportion of shared variants with their patient-matched primary than to other samples (Figure 21B).

Differential expression analyses revealed heterogeneous expression between the patient-matched samples, only uncovering 39 differentially expressed genes (Figure 22A, Data Supplement 5: S2). The most significantly upregulated gene in late ovarian cancer was *NTRK2*, showing upregulation in the majority of recurrences (Figure 22B). Other genes included a suite

of adipogenesis genes, such as *FABP4, ADIPOQ, APOD,* and upregulation of an ABC transporter, *ABCA6*.

Since resistance mechanisms in advanced cancers may be mutually exclusive, and thus would be missed by conventional differential expression analyses given the gene-level stringency, we performed a targeted analysis focusing on outlier expression gains and losses—particularly in genes that are clinically actionable (Data Supplement 5: S3)[195]. Four clinically actionable genes showed outlier increases in at least one-third of late disease samples versus their matched early disease lesion—*INHBA, IGF1 NTRK2* and *EPHA3* (Figure 22C).

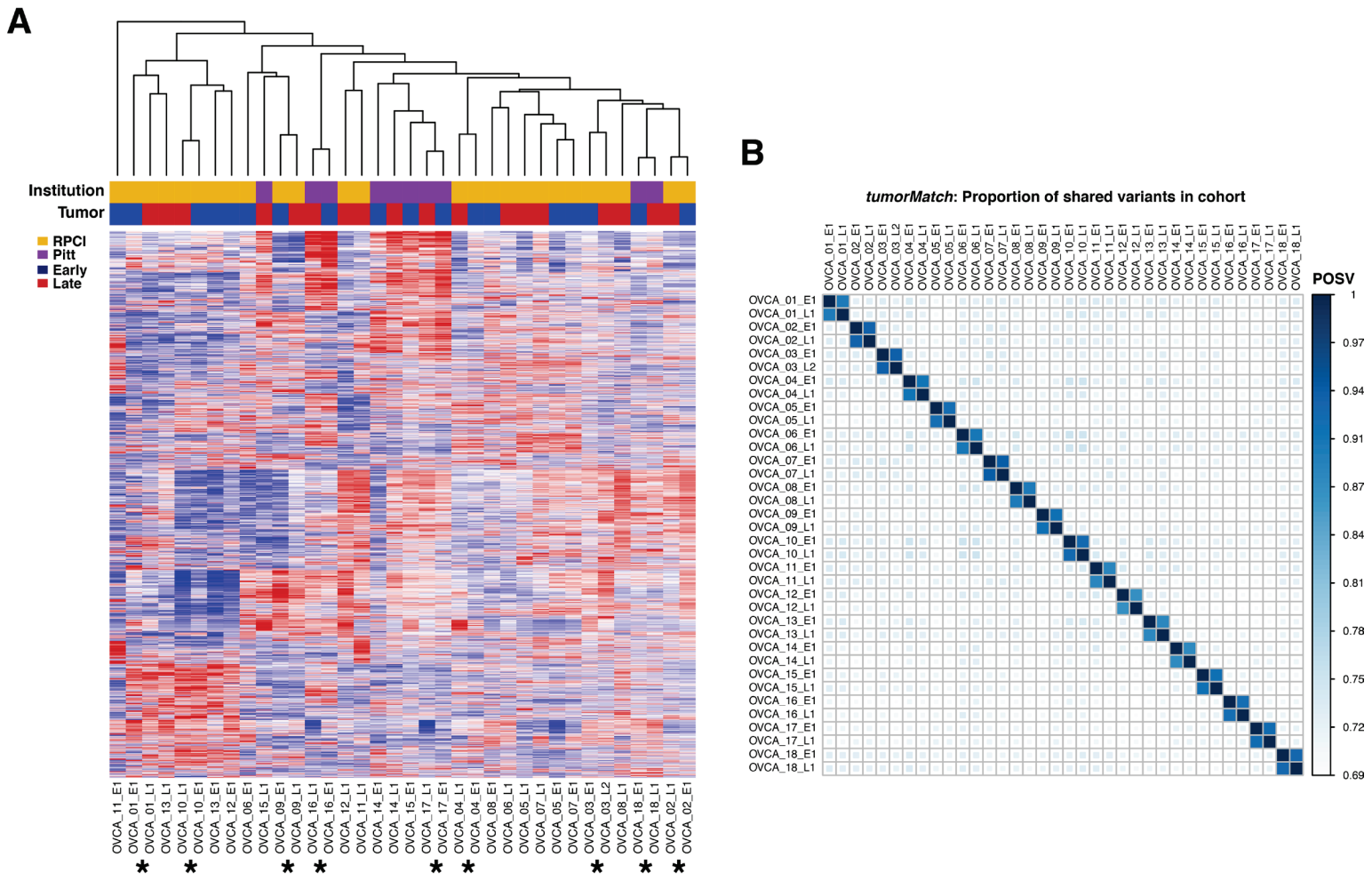

Figure 21: Unsupervised clustering and *tumorMatch* in ovarian cancer cohort
(A) Unsupervised hierarchical clustering on 20% most variable genes across the cohort (E1 = early disease, L1/2 = late disease). Institution (yellow = Roswell Park Cancer Institute; purple = University of Pittsburgh) and tumor type (blue = early disease; red = late disease) is indicated. Samples marked with an asterisk are early and late lesions that cluster together. **(B)** *tumorMatch* scores which represent the proportion of shared variants between samples. Darker blue and larger squares indicate a higher degree of genetic similarity between samples.

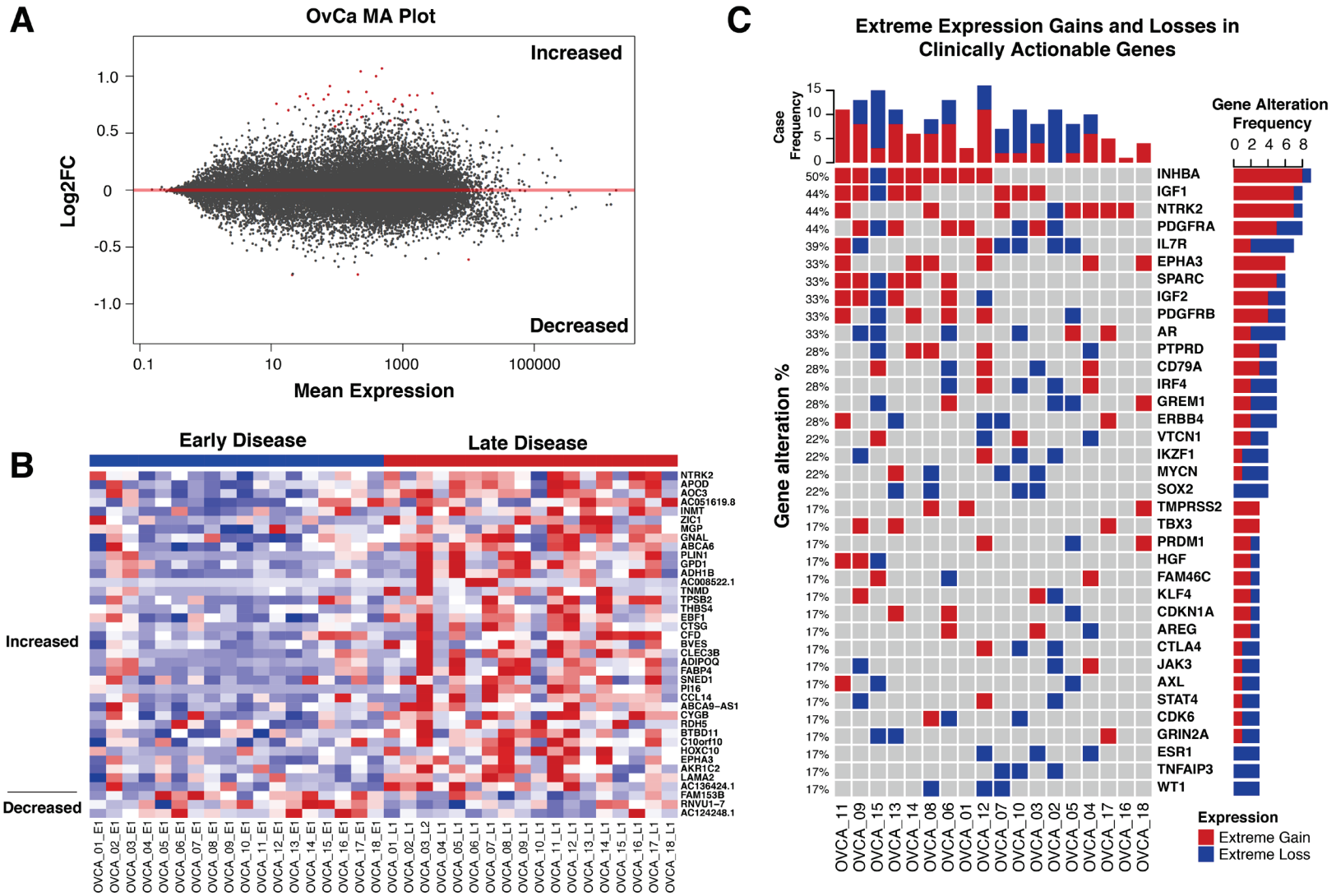

Figure 22: Differentially expressed genes and outlier expression events in ovarian cancer recurrences
(A) MA plot of genes interrogated for differential expression (log2FC [recurrence vs. primary] vs. mean of normalized counts). Significant genes (n = 39, padj < 0.10) are indicated in red. **(B)** Normalized expression heatmap of differentially expressed genes, ranked by adjusted p-value (top to bottom) and segregated by increased genes and decreased genes. **(C)** OncoPrint of outlier expression gains (red) and losses (blue) in patient-matched pairs, along with alteration case frequencies and recurrence percentage.

6.4.2 Preserved and acquired fusion RNAs in ovarian cancer recurrences

Given structural variation is a hallmark of ovarian cancer and preliminary data supports acquired fusions as potential mediators of chemoresistance in relapsed disease, we undertook an analysis of cancer-specific fusion RNAs. After excluding fusion RNAs found in a comprehensive panel of normals, a median of 7 cancer-specific fusion transcripts was acquired in each late disease sample. Nearly all recurrences also harbored "preserved" fusions—fusion transcripts detected in both the early and late lesion (Figure 23A). Given the low sensitivity observed in fusion finding algorithms[352], we selected three, bioinformatically called "acquired" fusion RNAs to validate using RT-PCR with primers flanking the breakpoints. All three were found to be either specific to the recurrence or highly enriched in the recurrence versus the matched primary (Appendix A.5, Figure 47).

152 fusions were predicted to produce an in-frame, chimeric protein—48 being acquired in late disease and 55 being preserved (Data Supplement 5: S4). Although no acquired fusions were present in more than one recurrence, fusions of particular interest included an acquired *WNT2-CTTNBP2* in case OVCA_04, which retained a Wnt signaling peptide in the N-terminal region of the hypothetical protein product, and a fusion involving *TOP2A* (chromosome 17) and *STAU1* (chromosome 20) in case OVCA_19. Given the latter fusion's involvement with a known chemoresistance mediator, *TOP2A,* we explored this fusion in more detail. The *TOP2A-STAU1* fusion, containing up to exon 19 in *TOP2A* and the 3' region of *STAU1* beginning at exon 6, carried a high degree of bioinformatic support with 19 unique reads spanning the breakpoint (Figure 23B). Visualization of the RNA-seq alignment also revealed increased coverage of *TOP2A* up until the breakpoint in only the late disease sample (Figure 23C). The fusion was then

validated with RT-PCR and Sanger sequencing using two separate PCR primer pairs spanning the breakpoint. Importantly, *TOP2A-STAU1* was not detected in the early lesion or in an unrelated sample—confirming its specificity to OVCA_19 and its acquisition in advanced disease (Figure 23D).

Because preserved fusions we found to be common in ovarian cancer recurrences, we searched for preserved fusion genes that were shared in multiple samples, which would increase their likelihood of being driver alterations. Two recurrent in-frame fusions were identified—*MED12-IRF2BPL* and *CCDC6-ANK3.* The bioinformatically called *MED12-IRF2BPL* fusion breakpoint was within highly homologous polyglutamine repeat regions of each fusion partner, suggesting this as a false positive fusion. *CCDC6-ANK3;* however, was found to harbor distinct breakpoints in each of the samples called—all of which produced a hypothetical, in-frame protein product. These breakpoints were confirmed with *RT-PCR* and another *CCDC6-ANK3* fusion was validated in the cisplatin-resistant *OVCAR3* cell line (Figure 24E)[353].

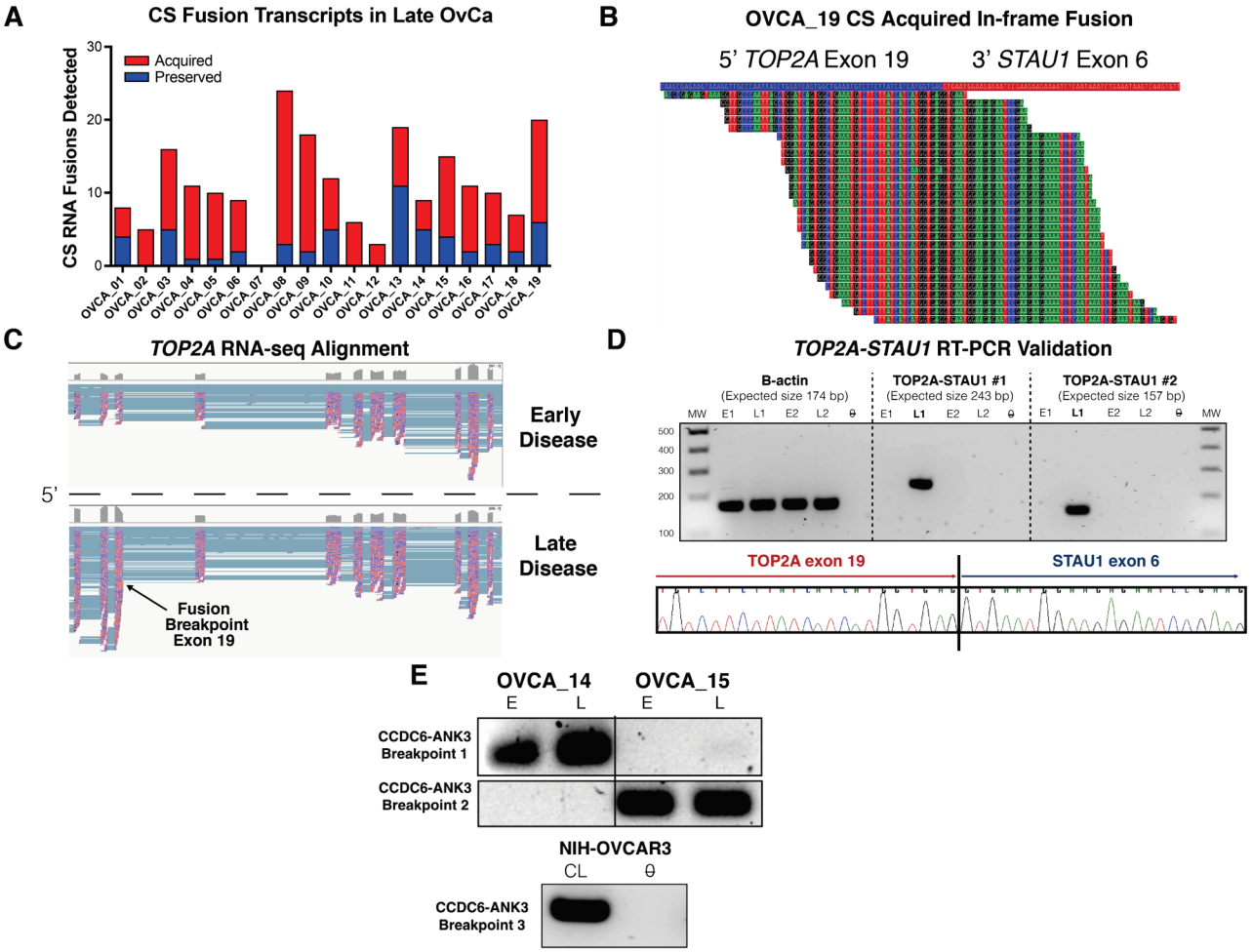

Figure 23: Fusion RNA landscape in recurrent ovarian cancer
(A) Landscape of cancer-specific (CS) fusion transcripts in late ovarian cancer. Frequency of cancer-specific fusions are shown for each case, with blue representing the number of preserved fusions (present in both early and late disease) and red representing the number of late disease acquired fusions. **(B)** Reads supporting the in-frame, late disease specific fusion involving *TOP2A* and *STAU1*. **(C)** STAR RNA-seq alignment showing enriched coverage of reads preceding *TOP2A* fusion breakpoint at exon 19. **(D)** RT-PCR of *ACTB* and *TOP2A-STAU1* in early and late disease samples of case OVCA_19 (E1, L1) and another, unrelated early and late disease pair (E2, L2). Sanger sequencing of PCR product showing fusion breakpoint sequence below gel image. **(E)** *CCDC6-ANK3* fusion validations. Top: Ovarian cancer cases, E = early disease, L = late disease sample. Bottom: OVCAR-3, CL = cell line, θ = no template control.

6.5 DISCUSSION

Recurrent ovarian cancer is generally incurable and presents a major clinical challenge. The molecular differences between primary tumors—which tend to be uniquely sensitive to initial rounds of chemotherapy—and relapsed disease are largely unknown. In this study, we performed a preliminary interrogation of molecular features acquired in advanced disease versus early. We identify differentially expressed genes consistently upregulated in late ovarian cancers—most notably *NTRK2*, a targetable tyrosine kinase. We subsequently analyze outlier expression gains in clinically actionable genes and find highly recurrent upregulations of *INHBA, IGF1* and *EPHA3*. Next, we define the landscape of cancer-specific fusion genes in relapsed ovarian cancer and although we observe minimal recurrent events, we discover both preserved and acquired cancer-specific fusion RNAs are quite common. This included an in-frame, relapse-specific *TOP2A-STAU1* fusion and a recurrent *CCDC6-ANK3* fusion harboring distinct breakpoints in three separate cancers—two cases with the fusion retained in both the primary and recurrent tumors and another fusion present in a cisplatin-resistant cell line. Collectively, these results represent the most comprehensive, expression-based characterization of relapsed ovarian cancer to date and identify novel targets for further preclinical investigation.

Like the previous studies (Chapters 2, 3, 4, 5), advanced ovarian cancers can be very similar to their patient-matched primaries transcriptionally, yet acquire recurrent alterations that make them distinct from earlier lesions. *NTRK2* is the most differentially expressed gene in ovarian cancer recurrences, harboring expression gains in the majority of relapsed samples. Tropomyosin kinases, such as *NTRK2*, have a recently appreciated role in oncogenesis, with preclinical evidence suggesting that fusions involving members of this gene family may serve as viable therapeutic targets (NCT02568267)[336–338]. Other differentially expressed genes included

the upregulation of adipogenesis pathway members, including *FABP4*. Fatty acids in the peritoneum, stemming largely from the omentum, have been proposed as fuel for ovarian cancer growth and therapy resistance with *FABP4* being a key regulator of this process and our results in late ovarian cancers further support this potential mediator of disease progression[354].

The most recurrent clinically actionable gains in ovarian cancer were *INHBA* (n = 8 [44%]), *IGF1* (n = 7 [39%]), *NTRK2* (n = 7 [39%]) and *EPHA3* (n = 6 [33%]). *INHBA* produces a protein product inhibin beta A—a subunit of both activin and inhibin that act as positive and negative regulators, respectively, of hormone secretion, particularly FSH. Dysregulated expression of these subunits has been shown to play a role in ovarian cancer pathogenesis and serum inhibins may serve as potential biomarkers to complement CA125 in ovarian cancer[355–357]. Given ovarian cancer recurrences have extreme gains of this particular subunit, the activin and inhibin axis in late, chemoresistant ovarian cancers should be explored in more detail. *IGF1* gains are also common in ovarian cancer recurrences, particularly relevant given IGF-signaling inhibitors are readily available. Indeed, IGF pathway members mediate chemotherapy resistance in ovarian cancer cell models and IGF growth factors and binding proteins, when upregulated, confer worse outcomes in ovarian cancer[358–362]. Finally, we find an ephrin receptor (*EPHA3*) to be upregulated in recurrences. Like IGF, expression of ephrin receptors in ovarian cancers is associated with shorter survival, with this receptor mediating many cellular functions including growth advantages, angiogenesis and cell adhesion phenotypes[363,364] Notably, ephrin and NTRK family members were found to be upregulated in chemotherapy treated breast cancer metastases and local recurrences as well (Chapters 3, 4, 5), suggesting these pathways may mediate disease progression and chemotherapy resistance across a broad range of cancer types.

