# Hotspot Mutations on Estrogen Receptor-α Are Multimodal and Contextual Drivers of Breast Cancer Endocrine Resistance and Metastasis

by

**Zheqi Li**

B.S., Shanghai Jiao Tong University, 2015

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This thesis or dissertation was presented

by

**Zheqi Li**

It was defended on

April 6, 2020

and approved by

Donald Defranco, PhD, Professor, Department of Molecular Pharmacology & Chemical Biology

Lin Zhang, PhD, Professor, Department of Molecular Pharmacology & Chemical Biology

Adrian Lee, PhD, Professor, Department of Molecular Pharmacology & Chemical Biology

Peter Lucas, MD, PhD, Professor, Department of Pathology

Dissertation Director:
Steffi Oesterreich, PhD, Professor, Department of Molecular Pharmacology & Chemical Biology

**Hotspot Mutations on Estrogen Receptor-α Are Multimodal and Contextual Drivers of Breast Cancer Endocrine Resistance and Metastasis**

Zheqi Li, PhD

University of Pittsburgh, 2020

Hotspot mutations in the estrogen receptor alpha (ERα) are frequently detected in ER+ metastatic breast cancer. There is increasing evidence that these mutations confer endocrine resistance and metastatic capacities to breast cancer patients. However, their functional role remains largely unknown.

In the first part, we report the generation of genome-edited MCF7 and T47D cell lines harboring Y537S and D538G *ESR1* mutations. *ESR1* mutations confer ligand-independent growth and endocrine resistance. Transcriptomic analysis revealed highly context-dependent gene expression profiles. I also characterized the critical role of enhanced IGF1R signaling in *ESR1* mutant cells through IRS1 upregulation and pointing towards a potential for co-targeting IGF1R and ERα in breast tumors with mutant *ESR1*.

In the second part, I addressed a critical question- whether these mutations contribute to metastatic process, or merely endocrine resistance. I show clinical evidence for the presence of *ESR1* mutations exclusively in distant but not in local recurrence. Consistent to transcriptomic profiling of *ESR1* mutant tumors, *ESR1* mutant cell models exhinit a reprogrammed cell adhesome, which functionally confers enhanced cell-cell contacts while decreasing cell-ECM adhesion. Contextual migratory phenotypes revealed druggable vulnerabilities, which could be exploited by combination of Wnt and ER targeting strategies. Analysis of global ER and FOXA1 binding sites

with accessible genome data uncovers loss of FOXA1 dependency of D538G mutated ER and novel FOXA1-driven chormatin. Collectively, these data serve as essential evidence for *ESR1* mutations-driven metastasis and provide guidance for future pre-clinical therapeutic strategies.

In the third part, I identified that basal markers were highly enriched in *ESR1* mutant breast cancers. This could be explained by dual mechanisms inducting basal cytokeratins: a CTCF-driven chromatin loop and progesterone receptor-mediated transactivation. Clinically, high basal cytokeratins are associated with enhanced immune response in *ESR1* mutant tumors attributing to the S100A8/9 signaling. Together, these observations show that activating ER mutations confer basal molecular feature and imply immune therapeutic vulnerabilities.

In summary, we deciphered the multimodal and contextual role of hotspot *ESR1* mutations in breast cancer endocrine resistance, metastasis and gain of basal features. Our study provides mechanistic and therapeutic insights to target these activating mutations in advanced breast cancer patients.

**Table of Contents**

## List of Tables

## List of Figures

## Preface

Over the past five years, I have been thinking for a thousand times of how my defense day will be like; I have been thinking for a thousand times of what kind of introduction I will receive from my lovely mentor before my formal defense presentation; I have been thinking for a thousand times of what kind of speech I will give in front of my significant other, my parents, my labmates and my friends, at the exciting moment when I could finally put "PhD" behind my name for the rest of my life. However, all these above things are no longer important to me, at least, for this moment, because of the COVID-19 pandemic crisis. What I really care now is how to protect my loved ones, how to minimize the risk of infection of people around me while keeping the project ongoing and how to spread proper scientific knowledge to educate more people around my community as my responsibility - these things are the original intentions of why I am so proud of choosing scientist as my career.

We are certainly in an unprecedented time - things are changing around us with an incredible speed as the ongoing spread of the coronavirus. I don't even know what sort of format my defense will be in two weeks. But anyhow, I have kept all my promise that I will be prepared to face and overcome all the known and unknown difficulties, with the courage and power in my deepest heart and given by the people I love most, just like what I have said to myself on the first day heading to the US five years ago.

I have been realizing that I have so many people to thank all the way during these amazing five years working at the Lee-Oesterreich lab as a graduate student. First, I would like to thank my

respectful mentors-Drs. Steffi Oesterreich and Adrian Lee. The first time I met Steffi was as an interviewee for my graduate program, her dedication, rigorousness and passion to science already attracted me so much after 30 minutes of in person talk. As I wished, I joined Steffi and Adrian's lab on December 2015, and started the journey of killing *ESR1* mutations. It was very fortunate for a graduate student to have the opportunity to be exposed to such a highly interdisciplinary research environment and work on such a fantastic and clinically significant project. Both Steffi and Adrian are incredibly supportive and helpful in every aspect of my research, career and life. I am thankful that Steffi and Adrian afforded me so many experiences that contributed to a well-rounded education, including but not limited to molecular and cell biology, animal study, bioinformatics and clinical research. More importantly, their passion and rigorousness to science really shaped and impacted my career decision and solidified my determination to become an independent principle investigator in the future to continuously make contributions to the field. They are truly my role models!

Secondly, I would like to thank Dr. Jennifer Xavier for her excellent input in facilitating to run the lab so smoothly. She has constantly not only offered me great scientific suggestions to my projects, but also supported all the essential external collaborations, instrument running and clinical specimen management. It was such as great fortune to work with Jenny in the past three years as colleague, mentee and friend.

Thirdly, I thank my thesis committee members Dr. Donald DeFranco, Dr. Li Zhang, Dr. Peter Lucas and Dr. Adrian Lee for their guidance and patience with this work. Their suggestions and support were always appreciated to my career and my research. I must also thank the entire Women's Cancer Research Center (WCRC) at the Magee-Womens Research Institute (MWRI)

and UPMC Hillman Cancer Center. They always invite experts from different areas and organized great seminars to ensure the high-quality education of my graduate training. It was a wise decision for me to select the Molecular Pharmacology program at my first year. Dr. Patrick Pagano, Dr. Bruce Freeman, Dr. Guillermo Romero, Dr. Tija Jacob and many others are to thank for their invitation to join this wonderful group, as well as their constant support during my training.

As a highly interdisciplinary study, this work would not have been possible without the extensive internal and external collaborative relationships that were developed. Dr. Amir Bahreini and Dr. Kevin Levine offered me great instructions and trainings on bioinformatics, and those experience constantly developed and polished my programming skills from zero background at the very beginning; Dr. Rekha Gyanchadani offered me rigorous trainings on clinical sample handling and processing; Dr. Ben Park and Dr. David Chu kindly provided us their robust MCF7 *ESR1* mutant cell models that has been used through all the study; Dr. Weiyi Toy and Dr. Sarat Chandarlapaty kindly offered their dox-inducible *ESR1* mutant cell models for additional validation; Dr. Maritza Montanez, Dr. Prithu Sundd, Mr. Callen Wallace and Dr. Simon Watkins in assisting the operation and quantification of qMFM technologies in Chapter 3; Mr. Yonatan Amzaleg and Dr. Min Yu provided the CTC ex vivo cell line aggregation data and suggestions in other ongoing vivo studies in Chapter 3; Dr. Jason Carroll, as the expert in ER and FOXA1 cistrome, kindly offered great suggestions in analyzing the ChIP-seq data sets in Chapter 3; Dr. Jason Gertz and Mr. Spencer Arnesen provided their in-house MCF7 and T47D CRISPR *ESR1* mutant models and ATAC-seq data sets in Chapter 3 and Chapter 4; Dr. Li Zhu and Dr. George Tseng helped with confirmation of some of the essential bioinformatics analysis and statistical methodologies in all the studies; Dr. Yu Jiang provided their ibidi pump system for our experiment

in Chapter 3; Dr. Simak Ali provided their in-house MCF7 cell models for the additional validation in Chapter 3; Dr. Peter Lucas provided professional pathological evaluations on all the clinical specimens and mouse tumor sections; Dr. Dorraya El-Ashry, Mr. Benjamin Troness and Dr. Ye Qin helped with the ongoing CTC *in vivo* experiments; Dr. Nikhil Wagle provided expanded RNA-seq data sets on *ESR1* mutant metastatic breast tumors.

Next, I would like to thank many lab members who kindly provided technical and intellectual inputs to this study. Particularly, I would like to point out the essential contributions from two particular colleagues. Dr. Amir Bahreini was a senior graduate student and my first mentor when I joined the lab. I worked with him for the first one and half years on this project. Not only has he contributed a lot to the progress of this study, particularly for Chapter 1, but also provided me with essential training on many basic wet bench and bioinformatics skills. His dedication and belief in my passion and abilities granted me more opportunities than I can ever thank him for. Mr. Yang Wu is a visiting scholar from Tsinghua University and joined the lab on September 2019 to work with me on this study. It was my great fortune to mentor such a brilliant and hardworking peer. He has made a significant contribution to Chapter 3 of this study particularly on the ibidi microfluidic experiments and will continue the journey to conquer *ESR1* mutations after I leave the lab. I will miss those sleepless nights working with him in the lab and wish him all the best in the future. In addition, I thank to Ms. Jian Chen, Ms. Kara Burlbaugh, Dr. Lan Cao, Ms. Meghan Moring, Ms. Corinne Farrell and Ms. Dorothy Carter for all their technical assistance in these works. I had the pleasure to mentor Meghan, Corinne, Dorothy and Lan, and I hope that their time in the lab was as fruitful and inspirational as it was for me. I would like to specially thank Dr. Jagmohan Hooda for overseeing and managing all the animal experiments in

this study, and all the efforts from Ms. Julie Scott, Mr. Kyle Biery and Ms. Christy Smolak at the animal aspect of this study.

I would also like to thank Dr. Nilgun Tasdemir, Mr. Kai Ding, Ms. Sayali Onkar, Mr. Tiantong Liu, Ms. Megan Yates, Mr. Osama Shah, Dr. Lyuqin Chen, Mr. Geoffrey Pecar, Ms. Beth Knapick, Ms. Ashu Elangovan, Mr. John Willis, Dr. Daniel Brown, Ms. Fangyuan Chen, Ms. Laura Savariau, Ms. Simeng Liu, Ms. Minji Chung, Dr. Ozy Nasrazadani, Ms. Lori Miller, Dr. Kevin Levine, Dr. Courtney Anderson, Dr. Gonghong Yan, Dr. Sreeja Sreekumar, Dr. Ye Qin, Dr. Peilu Wang, Dr. Nolan Priedigkeit, Dr. Ahmed Basudan, Dr. Emily Bossart, Dr. Rebecca Watters, Dr. Matthew Sikora, Dr. Karthik Kota, Dr. Tian Du, Dr. Alison Nagle, and Mr. Nick Smith as the best lab current and former lab members I have ever worked with. Thank you for all that you did to make this work possible.

Last but not least, I would like to thank many anonymous patients who graciously donated their tumor tissue and blood sample to this study and the clinicians running these translational programs. Improving breast cancer therapies and create more benefits to them is the driving force of all my efforts to this study.

On a personal note, I would be remiss if I did not thank my family, my loved one and every single friend who helped me through my graduate career. The entire contents may take hundreds of pages which I can't fit them here, but I will be eternally grateful for their support and accompany. To my parents, Ms. Minjuan Li and Mr. Zhun Li, my grandparents Ms. Tongying Yan and Mr. Hexue Li, I can't thank them enough for all the supports they offered all the time since I was a child. There were times that I was so depressed and their supports were the most effective "drug" to bring up my ethics and morals even they were not physically accompanying me on this

continent. I miss them every day and I promise I will keep all of them proud during the rest of my life. I have had the fortune to make a lot of friends either inside or outside the field of science. I thank Dr. Tian Du, Dr. Lan Cao, Dr. Huang Huang, Mr. Kai Ding and Dr. Qin Ye in fighting together with me in the lab at different stages of my graduate career, supporting with each other in both research and lives; I also thank Mr. Hao Chen, Mr. Yuetian Huang, Mr. Cheng Ye, Mr. Fang Xie, Mr. Han Xiao, Mr. Zhiwei Xu, Mr. Fan Xu, Mr. Liran Liu, Mr. Po Lam, Ms. Yize Lyu, Mr. Shi Chang, Mr. Zhenyu Yang, Dr. Xujie Liu, Mr. Xingjian Zhou, Mr. Xin Tan and all the current and previous members from University of Pittsburgh Chinese Students and Scholar's Association (Pitt CSSA) who have ever experienced an amazing time in my life outside academia.

Finally, I want to give special thanks to my lovely girlfriend Ms. Haoyang (Azure) Cui. You were always there for me all the time since we met with each other. There was numerous happies time of my life just because of you. And there were also numerous challenges we have faced together, and I know that I could not make it through without you by my side. The love and encouragement you gave were the irreversible enzymes to push our life forward and make it better and better. All these I did without you, they were my loss. I love you.

Zheqi (Vaciry) Li

03/16/2020 in Pittsburgh

Notes for funding information:

This work was supported by the Breast Cancer Research Foundation; Susan G. Komen Scholar awards; the Metastatic Breast Cancer Network Foundation; the National Cancer Institute; and the Fashion Footwear Association of New York, Magee-Women's Research Institute and

Foundation, UPMC Hillman Cancer Center, Nicole Meloche Foundation, Penguins Alumni Foundation, the Pennsylvania Breast Cancer Coalition and the Shear Family Foundation, the John S. Lazo Cancer Pharmacology Fellowship. This project used the University of Pittsburgh Pitt Biospecimen Core (PBC), supported in part by award NIH grant.

# 1.0 Background

## 1.1 Breast Cancer

### 1.1.1 Breast Cancer Epidemiology

Breast cancer is the most common cancer among women in the United States, and the second leading cause of cancer-related death. In the year of 2020, there will be estimated 276,480 new diagnosed cases of breast cancer, and approximately 63,220 will succumb to the disease in the United States[1]. The incidence of breast cancer has remained relatively stable over the past two decades, largely attributed to the improvement of environmental and physical factors that associated with disease progression[1, 2]. Epidemiologic studies have revealed a great number of factors impacting breast cancer risks, including but not limited to menopausal status[3], obesity[4], family history[5], genetic backgrounds[6], alcohol consumptions[7] and other lifestyles[8]. These disparate entities contribute to a disturbing reality: each woman in America faces a 1-in-8 risk in her lifetime of developing breast cancer.

### 1.1.2 Breast Cancer Subtype Classification

Breast cancer has been well characterized as a highly heterogeneous disease both inter- and intra- tumors. These tremendous complexities make the overarching nomenclature of "breast cancer" as an oversimplification for a collection of unique diseases with distinct biology.

Therefore, the identification and separation of these specific types of breast cancer are indispensable for informing and guiding treatment options for each individual patient in the era of precision medicine. There are two major types of classification methods: molecular intrinsic subtypes and histological subtypes.