A comprehensive analysis of cancer-specific (i.e. not present in normal tissue) fusion transcripts was then performed. Across each recurrence, a median of 7 acquired fusions were discovered. A particularly compelling acquisition was a relapse-acquired fusion transcript between the 5' region of *TOP2A* and the 3' region of *STAU1*. *STAU1*, an RNA-binding protein, post-transcriptionally regulates cell cycle mediators in cancer cells and the first 88 amino acids, notably absent in the identified fusion, are necessary for inhibiting cellular proliferation[365]. TOP2A is an essential protein for DNA replication and transcriptional regulation, mediates chemoresistance, is associated with poor prognosis in a variety of cancers and the ATPase, transducer and TOPRIM functional domains are importantly retained in the fusion product[366–370]. Additionally, inhibitors of TOP2A are used clinically with doxorubicin intriguingly showing selective efficacy in a subset of patients with recurrent OvCa[371,372]. Future functional studies of this fusion, as well as additional screening to see if *TOP2A* fusions are a common event in relapsed ovarian cancer, may be warranted.

Lastly, we identified recurrent, in-frame *CCDC6-ANK3* fusions preserved in both early and late cancers in two separate cases and present in the cisplatin resistant *OVCAR3* cell line. The three identified fusions harbored distinct breakpoints, implying a potential driver alteration in ovarian cancer. *CCDC6* and *ANK3* are relatively uncharacterized, although *CCDC6* is often a fusion partner with *RET* in lung cancers[373]. Intriguingly, the fact that these fusions, as well as others, can be preserved in both early and late disease suggest fusion transcripts can serve as early, "truncal" events—preserved in the majority of subclones in the cancer. This presents an interesting opportunity given fusion breakpoints are uniquely cancer-specific and 15 of 19 cases (79%) had preserved fusions. We have shown previously that quantifying fusion RNA transcripts in plasma can serve as personalized biomarkers of disease surveillance, as they correlate with

tumor burden as effectively as CA125[374]. Furthermore, recent methods have used CRISPR to exploit and target cancer-specific nucleotide sequences brought about through genomic rearrangements, which may imply fusion transcripts—even if only present in single patients—may serve as "no-of-one" biomarkers or therapy targets[375].

In summary, this study uncovers novel and potentially targetable acquisitions in advanced ovarian cancers that make them distinct from early tumors. We identify a suite of highly recurrent gains in more advanced disease, including druggable acquisitions of *NTRK2* and *IGF1*. Furthermore, we explore a previously unrecognized form of transcriptome evolution in advanced cancers, particularly the acquisition of fusion transcripts. Lastly, we define preserved, "truncal" fusion transcripts as common somatic events in ovarian cancer—the recurrent *CCDC6-ANK3* being a prime example—which may serve as unique, cancer-specific nucleotide targets and biomarkers in ovarian cancer.

7.0 CONCLUSIONS

An individual cancer serves as a microcosm of evolution. The cancer's intrinsic genetic toolkit, where it grows, the exposures and therapies it encounters—these complex and interacting pressures drive the evolution and ultimately define what the cancer becomes[376]. The final step in a cancer's evolution; however, is oftentimes killing its host—after the disease gains an ability to evade therapies and colonize vital organs. Surprisingly, these more advanced tumors are largely uncharacterized and the most poorly understood, tragically overshadowed by near stagnant survival gains of patients with advanced cancers in the past few decades[3].

In this collection of studies, we defined molecular features that make advanced cancers unique from the relatively benign, early lesions that they originate from. We use novel sequencing strategies and analyses to prove our hypothesis that advanced breast and ovarian cancers acquire distinct, recurrent and druggable molecular dependencies as they evolve towards therapy resistance and metastatic colonization—which may have profound implications for how we study, profile, understand and ultimately treat advanced disease.

In Chapter 2, to begin to explore this hypothesis, we performed a targeted expression analysis of breast cancer cells that colonize the brain. We find that the majority of breast cancer brain metastases retain their intrinsic transcriptional subtype; yet, nearly all cancers acquire expression features distinct from the primary lesion—some of which are readily druggable. The most recurrent expression alteration included gains of *ERBB2*, with brain metastases

significantly upregulating *ERBB2* in nearly 35% of cases. Importantly, we show approximately 20% of brain metastases acquire HER2-positivity in patients with originally HER2-negative primary tumors—which has immediate clinical implications since the majority of these patients can be offered a HER2-targeted therapy. Other significant expression gains included *FGFR4* and *FLT1* as well as recurrent losses in *ESR1*. Studies are now ongoing to determine how to preemptively identify which patients' tumors will switch to HER2 and if plasma biomarkers, such as circulating tumor cells or circulating free tumor DNA, can uncover patients who have switched without an invasive biopsy. Encouragingly, we found that primary breast cancers that switched to HER2 positive in the brain metastasis, exhibited increased *ERBB2* expression (data not shown), perhaps reflecting subclonal amplifications in the primary cancer which are subsequently selected for expansion during therapy or metastasis. Thus, it is possible that a comprehensive analysis of *ERBB2* levels and/or HER2-activation in primary cancers may identify a subset of HER2-negative breast cancers for whom metastasis may be prevented by early HER2 therapy.

Following this targeted study, we then performed a technical analysis to determine the efficacy of hybridization-based exome-capture RNA-sequencing (ecRNA-seq) on quantifying expression from samples that were formalin-fixed paraffin embedded (FFPE), up to a decade-old, decalcified and highly degraded. Using a particularly unique set of matched aged and frozen samples, we found minimal differences in expression values between highly degraded RNA from FFPE and intact RNA from frozen tissue. We then applied this technology to decalcified bone metastases—an historically challenging sample type to molecularly characterize given highly degraded nucleic acids. We discovered shifts in subtypes to more proliferative, HER2 and Luminal B profiles in bone metastases versus matched primary tumors, a significant temporal

influence on transcriptome evolution and like the previous Chapter, acquisitions of clinically actionable targets—particularly in the CDK-Rb-E2F and FGFR signaling pathways. Overall, this study defines a novel technology that can be used to perform expression-based characterizations on highly degraded RNA extracts and defines compelling targets for further investigation into bone metastasis mechanisms. Ongoing studies are determining the role of FGFR upregulation in recurrent breast cancers and functionally screening candidates identified in this study for metastatic or endocrine-resistance phenotypes.

In Chapter 4, we apply ecRNA-seq to the brain metastasis cohort, thereby obtaining a more global view of transcriptomic adaptations breast cancer cells make after colonizing the brain. We identify novel pathway-level gains in brain metastasis, create a signature of breast cancer-specific genes that can predict brain metastasis relapse, show that HER2-switching cases identified in Chapter 2 may be predicted by intermediate levels of a HER2-signature in the primary tumor and identify methylation of the *ESR1* locus as a mechanism of *ESR1* loss in advanced breast cancers—particularly important given *ESR1* loss is a shared feature of most of the patient-matched cohorts. Finally, we identify a suite of targetable kinases that are consistently upregulated in brain metastases—the two most recurrent being HER2 and RET. To test if these gains are genuine dependencies, we then applied *in vitro, ex vivo* patient-derived brain metastases and *in vivo* brain metastasis-derived xenograft models to show targeting HER2 or RET has significant antitumor activity, pointing towards RET as a novel target in patients with brain metastases. Ongoing studies are now using orthotopic brain metastasis mouse models to reinforce targeting these acquired kinases as viable therapeutic options. A particularly interesting observation that also demands future investigation is the majority of receptor kinases upregulated in brain metastases (*ERBB2*, *RET*, *ERBB4*) harbor ligands that are highly expressed in the brain

(neuregulins, GDNF-family ligands). This is suggestive that cancer subclones which express a particular receptor highly may selectively colonize a target organ that is enriched with the ligand of this receptor. Future studies should explore this "ligand-homing" hypothesis further and to see if this observation is preserved across other metastatic sites.

To determine if these changes are observed in non-metastatic, but therapy resistant lesions, in Chapter 5 we undertook another, more comprehensive targeted DNA and RNA expression analysis on long-term local recurrences that grew in an estrogen-depleted environment. We show an even more profound transcriptional profile conservation in local recurrences than in metastases. Likewise, we observed limited, pathogenic DNA-level enrichments in recurrences, with few exceptions including a recurrence-enriched suite of three *PIK3CA* mutations and a nonsense *ARID1A* mutation. In contrast to DNA-level changes; however, recurrent expression alterations were very common—the most notable being losses of *ESR1* in 42% of tumors and gains of *NTRK* family members, *TERT* and *FGFR4.* An analysis of tumors that became *ESR1* depleted uncovered consistent remodeling events in these tumors—most notably the acquisition of the stem-cell and luminal lineage marker *PROM1* along with a group of other genes usually expressed in basal-like breast cancers. Taken together, this study uncovered highly recurrent and targetable acquired transcriptional remodeling events in endocrine therapy resistant tumors and potentially identified a relatively common *ESR1*-depleted breast cancer subtype that gains basal-like transcriptional traits when exposed to estrogen deprivation. Future studies should focus on the more recurrent transcriptional events—such as *NTRK* gains and this *ESR1* depleted phenotype—particularly how these gained genes can confer estrogen independence or more malignant phenotypes.

Lastly, in Chapter 6, we prove these themes are conserved in ovarian cancer—a cancer with no effective targeted therapies. Undertaking the most comprehensive transcriptome-wide characterization of relapsed ovarian cancer to date, we identify recurrently acquired expression gains and losses when ovarian cancers become therapy resistant—specifically upregulations of *NTRK2* and *IGF1*; genes that are readily druggable. Moreover, we define a novel mechanism of tumor evolution in the acquisition of fusion transcripts, identifying late disease acquired fusion RNAs such as *TOP2A-STAU1*. Finally, we uncover "truncal" recurrent fusions that are preserved in both early and late disease, which could serve as unique, patient and cancer-specific nucleic acid or protein targets. Ongoing efforts are identifying strategies to detect and target these fusion transcripts through the use of *CRISPR* technology, or to use them as more exotic immunotherapy targets such as cancer-specific neoepitopes—especially given almost every case studied harbored these preserved fusions in both the primary and recurrent lesion.

Taken together, this body of work provides compelling evidence that transcriptome evolution should be considered a prime resource in identifying novel therapeutic strategies in advanced cancers. Nearly every cancer pair analyzed harbored a major expression gain or loss in a gene that is readily druggable—with sometimes highly recurrent gains in specific cancer subtypes and metastatic sites. Currently, clinical tumor profiling to identify actionable targets generally emphasizes largely static (i.e. one time point) DNA-level changes in the advanced setting. The results herein reinforce that longitudinal profiling of transcriptomic changes may be just as essential, if not more, in identifying precision therapeutic targets for cancers that have learned to evade traditional therapies and colonize distant organs.

APPENDIX A

SUPPLEMENTAL FIGURES

A.1 INTRINSIC SUBTYPE SWITCHING AND HER2 GAINS IN BREAST CANCER BRAIN METASTASES: SUPPLEMENTAL FIGURES

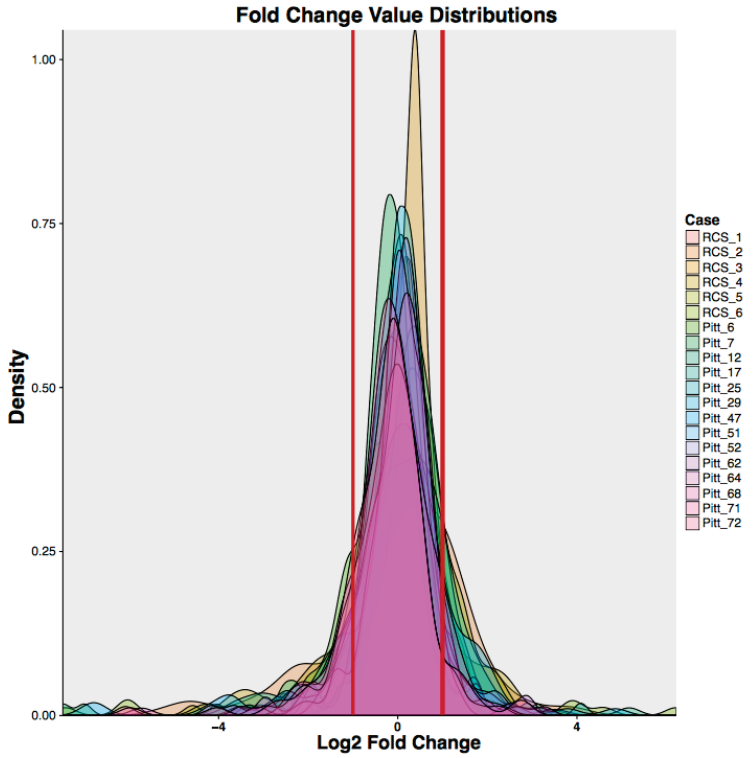

Figure 24: Fold change density distribution in patient-matched pairs
Fold-change value density plot for each case (i.e. Log2 brain metastasis normalized expression – Log2 primary metastasis normalized expression). Mean is -0.01 and 1 standard deviation above and below the mean are marked with vertical red lines. Expression alterations outside these lines were counted as 'expression alterations'.

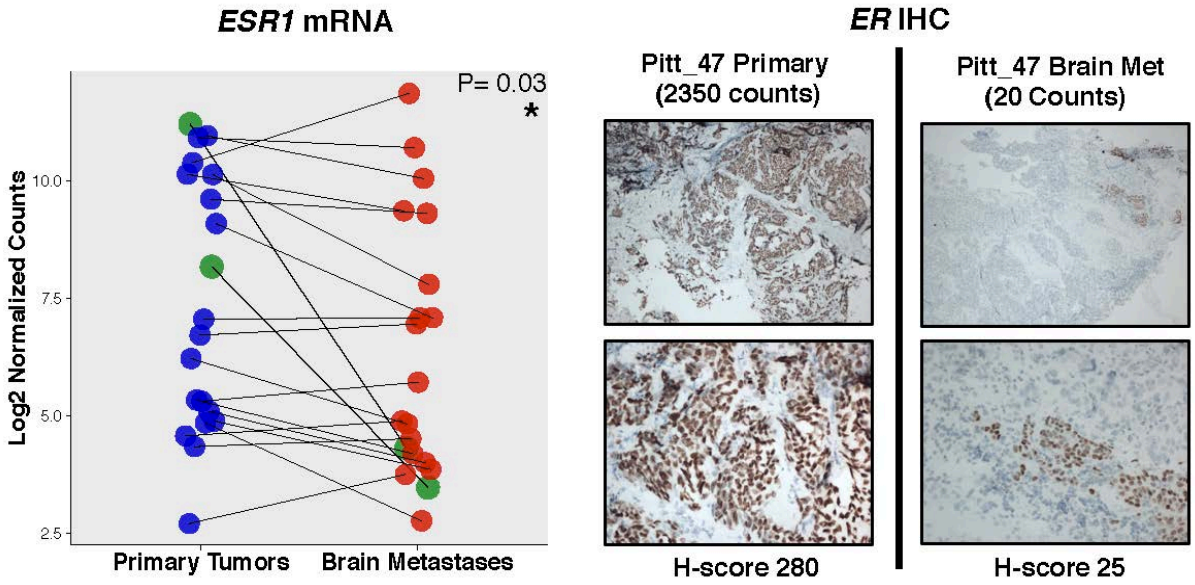

Figure 25: ER expression loss in breast cancer brain metastases
(A) Paired ladder plot of *ESR1* expression in patient-matched cases. Green dots represent samples with suspected hormone status switching, p-values (* p <= 0.05, ** p <= 0.01, *** p <= 0.001) shown are from Wilcoxon signed-rank tests (primaries vs. metastases). **(B)** Primary and metastatic IHC staining of ER from case Pitt_47, along with normalized NanoString expression counts and pathological H-score. Top images are low magnification, bottom images are high magnification.

A.2 EXOME-CAPTURE RNA-SEQUENCING OF DECADE-OLD BREAST CANCERS AND MATCHED DECALCIFIED BONE METASTASES IDENTIFIES CLINICALLY ACTIONABLE TARGETS: SUPPLEMENTAL FIGURES

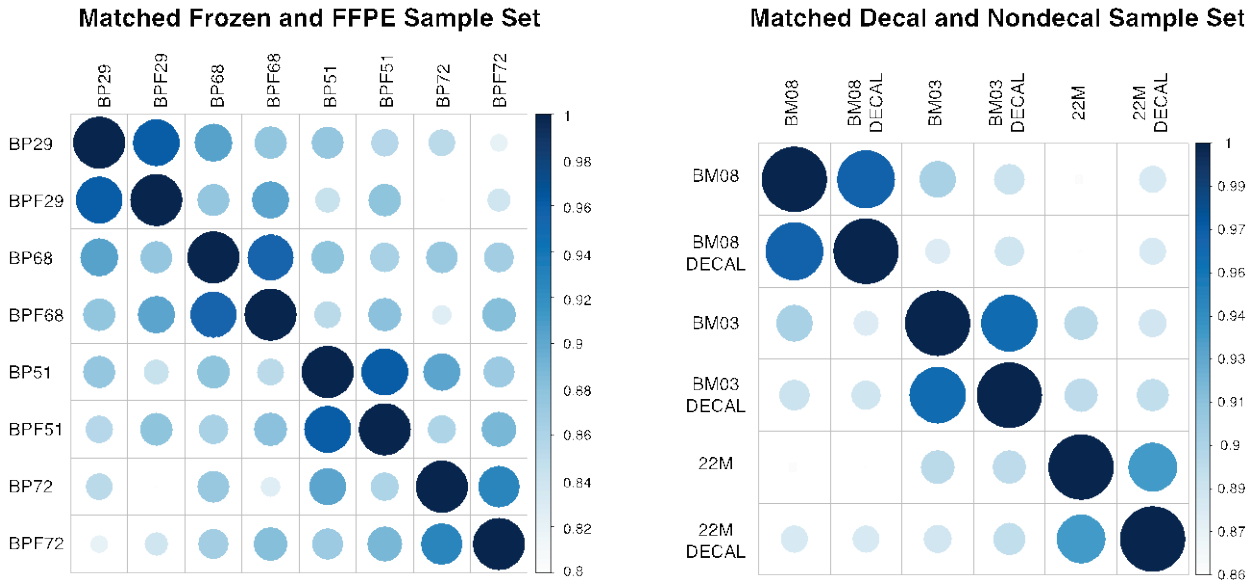

Figure 26: Expression correlation plots of ecRNA-seq sample sets
Correlation plots of matched flash-frozen vs. FFPE and matched FFPE-decalcified vs. FFPE-non-decalcified sample sets. Both size and shade of color represent Pearson *r* correlations between all samples within each sample set; larger circles and darker blue colors represent higher correlations.

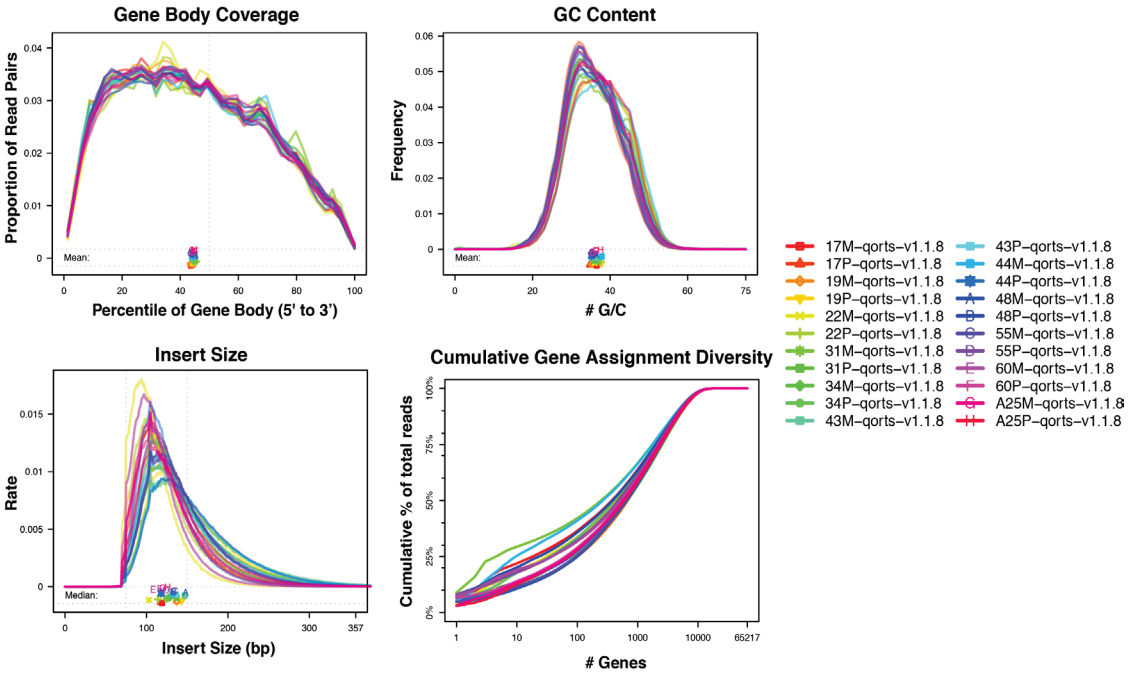

Figure 27: ecRNA-seq QC metrics for patient-matched sample cohort
ecRNA-sequencing gene body coverage, GC content and insert size distributions along with gene assignment diversity assignments for all 22 tumors in patient-matched cohort. Each tumor is plotted with a different color, legend on right.