#### 1.1.2.1 Molecular Intrinsic Subtypes

The former one has been established by Perou and colleagues in early 2000s[9]. They performed the first comprehensive expression analysis of sixty-five surgical specimens of breast tumors and were able to elucidate the samples subgroups according to the gene expression profile clustering. Further attempts were added by other groups in order to improve the initial classification method[10, 11]. There are mainly five molecular intrinsic subtypes: luminal A, luminal B, HER2-enriched, basal-like and normal-like. The gene expression profile of luminal tumor subset was largely driven by the expression of ERα and progesterone receptor (PR) and other related hormone receptor regulator such as *FOXA1* and *GATA3* [12]. Luminal A is associated with less aggressive breast cancers with lower Ki67 scores ($\leqslant$14%), while Luminal B is associated with more aggressive breast cancers and shows poorer clinical outcomes[12]. At the molecular level, luminal A tumors bear less frequent *TP53* mutations and cyclin D1 amplification, but harbor higher ER signaling due to either ER amplification or its pioneer factor redistribution [13, 14]. On the other hand, The HER2 subtype was found to have a specific *HER2/EGFR* gene expression signature that could separate the tumors into responders to HER2-targeted therapies[15]. The vast majority of them harbor *HER2* amplification [16]. The basal-like tumors are largely intersected with triple negative breast cancers (TNBCs, negative for ERα, PR and HER2 expression).

Clinically, they showed poorer prognosis and more invasive phenotypes than luminal subtypes[17]. They bear the highest rates of *TP53* mutations (80%) and their gene expressional profiles are more similar with serous ovarian cancers that are featured by widespread genomic instability[18]. Similarly, TNBC tumors harbor the highest mutation burdens among all molecular subtypes [19]. Of note, the clinical implications of these molecular subtypes were further examined and are now commonly called the PAM50-based subtype classifier for the original 50 genes used to delineate the intrinsic subtypes[20]. PAM50 gene panel is the first comprehensive test guiding clinical decision for personalized breast cancer diagnosis and therapy.

#### 1.1.2.2 Histological Subtypes

Breast cancers can also be classified by the histological subtypes-a method that has been developed for nearly one hundred years. There are two major subgroups: invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC)[21-23]. IDC typically accounts for 80% of all breast cancer cases, whereas ILC accounts for 10-15%. In addition, there are still approximately 20 rare histological subtypes[24]. IDC tumors typically show distinct masses with Hematoxylin and Eosin (H&E) staining, whereas ILC exhibits a unique growth pattern consisting of small, discohesive cellular growth throughout surrounding stroma[25]. It is caused by the loss of E-cadherin (*CDH1*) expression, a calcium-dependent transmembrane protein imperative for cell-cell interactions at adherens junctions, which has been recognized as the hallmark of ILC[26].

Besides, there are other clinical features that distinguish IDC and ILC. For instance, ILC are often found at unique metastatic sites such as ovaries or the gastrointestinal tract, whereas the common metastasis tropism towards bone, lung and brain remaining shared with IDC[27, 28]. At

the molecular level, ILC is known to be enriched for more unique genetic aberrations such as loss of *PTEN*, *FOXA1* mutations and hyperactivation of PI3K/AKT/mTOR signaling axis induced by mutations on components such as *PIK3CA* and *MTOR* [25]. In summary, ILC and IDC are distinct diseases within breast cancer, with some clear clinical, morphological, and molecular differences. However, molecular features between IDC and ILC still remain unclear.

### 1.1.3 Breast Cancer Metastasis

Breast cancer starts as a localized disease, but it can metastasize to the lymph nodes and distant organs. The most common metastatic sites for breast cancer include bone, lung, brain and liver[29]. Metastatic disease remains the underlying cause of death in the majority of patients with breast cancer who succumb to their disease. The different latency periods between initial treatment and eventual recurrence in different patients suggests breast cancer metastatic spread as a heterogeneous and multi-stage procedure [29, 30]. The metastatic spread of breast cancer typically involves a series of sequential steps: intravasation which allows single cells or collective cell clusters to enter the bloodstream, dissemination and survival of circulating tumor cells (CTCs) in unfavorable environments, followed by extravasation and colonization at a distant metastatic niche [31].

The hallmark for breast cancer metastatic initiation is epithelial-mesenchymal transition (EMT), a highly conserved cellular program that converts polarized, immotile epithelial cells to migratory mesenchymal cells[32, 33]. Cells with EMT gained pro-metastatic properties, including multiple phenotypic reprogramming such as enhanced totality, invasion, anoikis resistance, drug

resistance and cancer stem cell properties[34-36]. EMT is considered as a consequence of cellular plasticity, the dynamic interconversion status between epithelial and mesenchymal support cancer cells to survival at different microenvironments[37].

A small subset of cells occurring EMT are then able to escape from primary sites and enter the bloodstream. Circulating tumor cells are thus a pivotal population containing cells with the clonal capacity to initiate metastatic growth in distant organs, emphasizing the potential prognostic value of CTCs [38-40]. Recent studies have identified an essential role of CTC clusters in cancer metastasis, including an association with shorter metastasis-free survival [41]. Injection of clusters of breast cancer cells resulted in significantly increased lung metastasis in mice compared to single cell injection [41]. Mechanistically this was linked to increased cell-cell interaction, at least in part due to higher expression of plakoglobin, keratin 14 and CD44 [41-43]. In addition, a recent study suggest that cluster formation mediates unique epigenetic reprogramming in CTCs, including hypomethylation at genomic binding sites of various stemness and proliferation-associated transcription factors [44]. Clustering of CTCs is thought to help to overcome the loss of adherence and increased shear stress when traveling in the blood, thereby facilitating metastatic propensity. In this dissertation, I specifically focus on the clustering properties and how I examined this critical phenotype using multiple methodologies.

Components of the immune system are both heroes and villains in cancer metastasis. Despite their essential function in targeting and eliminating CTCs, immune cells can also aid metastatic spread and dissemination [45]. A recent study has uncovered a higher frequency of regulatory T cells in the peripheral blood of metastatic breast cancer patients with high CTC loads, suggesting that recruitment of Treg cells to CTCs might suppress the peripheral anti-tumor

immune response and thereby contribute to increased metastasis [46]. It has also been reported that metastatic propensity is positively correlated to the acute inflammatory state, and specifically the authors showed that CD4+ T cells were necessary at the site of metastasis in order to recruit CTCs to the lung [47]. Data from other studies suggest that CTCs could potentially "hijack" and "hitchhike" immune cells during extravasation process [48, 49]. For instance, heterotypic interactions between CTCs and neutrophils have been reported to support lung metastasis by promoting CTC adhesion to the pulmonary endothelium [50]. In sum, there is increasing evidence that heterotypic carcinoma-immune cell interactions can contribute to the metastatic process via diverse mechanisms.

The colonization step is less well-understood compared to the earlier steps. Recent studies have revealed that the colonization of breast cancer is closely linked to the microenvironment of the secondary remote site[51, 52]. Chen et al. have reported that breast cancer to brain metastasis requires the cancer cell-astrocytes interaction via gap junction formation via Connexin 43-PCDH7 interaction, which prompts cGAMP communication[53]. A recent study from Wang and colleagues uncovered the essential role of the osteogenic niche in promoting disseminated breast cancer cells colonized in bone[54]. In summary, metastasis is a systemic disease. The nature of the multimodal mechanisms behind metastatic spread implicates need for multiple target selection to overcome metastasis depending on the stage of breast cancer patients.

## 1.2 Estrogen Receptor and Endocrine Therapy

### 1.2.1 Estrogen Receptor Biology

Estrogen receptors (ER) belongs to the nuclear receptor family which are indispensable for sensing estrogen signaling and mediating development, metabolism, homeostasis and other essential functions of the organism[55]. There are two classes of estrogen receptor, ER-α and ER-β, mapping to chromosomes 6q and 14q and encoded by separate genes *ESR1* and *ESR2*, respectively. Of note, those two receptors exhibit distinct molecular signaling and biological functions[56]. ER-α has been widely reported as the key estrogen signaling receptor in ER+ breast cancer. As a typical nuclear receptor, ERα is comprised of five functional domains: AF-1 (N-terminal domain), DNA binding domain (DBD), hinge region, AF-2 (ligand binding domain) and the C-terminal domain[57].

The AF-1 domain located at N-terminal can be activated in a hormone independent manner and exists in an intrinsically disorder conformation[58]. It also binds to multiple co-regulators such as SRC1[59]. The DBD is mainly responsible for the ER interaction with the estrogen response element (ERE) palindromic sequence. In addition, ERα could also form heterodimers with other steroid hormone receptors such as ERβ, orphan receptor SHP and retinoid X receptor (RXR) [60-62]. DBD has two zinc-finger subdomains which play the roles in protein-DNA binding (P-Box) and forming ER dimerization interface (D-Box) respectively[63]. The hinge region is the shortest part of ER, it contains a nuclear localization signal sequence[64] and also serves in regulating AF-1 and AF-2 domain functions in a synergistic manner[63]. The AF2 domain serves to bind ligands

and contains twelve helix structures. In addition, it also facilitates ER dimerization and recruits other nuclear co-regulators[65]. Finally, the function of C-terminal domain remains largely unknown. Limited studies have reported its role in mediating tamoxifen-induced activation (A selective estrogen receptor modulator, seen details below) [66].

Upon activation by ligands, ER-α tends to form dimers and binds to the corresponding sites on genomic DNA to trigger the downstream gene expression. The ER-α global binding pattern has been firstly profiled by Carroll et al. in 2006 in breast cancer[67]. Binding sites of ER are dominantly determined by the pioneer factor FOXA1 in the context of breast cancer[68]. Approximately 50% of the entire active ER binding sites are co-localized with FOXA1. However, deregulation of FOXA1-ER axis has been characterized as one of the major mechanisms involving in the acquired endocrine resistance[69, 70]. In addition, recent studies deciphering higher dimensional chromatin interactions uncovered that ER-α functions by extensive 3D chromatin looping to bring genes together for coordinated transcriptional regulation[71, 72]. In summary, the mechanisms underlying how ER mediates global gene expressional profile are still not completed, and a large subset of those genes are highly likely regulated not limited to canonical transcriptional programs.

### 1.2.2 Endocrine Therapy

Breast cancer treatments have evolved from the primitive burning methods of medieval times to targeted therapeutics that are on the cusp of the personalized medicine movement[73]. From classic methodologies that have been utilized for centuries (e.g. surgery), to the class of

endocrine therapies that can be selected for patients' tumors expressing ERα, the therapeutic landscape for breast cancer patients is a constantly evolving or changing entity.

Endocrine therapies have been developed into one of the most successful targeted therapies in cancer history[74]. They are a class of therapies that targets ERα or its ligands. The major subclasses of endocrine therapies include Selective Estrogen Receptor Modulators (SERMs), Selective Estrogen Receptor Downregulators (SERDs), and Aromatase Inhibitors (AIs) (Fig.1). Clinically, the selection of endocrine therapeutic subtype is dependent on the patient menopausal status, prior treatments, disease progression and patient tolerances.

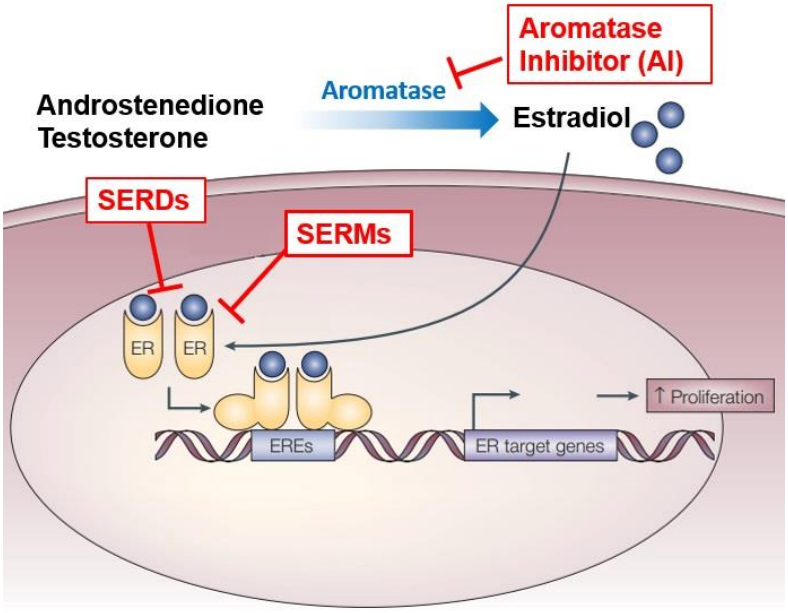



**Figure 1. The three subclasses of endocrine therapy for ER+ breast cancer.**

Estradiol binds to the estrogen receptor (ER), leading to dimerization, conformational change and binding to estrogen response elements (EREs) near estrogen-responsive genes including those responsible for proliferation. SERMs and SERDs directly bind to ER to block its downstream activity whereas aromatase inhibitors reduce the synthesis of estrogens from their androgenic precursors.

This figure is adapted from Johnston et al[75].

Among all three subclasses, AIs work by inhibiting aromatase activity to block the production of estradiol from testosterone. Currently, third-line AIs include nonsteroidal Letrozole or Anastrazole, and steroidal Exemestane[73]. Nonsteroidal AIs abolish the binding of the precursor testosterone to aromatase via a reversible process, while steroidal inhibitors irreversibly interact with aromatase. Of note, the cross-resistance between steroidal and nonsteroidal AIs has been documented as a mechanism for endocrine resistance[76]. The mechanism of AIs supports their utility in postmenopausal women in whom E2 is generated from androgens in extragonadal tissues.

SERDs is a class of compounds that can directly bind to ERα and trigger a rapid degradation thus are considered pure ER antagonists[77]. The most widely used SERD in clinic is fulvestrant (ICI 182, 780), and it has been approved as a first-line breast cancer therapy for metastatic breast cancer recently by FDA[78]. Recent pre-clinical and clinical studies are evaluating the effects of other novel oral SERDs in the treatment of ER+ breast cancer patients, such as AZD 9496[79], RAD 1901[80] and GDC-0810[81].

In contrast, SERMs modulate ER activity via direct interaction but could also serve as a partial agonist. SERMs typically competitively bind to the ligand binding domain of ER, and allosterically modulate the recruitment of other coregulators to ER[82], they exhibit partial agonist and partial antagonist activities depending on biological contexts such as tissue types and patient genetic background [83]. So far, FDA has approved 11 SERMs, whereas at least 20 more are current under clinical evaluation[82]. Despite this, Tamoxifen remains the mainstay of the choice of SERMs in ER+ breast cancer treatment. Tamoxifen is normally metabolized in liver by P450 enzymes into 4-hydroxytamoxifen (4OHT), afimoxifene, and endoxifen. Those metabolites

display higher binding affinity to ER than Tamoxifen itself, thus they are the primary compounds competing with estradiol for binding to ER[84].

### 1.2.3 Mechanisms Underlying Endocrine Resistance

Approximately 70% of advanced breast cancers are considered "hormone responsive" as defined by expression of the ER. Although endocrine therapy is always considered for advanced HR+ breast cancer, the emergence of resistance is inevitable over time and is present from the start in a proportion of patients and becomes a severe social and economic issue.

Over the past three decades, numerous studies have discovered multimodal mechanism that could cause resistance towards endocrine therapy in ER+ breast cancer[85-87]. The most straight forward interpretation is the aberrations on estrogen receptor itself which greatly impact the efficacy of anti-ER therapies. This includes the loss of ER in approximately 10% of endocrine resistant patients[88, 89] and the gained gene fusions with ER such as *ESR1-YAP1* fusion recently reported by Lei et al.[90]. In line with this, recent work has revealed that hotspot mutations on *ESR1* ligand-binding domain occurs in approximately 20%-40% endocrine resistant breast cancer patients[91]. A separate introduction is shown in the next section below, and this entire dissertation is focusing on these hotspot mutations on estrogen receptor.

The first major mechanism is associated with overexpression and amplification of various growth factor receptors including, but not limited to, *IGF1R, FGFR1, HER2, HER3, EGFR* [92-96]. Hyperactivation of these growth factor pathways likely provides alternative support to breast cancer cell proliferation and survival even under the exposure of endocrine therapies. These

bypassing growth factor receptor pathways frequently converge on the PI3K/AKT/mTOR and RAF/MEK/ERK axis[97]. Of note, aberrations in the PI3K intracellular signaling pathway occur in approximately 70% of breast cancers[98], thus the combination of PI3K signaling inhibitors with endocrine therapy are broadly considered as a strategy to overcome such resistance. This is exemplified by the recently BOLERO-2 trail combining mTOR inhibitor everolimus and exemestane to treat postmenopausal patients with ER+ advanced breast cancer[99].

Another series of mechanisms causing endocrine resistance relies on alterations in cell cycle regulators. Endocrine therapy typically leads to the G1 phase arrest of breast cancer cells as a consequence of proliferation suppression[100]. Thus, the acquired alterations of cell cycle regulator might facilitate cells to overcome such effects. Data from experimental model systems have identified the pivotal roles of amplification of multiple cell cycle progression promoters such as cyclin D1, cyclin E1, MYC[101, 102], and the loss-of-function mutations and decreased expression of cell cycle suppressors such as RB, p21 and p27 in mediating endocrine resistance[103, 104]. Clinically, the application of CDK4/6 inhibitors in combination of endocrine therapy has shown promising outcomes in overcoming acquired endocrine resistance[105].

Furthermore, other additional mechanisms might contribute to the development of endocrine resistance. For instance, inhibition of autophagy has been linked to restoration of endocrine sensitivity and promotion of apoptotic cell death in preclinical models of endocrine-resistant breast cancer [106]. A recent study by Achiger-Kawecka and colleagues has shown that the dynamic chromatin remodeling could lead to the differential accessibility of ER binding sites, which accompanies endocrine resistance[107].

### 1.2.4 Identification of Hotspot *ESR1* Mutations

A series of studies over the last seven years documented *ESR1* hotspot somatic mutations in endocrine resistant advanced breast cancer. *ESR1* mutations cluster in the ligand-binding domain, overlapping with activation function 2 (AF2)[108-111]. A number of groups, including ours, have documented the rare occurrence of *ESR1* mutations in primary tumors (<2%) but high mutation frequencies (20-40%) in metastatic lesions and circulation-free DNA (cfDNA), associated with significantly worse outcomes using the sensitive droplet digital PCR (ddPCR) [112-115] (Fig. 2). Moreover, other groups examined *ESR1* mutations using next generation sequencing, and could repeatedly identify those hotspot mutations in both primary tumors, metastatic lesions and cfDNA[108, 116-118]. Of note, Toy et al. have identified above 20 different somatic mutation subtypes at LBD of *ESR1* from MSK-IMPACT target panel DNA sequencing, among which Y537S (14%) and D538G (36%) showed the highest frequencies with majority of other mutations present only rarely [108]. Several other pre-clinical studies have elegantly shown that *ESR1* hotspot mutations at positions Y537 and D538 alter the position of helix 12 to favor an agonist conformation of the receptor, making the receptor constitutively active, and thereby resulting in resistance to endocrine therapy[119]. However, there is limited knowledge on whether mutant ER may directly play a role in conferring metastatic capacity, in addition to mediating endocrine resistance, potentially identifying therapeutic vulnerabilities in *ESR1* mutant tumors. In this dissertation, I sought to comprehensively use robust *in vitro*, *in vivo* and clinical resources to decipher the mechanism underlying the impact of *ESR1* mutations in breast cancer endocrine resistance and metastasis.

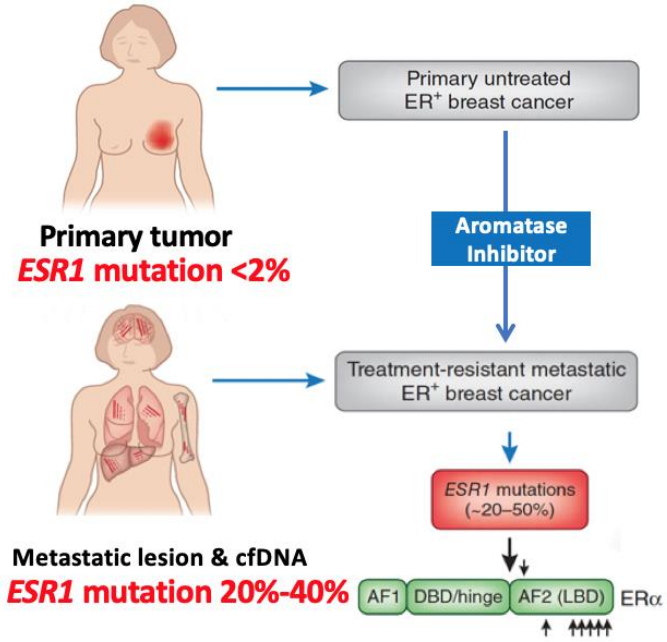



**Figure 2. Identification of *ESR1* Mutations in Metastatic Breast Cancer.**

Hotspot *ESR1* mutations are rare in primary breast tumors but identified in 20%-40% ER+ metastatic breast cancers. They typically occur after aromatase inhibitor treatment and cluster in the ligand binding domain of estrogen receptor.

This figure is adapted from Oesterreich et al.[120].

# 2.0 Hotspot *ESR1* Mutations Confer Endocrine Resistance to Breast Cancer

## 2.1 Introduction

### 2.1.1 Establishment of *ESR1* Mutant Cell Models

Gain-of-function mutations in *ESR1* are likely to play a key role in conferring endocrine therapy resistance in 20–40% of estrogen receptor-positive (ER+) metastatic breast cancer [120-122]. The majority of mechanistic studies have employed overexpression approaches, and results show that the mutant receptors cause ligand-independent growth and decreased sensitivity to antiestrogens[91, 123-126]. Reporter assays and gene expression analysis in transfected cell lines reveal ligand-independent activity of ER, associated with increased expression of classical ER target genes and some novel ER target genes[91, 123-126].

Two recent reports confirmed the ligand-independent activity of mutants in CRISPR generated cell lines[127, 128]. Harrod et al. generated a single Y537S MCF7 clone, in which ER was able to bind to DNA and regulate endogenous targets in a ligand-independent manner[109]. The study also showed that CDK7 is a promising target in *ESR1* mutant, endocrine-resistant disease. The study from Mao et al. focused on the potential role of increased unfolded protein response in *ESR1* mutant cells, and the interaction with progestins, which further promotes the proliferation of *ESR1* mutant cells due to increased expression of progesterone receptor[127].

In addition, various other studies have generated non-genome-edited cell models to mimic *ESR1* mutation in the context of ER+ breast cancer cell lines. Toy et al. and Jeselsohn et al. have reported the repeated observation of ligand-independent activation conferred by *ESR1* mutations both *in vitro* and *in vivo* using doxycycline-inducible cell models[108, 117]. Gelsomino et al. and Yu et al. have utilized stable overexpression *ESR1* mutant cell models and identified essential mechanisms associated with endocrine resistance via ER-IGF1R signaling mutual crosstalk and MYC upregulation[129, 130].

In the first section of this chapter, we set out to introduce the two most frequently identified *ESR1* mutations Y537S and D538G into two ER+ breast cancer cell lines, T47D and MCF7. Using multiple clones, we performed in-depth functional analysis that confirmed and expanded previous observations, and importantly identified mutation-specific and cell line-specific phenotypes, suggesting the need for the study of the individual mutations in a context-dependent manner. The genome-wide expression data and the models will be excellent resources for the research community studying endocrine resistance caused by *ESR1* mutations.

### 2.1.2 Mechanisms Underlying Endocrine Resistance Conferred by *ESR1* Mutations

Recent mechanistic studies performed by us and others, employing either transfected or genome-edited *ESR1* mutant cells, show that mutant ER has ligand-independent activity and diminished sensitivity towards antiestrogen drugs [109, 110, 127] . Clinical studies have documented rare *ESR1* mutations in primary breast cancers, but increased frequency in metastatic lesions and circulation-free DNA (cfDNA), suggesting a potential role of acquired *ESR1* mutations

in facilitating metastasis[131]. Given studies suggesting poor outcomes in patients harboring breast cancers with *ESR1* mutations, identification of new targets and design of novel therapeutic strategies have gained urgency.

Multiple recent preclinical studies have uncovered promising therapeutic targets in breast cancer cells harboring *ESR1* mutations. Harrod et al. highlighted the potential utility of a CDK7 inhibitor to block growth in MCF7 cells with CRISPR edited Y537S[128]. Mao et al. found increased unfolded protein response in CRIPSR-edited Y537S and D538G *ESR1* mutant cells[127]. These finding were further validated by Jeselsohn et al. in a separate study *in vitro* and *in vivo*, suggesting the potential application of CDK7 inhibitor THZ1 in treatment of *ESR1* mutant patients[117]. Mao et al. have reported that the ER mutation biomodulator BHPI, which hyperactivated the unfolded protein response (UPR), and blocked proliferation of both Y537S and D538G *ESR1* mutants cells. Recent findings from Gelsomino et al. identified enhanced signaling mutual activation between mutant ERα and IGF1R, proposing a role in tamoxifen resistance, indicating a potential for combination therapy by co-targeting ERα and mTOR in *ESR1* mutant tumors[130]. In addition, Yu et al. recently reported the MYC pathway hyperactivation is necessary and sufficient to cause ER mutant constitutive activation[129].

In the second section of this chapter, we focused on the enrichment of IGF1R pathway component gene expression in our genome-edited MCF7 and T47D cell models and proposed it as a potential mechanism leading to endocrine resistance. Our recent transcriptomic analysis of genome-edited MCF7 and T47D Y537S and D538G *ESR1* cell lines revealed mutation site and context-dependent gene expression changes compared to wild-type *ESR1*[110]. Similar to

Gelsomono's findings[130], RNA-seq analysis revealed altered expression of IGF1 pathway members in both *ESR1* mutant cell lines. Herein, we performed a preclinical study to characterize the mechanisms underlying the augmented IGF1 response in *ESR1* mutant cells, and evaluated the strategy of co-targeting ER and IGF1R for future therapeutic development.

## 2.2 Materials and Methods

### 2.2.1 Cell Culture

Original resources of T47D (HTB-133) and MCF7 (HTB-22) parental cells were obtained from ATCC. Both cell lines were authenticated at the University of Arizona Genetics Core. CRISPR-Cas9-edited T47D and rAAV-edited MCF7 *ESR1* mutant cells reported previously [110]. Cells were tested bianually for Mycoplasma contamination (Lonza, LT07-318). Individual clones (T47D: 3 WT, 2 Y537S and 3 D538G; MCF7: 2 WT, 2 Y537S and 2 D538G clones) were maintained in RPMI 1640 (T47D) and DMEM (MCF7) respectively, supplemented with 10 % FBS, 100 μg/mL penicillin and 100 mg/mL streptomycin, at 37 °C in a humidified incubator with 5% $CO_2$. Mutation allele frequencies were periodically confirmed using droplet digital PCR. Tamoxifen resistant (TamR) and long-term estradiol deprived (LTED) MCF7 and ZR75-1 cell lines were gifts from Dr. Rachel Schiff (Baylor College of Medicine, Houston, TX).

### 2.2.2 Compounds and Reagents

17β-estradiol (E2) and fulvestrant (Ful) were obtained from Sigma and Tocris, respectively. IGF1 was purchased from GroPep BioReagents (#AM001, Australia). IGF-II (#110-12), VEGF (#100-20), PDGF-BB (#100-14B) and FGF-basic (#100-18B) were from Peprotech (Rocky Hill, NJ). OSI-906 (S1091), Wortamannin (S2758), BX-795 (S1274) and BMS-754807 (S1124) were obtained from Selleck Chem (Houston, TX). U0126 (#1144) and Rapamycin (#1292) were purchased from Tocris Bioscience (Bristol, UK). AZD9496 was recently reported by Weir et al. was kindly provided by AstraZeneca.

### 2.2.3 Hormone Deprivation

For all the experiments, hormone deprivation was performed for all the experiments unless stated otherwise. Cells were regularly maintained in FBS as described above. Shortly before the experiment, cells were split into appropriate vessels with 80-90% cell confluency. Clones with the same genotypes were equally pooled. Start from the day next, cell culture medium was switched to phenol-red-free IMEM (Gibco, A10488) with 10% and 5% charcoal-stripped serum (CSS, Gemini, #100-119) for T47D and MCF7 respectively. Medium were changed twice a day for three consecutive days followed by an additional PBS wash each time.

### 2.2.4 Generation of Genome-edited *ESR1* Mutant Cell Line*

To select subgenomic RNAs (sgRNAs) (Appendix D Table 3) for CRISPR-Cas9 genome-editing of T47D cells [129, 132-134], we utilized a web tool (http://crispr.mit.edu) entering the sequence flanking Y537S and D538G mutations. The oligos were cloned into PX458 (www.addgene.com), also coding for Cas9, tracrRNA, green fluorescent protein (GFP), and the resulting plasmid was transfected along with the respective doublestranded 70 bp oligos into T47D cells. GFP+ cells were sorted by fluorescence-activated cell sorting (FACS), and the mutation was confirmed by digital droplet PCR (ddPCR) using previously described methods [135]. We obtained two clones for Y537S, three clones for D538G, and three clones for *ESR1* wild-type (WT), which were kept as individual clones, and pooled for experimental studies as indicated. Gene targeting of *ESR1* in MCF7 cells was carried out using recombinant adeno-associated virus (AAV) technology as previously described[136]. Clones were confirmed by ddPCR. Single-stranded cDNA was generated using the First Strand cDNA Synthesis Kit (Amersham Biosciences). Two clones and a targeted WT control for the *ESR1* exon 10 locus were isolated for each *ESR1* mutation. Primer sequences for PCR amplification, mutagenesis, targeting, and sequencing are shown in the Appendix D Table 4.

*T47D *ESR1* mutant CRISPR clones were generated by Amir Bahreini, PhD and Peilu Wang, MD. MCF7 *ESR1* mutant rAAV clones were established by David Chu, PhD from Ben Park MD's group.

### 2.2.5 Generation of overexpression *ESR1* mutant cell models

To generate *ESR1* mutant overexpression cell models, *ESR1* WT and mutant plasmids in pcDNA3.1 backbone were obtained from Addgene (*ESR1*-HA-WT #49498; *ESR1*-HA-Y537S #49499; *ESR1*-HA-D538G #49500, Empty vector #V790-20). MCF7 and T47D parental cells were transfected with each of the plasmid and subjected to 500 μg/ml G418 (Thermo Fisher, #10131035) for 3 weeks. G418-containing medium were changed every 3 days during the selection process. Overexpression of ER was further examined by immunoblot and ddPCR in pooled clones and used for further experiments.