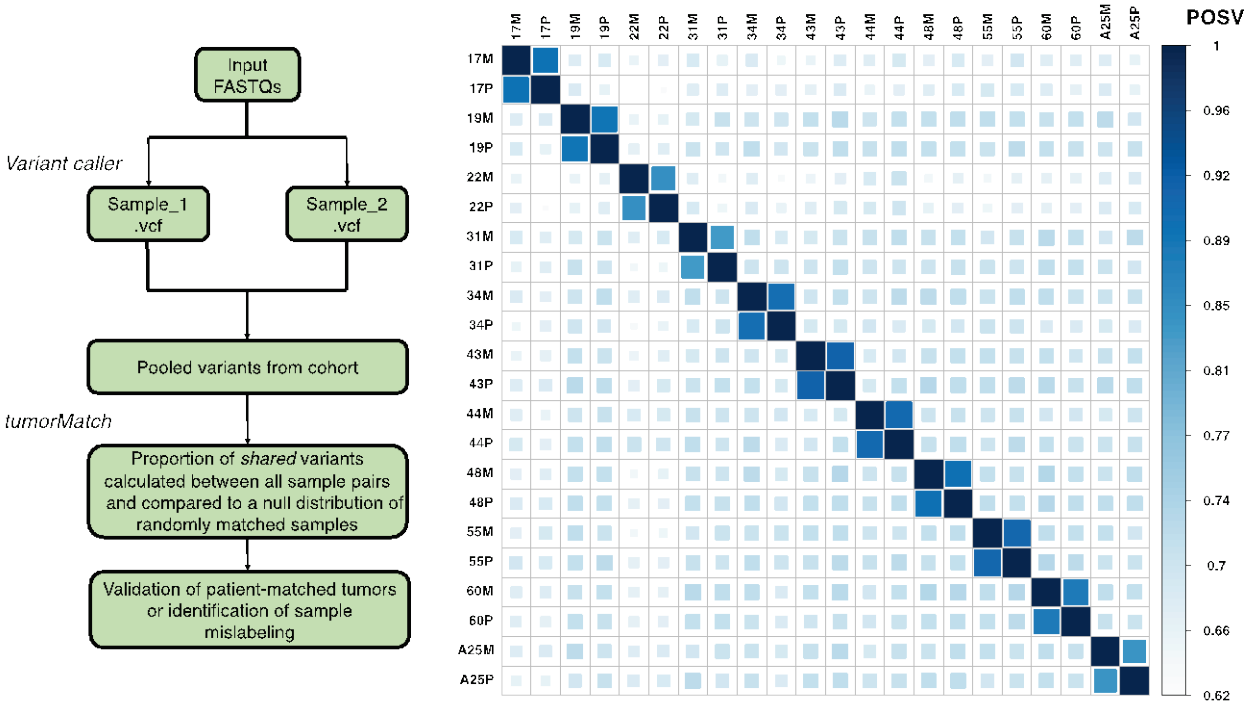

Figure 28: tumorMatch: Proportion of shared variants (POSV) between samples in patient-matched cohort
Left, diagram outlining *tumorMatch* method which identifies patient-matched tumor specimens or sample mislabeling. Right, correlation plot showing proportion of shared variants between all 22 tumors in the cohort; bigger squares and darker blue color represents a higher proportion of shared variants (POSV) between two samples.

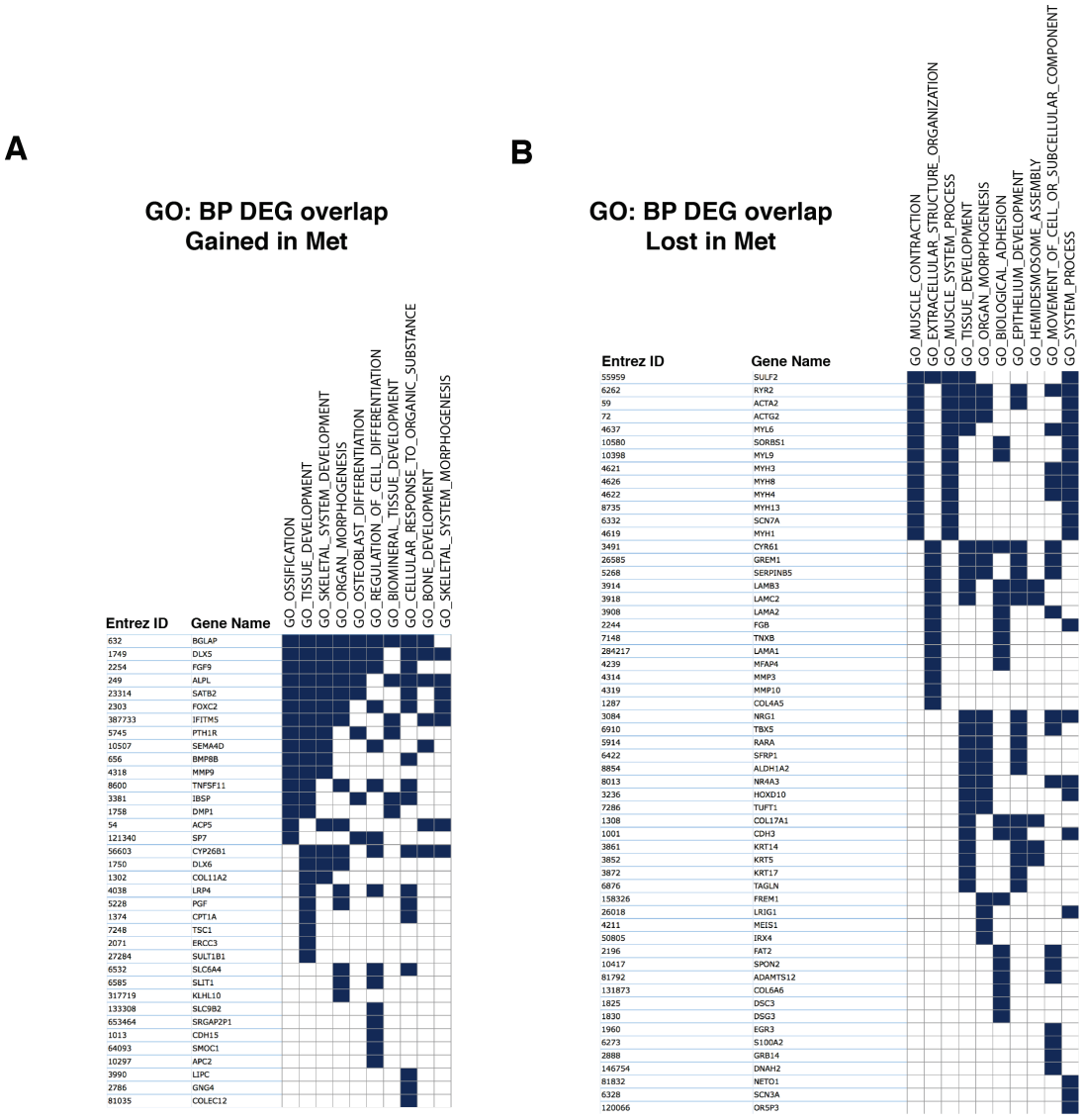

Figure 29: Gene Ontology: Biological Processes (GO:BP) gene overlaps for differentially expressed gene sets (A) GO:BP gene overlaps for genes with significant expression increases in bone metastases vs. patient-matched primaries. (B) GO:BP gene overlaps for genes with significant expression decreases in bone metastases vs. patient-matched primaries.

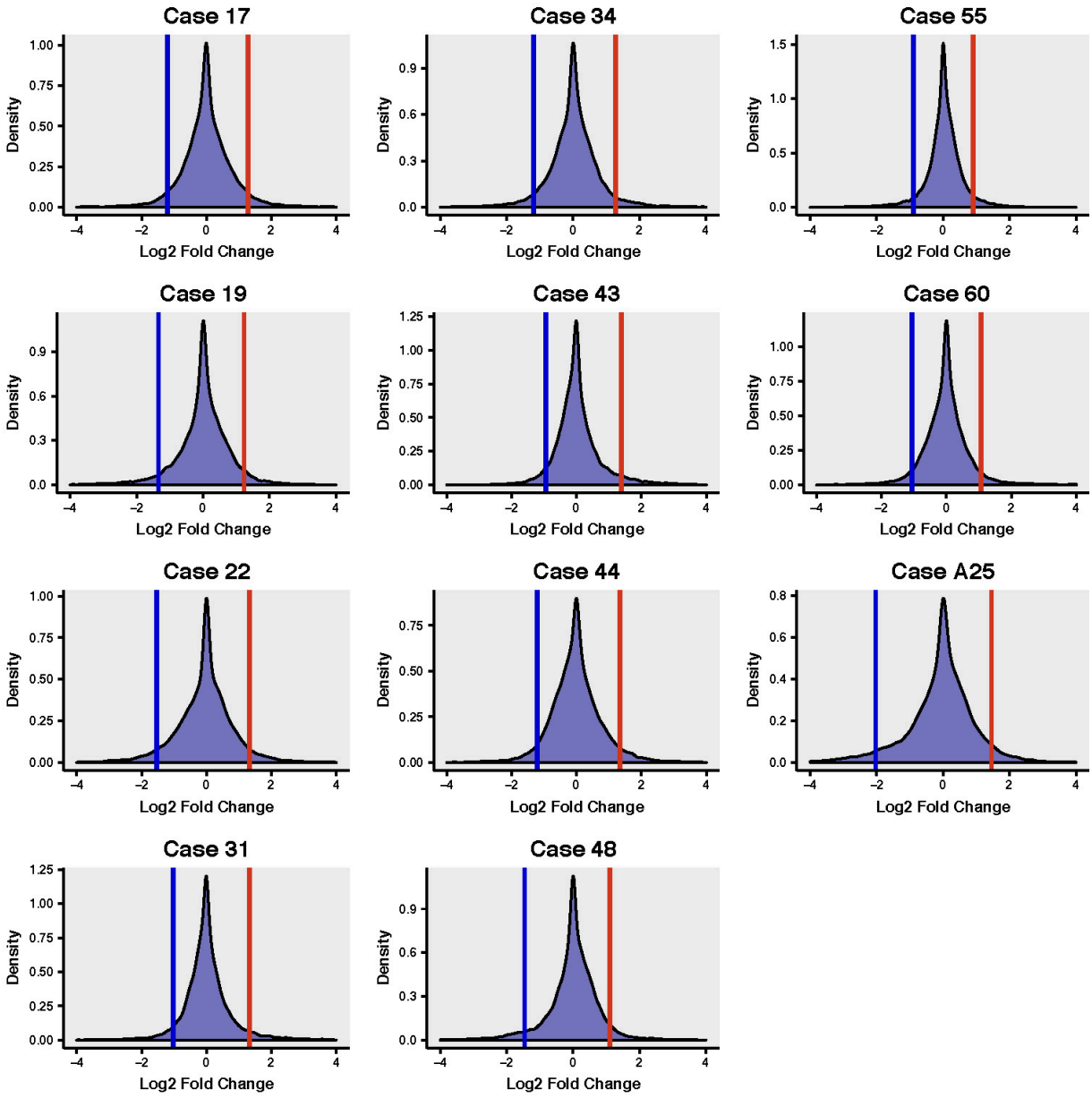

Figure 30: Case-specific expression fold-change distributions and expression alteration thresholds
Fold-change density plots using log2normCPM values (Metastasis log2normCPM – Primary log2normCPM) for all genes. Expression alteration thresholds for significant expression loss (marked in blue, 5th percentile) and significant expression gain (marked in red, 95th percentile) shown for each of the 11 patient-matched cases.

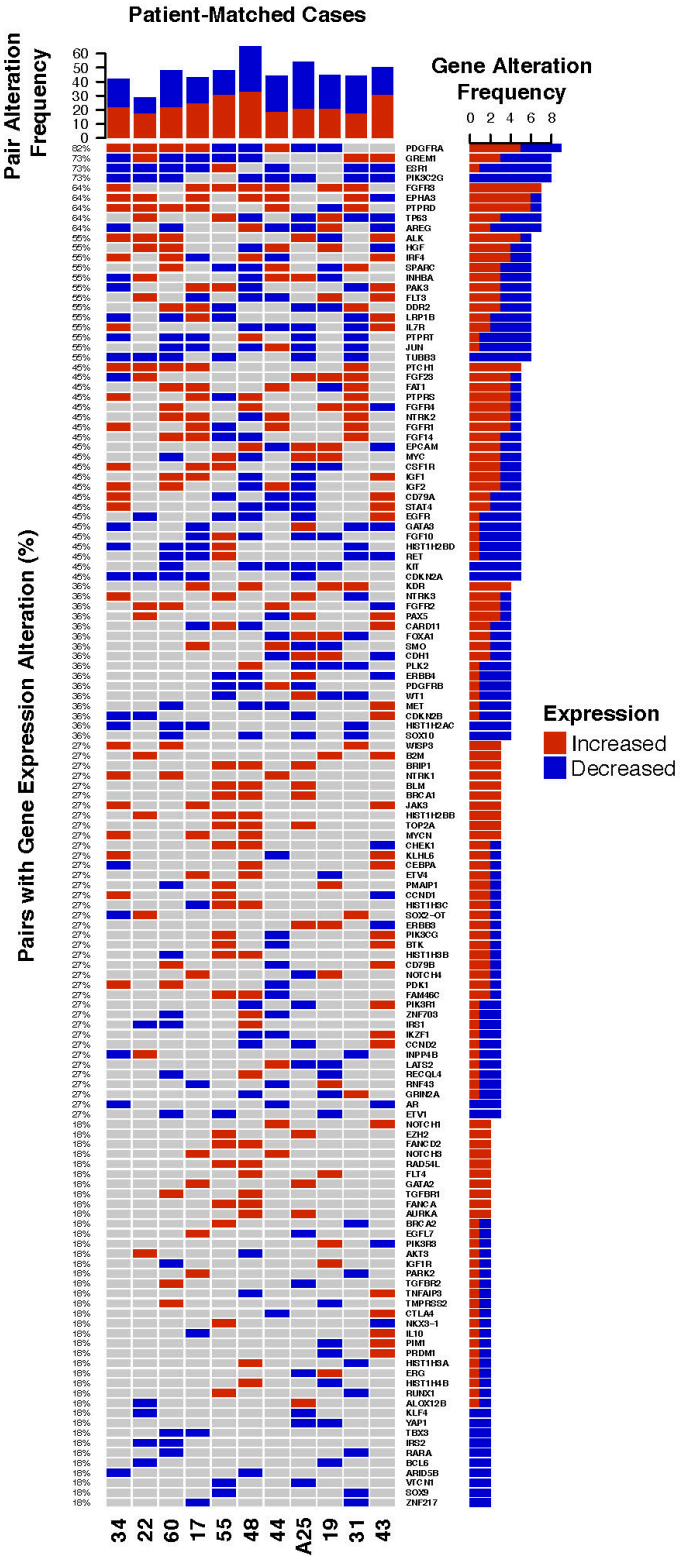

Figure 31: All recurrent expression alterations in clinically actionable genes
Oncoprint plot showing all recurrent (> 1 case) expression alterations in bone metastases with each column representing a patient-matched case. Pair alteration frequencies, gene-specific expression alteration percentages and gene alteration frequency shown. Red tiles represent significant expression gains and blue tiles represent significant expression losses (as defined by case-specific expression alteration thresholds). Genes ranked by gene alteration frequency.

A.3 TRANSCRIPTOME-WIDE IDENTIFICATION OF RET AND HER2 SIGNALING AS RECURRENTLY ENRICHED DEPENDENCIES IN BREAST CANCER BRAIN METASTASES: SUPPLEMENTAL FIGURES

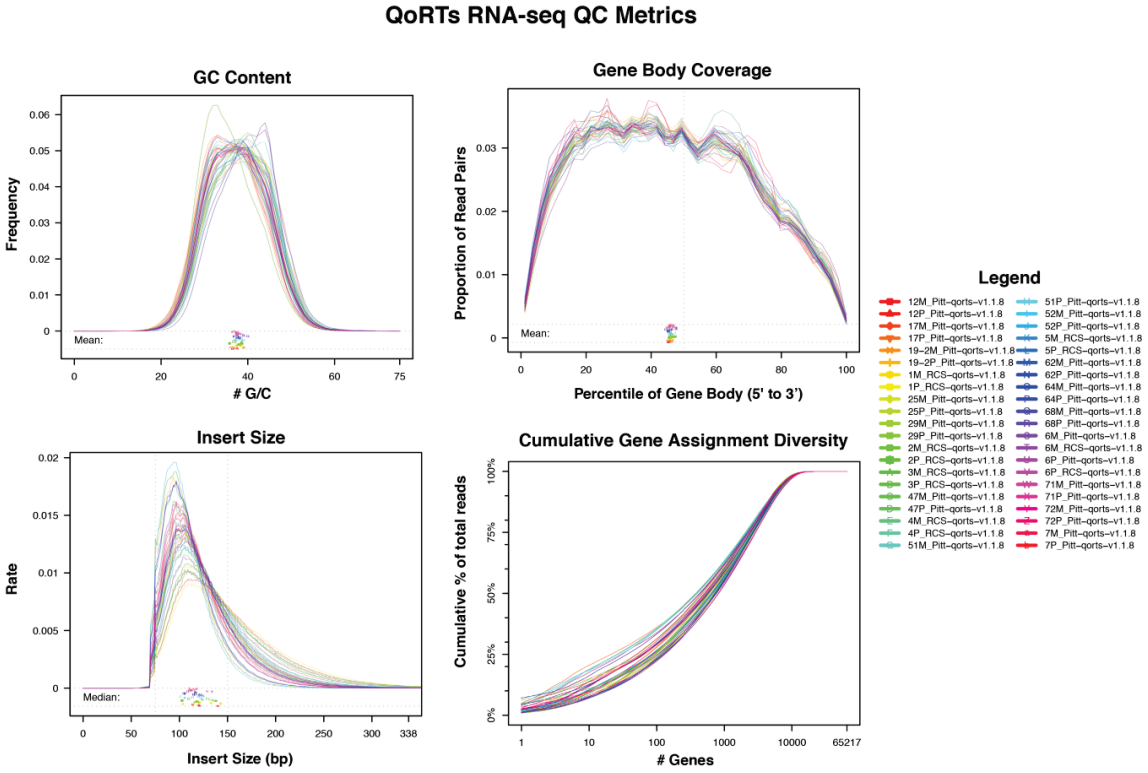

Figure 32: Brain metastasis cohort RNA-seq quality metrics.
Calculated and plotted using QoRTs (v1.1.8) following two-pass read alignment with STAR (v2.4.2a) for the 21 patient-matched cases. GC content, gene body coverage, insert size and cumulative gene assignment diversity are shown, with colors representing each of the 42 samples.

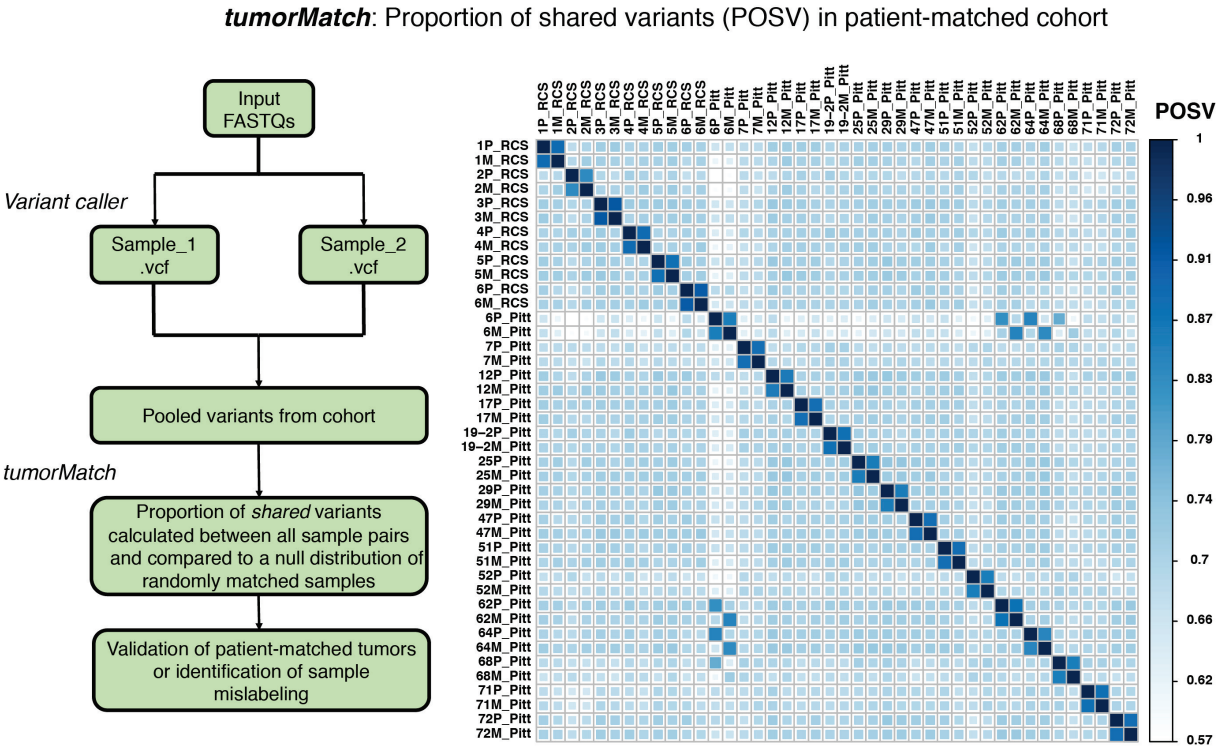

Figure 33: *tumorMatch* in brain metastasis cohort.
tumorMatch workflow diagram and correlation plot showing proportion of shared variants (POSV) values for each sample pairing. Larger and darker squares indicate higher POSV values and identify patient-matched samples.

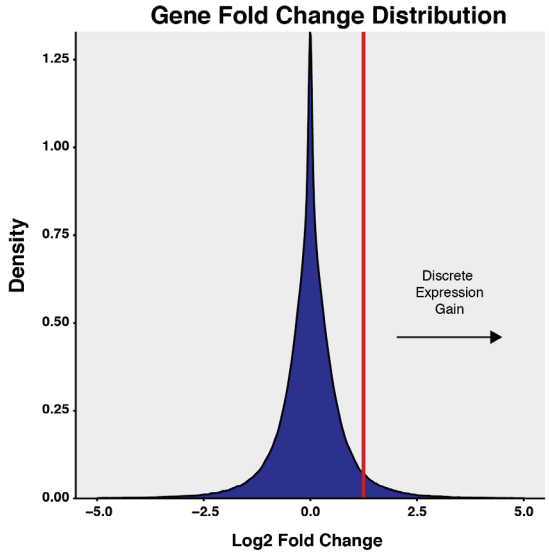

Figure 34: Gene-level fold change density distribution in brain metastasis cohort.
Fold-change value density plot for all cases (i.e. log2normCPM of brain metastases - log2normCPM of matched primary tumors). Genes with fold changes above the 95th percentile (marked with a red line) were considered a discrete, significant "expression gain."

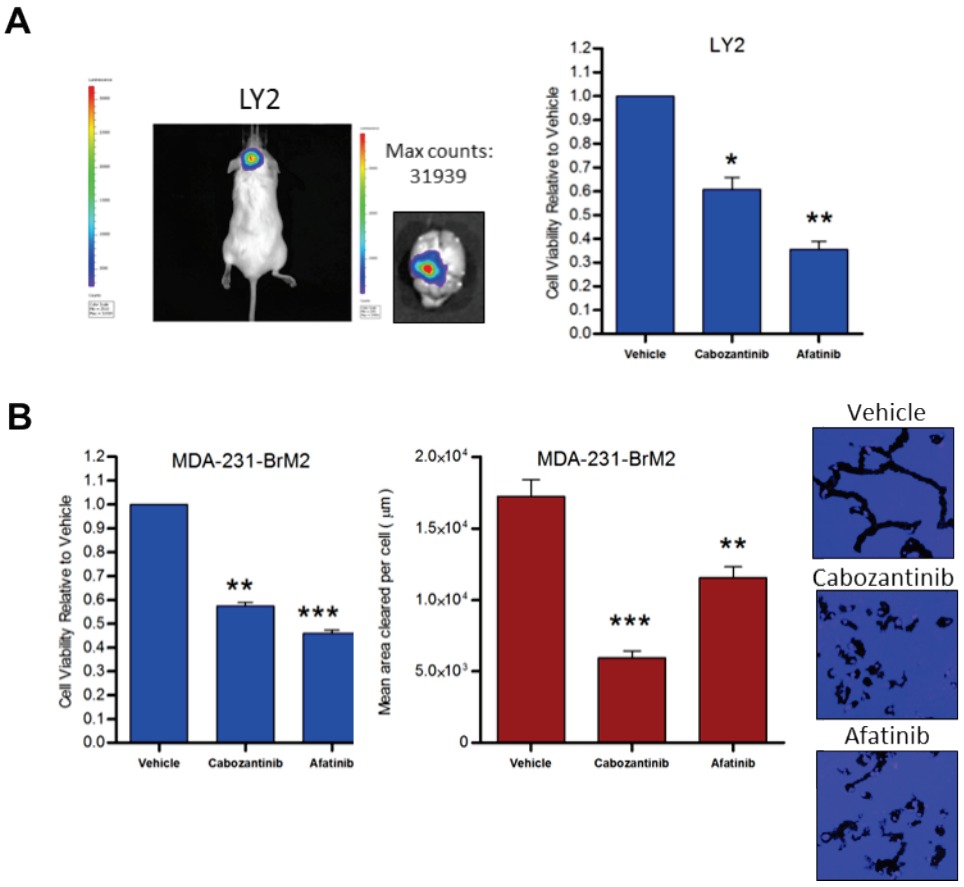

Figure 35: Cabozantinib and afatinib efficacy in MDA-231-BRM2 and LY2 models
(A) Representative *ex vivo* bioluminescence images of a mouse following intracardiac injection of luciferase expressing endocrine resistant breast cancer cells LY2. Bar chart displays significant effect of Cabozantinib (10nM, p=0.0159) and Afatinib (25nM, p=0.0027) on LY2 cell viability when compared to vehicle-treated samples (DMSO). **(B)** Bar chart displays significant effect of Cabozantinib (10nM, p=0.0012) and Afatinib (25n, p=0.0008) on MDA-231-BrM cell viability when compared to vehicle-treated samples (DMSO). Treatment with either Cabozantinib (10nm, p=0.0001) or Afatinib (25nM, p=0.0018) decreases cell motility of MDA-231-BrM cells. Histogram shows mean migratory area per cell (μm^2). Images are representative and show cells stained with DAPI and rhodamine-phalloidin. All error bars represent mean ± S.E.M., n=3 (***p<0.001, **p<0.01,*p<0.05). (Performed by Dr. Damir Vareslija)

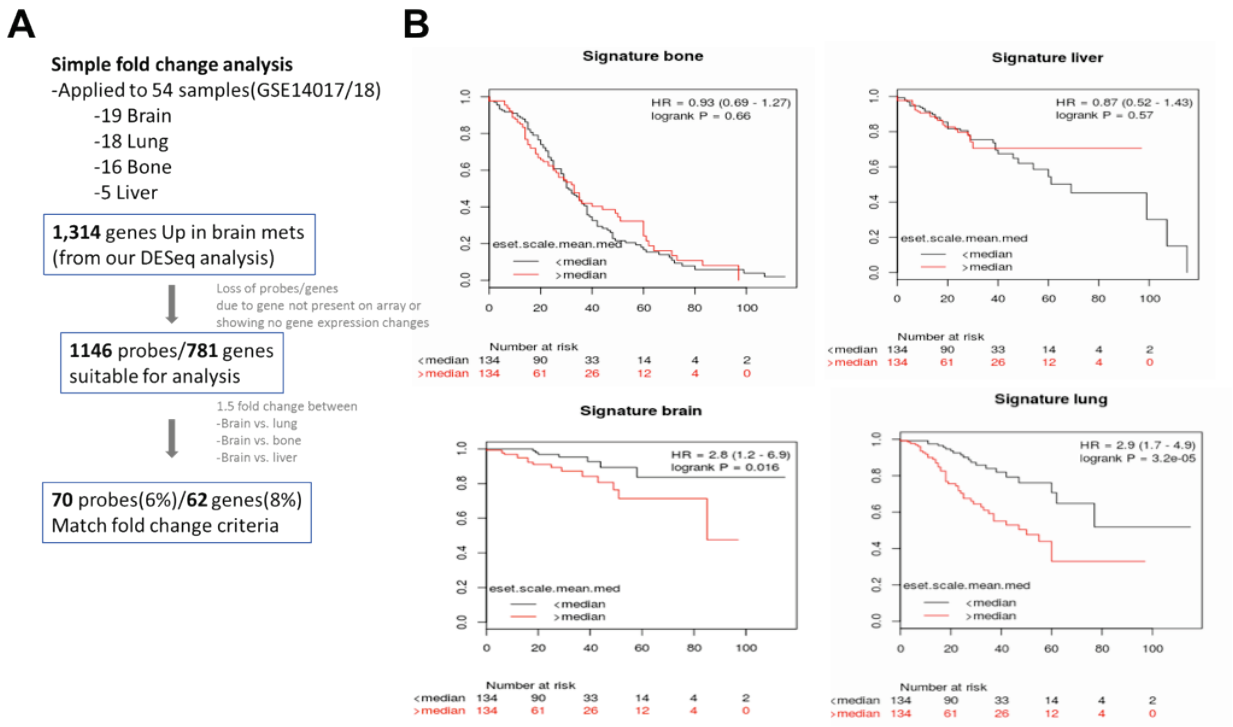

Figure 36: Identified brain metastasis genes predict brain-relapse free survival in primary breast tumors. **(A)** Schematic of the bioinformatic workflow used. Recurrent differentially up-regulated genes (n=1314) were screened in two merged public metastatic cohorts (GSE14017/18). **(B)** Kaplan–Meier curves for bone, liver, lung and brain metastasis-free survival on the basis of BrM-related gene set status in two breast primary cancer cohorts (n=268) (GSE12276/2034). p value based on log rank test. (Performed by Dr. Ailis Fagan)

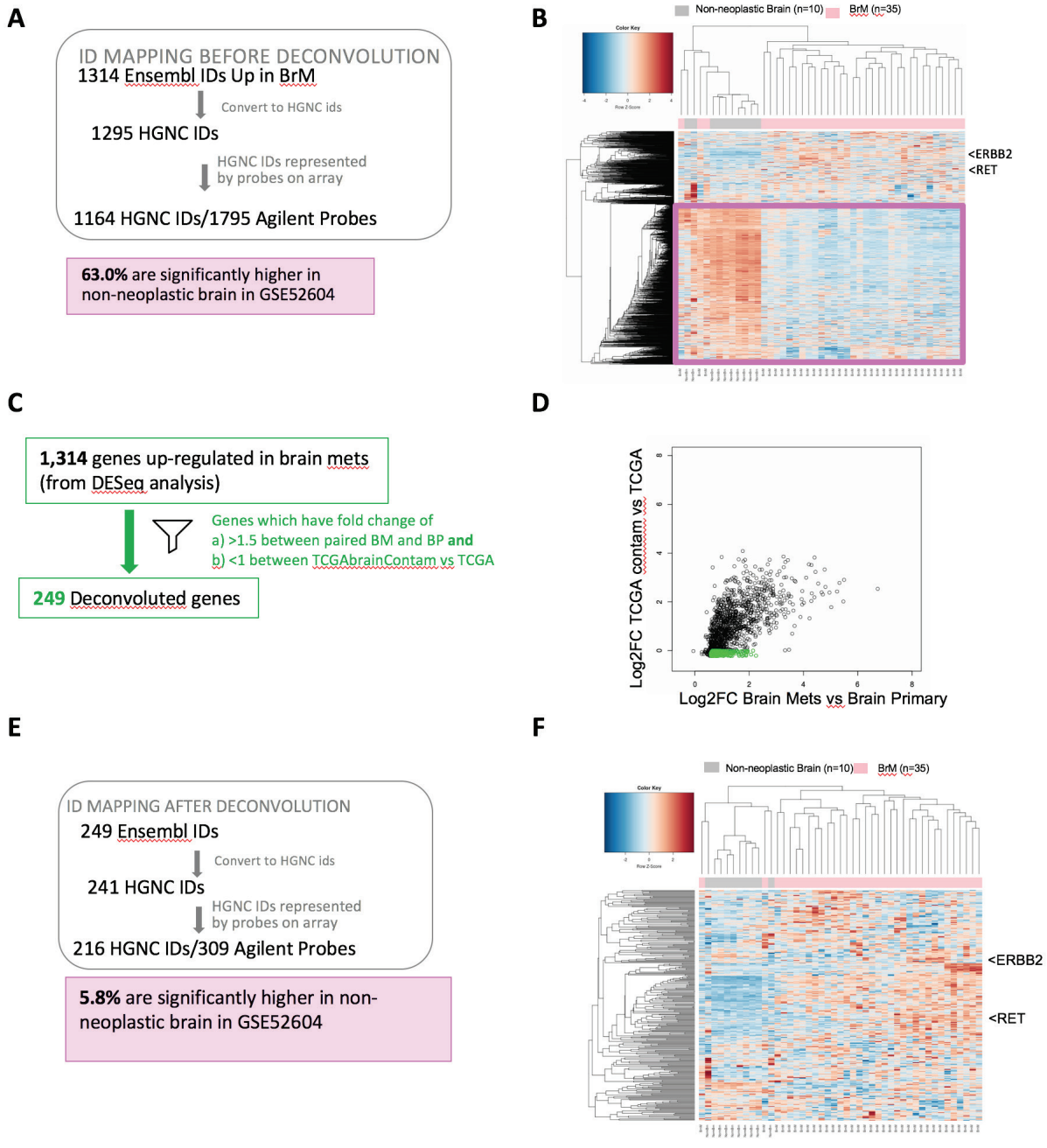

Figure 37: Brain metastasis gene deconvolution
(A) ID mapping workflow from HGNC IDs to GSE52604 probes for genes up regulated in brain metastasis before deconvolution. 63% of the genes up regulated in the BrM model were significantly higher in non-neoplastic brain in GSE52604. **(B)** Heatmap of the 1795 Agilent probes in the GSE52604 dataset with non-neoplastic brain and BrM samples. Highlighted in pink is a cluster of genes that are highly expressed in the non-neoplastic brain relative to brain metastasis. **(C)** Process of retrieving deconvolution genes using the brain contamination model. **(D)** 1314 genes plotted using the log2 fold changes from the experiment model vs. the log2 fold changes contamination model. Deconvoluted brain metastasis genes are highlighted in green. **(E)** Shows the ID mapping workflow for the deconvoluted brain metastasis genes. Of these probes, 5.8% of these were higher in the non-neoplastic brain relative to the brain metastasis in GSE52604. **(F)** Heatmap of the deconvoluted genes.

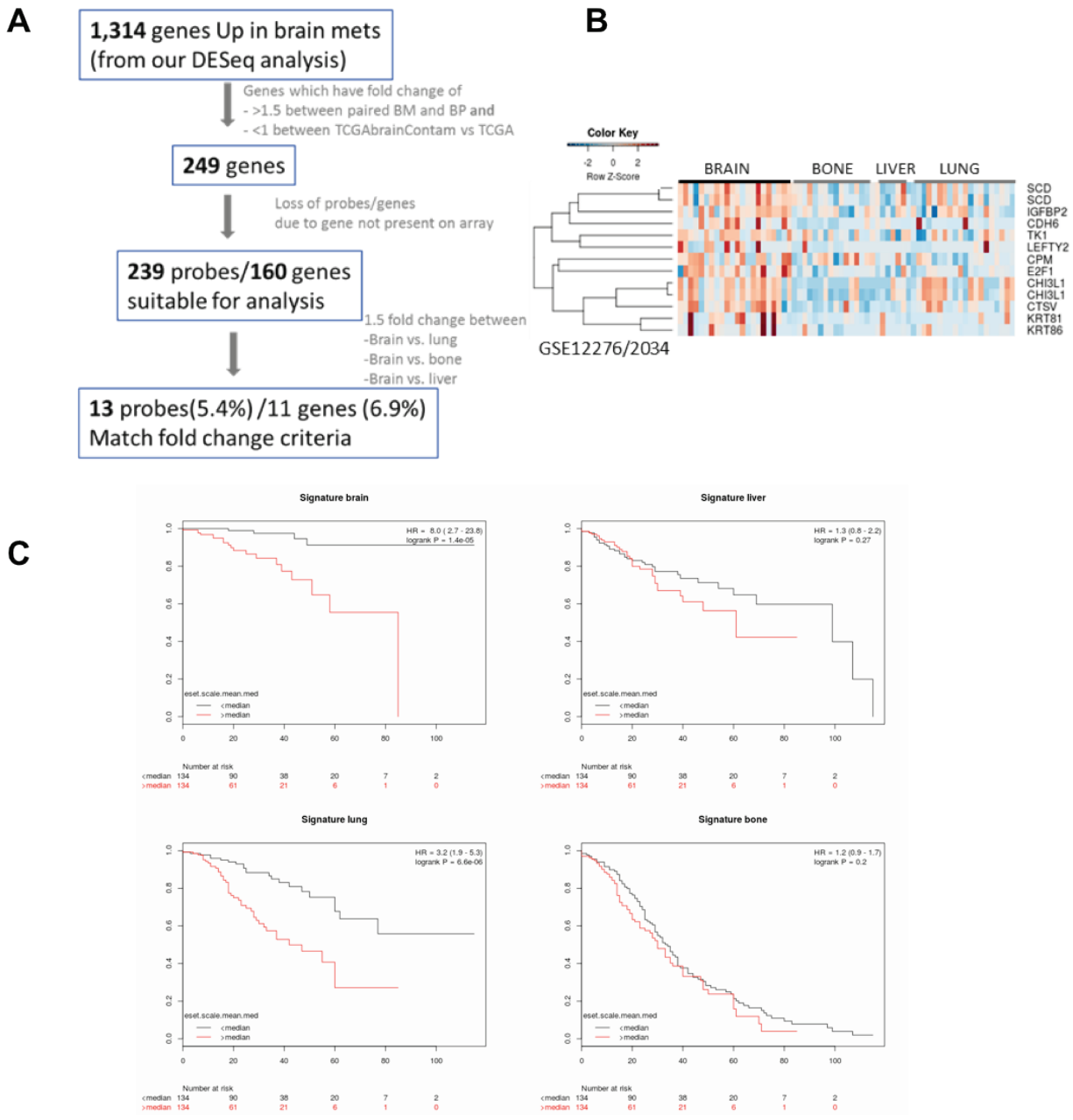

Figure 38: Metastasis-free survival with deconvoluted brain metastasis genes
(A) Recurrent differentially up-regulated genes (n=1314) were filtered to remove potential brain contaminating genes. **(B)** Deconvoluted genes upregulated in BrM were screened in two merged public metastatic cohorts (GSE14017/18). Heatmap displays 11 genes whose expression was 1.5 fold change higher in the mean of brain metastases relative to metastasis to lung, liver, or bone (deconvoluted BrM-related gene set). **(C)** Kaplan–Meier curves for brain, liver, lung and bone metastasis-free survival on the basis of BrM-related gene set expression in two merged breast cancer primary cohorts (n=268) (GSE12276/2034). P-value based on log rank test.