### 2.2.6 Droplet Digital PCR (ddPCR)*

Validation of *ESR1* mutation in cell models using ddCPR was performed following method described before[135]. Briefly, cell pellets were lysed in 2% SDS and sonicated in room temperature for 5 minutes. gDNA were then isolated using QIAquick PCR purification kit (Qiagen #28104). gDNA was then diluted to 10 ng/ μl. Oil droplet-DNA complex was generated using QX100 Bio-Rad automatic droplet generator with input of a mixture of DNA template, probes, amplification primes towards *ESR1* LBD region and PCR supermix. Oil droplet-DNA mixture were then amplified in thermal cycler and signals of each droplet towards WT and mutant probes were read using Bio-rad QX100 droplet reader. Allele frequencies were further calculated using Quanta Software (Bio-Rad). Sequences for ddPCR probes and amplified primers were shown in Appendix D Table 5.

*ddPCR for cell model validation was conducted by Amir Bahreini, PhD.

### 2.2.7 Immunoblotting

Protein was isolated after specific experimental treatment. For IGF1 stimulation, cells were starved in serum-free IMEM for 24 hours and then treated with IGF1 for 15 minutes. For siRNA-knockdown experiments, cells were reverse transfected with siRNA targeting IRS1 (L-003015, Dharmacon, Lafayette, CO) and/or IGF1R (L-003012, Dharmacon) for 24 hours and subjected to IGF1 stimulation. Protein were isolated using RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40 (Sigma Aldrich#21-3277 SAJ), 0.5% NaDeoxycholate, 0.1% SDS) supplemented with 1X protease and phosphatase inhibitor cocktail (Thermo Scientific #78430), sonicated in a 4°C cup horn sonicator at amplitude 100 for 5-10 minutes total with 1-minute pulses and 30 second pauses, and centrifuged for 10 minutes at 14,000 rpm at 4°C. All samples were quantified for protein concentration using BCA Assay (Thermo Scientific #23225) and 60-80 µg per sample were run on homemade SDS-PAGE gels (concentrations were dependent on target protein) with transfer to PVDF membranes (Millipore#IPFL00010). When utilizing the Odyssey system (LiCor), membranes were blocked using Odyssey PBS blocking buffer (LiCor#927-40000) for one hour and probed with primary antibodies listed Appendix D Table 6. After removal of primary antibodies, blots were washed with 1X PBS-Tween 20 (0.1%) for 10 minutes, three times, followed by incubation in 1:10,000 secondary antibodies (anti-mouse 800CW: LiCor#925-32210; anti-rabbit 800CW: LiCor#925-32211). Blots were again washed prior to imaging on the Odyssey Infrared Imaging system (LiCor). Quantifications were performed using ImageJ software [137].

### 2.2.8 Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

MCF7 and T47D cells were seeded into 6-well plate with 120, 000 and 90, 000 cells per well respectively with biological triplicates. After desired treatments, RNAs were extracted from each sample, and cDNA was synthesized using iScript kit (#1708890, BioRad, Hercules, CA). qRT-PCR reactions were performed with SybrGreen Supermix (#1726275, BioRad), and the ΔΔCt method was used to analyze relative mRNA fold changes and *RPLP0* levels were measured as the internal control. Relative mRNA levels between WT and mutant cells were compared using Dunnett's test. All primer sequences were shown in Appendix D Table 7.

### 2.2.9 Estrogen Response Element Transcriptional Reporter Assay

MCF7 and T47D cell models were hormone deprived and seeded into 24-well plate with $2x10^5$ cells per well in biological triplicates. After cells were adhered, 1 nM of 17β-estradiol (Sigma-Aldrich, #E8875) or 1% EtOH were added into corresponding wells. Cells were lysed with 100 µl 1x passive lysis buffer from Promega in room temperature for 20 minutes. 50 µl of cell lysis were then mixed with luciferin and renilla substrates, and their luminescence values were read after 10 seconds of reactions with Promega Glomax system. The relative light units (RLU) were calculated by normalizing the luciferin readouts to the corresponding renilla readouts. RLU levels between WT and mutant cells were compared using Dunnett's test.

### 2.2.10 Cell Growth Assay

Cells were seeded at 3,000 (MCF7) or 4,000 (T47D) cells/well in six replicate wells of a 96 welled 2D (Fisher#353072). Vehicle controls were always included, as well as media blank wells, and unused wells were given equal volume 1X DPBS to keep consistent surface tension across the plate, and to avoid evaporation of inner wells. Cells were seeded 1 to 1.5 days prior to treatment and typically collected at day 6 (MCF7) or day 7 (T47D) unless otherwise noted. Upon time point of collection, media was flicked off of 2D plates, extraneous media was wicked with a paper towel, and plates were promptly frozen at -80°C. Plates were then thawed to room temperature and osmotic pressure was applied across entire plates at 100 µL/well water, with incubation at 37°C for one hour, and freezing plates again at -80°C. Plates were again thawed, and then Hoechst dye was applied and measurements assessed for fluorescence per manufacturer's protocol of the FluoReporter Blue Fluorometric dsDNA Quantitation Kit (ThermoScientific#F2962) on a PerkinElmer 2030 Multilabel Reader with VictorX software. Data were corrected for each experiment by subtracting average background fluorescence from values, and normalizing to respective vehicle controls as "1." Error was displayed as standard deviation of the mean (STDEV). For dose response experiments, IC50s or EC50s were calculated using PRISM statistical package version 7.0. For analysis of drug synergy, the combination index values were calculated using the CalcuSync package version 2.0.

### 2.2.11 Chromatin Immunoprecipitation (ChIP)

For ERα ChIPs, MCF7 and T47D *ESR1* mutant cells were 80-90% confluence prior to deprivation. After deprivation, cells were treated with or without 1 nM E2 for 45 minutes. At the treatment time point, media were removed prior to further processing. At time of processing, cells were fixed in 1% Formaldehyde (Polysciences#18814) for 10 minutes while rocked at room temperature. Quenching was then performed with 0.125 M Glycine for 5 minutes with rocking at room temperature, samples were washed three times in cold 1X DPBS supplemented with 1X PPis, were scraped into tubes, and centrifuged at 2,500 rpm for 5 minutes at 4°C. Supernatant were aspirated and pellets were resuspended in 1 mL Nuclei Preparation Buffer (5 mM PIPES, 85 mM KCl, 0.5% NP-40, pH 8.0, 1X PPis) with rotating at 4°C for 30 minutes. Samples were centrifuged at 2,500 rpm for 5 minutes at 4°C. Supernatant were again aspirated, and pellets were resuspended in 300 µL TE Buffer (10 mM EDTA, 50 mM Tris-HCl, pH 8.1, 1X PPis, 1% final concentration SDS). Samples were sonicated at amplitude 100 for 25 minutes total with 1 minute pulses and 30 second pauses, and then centrifuged at 14,000 rpm for 10 minutes at 4°C. Supernatant were collected and portions of samples were checked for DNA shearing between 200-700 base pairs. Upon confirmation of shearing, portions of samples were saved for input use, and remaining samples were diluted 1:10 in Dilution Buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1, 1X PPis) and were precleared with 5 µg Rabbit Immunoglobulin G (IgG) and 135 µL Protein G-Sepharose (Thermo Fisher#10-1243) by rotating at 4°C for 2 hours. Beads were pelleted by centrifugation at 1,000 rpm for 1 minute at 4°C, and supernatant were collected. Samples were split into fractions for IgG or antibody of interest for immunoprecipitation (IP) or

chromatin immunoprecipitation (ChIP) with rotation overnight at 4°C. Samples were given 55 µL Protein G-Sepharose beads the next morning, and samples were rotated for 1 hour at 4°C prior to centrifugation at 1,000 rpm for 1 minute at 4°C. For IP sample verification, samples were washed three times in 1X DPBS supplemented with 1X PPis by rotating at 4°C for 5 minutes, and centrifugation at 1,000 rpm for 1 minute at 4°C; upon completion of washes, protein were eluted from beads and samples were processed as previously described by standard immunoblotting technique against 10% input. For ChIP samples, pelleted samples were washed sequentially by rotating at 4°C for 5 minutes each, followed by centrifugation as previously with the following: TSE I (1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl, 1X PPis), TSE II (1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl, 1X PPis), Buffer III (0.25 M LiCl, 1% NP-40, 1% Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1, 1X PPis) and twice with TE Buffer. ChIP samples were then combined with 100 µL Elution Buffer, vortexed, and incubated at 65°C for 30 minutes with rotation. Samples were centrifuged and supernatant were collected. Eluates were heated at 65°C for 6 hours along with input fractions that were combined with Elution Buffer. DNA was isolated using the Qiagen PCR Purification Kit protocol. ChIP qRT-PCR were performed using input samples diluted 1:5 and samples diluted 1:3, with 5 µL per technical replicate well. ChIP qRT-PCR Primers used are listed in Appendix D Table 8. ERα ChIPs were performed twice, independently, and displayed with STDEV for technical triplicates. Fold enrichment levels between WT and mutant cells were compared using Dunnett's test.

### 2.2.12 RNA-sequencing and Analysis*

Individual *ESR1* WT and mutant T47D and MCF7 clones were hormone-deprived in CSS for 3 days, pooled, and plated in quadruplicates in 6-well plates. The cells were treated with vehicle or 1 nM E2 for 24 h, RNA was isolated using Qiagen RNeasy kit. RNA-seq library was prepared using illumine TrueSeq RNA Access Library kit and sequencing was performed obtaining >15 M reads per sample with NextSeq 2000 at the Health Sciences Sequencing Core at Children's Hospital (Pittsburgh, PA). Salmon was used for quantification of the transcripts using default options and hg38 genome build as the reference [20]. Log2 (TPM+1) value of each gene was used for output and downstream analysis. The genes differentially expressed (DE) between WT and mutants were identified by the DEseq2 package using the contrast option to compare mutants to WT and to calculate the adjusted p value and fold change (FC) [21]. Genes with maximum transcripts per million (TPM) <1 across all samples were excluded from further analysis due to low gene expression. R was used for statistical analysis, and for plotting of the heatmaps. The chi-square test was used to assess the statistical significance of overlaps in venn diagrams.

**ESR1* mutant cell model RNA-sequencing was conducted by Amir Bahreini, PhD.

### 2.2.13 Calculation of IGF Activation Score*

The IGF activation score was calculated for each cell line as described previously[138]. Briefly, the activation score was calculated as the Pearson correlation between the IGF gene signature pattern (by using 1 and −1 for up and down, respectively) and the gene expression values of each cell line. The gene expression values were the log2 (TPM+1) of Salmon (v.0.6.0) mapped

reads to ENSEMBL gene-level IDs from Bahreini et al[110] . The microarray probes from the IGF up and down regulated genes were mapped to ENSEMBL IDs using the hgu133a2.db annotation package. hgu133a2.db: Affymetrix Human Genome U133A 2.0 Array annotation data (chip hgu133a2). This was performed using R version 3.2.2.

*This analysis was conducted by Kevin Levine, MD, PhD.

## 2.3 Results

### 2.3.1 Characterization of *ESR1* Mutant Cell Model Reveals Mutation Sites and Context Dependent Transcriptomic Reprogramming

#### 2.3.1.1 Molecular Characterization of Genome-edited Y537S and D538G ER Mutant Breast Cancer Cell Model

Successful genome editing was confirmed by sequencing multiple clones of Y537S and D538G in T47D and MCF7 cells. The mutation allele frequency was 50%, reflecting heterozygous targeting in all clones except the T47D Y537S#2 clone in which it was 22%. These frequencies correlated well with mRNA expression of WT and mutant ER (Fig. 3A). At the protein level, the pooled clones showed minimal variation at baseline levels with slightly higher expression of D538G than WT, and Y537S slightly lower in both cell lines (Fig. 3B). Fulvestrant decreased protein levels in all clones, although the residual ER protein levels were higher in D538G. Mutant ER displayed two to three folds higher constitutive phosphorylation compared to WT ER in both cell lines (Appendix A Fig. 63), although not to the level previously observed upon overexpression

[91]. E2 treatment inhibited phosphorylation in *ESR1*-mutant MCF7 cells, which was not observed in T47D mutant cells, again suggesting some cell-line-specific effects of mutant ER. Similar data were obtained when using the individual clones (Appendix A Fig. 64).

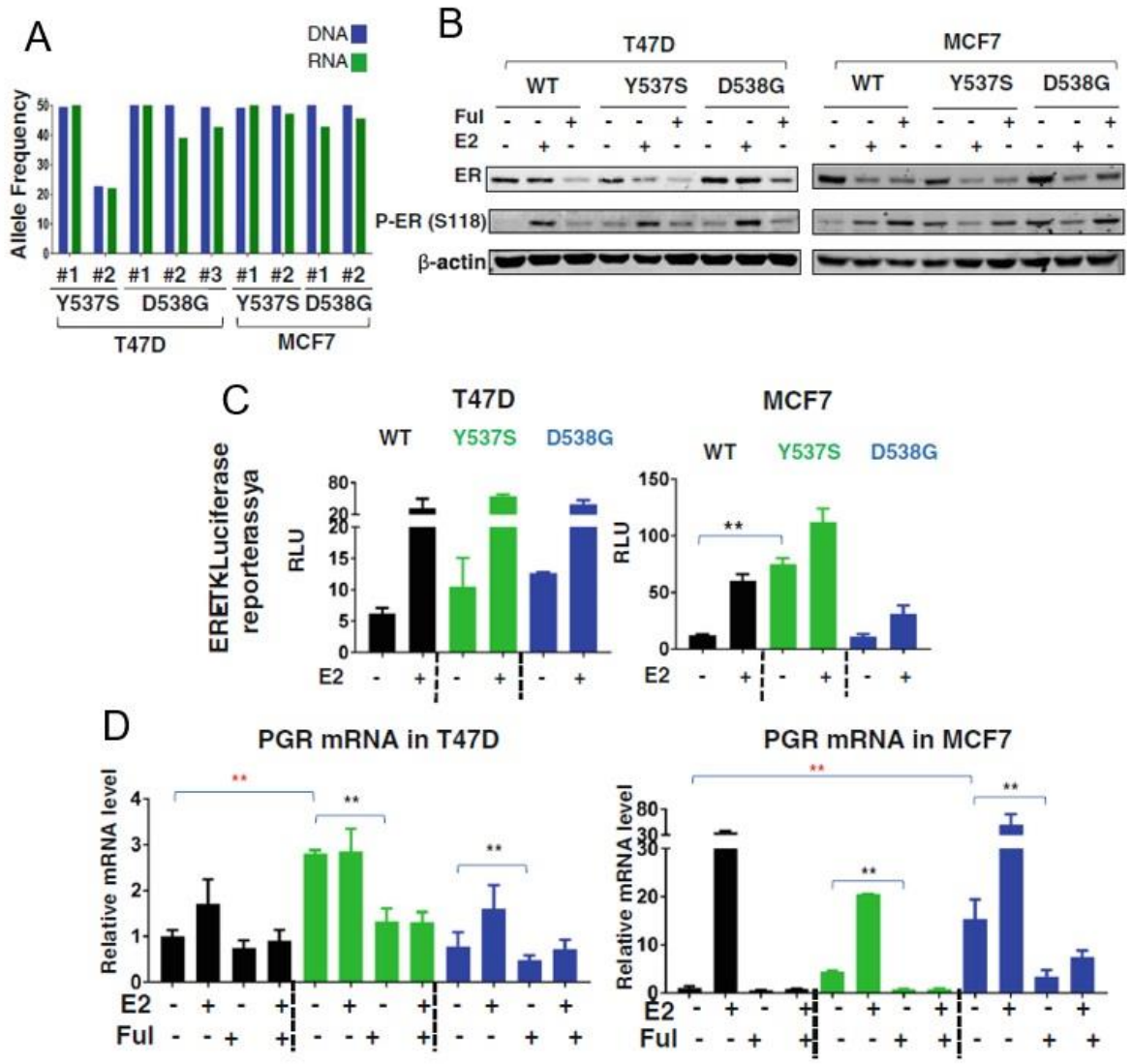



**Figure 3. Generation and characterization of *ESR1* mutant, genome-edited MCF7 and T47D cell line models.**

A. *ESR1* mutation allele frequency in DNA and RNA was determined by digital droplet PCR. This experiment was done once. B. T47D and MCF7 wild-type (WT) or mutant clones were pooled and treated with vehicle, 1 nM estradiol (E2) or 1 μM of fulvestrant (Ful) for 24 h, and lysates were immunoblotted as indicated. The blot is representative of three independent experiments. ER estrogen receptor. This experiment was reproduced twice. C. T47D and MCF7 clones were pooled after hormone deprivation, transfected with ERE-TK, and relative light units (RLU) were determined (Dunnett's test, $^{**}p < 0.01$). The experiment was repeated three times and the figure shows one representative experiment with two biological replicates. D. Hormone-deprived T47D and MCF7 cells were treated with vehicle, 1 nM E2, 1 μM fulvestrant or 1 nM

E2 with 1 μM fulvestrant for 12 h, and RNA was isolated, and RT-qPCR was performed (Dunnett's test for comparison of basal level, Student's t test for comparison of fulvestrant response in the presence of E2, *$p < 0.05$, **$p < 0.01$). The experiment was repeated three times and the figure shows one representative experiment with three biological replicates.