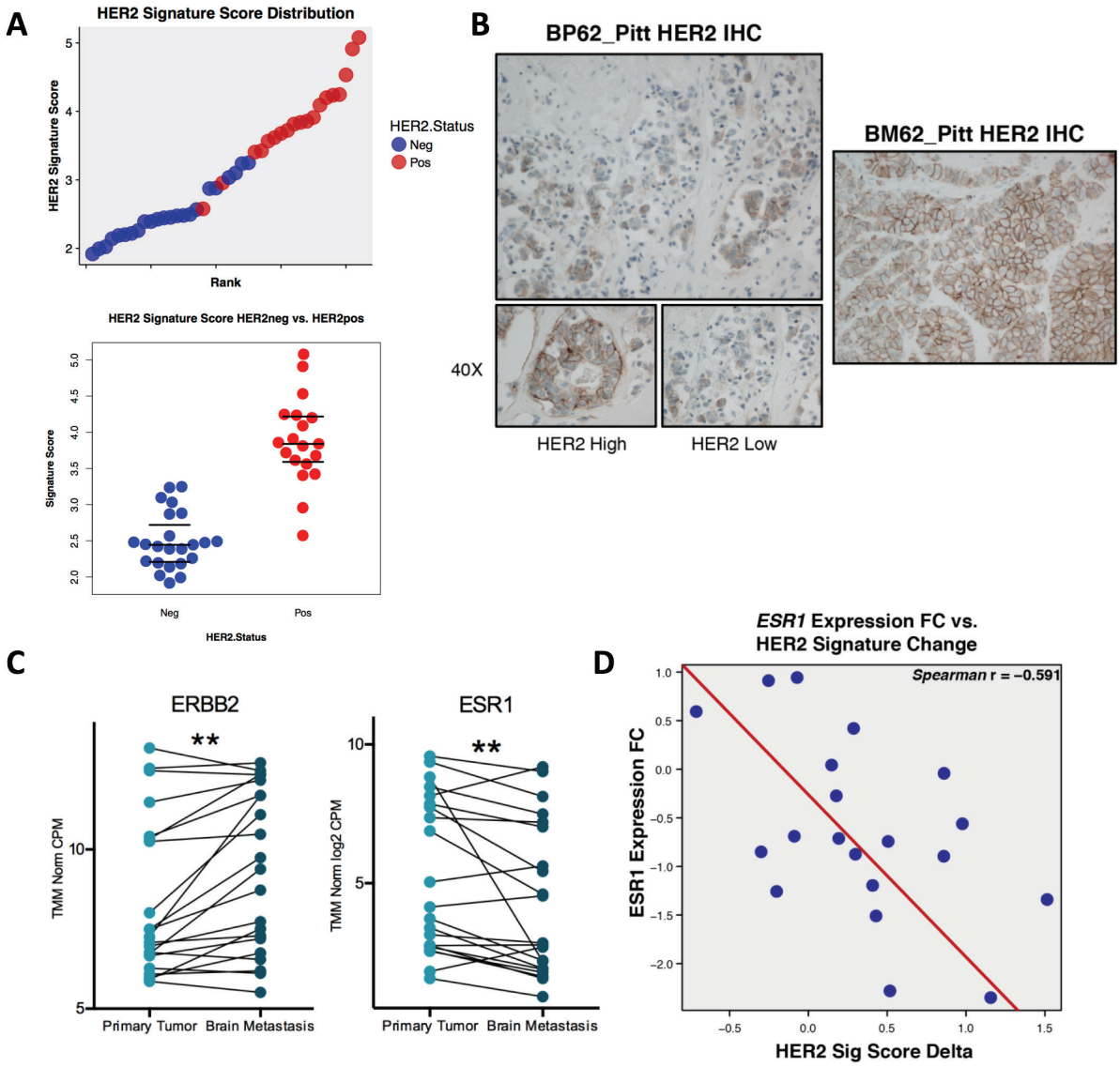

Figure 39: HER2 and ER in breast cancer brain metastases.
(A) HER2 signature score distribution in HER2 positive (red dots) and HER2 negative (blue dots). Scatter plot of HER2 signature compared in HER positive (red dots) and HER2 negative (blue dots). **(B)** Immunohistochemistry staining of HER2 protein case 62_Pitt showing highly heterogeneous areas of HER2 high and low positivity. Also shown is HER2 positivity gained in brain metastases of 62_Pitt. **(C)** Paired ladder plot of ESR1 and ERBB2 expression levels in patient-matched primary and brain metastases (Wilcoxon signed-rank test). Light green dots represent primary tumor expression values and dark green dots represent metastatic tumor expression values. **(D)** Correlation analyses of HER2 signature score versus ESR1 fold change expression (Spearman rho= -0.591).

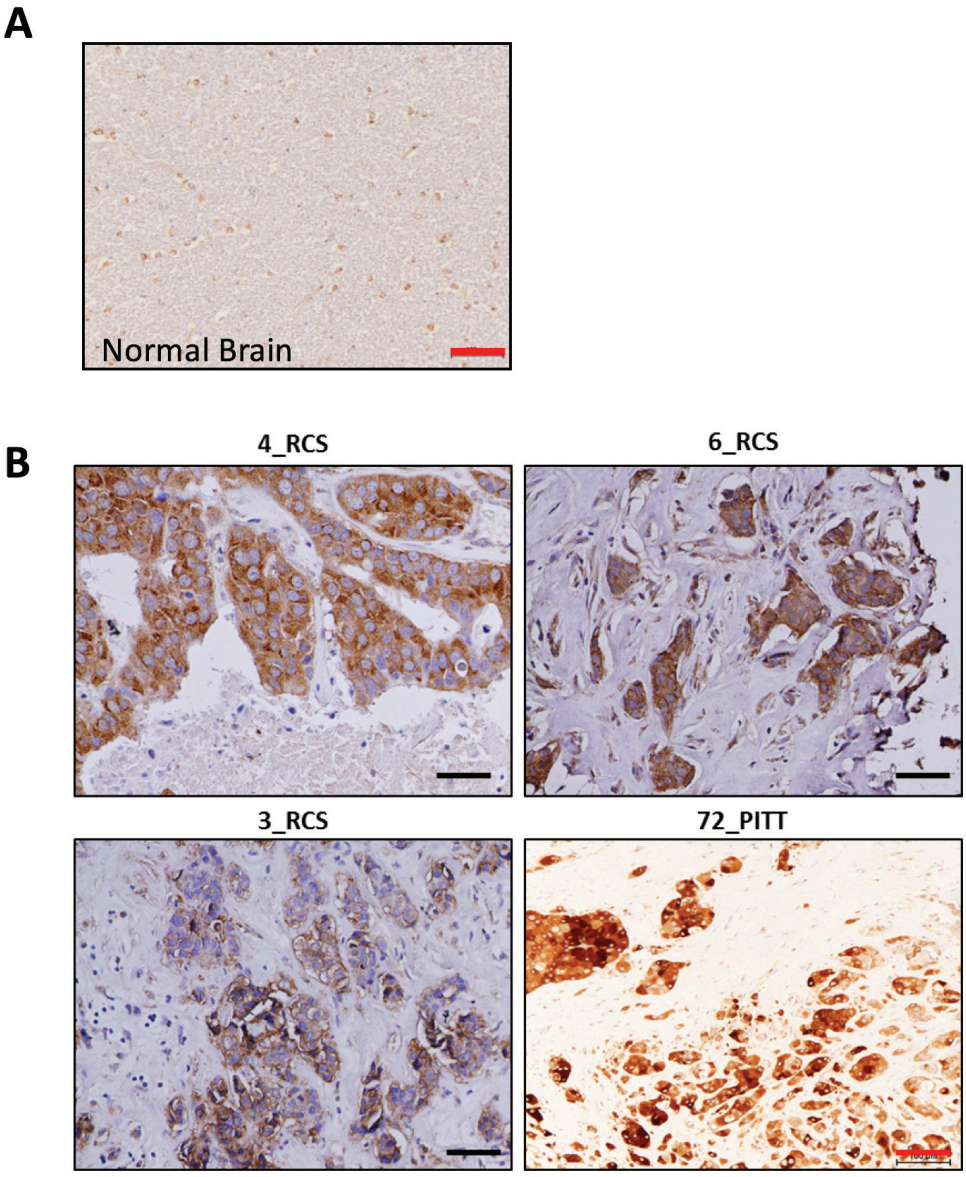

Figure 40: RET protein expression in brain parenchyma and metastases.
(A) RET immunohistochemistry image showing RET protein expression in normal brain. **(B)** 20x images of RET protein expression by immunohistochemistry in representative breast to brain metastases. Images shown are 20x; black scale bars correspond to 50μm and red scale bars correspond to 100μm.

A.4 RECURRENT TRANCRIPTIONAL REMODELING EVENTS IN LONG-TERM ESTROGEN-DEPRIVED BREAST CANCER RECURRENCES: SUPPLEMENTARY FIGURES

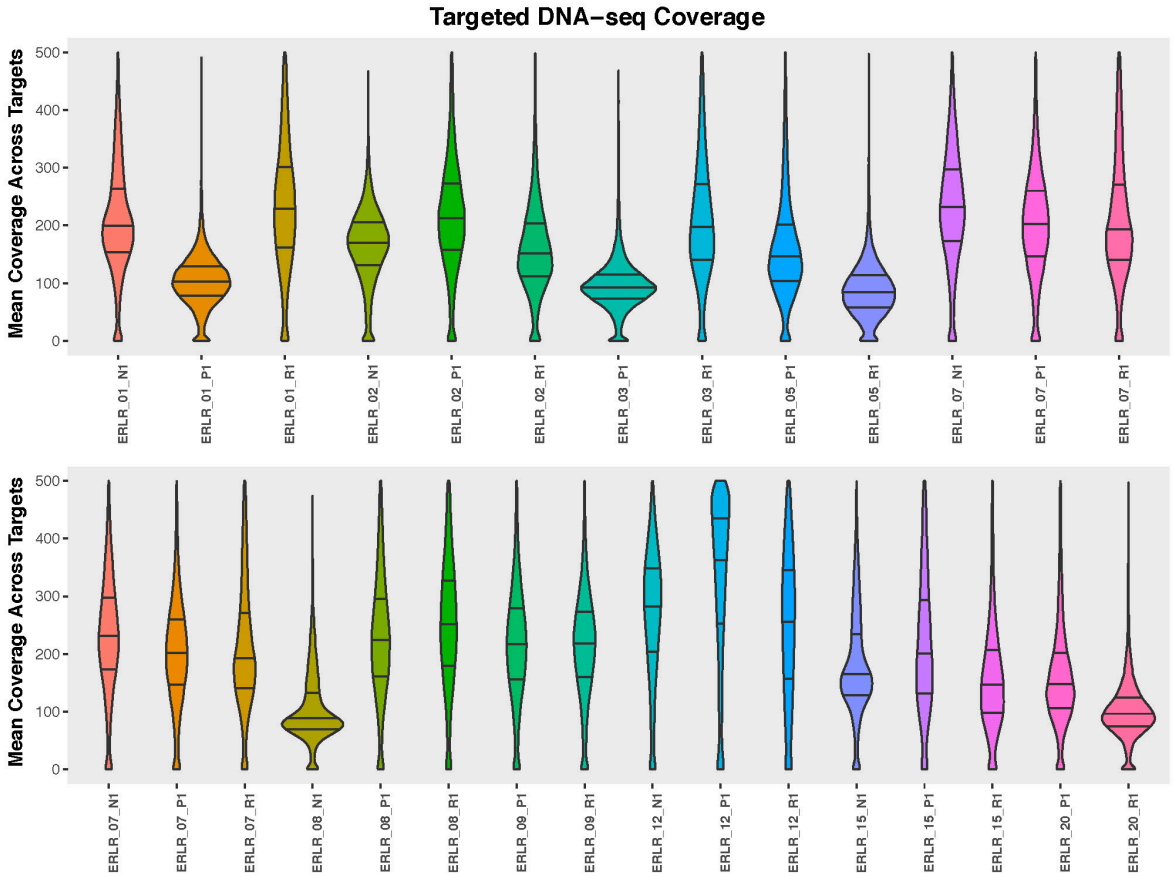

Figure 41: DNA-seq target interval coverages
Violin plots showing the distribution of mean DNA-sequencing coverage across all targeted intervals. 25th, 50th and 75th percentiles are indicated with horizontal black lines. To better visualize distributions, y-axis limit was set at 500.

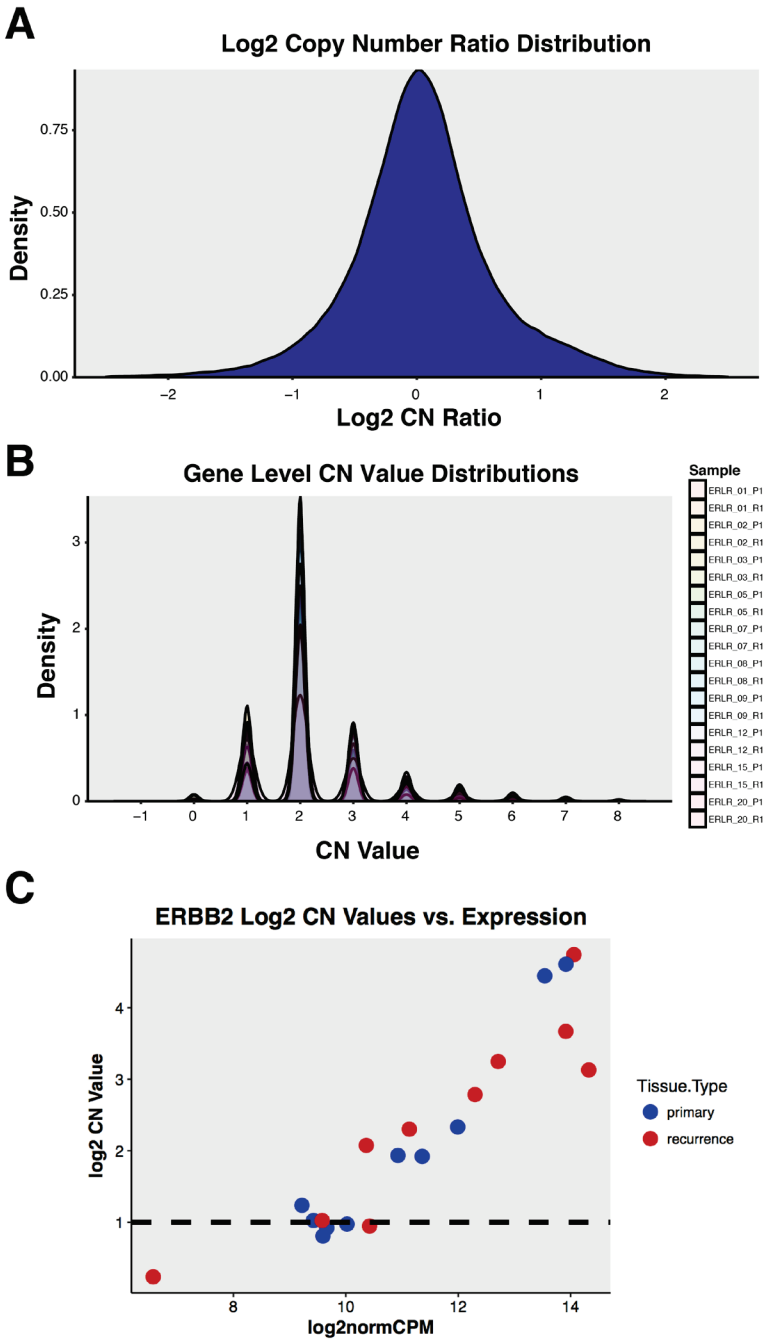

Figure 42: Copy number call distribution and correlation with expression values
(A) Log2 copy number ratio distribution, derived from *CNVkit*, for all samples in cohort. **(B)** Distribution of discrete, gene-level copy number calls with gene-level values representing the mean of discrete calls across all probed regions covering the gene. **(C)** Log2 CN values versus log2normCPM values for *ERBB2* using all samples in the cohort, revealing high correlation (R = 0.924, p-value < 0.001) between calculated CNA calls and expression.

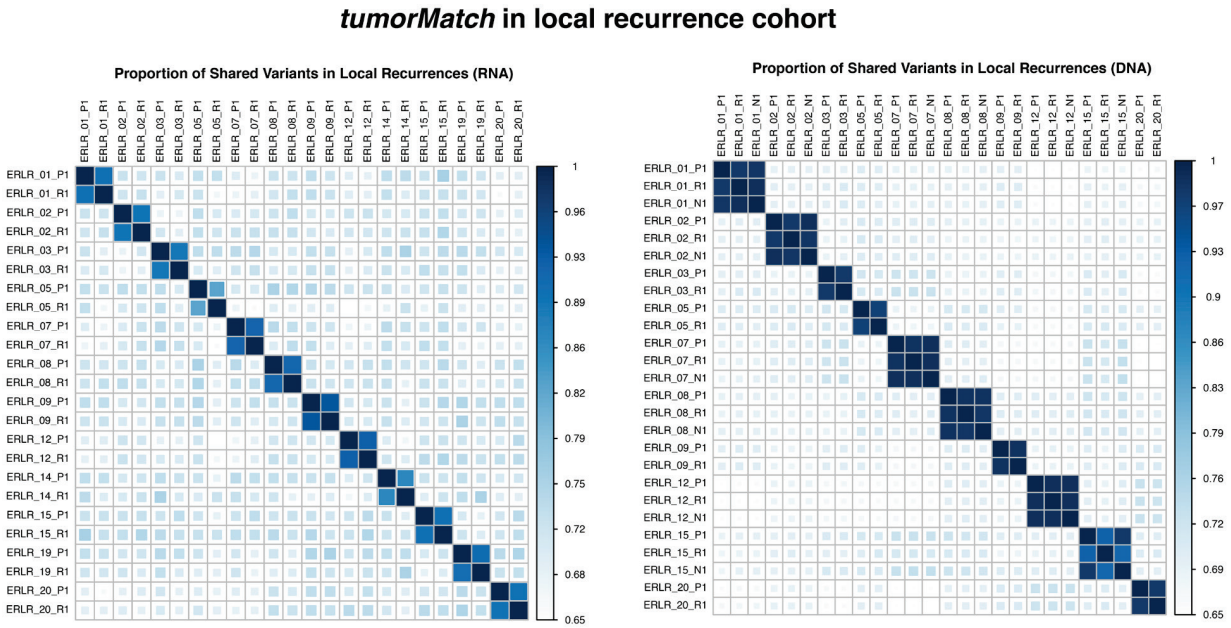

Figure 43: *tumorMatch* in local recurrences cohort
tumorMatch proportion of shared variants correlation plots for both RNA- (left) and DNA-sequencing, showing all paired specimens, including trios that include normal, are patient-matched.

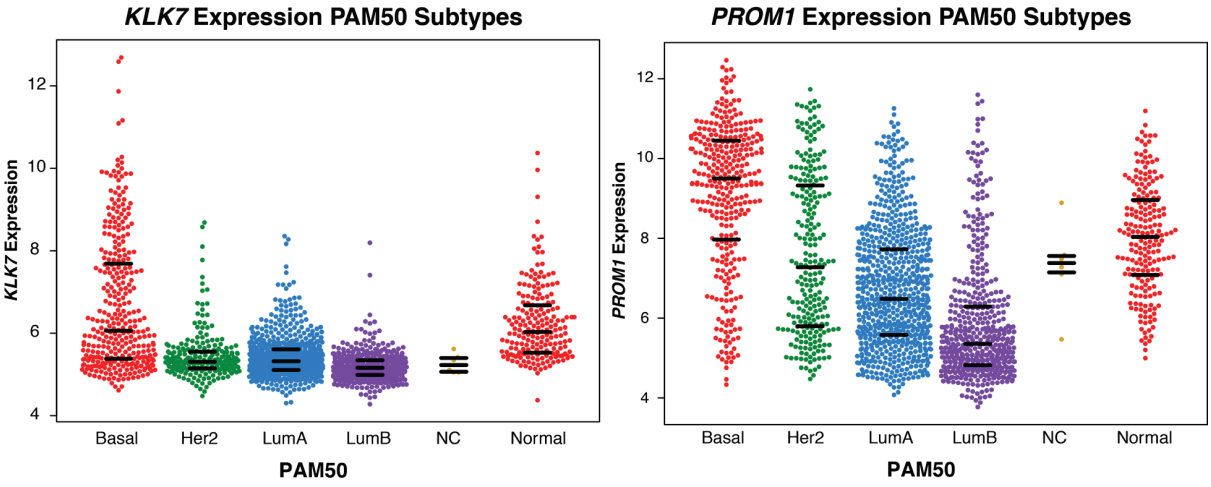

Figure 44: *KLK7* and *PROM1* basal breast carcinoma expression
Normalized microarray expression values (METABRIC) of *KLK7* and *PROM1*, segregated by PAM50 subtypes. Horizontal black bars indicate 25th, 50th and 75th percentile values.

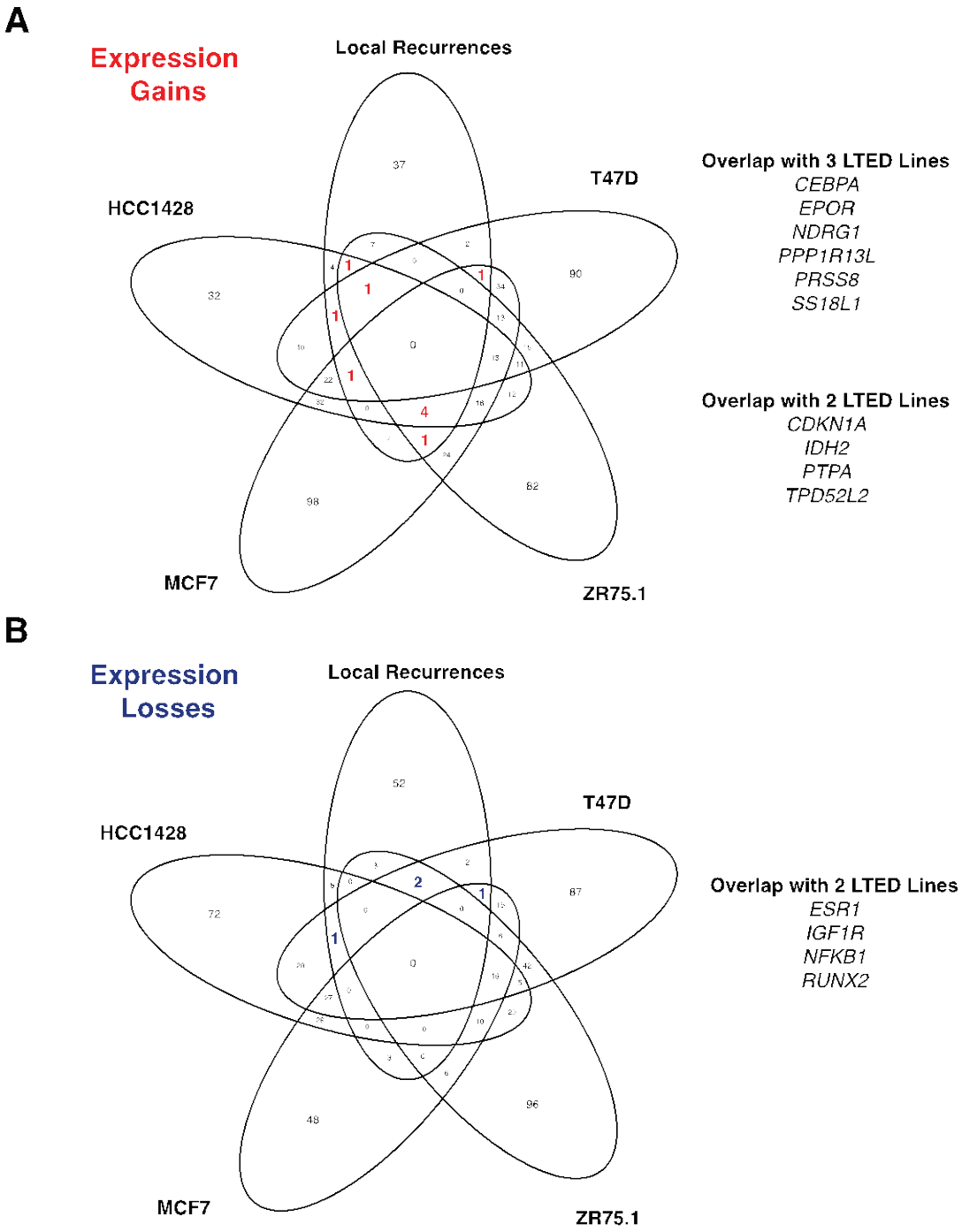

Figure 45: Overlap of differentially expressed genes between local recurrences and ER+ LTED lines
(A) Genes significantly upregulated in both local recurrences vs. primaries and LTED vs. parental lines. **(B)** Genes significantly downregulated in both local recurrences vs. primaries and LTED vs. parental lines.

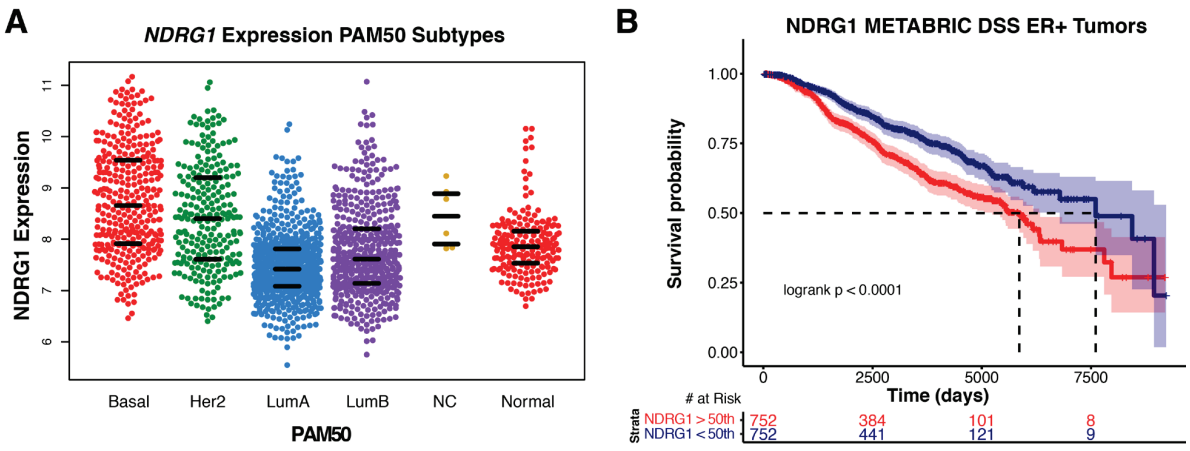

Figure 46: NDRG1 expression in PAM50 subtypes and survival influence in ER-positive breast cancer
(A) *NDRG1* expression in PAM50 subtypes. **(B)** Disease-specific survival in METABRIC of patients with ER-positive primary tumors that express *NDRG1* highly (>50th percentile, red) or lowly (<50th percentile, blue).

A.5 ACQUIRED MOLECULAR FEATURES OF RECURRENT CHEMORESISTANT OVARIAN CANCERS: SUPPLEMENTARY FIGURES

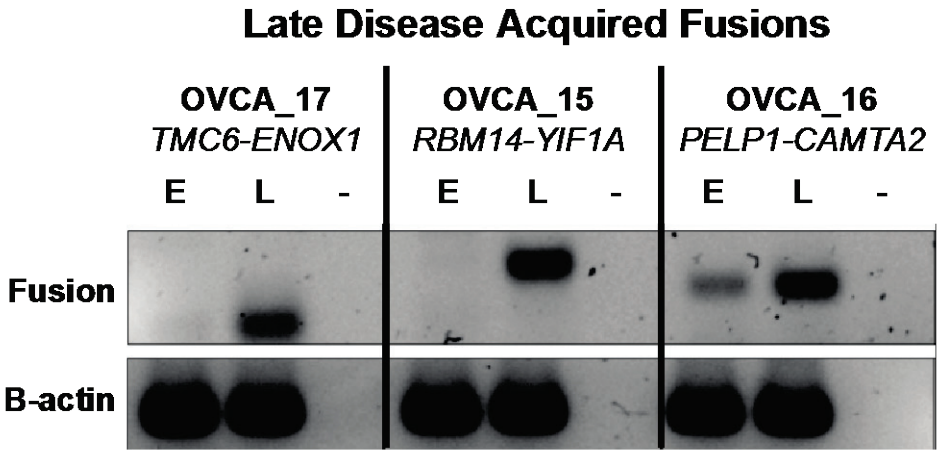

Figure 47: RT-PCR validation of late-disease acquired fusions.
Three bioinformatically called fusion RNAs validated with RT-PCR using fusion breakpoint flanking primer pairs. Case and fusion are indicated, E = early disease sample, L = late disease sample.

APPENDIX B

SUPPLEMENTAL TABLES

Table 8: Multi-gene test classifications in patient-matched pairs
Case along with PAM50 subtype calls, inferred OncoTypeDX score and corresponding clinical risk value. Discordant pairs are marked with an asterisks.

Case	PAM50 Subtype	OncoTypeDX Score	OncoTypeDX Risk
BP_RCS_1	Her2	41.2	1
BM_RCS_1	Her2	55.2	1
BP_RCS_2	LumA	15.6	0
BM_RCS_2	Her2*	60.9	1*
BP_RCS_3	LumB	40.2	1
BM_RCS_3	LumB	56.4	1
BP_RCS_4	LumB	9.8	0
BM_RCS_4	LumA*	3.3	0
BP_RCS_5	Basal	58.6	1
BM_RCS_5	Basal	66.3	1
BP_RCS_6	LumA	-4.9	0
BM_RCS_6	LumA	12.6	0
BP_Pitt_6	Basal	29.9	0.5
BM_Pitt_6	Basal	12.8	0*
BP_Pitt_7	Her2	55.1	1
BM_Pitt_7	Her2	60.0	1
BP_Pitt_12	Basal	32.1	1
BM_Pitt_12	Basal	52.3	1
BP_Pitt_17	LumA	7.5	0
BM_Pitt_17	LumA	25.1	0.5*
BP_Pitt_25	Basal	50.0	1
BM_Pitt_25	Basal	36.9	1

BP_Pitt_29	Basal	66.5	1
BM_Pitt_29	Basal	59.0	1
BP_Pitt_47	LumA	-9.3	0
BM_Pitt_47	Her2*	13.8	0
BP_Pitt_51	LumB	33.5	1
BM_Pitt_51	LumB	28.9	0.5*
BP_Pitt_52	Her2	39.5	1
BM_Pitt_52	Her2	39.4	1
BP_Pitt_62	LumB	14.6	0
BM_Pitt_62	LumB	31.4	1*
BP_Pitt_64	Basal	51.5	1
BM_Pitt_64	Basal	46.5	1
BP_Pitt_68	Basal	53.0	1
BM_Pitt_68	Basal	45.2	1
BP_Pitt_71	Basal	48.9	1
BM_Pitt_71	Basal	53.5	1
BP_Pitt_72	LumA	-2.3	0
BM_Pitt_72	LumA	0.1	0

Table 9: Fusion validation primers

Primer	Sequence
ACTB_F1	AGCCTCGCCTTTGCCGA
ACTB_R1	CTGGTGCCTGGGGCG
TOP2A_STAU1_F1	TCCCTTCTATGGTGGATGGT
TOP2A_STAU1_R1	CCACCTCGAAATTCACAGGC
TOP2A_STAU1_F2	CGAGAAGTAAAGGTTGCCCAAT
TOP2A_STAU1_R2	TCACAGGCAAGTTCCGTTTAAG
TMC6_ENOX1_F1	CGAGACCTCAGTTCCCGG
TMC6_ENOX1_R1	TGTTCTCTCGGTCTTGGTTGA
RBM14_YIF1A_F1	GATTTTCGTGGGCAATGTGTCG
RBM14_YIF1A_R1	CTCCTGTGGCTGGGTATCC
PELP1_CAMTA2_F1	TTTGCAGACTGGGAAGCCTA
PELP1_CAMTA2_R1	CTGCTCCAGTCGCTCTAGTAT
CCDC6_ANK3_F1	CAAGAGAACAAGGTGCTGAAGA
CCDC6_ANK3_R1	TACGAGTGGCTCTTCTTTTCCA
CCDC6_ANK3_F2	AAAGCCGAACTAGAACAGCATC
CCDC6_ANK3_R2	TCCGAGACTAAAGCCCATGTAA
CCDC6_ANK3_F3	AGCTGGAGACCTACAAACTGAA

APPENDIX C

DATA SUPPLEMENTS

Data Supplement 1: INTRINSIC SUBTYPE SWITCHING AND HER2 GAINS IN BREAST CANCER BRAIN METASTASES

Data Supplement 2: EXOME-CAPTURE RNA-SEQUENCING OF DECADE-OLD BREAST CANCERS AND MATCHED DECALCIFIED BONE METASTASES IDENTIFIES CLINICALLY ACTIONABLE TARGETS

Data Supplement 3: TRANSCRIPTOME-WIDE IDENTIFICATION OF RET AND HER2 SIGNALING AS RECURRENTLY ENRICHED DEPENDENCIES IN BREAST CANCER BRAIN METASTASES

Data Supplement 4: RECURRENT TRANSCRIPTIONAL REMODELING EVENTS IN LONG-TERM ESTROGEN-DEPRIVED BREAST CANCER RECURRENCES

Data Supplement 5: ACQUIRED MOLECULAR FEATURES IN RECURRENT CHEMORESISTANT OVARIAN CANCERS

REFERENCES

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA Cancer J Clin* **64,** 9–29 (2014).

2. Cancer Statistics Review, 1975-2014 - SEER Statistics. at <https://seer.cancer.gov/csr/1975_2014/>

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

4. Jensen, E. V., Block, G. E., Smith, S., Kyser, K. & DeSombre, E. R. Estrogen receptors and breast cancer response to adrenalectomy. *Natl Cancer Inst Monogr* **34,** 55–70 (1971).

5. Slamon, D. J. *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244,** 707–712 (1989).

6. Redig, A. J. & Jänne, P. A. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J Clin Oncol* **33,** 975–977 (2015).

7. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98,** 10869–10874 (2001).

8. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100,** 8418–8423 (2003).

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

10. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27,** 1160–1167 (2009).

11. Sparano, J. A. *et al.* Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med* **373,** 2005–2014 (2015).

12. Cardoso, F. *et al.* 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* **375,** 717–729 (2016).

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

14. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–352 (2012).

15. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7,** 11479 (2016).

16. Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121,** 2750–2767 (2011).

17. Prat, A. *et al.* Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* **18,** 123–133 (2013).

18. Giordano, S. H. *et al.* Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol* **32,** 2078–2099 (2014).

19. Denduluri, N. *et al.* Selection of Optimal Adjuvant Chemotherapy Regimens for Human Epidermal Growth Factor Receptor 2 (HER2) -Negative and Adjuvant Targeted Therapy for HER2-Positive Breast Cancers: An American Society of Clinical Oncology Guideline Adaptation of the Cancer Care Ontario Clinical Practice Guideline. *J Clin Oncol* **34,** 2416–2427 (2016).

20. Burstein, H. J. *et al.* Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: american society of clinical oncology clinical practice guideline focused update. *J Clin Oncol* **32,** 2255–2269 (2014).

21. Van Poznak, C. *et al.* Use of biomarkers to guide decisions on systemic therapy for women with metastatic breast cancer: american society of clinical oncology clinical practice guideline. *J Clin Oncol* **33,** 2695–2704 (2015).

22. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38,** 1289–1297 (2006).

23. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103,** 843–852 (2000).

24. Oñate, S. A., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270,** 1354–1357 (1995).

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

26. Porter, W., Saville, B., Hoivik, D. & Safe, S. Functional synergy between the transcription

factor Sp1 and the estrogen receptor. *Mol Endocrinol* **11,** 1569–1580 (1997).

27. Migliaccio, A. *et al.* Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J* **15,** 1292–1300 (1996).

28. Kahlert, S. *et al.* Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. *J Biol Chem* **275,** 18447–18453 (2000).

29. Castoria, G. *et al.* PI3-kinase in concert with Src promotes the S-phase entry of oestradiol-stimulated MCF-7 cells. *EMBO J* **20,** 6050–6059 (2001).

30. Song, R. X.-D. *et al.* Linkage of rapid estrogen action to MAPK activation by ERalpha-Shc association and Shc pathway activation. *Mol Endocrinol* **16,** 116–127 (2002).

31. Kato, S. *et al.* Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270,** 1491–1494 (1995).

32. Bunone, G., Briand, P. A., Miksicek, R. J. & Picard, D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J* **15,** 2174–2183 (1996).

33. Jordan, V. C. Effect of tamoxifen (ICI 46,474) on initiation and growth of DMBA-induced rat mammary carcinomata. *European Journal of Cancer (1965)* **12,** 419–424 (1976).

34. Harper, M. J. & Walpole, A. L. A new derivative of triphenylethylene: effect on implantation and mode of action in rats. *J Reprod Fertil* **13,** 101–119 (1967).

35. Turner, R. T., Wakley, G. K., Hannon, K. S. & Bell, N. H. Tamoxifen prevents the skeletal effects of ovarian hormone deficiency in rats. *J Bone Miner Res* **2,** 449–456 (1987).

36. Gottardis, M. M., Robinson, S. P., Satyaswaroop, P. G. & Jordan, V. C. Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. *Cancer Res* **48,** 812–815 (1988).

37. McDonnell, D. P., Clemm, D. L., Hermann, T., Goldman, M. E. & Pike, J. W. Analysis of estrogen receptor function in vitro reveals three distinct classes of antiestrogens. *Mol Endocrinol* **9,** 659–669 (1995).

38. Fawell, S. E. *et al.* Inhibition of estrogen receptor-DNA binding by the "pure" antiestrogen ICI 164,384 appears to be mediated by impaired receptor dimerization. *Proc Natl Acad Sci U S A* **87,** 6883–6887 (1990).

39. Dauvois, S., White, R. & Parker, M. G. The antiestrogen ICI 182780 disrupts estrogen receptor nucleocytoplasmic shuttling. *J Cell Sci* **106 (Pt 4),** 1377–1388 (1993).

40. Osborne, C. K. *et al.* Comparison of the Effects of a Pure Steroidal antiestrogen With

Those of Tamoxifen in a Model of Human Breast Cancer. *JNCI Journal of the National Cancer Institute* **87,** 746–750 (1995).

41. 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,** 146–159 (2016).

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

43. Bodai, B. I. & Tuso, P. Breast cancer survivorship: a comprehensive review of long-term medical issues and lifestyle recommendations. *Perm J* **19,** 48–79 (2015).

44. Osborne, C. K. & Schiff, R. Mechanisms of endocrine resistance in breast cancer. *Annu Rev Med* **62,** 233–247 (2011).

45. Ades, F. *et al.* Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *J Clin Oncol* **32,** 2794–2803 (2014).

46. Kuukasjärvi, T., Kononen, J., Helin, H., Holli, K. & Isola, J. Loss of estrogen receptor in recurrent breast cancer is associated with poor response to endocrine therapy. *J Clin Oncol* **14,** 2584–2589 (1996).

47. Oh, A. S. *et al.* Hyperactivation of MAPK induces loss of ERalpha expression in breast cancer cells. *Mol Endocrinol* **15,** 1344–1359 (2001).

48. Creighton, C. J. *et al.* Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res* **66,** 3903–3911 (2006).

49. Shou, J. *et al.* Mechanisms of Tamoxifen Resistance: Increased Estrogen Receptor-HER2/neu Cross-Talk in ER/HER2-Positive Breast Cancer. *JNCI Journal of the National Cancer Institute* **96,** 926–935 (2004).

50. Turner, N. *et al.* FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* **70,** 2085–2094 (2010).

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

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

53. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol* **12,** 573–583 (2015).

54. Fribbens, C. *et al.* Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *J Clin Oncol* **34,** 2961–2968 (2016).

55. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* **4,** 1116–1130 (2013).

56. Ma, C. X., Reinert, T., Chmielewska, I. & Ellis, M. J. Mechanisms of aromatase inhibitor resistance. *Nat Rev Cancer* **15,** 261–275 (2015).

57. Gutierrez, C. & Schiff, R. HER2: biology, detection, and clinical implications. *Arch Pathol Lab Med* **135,** 55–62 (2011).

58. Hudis, C. A. Trastuzumab--mechanism of action and use in clinical practice. *N Engl J Med* **357,** 39–51 (2007).

59. Carter, P. *et al.* Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci U S A* **89,** 4285–4289 (1992).

60. 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,** 4744–4749 (2001).

61. Yakes, F. M. *et al.* Herceptin-induced inhibition of phosphatidylinositol-3 kinase and Akt Is required for antibody-mediated effects on p27, cyclin D1, and antitumor action. *Cancer Res* **62,** 4132–4141 (2002).

62. Piccart-Gebhart, M. J. *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* **353,** 1659–1672 (2005).

63. Klapper, L. N., Waterman, H., Sela, M. & Yarden, Y. Tumor-inhibitory antibodies to HER-2/ErbB-2 may act by recruiting c-Cbl and enhancing ubiquitination of HER-2. *Cancer Res* **60,** 3384–3388 (2000).

64. Adams, C. W. *et al.* Humanization of a recombinant monoclonal antibody to produce a therapeutic HER dimerization inhibitor, pertuzumab. *Cancer Immunol Immunother* **55,** 717–727 (2006).

65. Franklin, M. C. *et al.* Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5,** 317–328 (2004).

66. Scheuer, W. *et al.* Strongly enhanced antitumor activity of trastuzumab and pertuzumab combination treatment on HER2-positive human xenograft tumor models. *Cancer Res* **69,** 9330–9336 (2009).

67. Swain, S. M. *et al.* Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic

breast cancer. *N Engl J Med* **372,** 724–734 (2015).

68. 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).

69. Rabindran, S. K. *et al.* Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. *Cancer Res* **64,** 3958–3965 (2004).

70. Pivot, X. *et al.* CEREBEL (EGF111438): A Phase III, Randomized, Open-Label Study of Lapatinib Plus Capecitabine Versus Trastuzumab Plus Capecitabine in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer. *J Clin Oncol* **33,** 1564–1573 (2015).