*Fig. 3A is provided by Amir Bahreini, PhD.

We then tested ER transcriptional activity using reporter assays and observed a trend towards increased activity in both T47D mutants, and a significant increase in MCF7 Y537S cells (Fig. 3C). Expression of *PGR* mRNA, a classical ER target gene, was significantly increased in the absence of ligand in T47D Y537S cells (Fig. 3D), and similar data were observed when measuring expression in individual clones (Appendix A Fig. 64). In MCF7 cells, *PGR* was significantly increased in D538G cells. Ligand-independent activation of *PGR* was inhibited with Ful, confirming ER-dependency of the effect. Collectively, these data show overall utility of the models for studying ligand-independent activity of ER mutants, but also provide some evidence for mutation site and cell context-dependent activities.

In line with this, I also generated MCF7 and T47D pcDNA overexpression Y537S and D538G *ESR1* mutant cell models (Appendix A Fig. 65). Immunoblot showed robust expression of both HA-tagged WT and mutant ER. Droplet digital PCR further confirmed the mutation allele frequencies (MAFs) at both DNA and RNA levels (Appendix A Fig. 65A). Although there were above 60% mutant genomic DNA MAFs in all four cell types, I surprisingly found the RNA level MAFs were only below 30% (MCF7) and 20% (T47D), indicating that the sustained plasmids in cells failed to be fully transcripted in mRNA (Appendix A Fig. 65B).

#### 2.3.1.2 *ESR1* Mutant Cells Exhibit Resistance Towards Anti-estrogen/ER Therapies

Y537S and D538G mutant cells showed higher ligand-independent growth compared to WT in both cell lines (Fig. 4A and 4B). The T47D D538G cells showed an additional strong E2 growth response, not seen in Y537S, or in the MCF7 cells. We had recently reported that growth effects can vary dependent on the source of the charcoal-stripped serum[139], and we therefore tested growth in a second CSS lot. We again observed ligand-independent growth of the ER mutant cells, except for T47D-Y537S (Appendix A Fig. 67), suggesting that there is a factor in serum yet to be identified that contributes to ligand-independent growth, and that varies in CSS lots. In addition, the ligand-independent growth in CSS was not discerned in both MCF7 and T47D pcDNA overexpression cell models (Appendix A Fig. 66), likely due to the low amount of mutant ER transcripts shown by ddPCR results in Appendix A Fig. 65.

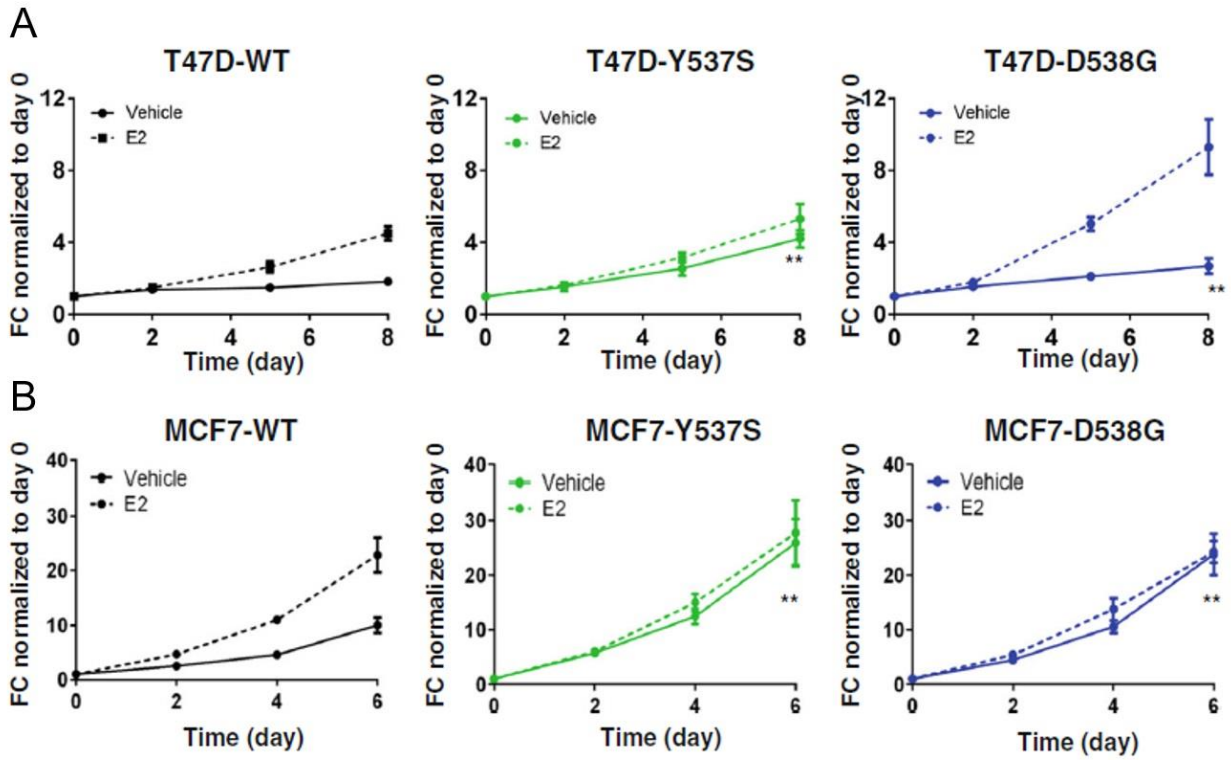
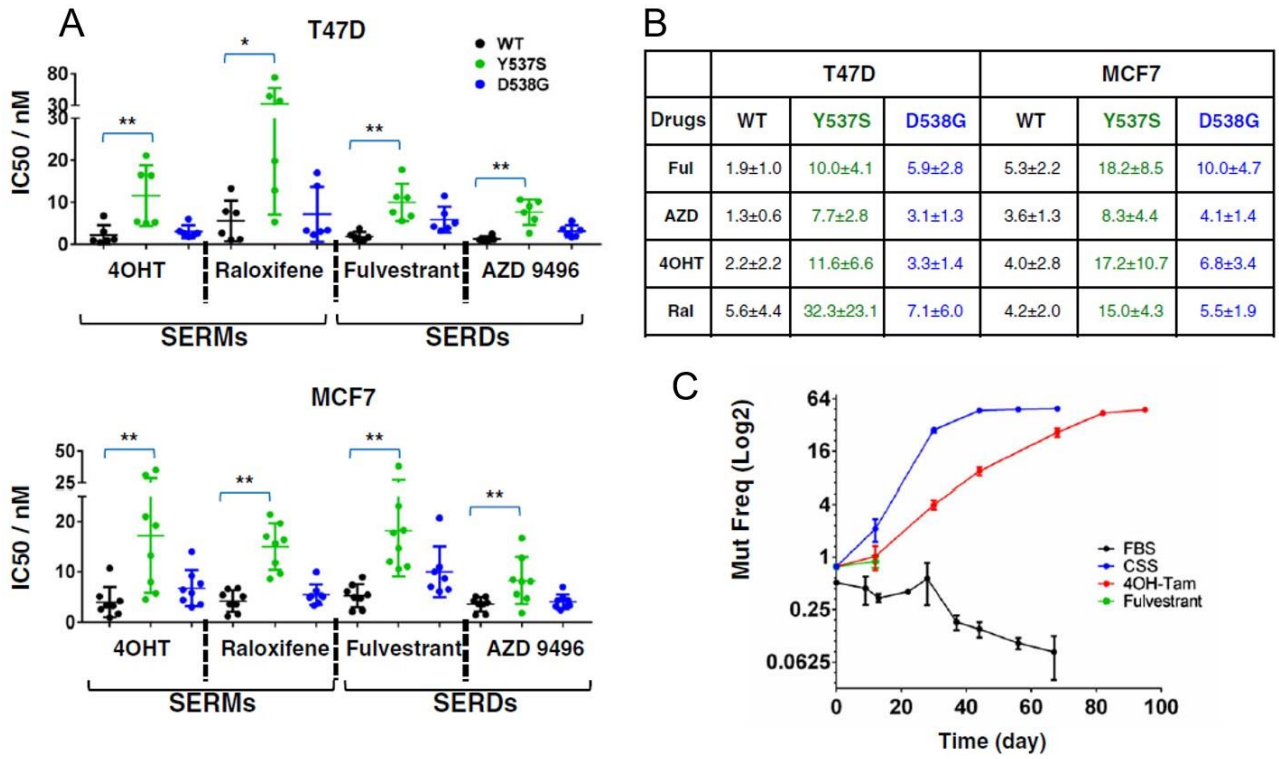



**Figure 4. *ESR1* mutant cells exhibit ligand-independent growth.**

T47D (A) and MCF7 (B) wild-type (WT) or mutant clones were hormone-deprived for 3 days, pooled, treated with vehicle or 1 nM estradiol (E2) for up to 8 days, and cell numbers were quantified by the FluoResporter kit. Growth fold change (FC) was normalized to day 0: **p < 0.01, Dunnett's test, comparison of FC growth between WT and mutant cells on the last day. The experiment was repeated three times with six biological replicates, and similar results were obtained.

Dose–response studies in 2D growth assays with SERMs and SERDs revealed antiestrogen resistance: cells with mutant ER had higher IC50 for the SERMs 4OHT and raloxifene, and the SERDs fulvestrant and AZD9496 compared to WT (Fig. 5A and 5B; Appendix A Fig. 68). We again observed differences between the mutants, with Y537S displaying increased resistance compared to D538G. In addition, AZD9496 was more growth-impeding compared to the other antiestrogens, which was especially obvious in Y537S cells.

Finally, we performed competitive outgrowth experiments in which T47D WT cells were mixed with D538G cells (99:1), and WT-mutant ratio changes overtime were followed by measuring mutant allele frequency using ddPCR (Fig. 5C). In the absence of E2, the mutation frequency of D538G increased until it plateaued at 50% (which represents maximal frequency in the heterozygous D538G clone). A similar competitive advantage of the mutant clone was observed in the presence of 4OHT. In contrast, there was a competitive disadvantage for D538G cells in FBS. In the presence of fulvestrant, all cells died after 2 weeks. In addition, the same trends under CSS, 4OHT+E2 and Fulvestrant+E2 conditions were also reproducible in MCF7 *ESR1* mutant and WT cell coculture experiment, whereas the growth disadvantages in FBS was only discerned in MCF7-D538G cells (Appendix A Fig. 69). Collectively, these data support the previously raised notion [140-142] that SERDs might be more effective against mutant ER compared to SERMs.




| | T47D | | | MCF7 | | |
|---|---|---|---|---|---|---|
| Drugs | WT | Y537S | D538G | WT | Y537S | D538G |
| Ful | 1.9±1.0 | 10.0±4.1 | 5.9±2.8 | 5.3±2.2 | 18.2±8.5 | 10.0±4.7 |
| AZD | 1.3±0.6 | 7.7±2.8 | 3.1±1.3 | 3.6±1.3 | 8.3±4.4 | 4.1±1.4 |
| 4OHT | 2.2±2.2 | 11.6±6.6 | 3.3±1.4 | 4.0±2.8 | 17.2±10.7 | 6.8±3.4 |
| Ral | 5.6±4.4 | 32.3±23.1 | 7.1±6.0 | 4.2±2.0 | 15.0±4.3 | 5.5±1.9 |

**Figure 5. *ESR1* mutant-cells display resistance against selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders.**

Graphical (A) and tabular (B) presentation of half maximal inhibitory concentration (IC50) values that were determined in dose–response curves in wild-type (WT), Y537S and D538G cells treated with 20 pM estradiol (E2) plus varying doses of 4OHT, raloxifene (Ral), fulvestrant (Ful), and AZD9496 in T47D and MCF7 cell lines. Dunnett's test of variance was performed to compare the IC50 values of mutants to WT within each cell line and drug (*$p < 0.05$, **$p < 0.01$). Each dot is representative of the mean of a single experiment with six biological replicates. The experiments were performed six times (T47D) or eight times (MCF7). C. Pooled T47D-WT and T47D-D538G cells were mixed at a ratio of 99:1 and grown in 10% FBS, 10% CSS, 10% CSS + 1 nM E2 + 100 nM 4OHT, or 10% CSS + 1 nM E2 + 30 nM fulvestrant. The mutation allele frequency was analyzed at each passage using digital droplet PCR. This experiment was done once.

*Fig. 5C is provided by Amir Bahreini, PhD.

#### 2.3.1.3 Transcriptomic Analysis Reveals Regulation of Ligand-independent ER Targets and Novel Target Genes

RNA-seq analysis was performed to determine the effect of the mutations on endogenous target gene expression. Analysis of variable genes confirmed that the biological replicates clustered together (Appendix A Fig. 70), and that the mutants are very different from WT in the vehicle setting (Appendix A Fig. 71). A total of 1,198 and 1,327 genes were differentially regulated comparing WT and mutant cells in the absence of ligand in T47D and in MCF7 cells, respectively (FC >2, $p < 0.005$) (Fig. 6A). The majority of the differentially expressed genes were estrogen-regulated in WT clones, supporting the ligand-independent activity of the mutant receptor.

Among the ligand-independent regulated genes were the classic ER target genes *GREB1* and *IGFBP4*. Ligand-independent expression was confirmed in pooled (Fig. 6B) and in individual (Appendix A Fig. 72) mutant MCF7 and T47D cell clones, although we again observed mutation site-specific and cell-line-specific differences in the effects. ChIP analysis revealed increased ER binding to the *GREB1* and *IGBP4* promoters in the absence of ligand in T47D and MCF7 cells (Fig. 6C). *IGFBP4* transcript levels were not increased significantly in T47D-D538G and MCF7-Y537S despite ER recruitment as observed by ChIP, suggesting that promoter occupancy is not sufficient to initiate transcription. Ability to inhibit the ligand-independent expression with fulvestrant (Fig. 6B), and *ESR1* knockdown using small interfering RNA (siRNA) (Appendix A Fig. 73), confirms ER-dependency of such ligand-independent regulation of expression.