71. Park, J. W. *et al.* Adaptive randomization of neratinib in early breast cancer. *N Engl J Med* **375,** 11–22 (2016).

72. Venur, V. A. & Leone, J. P. Targeted Therapies for Brain Metastases from Breast Cancer. *Int J Mol Sci* **17,** (2016).

73. Scaltriti, M. *et al.* Clinical benefit of lapatinib-based therapy in patients with human epidermal growth factor receptor 2-positive breast tumors coexpressing the truncated p95HER2 receptor. *Clin Cancer Res* **16,** 2688–2695 (2010).

74. Mitra, D. *et al.* An oncogenic isoform of HER2 associated with locally disseminated breast cancer and trastuzumab resistance. *Mol Cancer Ther* **8,** 2152–2162 (2009).

75. Hanker, A. B. *et al.* An Acquired HER2(T798I) Gatekeeper Mutation Induces Resistance to Neratinib in a Patient with HER2 Mutant-Driven Breast Cancer. *Cancer Discov* (2017). doi:10.1158/2159-8290.CD-16-1431

76. 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* (2017). doi:10.1158/1078-0432.CCR-16-2191

77. Majewski, I. J. *et al.* PIK3CA mutations are associated with decreased benefit to neoadjuvant human epidermal growth factor receptor 2-targeted therapies in breast cancer. *J Clin Oncol* **33,** 1334–1339 (2015).

78. 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).

79. Dave, B. *et al.* Loss of phosphatase and tensin homolog or phosphoinositol-3 kinase activation and response to trastuzumab or lapatinib in human epidermal growth factor receptor 2-overexpressing locally advanced breast cancers. *J Clin Oncol* **29,** 166–173 (2011).

80. Ritter, C. A. *et al.* Human breast cancer cells selected for resistance to trastuzumab in vivo overexpress epidermal growth factor receptor and ErbB ligands and remain dependent on the ErbB receptor network. *Clin Cancer Res* **13,** 4909–4919 (2007).

81. Lu, Y., Zi, X., Zhao, Y., Mascarenhas, D. & Pollak, M. Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). *J Natl Cancer Inst* **93,** 1852–1857 (2001).

82. Sergina, N. V. *et al.* Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445,** 437–441 (2007).

83. Huang, X. *et al.* Heterotrimerization of the growth factor receptors erbB2, erbB3, and insulin-like growth factor-i receptor in breast cancer cells resistant to herceptin. *Cancer Res* **70,** 1204–1214 (2010).

84. Loi, S. *et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol* **25,** 1544–1550 (2014).

85. Perez, E. A. *et al.* Genomic analysis reveals that immune function genes are strongly linked to clinical outcome in the North Central Cancer Treatment Group n9831 Adjuvant Trastuzumab Trial. *J Clin Oncol* **33,** 701–708 (2015).

86. Wang, Y.-C. *et al.* Different mechanisms for resistance to trastuzumab versus lapatinib in HER2-positive breast cancers--role of estrogen receptor and HER2 reactivation. *Breast Cancer Res* **13,** R121 (2011).

87. Goel, S. *et al.* Overcoming Therapeutic Resistance in HER2-Positive Breast Cancers with CDK4/6 Inhibitors. *Cancer Cell* **29,** 255–269 (2016).

88. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) *et al.* Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *The Lancet* **379,** 432–444 (2012).

89. Sledge, G. W. *et al.* Past, present, and future challenges in breast cancer treatment. *J Clin Oncol* **32,** 1979–1986 (2014).

90. Baselga, J. *et al.* Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* **366,** 520–529 (2012).

91. André, F. *et al.* Everolimus for women with trastuzumab-resistant, HER2-positive, advanced breast cancer (BOLERO-3): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Oncol* **15,** 580–591 (2014).

92. Piccart, M. *et al.* Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* **25,** 2357–2362 (2014).

93. André, F. *et al.* Molecular Alterations and Everolimus Efficacy in Human Epidermal Growth Factor Receptor 2-Overexpressing Metastatic Breast Cancers: Combined Exploratory Biomarker Analysis From BOLERO-1 and BOLERO-3. *J Clin Oncol* **34,** 2115–2124 (2016).

94. Abramson, V. G. *et al.* Phase Ib Study of Safety and Pharmacokinetics of the PI3K Inhibitor SAR245408 with the HER3-Neutralizing Human Antibody SAR256212 in Patients with Solid Tumors. *Clin Cancer Res* (2016). doi:10.1158/1078-0432.CCR-16-1764

95. Mayer, I. A. *et al.* A Phase Ib Study of Alpelisib (BYL719), a PI3Kα-Specific Inhibitor, with Letrozole in ER+/HER2- Metastatic Breast Cancer. *Clin Cancer Res* **23,** 26–34 (2017).

96. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

97. O'Leary, B., Finn, R. S. & Turner, N. C. Treating cancer with selective CDK4/6 inhibitors. *Nat Rev Clin Oncol* **13,** 417–430 (2016).

98. Gillett, C. *et al.* Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. *Cancer Res* **54,** 1812–1817 (1994).

99. 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,** 2069–2076 (1999).

100. Zwijsen, R. M. *et al.* CDK-independent activation of estrogen receptor by cyclin D1. *Cell* **88,** 405–415 (1997).

101. Castro-Rivera, E., Samudio, I. & Safe, S. Estrogen regulation of cyclin D1 gene expression in ZR-75 breast cancer cells involves multiple enhancer elements. *J Biol Chem* **276,** 30853–30861 (2001).

102. Neuman, E. *et al.* Cyclin D1 stimulation of estrogen receptor transcriptional activity independent of cdk4. *Mol Cell Biol* **17,** 5338–5347 (1997).

103. Zwijsen, R. M., Buckle, R. S., Hijmans, E. M., Loomans, C. J. & Bernards, R. Ligand-independent recruitment of steroid receptor coactivators to estrogen receptor by cyclin D1. *Genes Dev* **12,** 3488–3498 (1998).

104. Kim, S. *et al.* Abstract PR02: LEE011: An orally bioavailable, selective small molecule

inhibitor of CDK4/6- Reactivating Rb in cancer. *Mol Cancer Ther* **12,** PR02–PR02 (2013).

105. 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).

106. Turner, N. C. *et al.* Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer. *N Engl J Med* **373,** 209–219 (2015).

107. Finn, R. S. *et al.* Palbociclib and letrozole in advanced breast cancer. *N Engl J Med* **375,** 1925–1936 (2016).

108. Hortobagyi, G. N. *et al.* Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N Engl J Med* **375,** 1738–1748 (2016).

109. Herrera-Abreu, M. T. *et al.* Early Adaptation and Acquired Resistance to CDK4/6 Inhibition in Estrogen Receptor-Positive Breast Cancer. *Cancer Res* **76,** 2301–2313 (2016).

110. Brewster, A. M. *et al.* Residual risk of breast cancer recurrence 5 years after adjuvant therapy. *J Natl Cancer Inst* **100,** 1179–1183 (2008).

111. Davies, C. *et al.* Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *The Lancet* **381,** 805–816 (2013).

112. Weigelt, B., Peterse, J. L. & van 't Veer, L. J. Breast cancer metastasis: markers and models. *Nat Rev Cancer* **5,** 591–602 (2005).

113. Tevaarwerk, A. J. *et al.* Survival in patients with metastatic recurrent breast cancer after adjuvant chemotherapy: little evidence of improvement over the past 30 years. *Cancer* **119,** 1140–1148 (2013).

114. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9,** 274–284 (2009).

115. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–715 (2008).

116. Gonzalez, D. M. & Medici, D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal* **7,** re8 (2014).

117. Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nat Rev Cancer* **9,** 239–252 (2009).

118. Nieswandt, B., Hafner, M., Echtenacher, B. & Männel, D. N. Lysis of tumor cells by natural killer cells in mice is impeded by platelets. *Cancer Res* **59,** 1295–1300 (1999).

119. Palumbo, J. S. *et al.* Platelets and fibrin(ogen) increase metastatic potential by impeding natural killer cell-mediated elimination of tumor cells. *Blood* **105,** 178–185 (2005).

120. Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20,** 576–590 (2011).

121. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339,** 580–584 (2013).

122. Reymond, N., Água, B. B. d' & Ridley, A. J. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* **13,** 858–870 (2013).

123. Spiegel, A. *et al.* Neutrophils Suppress Intraluminal NK Cell-Mediated Tumor Cell Clearance and Enhance Extravasation of Disseminated Carcinoma Cells. *Cancer Discov* **6,** 630–649 (2016).

124. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351,** 781–791 (2004).

125. Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* **32,** 479–484 (2014).

126. De Luca, F. *et al.* Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. *Oncotarget* **7,** 26107–26119 (2016).

127. Fidler, I. J. Metastasis: quantitative analysis of distribution and fate of tumor emboli labeled with 125 I-5-iodo-2'-deoxyuridine. *J Natl Cancer Inst* **45,** 773–782 (1970).

128. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* **153,** 865–873 (1998).

129. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. *Nature* **529,** 298–306 (2016).

130. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481,** 85–89 (2011).

131. Braun, S. *et al.* A pooled analysis of bone marrow micrometastasis in breast cancer. *N Engl J Med* **353,** 793–802 (2005).

132. Zhang, X. H.-F. *et al.* Latent bone metastasis in breast cancer tied to Src-dependent survival signals. *Cancer Cell* **16,** 67–78 (2009).

133. Gao, H. *et al.* Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling. *Cell* **166,** 47–62 (2016).

134. Ghajar, C. M. *et al.* The perivascular niche regulates breast tumour dormancy. *Nat Cell Biol* **15,** 807–817 (2013).

135. Wang, H. *et al.* The osteogenic niche promotes early-stage bone colonization of disseminated breast cancer cells. *Cancer Cell* **27,** 193–210 (2015).

136. Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *J Clin Oncol* **28,** 3271–3277 (2010).

137. Smid, M. *et al.* Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68,** 3108–3114 (2008).

138. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436,** 518–524 (2005).

139. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459,** 1005–1009 (2009).

140. Padua, D. *et al.* TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133,** 66–77 (2008).

141. Sethi, N., Dai, X., Winter, C. G. & Kang, Y. Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells. *Cancer Cell* **19,** 192–205 (2011).

142. Nam, D.-H. *et al.* Activation of notch signaling in a xenograft model of brain metastasis. *Clin Cancer Res* **14,** 4059–4066 (2008).

143. Chen, Q. *et al.* Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature* **533,** 493–498 (2016).

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

145. Cancer of the Ovary - Cancer Stat Facts. at <https://seer.cancer.gov/statfacts/html/ovary.html>

146. Ozols, R. F. *et al.* Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. *J Clin Oncol* **21,** 3194–3200 (2003).

147. Armstrong, D. K. Relapsed ovarian cancer: challenges and management strategies for a chronic disease. *Oncologist* **7 Suppl 5,** 20–28 (2002).

148. Bowtell, D. D. *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* **15,** 668–679 (2015).

149. Kurman, R. J. & Shih, I.-M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* **34,** 433–443 (2010).

150. Kim, J. *et al.* High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc Natl Acad Sci U S A* **109,** 3921–3926 (2012).

151. Klinkebiel, D., Zhang, W., Akers, S. N., Odunsi, K. & Karpf, A. R. DNA Methylome Analyses Implicate Fallopian Tube Epithelia as the Origin for High-Grade Serous Ovarian Cancer. *Mol Cancer Res* **14,** 787–794 (2016).

152. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).

153. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45,** 1127–1133 (2013).

154. Patch, A.-M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521,** 489–494 (2015).

155. Ledermann, J., Harter, P. & Gourley, C. Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer — NEJM. *... England Journal of ...* (2012).

156. Mirza, M. R. *et al.* Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* **375,** 2154–2164 (2016).

157. Garcia, A. A. *et al.* A phase II evaluation of lapatinib in the treatment of persistent or recurrent epithelial ovarian or primary peritoneal carcinoma: a gynecologic oncology group study. *Gynecol Oncol* **124,** 569–574 (2012).

158. Schilder, R. J. *et al.* Phase II evaluation of dasatinib in the treatment of recurrent or persistent epithelial ovarian or primary peritoneal carcinoma: a Gynecologic Oncology Group study. *Gynecol Oncol* **127,** 70–74 (2012).

159. Aghajanian, C. *et al.* OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *J Clin Oncol* **30,** 2039–2045 (2012).

160. Pujade-Lauraine, E. *et al.* Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* **32,** 1302–1308 (2014).

161. Argenta, P. A. *et al.* A phase II study of fulvestrant in the treatment of multiply-recurrent epithelial ovarian cancer. *Gynecol Oncol* **113,** 205–209 (2009).

162. Matulonis, U. A. *et al.* Ovarian cancer. *Nat Rev Dis Primers* **2,** 16061 (2016).

163. McGuire, W. P. *et al.* Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med* **334,** 1–6 (1996).

164. Armstrong, D. K. *et al.* Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *N Engl J Med* **354,** 34–43 (2006).

165. Katsumata, N. *et al.* Long-term results of dose-dense paclitaxel and carboplatin versus conventional paclitaxel and carboplatin for treatment of advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer (JGOG 3016): a randomised, controlled, open-label trial. *Lancet Oncol* **14,** 1020–1026 (2013).

166. Doyle, L. A. & Ross, D. D. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). *Oncogene* **22,** 7340–7358 (2003).

167. Duan, Z., Brakora, K. A. & Seiden, M. V. Inhibition of ABCB1 (MDR1) and ABCB4 (MDR3) expression by small interfering RNA and reversal of paclitaxel resistance in human ovarian cancer cells. *Mol Cancer Ther* **3,** 833–838 (2004).

168. Hu, L., McArthur, C. & Jaffe, R. B. Ovarian cancer stem-like side-population cells are tumourigenic and chemoresistant. *Br J Cancer* **102,** 1276–1283 (2010).

169. Samimi, G. *et al.* Increase in expression of the copper transporter ATP7A during platinum drug-based treatment is associated with poor survival in ovarian cancer patients. *Clin Cancer Res* **9,** 5853–5859 (2003).

170. Samimi, G. *et al.* Increased expression of the copper efflux transporter ATP7A mediates resistance to cisplatin, carboplatin, and oxaliplatin in ovarian cancer cells. *Clin Cancer Res* **10,** 4661–4669 (2004).

171. Ding, X. *et al.* MLH1 expression sensitises ovarian cancer cells to cell death mediated by XIAP inhibition. *Br J Cancer* **101,** 269–277 (2009).

172. Sakai, W. *et al.* Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451,** 1116–1120 (2008).

173. Barber, L. J. *et al.* Secondary mutations in BRCA2 associated with clinical resistance to a PARP inhibitor. *J Pathol* **229,** 422–429 (2013).

174. Hennessy, B. T. *et al.* Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res*

69, 4116–4124 (2009).

175. Bapat, S. A., Mali, A. M., Koppikar, C. B. & Kurrey, N. K. Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. *Cancer Res* **65,** 3025–3029 (2005).

176. Steg, A. D. *et al.* Stem Cell Pathways Contribute to Clinical Chemoresistance in Ovarian Cancer. *Clin Cancer Res* **18,** 869–881 (2012).

177. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* **13,** 714–726 (2013).

178. Castells, M., Thibault, B., Delord, J.-P. & Couderc, B. Implication of tumor microenvironment in chemoresistance: tumor-associated stromal cells protect tumor cells from cell death. *Int J Mol Sci* **13,** 9545–9571 (2012).

179. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov* **5,** 1164–1177 (2015).

180. Hugo, W. *et al.* Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell* **162,** 1271–1285 (2015).

181. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* (2017). doi:10.1038/nm.4333

182. Lin, N. U., Bellon, J. R. & Winer, E. P. CNS metastases in breast cancer. *J. Clin. Oncol.* **22,** 3608–3617 (2004).

183. Dawood, S. *et al.* Defining prognosis for women with breast cancer and CNS metastases by HER2 status. *Ann Oncol* **19,** 1242–1248 (2008).

184. Bachelot, T. *et al.* Lapatinib plus capecitabine in patients with previously untreated brain metastases from HER2-positive metastatic breast cancer (LANDSCAPE): a single-group phase 2 study. *Lancet Oncol* **14,** 64–71 (2013).

185. Krop, I. E. *et al.* Trastuzumab emtansine (T-DM1) versus lapatinib plus capecitabine in patients with HER2-positive metastatic breast cancer and central nervous system metastases: a retrospective, exploratory analysis in EMILIA. *Ann Oncol* **26,** 113–119 (2015).

186. Partridge, A. H. *et al.* Chemotherapy and Targeted Therapy for Women With Human Epidermal Growth Factor Receptor 2–Negative (or unknown) Advanced Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **32,** JCO.2014.56.7479–3329 (2014).

187. McCreery, M. Q. *et al.* Evolution of metastasis revealed by mutational landscapes of

chemically induced skin cancers. *Nat Med* **21,** 1514–1520 (2015).

188. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520,** 353–357 (2015).

189. 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).

190. Hammond, M. E. H. *et al.* American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* **28,** 2784–2795 (2010).

191. 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).

192. Patil, P., Bachant-Winner, P.-O., Haibe-Kains, B. & Leek, J. T. Test set bias affects reproducibility of gene signatures. *Bioinformatics (Oxford, England)* **31,** 2318–2323 (2015).

193. Gyanchandani, R. *et al.* Intra-tumor heterogeneity affects gene expression profile test prognostic risk stratification in early breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* clincanres.2889.2015 (2016).

194. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)* btw313 (2016).

195. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res* **44,** D1036–44 (2016).

196. Gloyeske, N. C. *et al.* Immunohistochemical profile of breast cancer with respect to estrogen receptor and HER2 status. *Applied immunohistochemistry & molecular morphology : AIMM / official publication of the Society for Applied Immunohistochemistry* **23,** 202–208 (2015).

197. Nilsen, G. *et al.* Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13,** 591 (2012).

198. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25,** 1754–1760 (2009).

199. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25,** 2078–2079 (2009).

200. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20,** 1297–1303 (2010).

201. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12,** e1004873 (2016).

202. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29,** 24–26 (2011).

203. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **31,** 1023–1031 (2013).

204. Chmielecki, J. *et al.* Oncogenic alterations in ERBB2/HER2 represent potential therapeutic targets across tumors from diverse anatomic sites of origin. *Oncologist* **20,** 7–12 (2015).

205. Weigelt, B. *et al.* Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A* **100,** 15901–15905 (2003).

206. Weigelt, B. *et al.* Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* **65,** 9155–9158 (2005).

207. Brannon, A. R. *et al.* Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol* **15,** 454 (2014).

208. Vignot, S. *et al.* Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: evaluation of concordance between genomic and transcriptional profiles. *European journal of cancer (Oxford, England : 1990)* **51,** 791–799 (2015).

209. Zhang, L. *et al.* The identification and characterization of breast cancer CTCs competent for brain metastasis. *Sci Transl Med* **5,** 180ra48 (2013).

210. Palmieri, D. *et al.* Her-2 overexpression increases the metastatic outgrowth of breast cancer cells in the brain. *Cancer Res* **67,** 4190–4198 (2007).

211. Duchnowska, R. *et al.* Conversion of epidermal growth factor receptor 2 and hormone receptor expression in breast cancer metastases to the brain. *Breast Cancer Res* **14,** R119 (2012).

212. Thomson, A. H. *et al.* Changing molecular profile of brain metastases compared with matched breast primary cancers and impact on clinical outcomes. *Br J Cancer* **114,** 793–800 (2016).

213. Saunus, J. M. *et al.* Integrated genomic and transcriptomic analysis of human brain metastases identifies alterations of potential clinical significance. *J Pathol* **237,** 363–378 (2015).

214. Bose, R. *et al.* Activating HER2 mutations in HER2 gene amplification negative breast

cancer. *Cancer Discov* **3,** 224–237 (2013).

215. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* **34,** 547–555 (2016).

216. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6,** 8839 (2015).

217. Wang, Y. *et al.* Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc Natl Acad Sci U S A* **112,** 9704–9709 (2015).

218. Karantza, V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene* **30,** 127–138 (2011).

219. Joosse, S. A. *et al.* Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18,** 993–1003 (2012).

220. Shiino, S. *et al.* Prognostic impact of discordance in hormone receptor status between primary and recurrent sites in patients with recurrent breast cancer. *Clin Breast Cancer* **16,** e133–40 (2016).

221. Ibrahim, T., Mercatali, L. & Amadori, D. A new emergency in oncology: Bone metastases in breast cancer patients (Review). *Oncol Lett* **6,** 306–310 (2013).

222. Weilbaecher, K. N., Guise, T. A. & McCauley, L. K. Cancer to bone: a fatal attraction. *Nat Rev Cancer* **11,** 411–425 (2011).

223. Moos, R. von & Haynes, I. Where Do Bone-Targeted Agents RANK in Breast Cancer Treatment? *J Clin Med* **2,** 89–102 (2013).