Given our observations of mutation site-specific and cell-line-dependent effects on phenotypes and candidate target genes, we quantified the overlap of ligand-independent target

genes between the mutants (within one cell line), and between the cell lines (within one mutant). While there was significant overlap of the ligand-independent target genes when comparing the two mutations (Y537S and D538G) within the individual cell lines (Fig. 6D), there were some unique target genes for both mutants. Despite significant overlap of E2 target genes regulated by WT ER when comparing T47D and MCF7 cells, there was limited overlap when comparing the ER mutant ligand-independent target genes between the two cell lines (Fig. 6D).

The RNA-seq analysis also led to the identification of a set of "novel" target genes ($n = 425$ in MCF7, and $n = 570$ in T47D) that were not E2-regulated in WT cells, but instead were differentially expressed in the *ESR1* mutant clones in the absence of E2 (Fig. 6A). There was significant overlap of these novel target genes between the two mutants within each cell line ($p$ value <2E-16), but there was limited overlap between the different cell lines (Appendix A Fig. 74). Despite the limited overlap, Ingenuity Pathway Analysis (IPA) of the novel genes showed enrichment of metastatic associated phenotypes including "cell movement" (Appendix D Table 9). Genes from these pathways will be candidates for future studies when addressing mechanisms for the metastatic propensity of *ESR1* mutant cells.

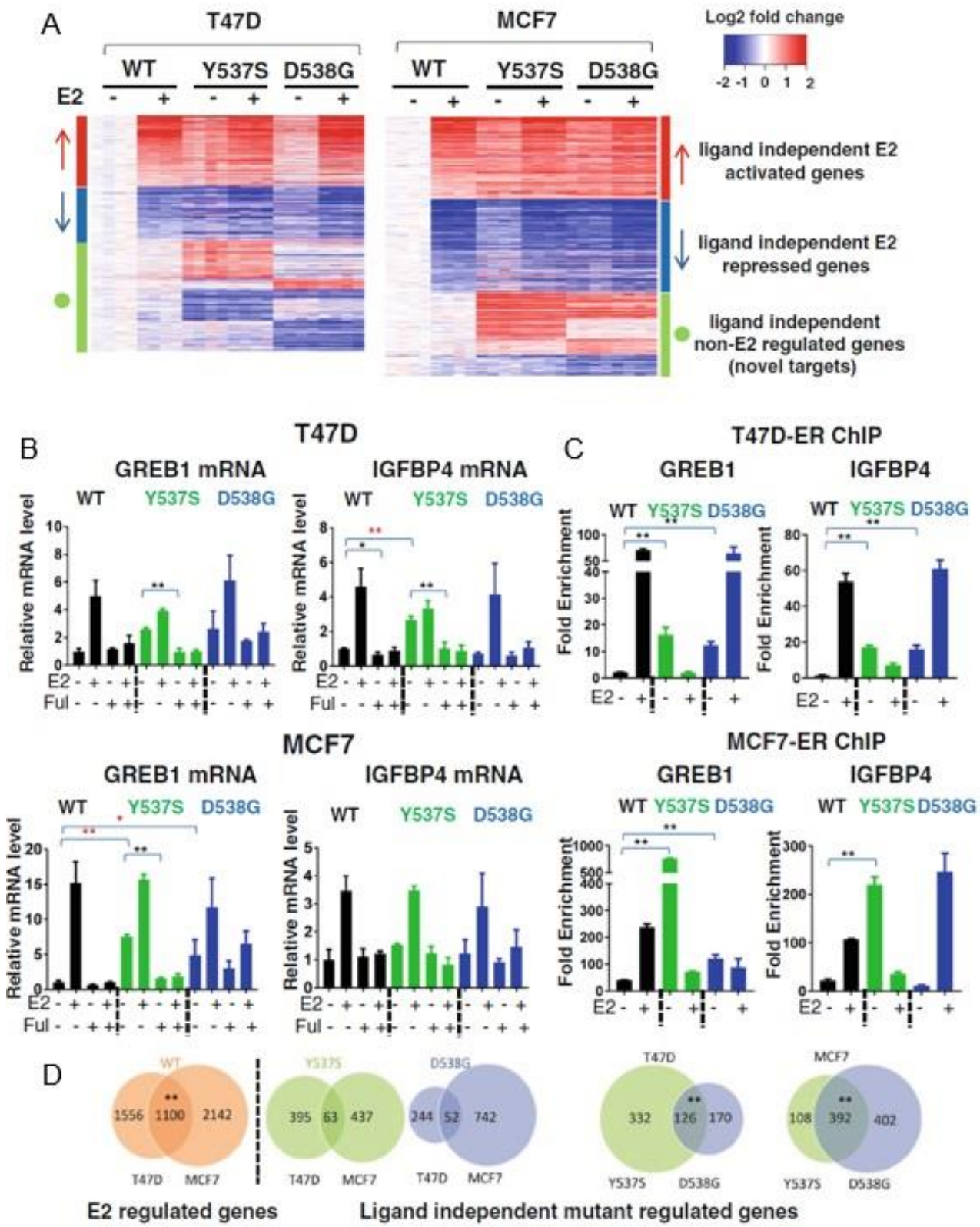

A
T47D
MCF7
WT
Y537S
D538G
E2
Log2 fold change
-2 -1 0 1 2
ligand independent E2 activated genes
ligand independent E2 repressed genes
ligand independent non-E2 regulated genes (novel targets)
B
T47D
GREB1 mRNA
IGFBP4 mRNA
Relative mRNA level
Ful
C
T47D-ER ChIP
GREB1
IGFBP4
Fold Enrichment
MCF7
MCF7-ER ChIP
D
1556 1100 2142
395 63 437
244 52 742
332 126 170
108 392 402
T47D
MCF7
Y537S
D538G
E2 regulated genes
Ligand independent mutant regulated genes

**Figure 6. Genome-wide transcriptomic analysis reveals regulation of ligand-independent estrogen receptor (ER) targets, and of novel target genes.**

A. T47D and MCF7 cell lines were hormone-deprived for 3 days, treated with vehicle (veh) or 1 nM of estradiol (E2) for 24 h, RNA was isolated and RNA sequencing analysis was performed. The heat map shows normalized log2 fold change (FC) of genes differentially regulated in mutants vs. wild-type (WT) in the absence of ligand (FC >2, p value <0.005). The genes are sorted based on E2 regulation in WT (red arrow ligand-independent E2 activated genes, blue arrow ligand-independent E2 downregulated genes, green circle ligand-independent non-E2 regulated genes, i.e. "novel target genes"). B. Hormone-deprived T47D and MCF7 cells were treated with veh, 1 Nm E2, 1 μM of fulvestrant (Ful) or 1 nM E2 plus 1 μM of Ful for 24 h. RNA was isolated, and GREB1 or insulin-like growth factor-binding protein 4 (IGFBP4) expression was assessed by quantitative RT-qPCR (Dunnett's test) for comparison of basal level, Student's t test for comparison of Ful response in the presence of E2, *p < 0.05, **p < 0.01). The experiment was repeated twice with three biological replicates each time. C. Cells were hormone-deprived, treated with 1 nM of E2 for 45 minutes, and chromatin-immunoprecipitation (ChIP) assays were performed on the ER binding sites on GREB1 and IGFBP4 promoters. The data are presented as fold enrichment compared to IgG control (Dunnett's test, **p < 0.01). The experiment was repeated twice with three biological replicates each time and the figure shows a representative experiment. D. The chi-square test was used to assess the statistical significance of overlaps in venn diagrams. Left panel overlap of E2-regulated genes in WT cells between the cell lines (chi-square test, **p < 0.01). Right panel overlap of ligand-independent target genes between different mutations within each cell line and between the two cell lines (chi-square test, **p< 0.01).

*Fig. 6A and 6D are provided by Amir Bahreini, PhD.

### 2.3.2 Upregulation of IRS1 Expression Enhances IGF-1 Response in *ESR1* Mutant Breast Cancer Cells

#### 2.3.2.1 Y537S and D538G Mutant Cells Exhibit Upregulation of IRS1

I next investigated the potential pathways that drive endocrine resistance in *ESR1* mutant cell models. We have previously shown a significant similarity between genes whose expression is altered by estradiol or IGF1[138]. Here, we investigated levels of these same IGF1-stimulated genes in *ESR1* mutant cells, by performing a correlation analysis between the RNA-seq data sets

of *ESR1* mutant cells and the previously reported IGF1 signature panel of 976 transcripts (814 genes)[138]. Replicates from each group of RNA-seq results were compared to the IGF1 signature individually, and pearson correlation coefficient values were calculated to indicate the similarities (Fig. 7A). In the absence of estradiol (E2), gene expression in both MCF7 and T47D *ESR1* mutant cells showed a significantly increased similarity to the IGF gene signature pattern compared to WT groups. Consistent with previous results[138], E2-stimulation led to a higher similarity to the IGF gene signature in all cell line models. I then directly examined genes in the IGF1 signaling pathway (n=39) that showed a fold change > 1.5 ($p < 0.005$) in *ESR1* mutant cells (Fig. 7B). Of note, I selected genes with fold change above 1.5 here but not 2 as described in Fig. 6, in order to obtain a broader target gene range for the downstream functional study selection. I found that *IGF2, IGFBP6, IRS1, IGFBP4* and *INSR* were consistently altered in both *ESR1* mutant cell line models compared to WT cells. *IRS1,* which is an essential adaptor in the IGF1 pathway, was the most consistently upregulated gene in both mutations and cell lines in the entire panel. IGF1R levels were also increased in MCF7 *ESR1* mutant cells. Interestingly, the *INSR* gene exhibited decreased levels in both cell lines (Appendix A Fig. 75A and 75B), perhaps reflecting a secondary negative feedback regulation following increased IGF1R/IRS1 signaling. Moreover, the transcript reads from RNA-seq suggest no difference of IR isoform ratios (IR-A/IR-B) between different mutants of each cell line (Appendix A Fig. 75C). Upregulation of *IRS1* was validated using immunoblotting (Fig. 7C) and qRT-PCR (Fig. 7D). Increased protein levels of *IRS1* were observed in both *ESR1* mutant cell lines under hormone deprived condition, whereas higher IGF1R expression was only seen in MCF7 mutant cells, a result that correlated well with the RNA-seq data. The qRT-PCR results also demonstrated higher basal level of IRS1 in *ESR1* mutant cell lines,

with the exception of T47D-D538G clones where the increase was not significant. Similar data were observed when measuring IRS1 mRNA in individual clones that constitute the pool (Appendix A Fig. 76A). qRT-PCR results indicated that *IRS1* levels were induced by E2 and inhibited by fulvestrant treatment, which is consistent to previous studies characterizing *IRS1* as an ER-induced gene[143, 144]. Decreasing *ESR1* expression with siRNA attenuated the higher *IRS1* levels in *ESR1* mutant cells, confirming the dependency of the *IRS1* upregulation on the genome-edited mutant *ESR1* (Appendix A Fig. 76B).

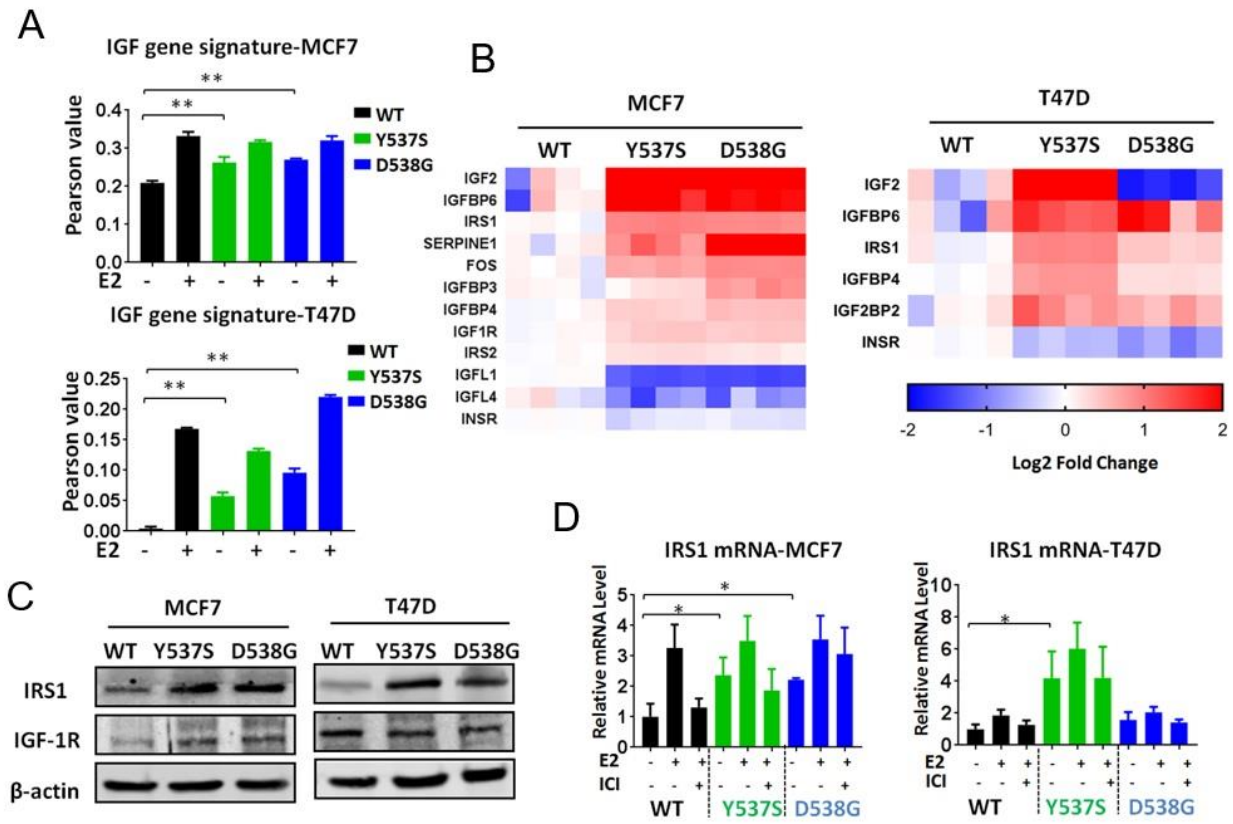



**Figure 7. MCF7 and T47D cells with CRISPR-edited *ESR1* Y537S and D538G mutations exhibit upregulation of IGF signaling components and downstream activity.**

A. RNA-sequencing results from all four replicates of each group were aligned and compared to the IGF gene signature panel. Pearson values from vehicle groups were compared between WT and two mutants with Dunnett's test. B. The heat maps from our previous RNA-sequencing shows normalized log2 FC of IGF1 pathway related genes differentially regulated in mutants vs WT (FC >1.5, p-value<0.005). C. Post-hormone deprived cells were pooled and protein samples were extracted and subjected for immunoblotting with β-actin as the internal control. This experiment was repeated twice and representative data was shown here. D. Hormone-deprived MCF7 and T47D cells were treated with vehicle, 1 nM of E2 or combination of 1nM E2 and 1 µM of fulvestrant for 24 hours. RNA was isolated, and qRT-PCR was performed (Dunnett's test for comparison of basal level * p<0.05, **p<0.01). Representative data were shown here from two independent experiments with three biological replicates.

*Fig. 7A was analyzed by Kevin Levine, MD, PhD.

#### 2.3.2.2 Y537S and D538G *ESR1* Mutant Cells Display Enhanced Proliferation in Response to IGF1

Given the alteration in IGF signaling components, including the constitutive expression of *IRS1*, I hypothesized that *ESR1* mutant cells would show an elevated response to IGF1. I therefore performed time course proliferation assays with or without IGF1 after hormone deprivation and serum-free starvation (Fig. 8A). A hallmark of *ESR1* mutant cells is the greater estrogen-independent growth in charcoal stripped serum (CSS)[110]. In the complete absence of serum, however, the ligand-independent growth of T47D mutant cells was not discerned (Fig. 8A), whereas MCF7 mutant cells still maintained a stronger proliferation rate compared to WT controls, suggesting that the ligand-independent growth of T47D mutant cells depends on additional factors in CSS. When exposed to IGF1, T47D-Y537S cells exhibited a significantly ($p<0.0001$) ~2-fold stronger IGF1 response compared to WT control cells, whereas D538G mutant cells showed only a weak IGF1-induced proliferation that did not reach significance (Fig. 8A, lower panel). In MCF7, both mutations resulted in significant IGF1-mediated induction of proliferation, with D538G mutant cells showing a stronger response (Fig. 8A, top panel).

Dose response studies were performed to evaluate the sensitivity and potency of IGF1 (Fig. 8B). Both mutations resulted in increased potency of IGF1-stimulated proliferation, particularly at concentrations above 1 ng/ml of IGF1 (Fig. 8B, top panel), and this was observed in both cell line models. Examining sensitivity, $EC_{50}$ calculated from the dose response curves suggest no difference between T47D WT and mutant cells, whereas MCF7 Y537S and D538G mutants both showed a 3 to 4 folds decrease in $EC_{50}$ (Fig. 8B, down panel).

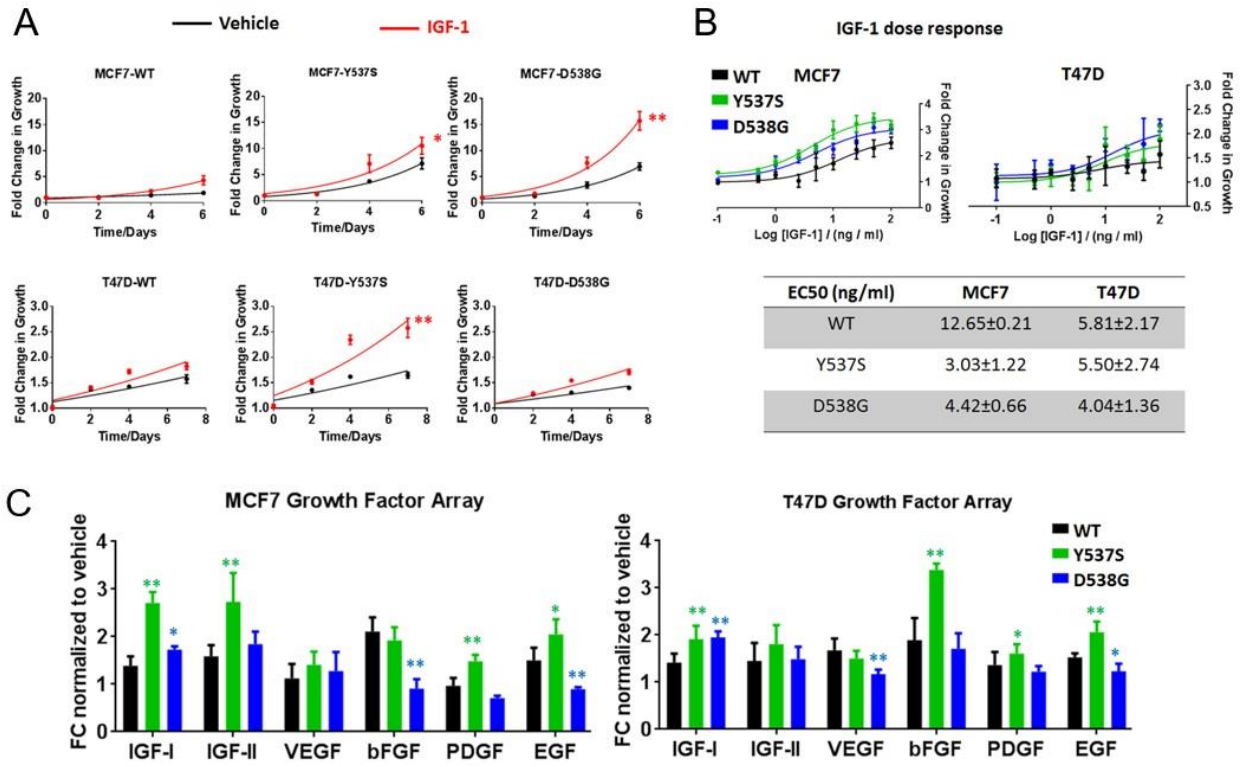



**Figure 8. *ESR1* mutant cells display enhanced IGF1-stimulated growth response.**

A. MCF7 cells and T47D cells were hormone deprived for 3 days, followed by incubation in serum-free medium for 24 hours. Cells were then seeded into 96-well plates and treated with or without 100 ng/ml of IGF1 for up to 8 days. Cell numbers were quantified by FluoReporter kit. Growth was normalized to day 0, and presented as fold change (FC). Two-way ANOVA was performed to compare the time course growth rates between WT and mutant cells. Representative data were shown here from two independent experiments. B. $EC_{50}$ values were determined by IGF1 dose response growth curves following the identical procedure as in A) with IGF1 concentrations as indicated and 20 pM of E2. Mean EC50s were shown on the table with three (T47D) or two (MCF7) independent experiments. Dunnett's test was performed to compare the EC50s of WT and mutant cells. Representative data were shown here from three (T47D) and two (MCF7) independent experiments. C. Hormone-deprived and serum-free starved MCF7 and T47D were treated with 100 ng /ml IGF1, 100 ng/ml IGF-II, 10 ng/ml VEGF, 10 ng/ml FGF-basic (bFGF), 10 ng/ml PDGF-BB or 50 ng/ml EGF for six (MCF7) or seven (T47D) days. Fold changes induced by growth factors were obtained by normalizing to the average of vehicle controls. Dunnett's test was applied to compare the growth factor effects between WT and mutant cells (* $p<0.05$, ** $p<0.01$). Representative data were shown here from two independent experiments.

We next determined whether the enhanced IGF1 response was a unique phenotype in *ESR1* mutant-cells, or whether it is present in other endocrine resistant models. We therefore tested IGF1

response in tamoxifen-resistant breast cancer cells (TamR), and in long-term estrogen deprived (LTED) models, mimicking resistance to aromatase inhibitors (Appendix A Fig. 77A). These analyses showed that there was no significant difference in IGF1 response between parental and the endocrine resistant TamR and LTED models (Appendix A Fig. 77A and 77B). Furthermore, the dose response curves in MCF7 cells indicated that both TamR and LTED cells exhibited reduced sensitivity towards IGF1 exposure – a 4-fold and 18-fold increase in $EC_{50}$ values were detected in the MCF7-LTED and TamR models, respectively (Appendix A Fig. 77C). Together, these results suggest that the enhanced IGF1 response might be a unique property in endocrine-resistant cells with *ESR1* mutations.

I next determined the effect of other growth factors known to play a role in breast cancer progression, in order to determine whether the IGF1 effects are unique, or whether the *ESR1* mutations sensitize the cells to activation of other growth factor receptors. Therefore, I measured growth response of *ESR1* WT and mutant cells to IGF-II, VEGF, PDGF, bFGF and EGF (Fig. 8C). Although I did observe increased response of the Y537S mutant cells to IGF-II, bFGF, PDGF and EGF, consistent increase in both cell lines for both mutations was only observed for IGF1.

#### 2.3.2.3 PI3K-Akt Axis Mediates the Enhanced IGF1 Response in *ESR1* Mutant Cells

I next investigated which downstream signaling pathways mediate the enhanced IGF1 response in *ESR1* mutant cells (Fig. 9). Cells were stimulated with increasing concentrations of IGF1 for 15 minutes, and the phosphorylation status of IGF1R (Y1135) and canonical IGF1 downstream substrates Akt (T308 and S473) and ERK1/2 (T202/T204) were measured (Fig. 9). Consistent with the growth response data, greater IGF1R phosphorylation at Y1135 was observed

in both MCF7 (Fig. 9A, B) and T47D (Fig. 9C, D) and Y537S and D538G mutant cells, particularly at higher concentrations of IGF1 stimulation (80 or 100 ng/ml). The two phosphorylation sites (T308 and S473) of Akt are phosphorylated by different kinases and involved in distinct biological functions[145, 146]. I noted that only pT308-Akt was increased in *ESR1* mutant cells, whereas phosphorylation at pS473-Akt displayed higher basal levels but weaker or no IGF1-induced stimulation. In contrast, p-ERK1/2 (T202/T204) showed a cell line-specific induction, with greater induction in T47D but not in MCF7 *ESR1* mutants compared to WT.

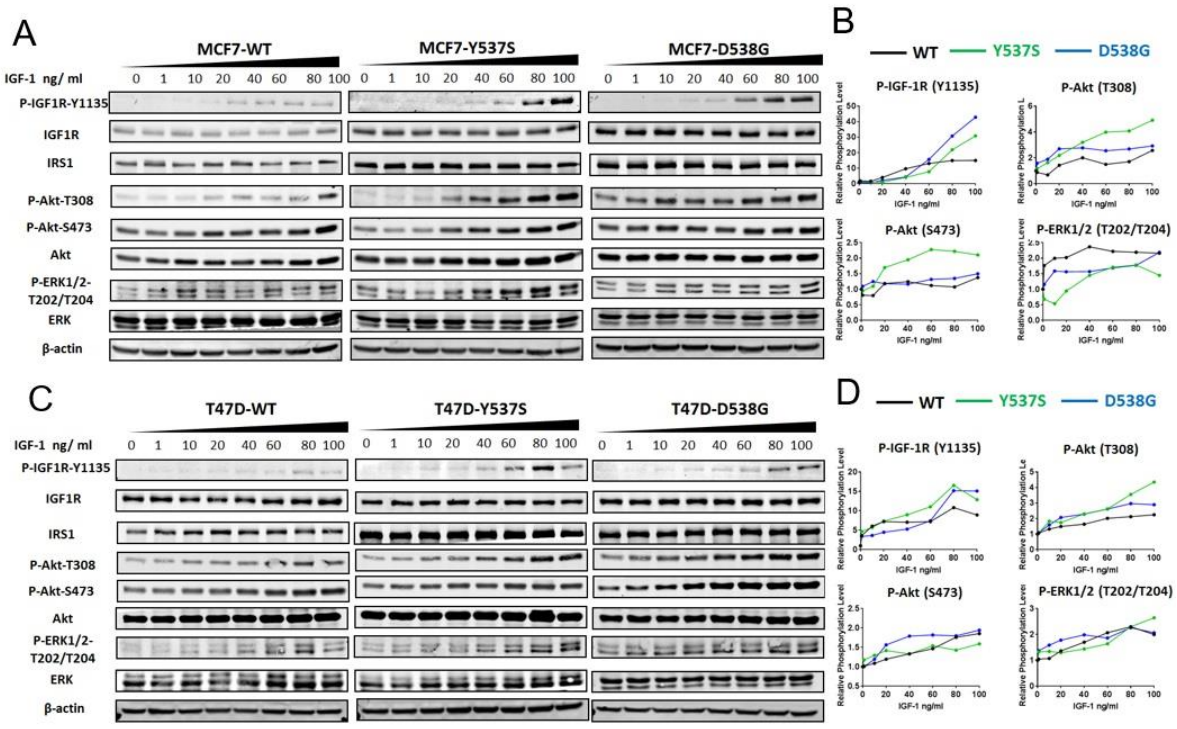



**Figure 9. PI3K-Akt axis mediates the enhanced IGF1 response in *ESR1* mutant cells.**

A) and C) MCF7 and T47D cells were hormone deprived for 3 days and starved in serum-free medium for 24 hours. Cells were then stimulated with various concentrations of IGF1 for 15 minutes. Protein were isolated and immune-blots were performed with antibodies as indicated, and β-actin as loading control. B) and D) The intensities of bands were quantified by ImageJ, and relative phosphorylation levels were calculated by correcting phosphorylation levels to total protein, and then normalizing to vehicle controls. Representative data were shown here from two independent experiments.

A parallel set of time-dependent IGF1 stimulation experiments were performed to compare the rates of IGF1 downstream substrate phosphorylation response in the various cell lines (Appendix A Fig. 78). In T47D WT cells, p-IGF1R (Y1135) was maximal after 45 minutes and declined at 60 minutes, whereas in both Y537S and D538G mutant cells p-IGF1R increased rapidly with a maximal response within only 10 minutes (Appendix A Fig. 78B). In MCF7 cells, similar time-dependent effects were seen, but differences in p-IGF1R were found in *ESR1* mutant cells compared to WT cells (Appendix A Fig. 78A). Consistently, pT308-Akt not pS473 site was differentially affected in cells with both types of *ESR1* mutations. pT308-Akt reached the peak intensity within 5 and 10 minutes of IGF1 stimulation in T47D and MCF7 mutant cells, whereas 10 and 30 minutes were required for the corresponding WT cells. Consistently, phosphorylation of S6 kinase also exhibited a stronger response in both *ESR1* mutant-cells, with a higher basal level in T47D cell line. I also discerned a more rapid ERK response in MCF7-Y537S mutant-cell line but not in other mutant-cells.

Given the observed effects on IGF sensitivity and potency, I next utilized a number of inhibitors to test the dependency of the observed increased IGF1-signaling on the PI3K-Akt axis (Appendix A Fig. 79). Specifically, I used Wortamannin, BX-579, UO126 and Rapamycin to inhibit PI3K, PDK1, MEK and mTOR, respectively. In addition, as a control, I inhibited IGF1R with OSI-906 that, as expected, completely abrogated IGF1-induced growth. Effects of inhibition of PI3K, mTOR, and MEK were cell line dependent, with mTOR inhibitors blocking IGF1-induced growth in *ESR1* mutant T47D cells, but not MCF7 cells. In contrast, inhibition of PI3K and MEK was effective in *ESR1* mutant MCF7, but not in T47D cells. The only inhibitor that consistently resulted in inhibition of IGF1-induced growth in both cell line models was BX-579,

an inhibitor of PDK1, a kinase that specifically catalyzes T308-Akt phosphorylation, which is consistent to the western blot results in Fig. 9, where T308-Akt uniquely showed enhanced response in *ESR1* mutant cells under IGF1 stimulation. .

In addition, to further confirm this signaling enhancement is a unique mechanism for *ESR1* mutation, I examined the status of IGF-1 downstream regulators in TamR and LTED models. TamR and LTED cell models showed decreased IGF-1 induced phosphorylation of IGF1R (Y1135), IRS1 (S320) and Akt (T308 and S473) compared to parental controls (Appendix A Fig. 80A), which is consistent with previous studies using these cells and tamoxifen-resistant breast cancers[147, 148]. In contrast, an identical IGF1 stimulation triggered consistent and robust enhanced IGF1 downstream substrate phosphorylation in *ESR1*-mutant cells (Appendix A Fig. 80A).

#### 2.3.2.4 Knockdown of IRS1 Attenuates the Enhanced IGF1 Response in *ESR1* Mutant Cells

Given the consistent and high upregulation of IRS1 in the *ESR1* mutant cells, I next performed knockdown studies to test the functional significance of IRS1 in the increased IGF1 response. Growth curves demonstrated that loss of IRS1 reduced IGF1-induced proliferation in both WT and mutant cell lines, with stronger effects in the mutant cells (Fig. 10A).