224. James, J. J. *et al.* Bone metastases from breast carcinoma: histopathological - radiological correlations and prognostic features. *Br J Cancer* **89,** 660–665 (2003).

225. Zhang, X. H.-F., Giuliano, M., Trivedi, M. V., Schiff, R. & Osborne, C. K. Metastasis dormancy in estrogen receptor-positive breast cancer. *Clin Cancer Res* **19,** 6389–6397 (2013).

226. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3,** 537–549 (2003).

227. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J Clin Invest* **115,** 44–55 (2005).

228. Lu, X. *et al.* VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging α4β1-positive osteoclast progenitors. *Cancer Cell* **20,** 701–714 (2011).

229. Rosol, T. J. Pathogenesis of bone metastases: role of tumor-related proteins. *J Bone Miner Res* **15,** 844–850 (2000).

230. Priedigkeit, N. *et al.* Intrinsic subtype switching and acquired ERBB2/HER2 amplifications and mutations in breast cancer brain metastases. *JAMA Oncology* (2016). doi:10.1001/jamaoncol.2016.5630

231. Yamamoto-Fukuda, T. *et al.* Effects of Various Decalcification Protocols on Detection of DNA Strand Breaks by Terminal DUTP Nick End Labelling. *The Histochemical Journal* at <http://link.springer.com/article/10.1023/A:1004171517639>

232. Gertych, A. *et al.* Effects of tissue decalcification on the quantification of breast cancer biomarkers by digital image analysis. *Diagn Pathol* **9,** 213 (2014).

233. Schrijver, W. A. M. E. *et al.* Influence of decalcification procedures on immunohistochemistry and molecular pathology in breast cancer. *Mod Pathol* **29,** 1460–1470 (2016).

234. Srinivasan, M., Sedmak, D. & Jewell, S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* **161,** 1961–1971 (2002).

235. Van Allen, E. M. *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med* **20,** 682–688 (2014).

236. Munchel, S. *et al.* Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics. *Oncotarget* **6,** 25943–25961 (2015).

237. Greytak, S. R., Engel, K. B., Bass, B. P. & Moore, H. M. Accuracy of Molecular Data Generated with FFPE Biospecimens: Lessons from the Literature. *Cancer Res* **75,** 1541–1547 (2015).

238. Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res* **25,** 1372–1381 (2015).

239. Graw, S. *et al.* Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. *Sci Rep* **5,** 12335 (2015).

240. Cabanski, C. R. *et al.* cDNA hybrid capture improves transcriptome analysis on low-input and archived samples. *J Mol Diagn* **16,** 440–451 (2014).

241. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast

and bias-aware quantification of transcript expression. *Nat Methods* (2017). doi:10.1038/nmeth.4197

242. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. [version 2; referees: 2 approved]. *F1000Res* **4,** 1521 (2015).

243. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11,** R25 (2010).

244. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

245. Hartley, S. W. & Mullikin, J. C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* **16,** 224 (2015).

246. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29,** 15–21 (2012).

247. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **11,** 11.10.1–11.10.33 (2013).

248. Wei, T. & Simko, V. *corrplot*. (R, 2016). at <https://CRAN.R-project.org/package=corrplot>

249. 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).

250. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102,** 15545–15550 (2005).

251. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).

252. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14,** 7 (2013).

253. Bland, J. M. & Altman, D. G. The logrank test. *BMJ* **328,** 1073 (2004).

254. Kassambara, A. & Kosinski, M. *survminer: Drawing Survival Curves using "ggplot2."* (CRAN, 2016). at <https://CRAN.R-project.org/package=survminer>

255. Singh, V. M. *et al.* Analysis of the effect of various decalcification agents on the quantity

and quality of nucleic acid (DNA and RNA) recovered from bone biopsies. *Ann Diagn Pathol* **17,** 322–326 (2013).

256. Zheng, G. *et al.* Clinical mutational profiling of bone metastases of lung and colon carcinoma and malignant melanoma using next-generation sequencing. *Cancer* **124,** 744–753 (2016).

257. Cejalvo, J. M. *et al.* Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer. *Cancer Res* (2017). doi:10.1158/0008-5472.CAN-16-2717

258. Chu, G. C.-Y. & Chung, L. W. K. RANK-mediated signaling network and cancer metastasis. *Cancer Metastasis Rev* **33,** 497–509 (2014).

259. Loh, Y. N. *et al.* The Wnt signalling pathway is upregulated in an in vitro model of acquired tamoxifen resistant breast cancer. *BMC Cancer* **13,** 174 (2013).

260. Yu-Rice, Y. *et al.* FOXC1 is involved in ERα silencing by counteracting GATA3 binding and is implicated in endocrine resistance. *Oncogene* **35,** 5400–5411 (2016).

261. Doyle, L. A. *et al.* A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc Natl Acad Sci U S A* **95,** 15665–15670 (1998).

262. Mo, W. & Zhang, J.-T. Human ABCG2: structure, function, and its role in multidrug resistance. *Int J Biochem Mol Biol* **3,** 1–27 (2012).

263. Nobori, T. *et al.* Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature* **368,** 753–756 (1994).

264. Furuta, S. *et al.* Removal of BRCA1/CtIP/ZBRK1 repressor complex on ANG1 promoter leads to accelerated mammary tumor growth contributed by prominent vasculature. *Cancer Cell* **10,** 13–24 (2006).

265. Liu, F. & Lee, W.-H. CtIP activates its own and cyclin D1 promoters via the E2F/RB pathway during G1/S progression. *Mol Cell Biol* **26,** 3124–3134 (2006).

266. Wu, M. *et al.* CtIP silencing as a novel mechanism of tamoxifen resistance in breast cancer. *Mol Cancer Res* **5,** 1285–1295 (2007).

267. Yun, M. H. & Hiom, K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* **459,** 460–463 (2009).

268. Wang, J. *et al.* Loss of CtIP disturbs homologous recombination repair and sensitizes breast cancer cells to PARP inhibitors. *Oncotarget* **7,** 7701–7714 (2016).

269. Yan, W., Cao, Q. J., Arenas, R. B., Bentley, B. & Shao, R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* **285,**

14042–14051 (2010).

270. Chou, J. *et al.* GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nat Cell Biol* **15,** 201–213 (2013).

271. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4,** 2612 (2013).

272. Li, Y. & Xie, X. A mixture model for expression deconvolution from RNA-seq in heterogeneous tissues. *BMC Bioinformatics* **14 Suppl 5,** S11 (2013).

273. Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* **29,** 1083–1085 (2013).

274. Onuchic, V. *et al.* Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types. *Cell Rep* **17,** 2075–2086 (2016).

275. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. [version 2; referees: 2 approved]. *F1000Res* **4,** 1521 (2016).

276. Culhane, A. C., Thioulouse, J., Perrière, G. & Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21,** 2789–2790 (2005).

277. Desmedt, C. *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14,** 5158–5165 (2008).

278. Taminau, J. *et al.* Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinformatics* **13,** 335 (2012).

279. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).

280. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4,** 1184–1191 (2009).

281. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43,** e47 (2015).

282. Liu, Y., Siegmund, K. D., Laird, P. W. & Berman, B. P. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol* **13,** R61 (2012).

283. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide

DNA methylation profiles. *Genome Biol* **13,** R87 (2012).

284. Varešlija, D. *et al.* Adaptation to AI Therapy in Breast Cancer Can Induce Dynamic Alterations in ER Activity Resulting in Estrogen-Independent Metastatic Tumors. *Clin Cancer Res* **22,** 2765–2777 (2016).

285. Dowsett, M. *et al.* Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst* **103,** 1656–1664 (2011).

286. McBryan, J. *et al.* Transcriptomic Profiling of Sequential Tumors from Breast Cancer Patients Provides a Global View of Metastatic Expression Changes Following Endocrine Therapy. *Clin Cancer Res* **21,** 5371–5379 (2015).

287. Dean, J. L. *et al.* Therapeutic response to CDK4/6 inhibition in breast cancer defined by ex vivo analyses of human tumors. *Cell Cycle* **11,** 2756–2761 (2012).

288. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* **365,** 671–679 (2005).

289. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165,** 45–60 (2016).

290. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).

291. Chamberlain, M. C., Baik, C. S., Gadi, V. K., Bhatia, S. & Chow, L. Q. M. Systemic therapy of brain metastases: non-small cell lung cancer, breast cancer, and melanoma. *Neuro Oncol* **19,** i1–i24 (2017).

292. Valiente, M. *et al.* Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156,** 1002–1016 (2014).

293. Ni, J. *et al.* Combination inhibition of PI3K and mTORC1 yields durable remissions in mice bearing orthotopic patient-derived xenografts of HER2-positive breast cancer brain metastases. *Nat Med* **22,** 723–726 (2016).

294. Zhang, L. *et al.* Microenvironment-induced PTEN loss by exosomal microRNA primes brain metastasis outgrowth. *Nature* **527,** 100–104 (2015).

295. Adamo, B. *et al.* Phosphatidylinositol 3-kinase pathway activation in breast cancer brain metastases. *Breast Cancer Res* **13,** R125 (2011).

296. Hurtado, A. *et al.* Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* **456,** 663–666 (2008).

297. Plaza-Menacho, I. *et al.* Targeting the receptor tyrosine kinase RET sensitizes breast cancer cells to tamoxifen treatment and reveals a role for RET in endocrine resistance. *Oncogene* **29,** 4648–4657 (2010).

298. Morandi, A. *et al.* GDNF-RET signaling in ER-positive breast cancers is a key determinant of response and resistance to aromatase inhibitors. *Cancer Res* **73,** 3783–3795 (2013).

299. Mulligan, L. M. RET revisited: expanding the oncogenic portfolio. *Nat Rev Cancer* **14,** 173–186 (2014).

300. Tolaney, S. M. *et al.* Cabozantinib for metastatic breast carcinoma: results of a phase II placebo-controlled randomized discontinuation study. *Breast Cancer Res Treat* **160,** 305–312 (2016).

301. Tolaney, S. M. *et al.* Phase II and Biomarker Study of Cabozantinib in Metastatic Triple-Negative Breast Cancer Patients. *Oncologist* **22,** 25–32 (2017).

302. Dar, A. C., Das, T. K., Shokat, K. M. & Cagan, R. L. Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. *Nature* **486,** 80–84 (2012).

303. Gild, M. L. *et al.* Targeting mTOR in RET mutant medullary and differentiated thyroid cancer cells. *Endocr Relat Cancer* **20,** 659–667 (2013).

304. Kramer, E. R. *et al.* Absence of Ret signaling in mice causes progressive and late degeneration of the nigrostriatal system. *PLoS Biol* **5,** e39 (2007).

305. Jain, S. *et al.* RET is dispensable for maintenance of midbrain dopaminergic neurons in adult mice. *J Neurosci* **26,** 11230–11238 (2006).

306. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) *et al.* Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *The Lancet* **378,** 771–784 (2011).

307. Goss, P. E. *et al.* Extending Aromatase-Inhibitor Adjuvant Therapy to 10 Years. *N Engl J Med* **375,** 209–219 (2016).

308. Jacobson, J. A. *et al.* Ten-year results of a comparison of conservation with mastectomy in the treatment of stage I and II breast cancer. *N Engl J Med* **332,** 907–911 (1995).

309. Fredriksson, I. *et al.* Local recurrence in the breast after conservative surgery--a study of prognosis and prognostic factors in 391 women. *Eur J Cancer* **38,** 1860–1870 (2002).

310. Voogd, A. C. *et al.* Long-term prognosis of patients with local recurrence after conservative surgery and radiotherapy for early breast cancer. *Eur J Cancer* **41,** 2637–2644 (2005).

311. Shi, H. *et al.* Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov* **4,** 80–93 (2014).

312. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (2012).

313. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* **2,** e73 (2005).

314. Miller, W. R. *et al.* Changes in breast cancer transcriptional profiles after treatment with the aromatase inhibitor, letrozole. *Pharmacogenet Genomics* **17,** 813–826 (2007).

315. Mackay, A. *et al.* Molecular response to aromatase inhibitor treatment in primary breast cancer. *Breast Cancer Res* **9,** R37 (2007).

316. 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).

317. Miller, C. A. *et al.* Aromatase inhibition remodels the clonal architecture of estrogen-receptor-positive breast cancers. *Nat Commun* **7,** 12498 (2016).

318. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22,** 568–576 (2012).

319. Obenchain, V. *et al.* VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* **30,** 2076–2078 (2014).

320. Wong, S. Q. *et al.* Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics* **7,** 23 (2014).

321. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38,** e164 (2010).

322. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

323. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).

324. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15,** R29 (2014).

325. Simigdala, N. *et al.* Cholesterol biosynthesis pathway as a novel mechanism of resistance to estrogen deprivation in estrogen receptor-positive breast cancer. *Breast Cancer Res* **18,** 58 (2016).

326. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43,** D805–11 (2015).

327. Wang, C., Christin, J. R., Oktay, M. H. & Guo, W. Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. *Cell Rep* **18,** 2825–2835 (2017).

328. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168,** 670–691 (2017).

329. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* **48,** 1119–1130 (2016).

330. Wahner-Roedler, D. L., Reynolds, C. A. & Boughey, J. C. Collision tumors with synchronous presentation of breast carcinoma and lymphoproliferative disorders in the axillary nodes of patients with newly diagnosed breast cancer: a case series. *Clin Breast Cancer* **11,** 61–66 (2011).

331. Liu, L., Kimball, S., Liu, H., Holowatyj, A. & Yang, Z.-Q. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget* **6,** 2466–2482 (2015).

332. Manso, L. *et al.* Analysis of Paired Primary-Metastatic Hormone-Receptor Positive Breast Tumors (HRPBC) Uncovers Potential Novel Drivers of Hormonal Resistance. *PLoS ONE* **11,** e0155840 (2016).

333. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330,** 228–231 (2010).

334. Guan, B., Wang, T.-L. & Shih, I.-M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res* **71,** 6718–6727 (2011).

335. Zhang, X. *et al.* Frequent low expression of chromatin remodeling gene ARID1A in breast cancer and its clinical significance. *Cancer Epidemiol* **36,** 288–293 (2012).

336. Tognon, C. *et al.* Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* **2,** 367–376 (2002).

337. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat Commun* **5,** 4846 (2014).

338. Doebele, R. C. *et al.* An Oncogenic NTRK Fusion in a Patient with Soft-Tissue Sarcoma with Response to the Tropomyosin-Related Kinase Inhibitor LOXO-101. *Cancer Discov* **5,** 1049–1057 (2015).

339. Menichincheri, M. *et al.* Discovery of Entrectinib: A New 3-Aminoindazole As a Potent Anaplastic Lymphoma Kinase (ALK), c-ros Oncogene 1 Kinase (ROS1), and Pan-Tropomyosin Receptor Kinases (Pan-TRKs) inhibitor. *J Med Chem* **59,** 3392–3408 (2016).

340. Hill, V. K. *et al.* Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer Res* **71,** 2988–2999 (2011).

341. Veeck, J. *et al.* Promoter hypermethylation of the SFRP2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. *Mol Cancer* **7,** 83 (2008).

342. 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).

343. Touat, M., Ileana, E., Postel-Vinay, S., André, F. & Soria, J.-C. Targeting FGFR signaling in cancer. *Clin Cancer Res* **21,** 2684–2694 (2015).

344. Sansone, P. *et al.* Self-renewal of CD133(hi) cells by IL6/Notch3 signalling regulates endocrine resistance in metastatic breast cancer. *Nat Commun* **7,** 10442 (2016).

345. Sansone, P. *et al.* Evolution of Cancer Stem-like Cells in Endocrine-Resistant Metastatic Breast Cancers Is Mediated by Stromal Microvesicles. *Cancer Res* **77,** 1927–1941 (2017).

346. Miyabayashi, T. *et al.* Vaccination with CD133(+) melanoma induces specific Th17 and Th1 cell-mediated antitumor reactivity against parental tumor. *Cancer Immunol Immunother* **60,** 1597–1608 (2011).

347. Calinescu, A.-A. *et al.* Overview of current immunotherapeutic strategies for glioma. *Immunotherapy* **7,** 1073–1104 (2015).

348. Guan, R. J. *et al.* Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer. *Cancer Res* **60,** 749–755 (2000).

349. Kyrgiou, M., Salanti, G., Pavlidis, N., Paraskevaidis, E. & Ioannidis, J. P. A. Survival benefits with diverse chemotherapy regimens for ovarian cancer: meta-analysis of multiple treatments. *J Natl Cancer Inst* **98,** 1655–1663 (2006).

350. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).

351. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).

352. Liu, S. *et al.* Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res* **44,** e47 (2016).

353. Hamilton, T. C. *et al.* Characterization of a human ovarian carcinoma cell line (NIH:OVCAR-3) with androgen and estrogen receptors. *Cancer Res* **43,** 5379–5389 (1983).

354. Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat Med* **17,** 1498–1503 (2011).

355. Zheng, W. *et al.* Imbalanced expression of inhibin and activin subunits in primary epithelial ovarian cancer. *Gynecol Oncol* **69,** 23–31 (1998).

356. Cooke, I., O'Brien, M., Charnock, F. M., Groome, N. & Ganesan, T. S. Inhibin as a marker for ovarian cancer. *Br J Cancer* **71,** 1046–1050 (1995).

357. Walentowicz, P. *et al.* Serum inhibin A and inhibin B levels in epithelial ovarian cancer patients. *PLoS ONE* **9,** e90575 (2014).

358. Eckstein, N. *et al.* Hyperactivation of the insulin-like growth factor receptor I signaling pathway is an essential event for cisplatin resistance of ovarian cancer cells. *Cancer Res* **69,** 2996–3003 (2009).

359. Sayer, R. A. *et al.* High insulin-like growth factor-2 (IGF-2) gene expression is an independent predictor of poor survival for patients with advanced stage serous epithelial ovarian cancer. *Gynecol Oncol* **96,** 355–361 (2005).

360. Singh, R. K. *et al.* IGF-1R inhibition potentiates cytotoxic effects of chemotherapeutic agents in early stages of chemoresistant ovarian cancer cells. *Cancer Lett* **354,** 254–262 (2014).

361. Huang, G. S. *et al.* Insulin-like growth factor 2 expression modulates Taxol resistance and is a candidate biomarker for reduced disease-free survival in ovarian cancer. *Clin Cancer Res* **16,** 2999–3010 (2010).

362. Lau, M.-T. & Leung, P. C. K. The PI3K/Akt/mTOR signaling pathway mediates insulin-like growth factor 1-induced E-cadherin down-regulation and cell proliferation in ovarian cancer cells. *Cancer Lett* **326,** 191–198 (2012).

363. Herath, N. I. *et al.* Over-expression of Eph and ephrin genes in advanced ovarian cancer: ephrin gene expression correlates with shortened survival. *BMC Cancer* **6,** 144 (2006).

364. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat Rev Cancer* **10,** 165–180 (2010).

365. Boulay, K. *et al.* Cell cycle-dependent regulation of the RNA-binding protein Staufen1. *Nucleic Acids Res* **42,** 7867–7883 (2014).

366. Harris, L. N. *et al.* Induction of topoisomerase II activity after ErbB2 activation is associated with a differential response to breast cancer chemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **7,** 1497–1504 (2001).

367. Chekerov, R. *et al.* Altered expression pattern of topoisomerase IIalpha in ovarian tumor epithelial and stromal cells after platinum-based chemotherapy. *Neoplasia (New York, N.Y.)* **8,** 38–45 (2006).

368. Burgess, D. J. *et al.* Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc Natl Acad Sci U S A* **105,** 9053–9058 (2008).

369. Wong, N. *et al.* TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance. *International journal of cancer. Journal international du cancer* **124,** 644–652 (2009).

370. de Resende, M. F. *et al.* Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer. *J Transl Med* **11,** 36 (2013).

371. Gordon, A. N. *et al.* Recurrent epithelial ovarian carcinoma: a randomized phase III study of pegylated liposomal doxorubicin versus topotecan. *J. Clin. Oncol.* **19,** 3312–3322 (2001).

372. Green, A. E. & Rose, P. G. Pegylated liposomal doxorubicin in ovarian cancer. *Int J Nanomedicine* **1,** 229–239 (2006).

373. Takeuchi, K. *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat Med* **18,** 378–381 (2012).

374. Martignetti, J. A. *et al.* Personalized Ovarian Cancer Disease Surveillance and Detection of Candidate Therapeutic Drug Target in Circulating Tumor DNA. *Neoplasia (New York, N.Y.)* **16,** 97–W29 (2014).

375. Chen, Z.-H. *et al.* Targeting genomic rearrangements in tumor cells through Cas9-mediated insertion of a suicide gene. *Nat Biotechnol* (2017). doi:10.1038/nbt.3843

376. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* **168,** 613–628 (2017).