To investigate the corresponding signaling changes, I studied IGF1-induced phosphorylation of downstream targets in the absence and presence of IRS1 (Fig. 10B). The immunoblotting of scramble controls correlated strongly to the concentration and time-gradient stimulation results in Figure 9. As a positive control, knockdown of IGF1R markedly reduced

IGF1-induced activation of IRS1 and Akt. Intriguingly, knockdown of IRS1 resulted in diminished IGF1R phosphorylation in all groups, indicating a yet to be identified feedback mechanism (Fig. 10C). As a consequence, the downstream activation of Akt (both S473 and T308) and ERK1/2 (T202/T204) were blocked, albeit the effect on p-ERK1/2 was weaker in MCF7 normalized to scramble control groups. The enhanced phosphorylation of AKT (T308) was also diminished in MCF7 cell line after IRS1 knockdown.

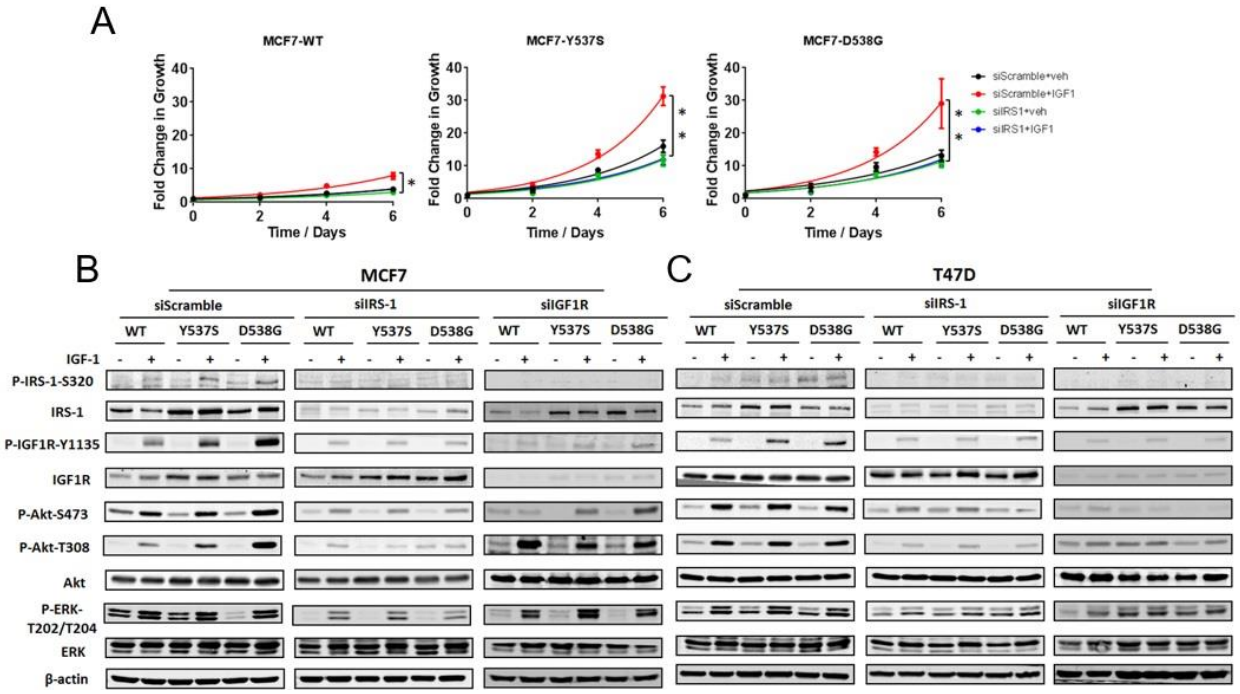



**Figure 10. Knockdown of IRS1 attenuates the IGF1 response in *ESR1* mutant cells.**

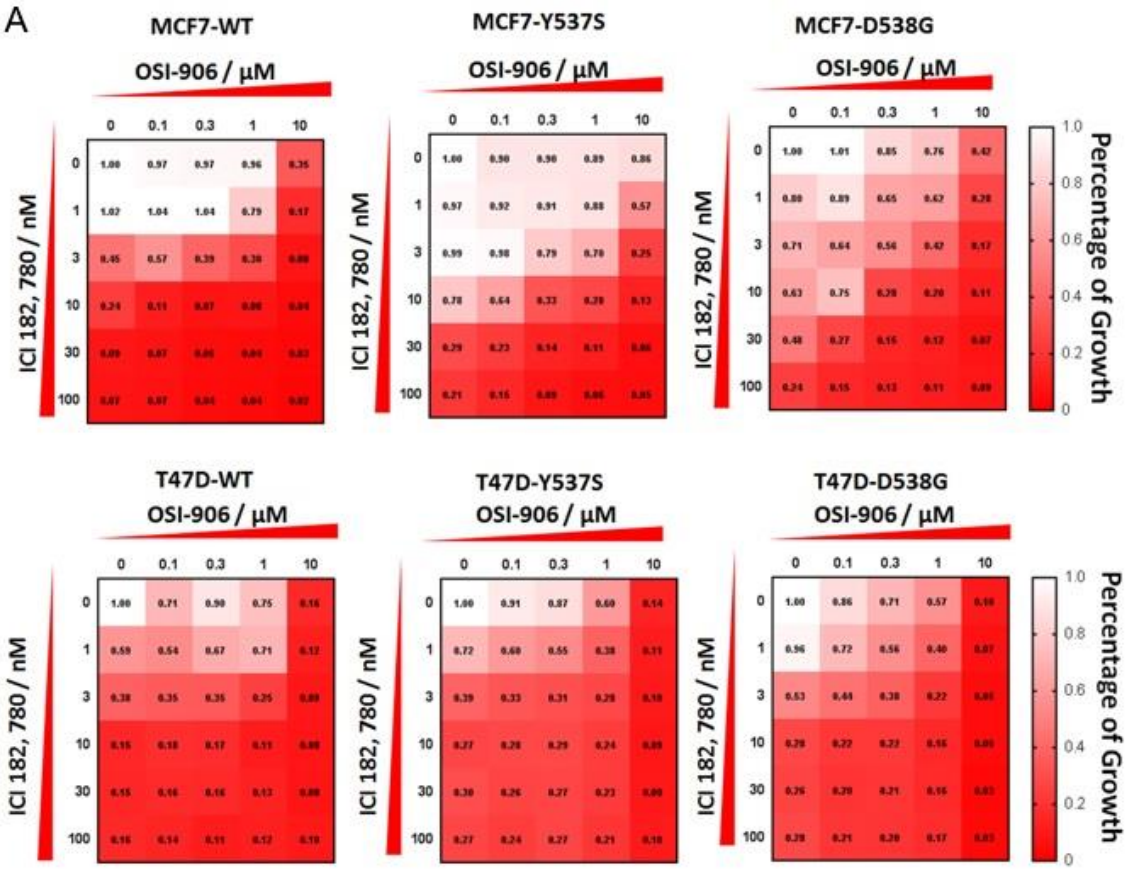
A) IRS1 or scramble siRNAs were reverse transfected into hormone-deprived MCF7 for 24 hours. Cells were then starved in serum-free medium and seeded into 96 well plates with 2500 (MCF7) cells per well and treated with or without 100 ng/ml IGF1 for up to 6 days. Cell numbers were quantified by FluoReporter kit. Cell numbers in scramble and IRS1 knockdown groups with IGF1 simulation were compared by student's t-test (* $p<0.05$, ** $p<0.01$). This experiment was done once. B) MCF7 and T47D cells were hormone deprived and transfected with IRS1, IGF1R or scramble siRNA. Cells were then starved in serum-free medium and treated with 100 ng/ml IGF1 for 15 minutes. Immuno-blots were performed with antibodies as indicated. Representative data were shown here from two independent experiments.

#### 2.3.2.5 Combination of Anti-IGF1R (OSI-906) and Anti-ER (fulvestrant) Exhibits Stronger Synergism in *ESR1* Mutant Cells

Finally, I asked whether activation of the IGF1 pathway altered *ESR1* mutant cells response to small molecule inhibitors targeting IGF1R. First, I examined the effect of BMS-754807 and OSI-906 on 2D growth of *ESR1* mutant cells (Appendix A Fig. 81). MCF7 cells were more sensitive to IGF1R inhibition compared to T47D. Both MCF7 Y537S and D538G mutant cells exhibited greater $IC_{50}$ towards both IGF1R inhibitors, with Y537S showing a stronger resistance. The similar resistance trend was also detected in response of T47D cells to BMS-754807, however, no obvious differences were detected with OSI-906.

Combination therapies co-targeting IGF1R and ERα in breast cancer have been developed and evaluated. With the respect of this therapeutic strategy, I examined the combined effect of OSI-906 and fulvestrant (Fig. 11A). Therefore, I treated WT and *ESR1* mutant T47D and MCF7 cells with increasing concentrations of fulvestrant (1 nM to 100 nM) and OSI-906 (0.1 μM to 10 μM). Effects on growth rates are shown in Figure 5A, and combination index values are shown in Fig. 11B. These data show that under half effective concentrations ($EC_{50}$), OSI-906 and fulvestrant display synergistic effects (CI<1) in both mutant cell line models, in both MCF7 and T47D, with the exception of T47D-Y537S cells that showed no effect (CI=1). In contrast, antagonistic effects (CI>1) were observed in both WT cell line models. At $ED_{75}$ and $ED_{90}$, the drugs displayed synergistic activities in all mutant cell line models, with consistently stronger effects compared to WT cells.

A




B

| Combination index | ED50 | ED75 | ED90 |
|---|---|---|---|
| MCF7-WT | 9.5 ± 3.04 | 1.39 ± 0.49 | 0.21 ± 0.08 |
| MCF7-Y537S | 0.4 ± 0.22 | 0.34 ± 0.15 | 0.29 ± 0.1 |
| MCF7-D538G | 0.39 ± 0.23 | 0.3 ± 0.19 | 0.25 ±0.13 |
| T47D-WT | 1.54 ± 0.69 | 0.82 ± 0.24 | 0.64 ± 0.2 |
| T47D-Y537S | 1.02 ± 0.48 | 0.71 ± 0.32 | 0.64 ± 0.06 |
| T47D-D538G | 0.34 ± 0.08 | 0.34 ± 0.01 | 0.4 ± 0.1 |

**Figure 11. Combination treatment with IGF1R inhibitors and fulvestrant exhibits synergistic effects in *ESR1* mutant cells.**

A. Hormone-deprived MCF7 and T47D clones were pooled and treated with the corresponding combined concentrations of OSI-906 and fulvestrant in the presence of 20 pM of estradiol. Cell numbers were quantified with FluoReporter kit after six (MCF7) and seven (T47D) days. Percentages of growth were calculated by normalizing to the averages of the vehicle control groups. The data were calculated by combining two independent experiments. B. Combination index mean values at $ED_{50}$, $ED_{75}$ and $ED_{90}$ were calculated based on two independent experiments.

## 2.4 Discussion

In the first section of this chapter, we report the generation, characterization and transcriptome analysis of genome-edited "knock-in" models of the most frequent *ESR1* mutations, Y537S and D538G. As recently reported by others[91, 123-126], our data show that the mutant receptors gain ligand-independent transcriptional activity, and this is associated with ligand-independent growth and endocrine resistance. Our study is the first comparing the effect of two mutations, in two different genome-edited breast cancer cell lines, allowing us to conclude that there are mutation-dependent and context-dependent differences.

The majority of previous reports have employed cell lines transfected with ER constructs, potentially resulting in effects associated with non-physiological overexpression of the receptor. An example is ER phosphorylation, which we observed in our models; however, this was not at the high levels previously described in cells transiently transfected with mutant ER[91, 123-126]. Harrod et al. recently reported a Y537S clone generated with CRISPR technology[128], and similarly observed an increased ligand-independent Ser118 phosphorylation. However, the effect was weaker than estrogen-induced phosphorylation in the WT control cells, and there was no significant difference between WT and mutant cells in response to a drug inhibiting the one of the kinase signaling pathways causing Ser118 phosphorylation. Thus, additional studies are necessary to understand whether there is a causative role of Ser118 phosphorylation in the *ESR1* mutant-associated phenotypes. In addition, we observed increased Ser118 phosphorylation after E2 treatment in T47D *ESR1* mutant cells, which is decreased in MCF7 models. Given S118 site phosphorylation is induced by ERK1/2 or CDK7 [149], it is possible that E2 treatment causes

differential kinase activities in the two cell lines with *ESR1* mutations. Of note, despite the differential S118 phosphorylation, E2 treatment does not lead to increased growth ratios in all *ESR1* mutant cell models except T47D-D538G as shown in Fig. 4.

We observed ligand-independent transcriptional activity of ER in reporter assays, in expression analysis of candidate genes, and in our genome-wide transcriptomic study. Under our experimental conditions, the magnitude of the effect was larger on expression of endogenous candidate genes, such as *PGR* and *GREB1*, compared to effects using the ERE-TK reporter plasmid. This was especially obvious for the D538G mutant in MCF7 cells, where we failed to observe reproducible effects on the ERE-TK reporter, while the same experimental conditions revealed strong induction of endogenous target genes. These data might have implications for the assay design for identification of drugs targeting mutant ER.

The transcriptomic studies identified a number of growth factors and cytokines that were regulated in a ligand-independent manner in the *ESR1* mutant cell lines. These included insulin-like growth factor 2 (*IGF2*), a number of wnt ligands, *CXCL12*, and *IL20*. Future studies will address if and how these factors can contribute to ligand-independent growth through autocrine signaling. Of note, the gene expression analysis also revealed novel target genes that were not regulated by E2 in WT control cells. The number of novel genes was significantly higher in the MCF7-Y537S clone described by Harrod et al.[128]. Many factors might induce the different amount of novel target genes between the two models including but not limited to original of parental cell strains, number of clones generated and sequenced and RNA-seq data process pipelines. Additional studies are necessary to decipher whether these genes are genuine ER target genes, as a result of potential gain-of-function of the mutant receptor.

Our studies show partial resistance of the mutant ER cells to SERMs and SERDs, as measured by IC50 in growth assays. Of note, the magnitude of resistance was dependent on the cell line and mutation site, with Y537S having significantly stronger resistance compared to D538G, similar to that recently reported by Mao et al.[127]. In general, SERDs were more effective than SERMs, with the novel oral SERD AZ9496 having the highest efficacy when comparing the drugs. Supporting the notion of relative SERD efficacy in *ESR1* mutant disease are our mixing experiments in which WT:mutant cells (99:1) do not survive in the presence of fulvestrant, while the mutant cells outgrow the WT cells in the presence of tamoxifen, or in the absence of ligand, in CSS. This is further supported by retrospective analysis of clinical trial samples, recently reported in two independent studies[141, 142]. We have recently opened a trial in which this question will be addressed in a prospective study (NCT02913430) by comparing the response of *ESR1* mutant patients towards CDK4/6 inhibitor plus fulvestrant or tamoxifen.

Finally, we observed significant differences in the effects of mutant ERs between Y537S and D538G, and between T47D and MCF7 cells. For example, fulvestrant-mediated degradation of D538G was less pronounced and E2-induced transcriptional effects and growth response were stronger in D538G, compared to that seen in Y537S. In general, Y537S had stronger endocrine resistance than D538G, in line with clinical data reported from the BOLERO trial in which patients with Y537S mutant tumors had shorter overall survival compared to those with the D538G mutation[150]. Phenotypic differences between the mutants could, at least in part, explain the co-existence of more than one mutation within the same tumor, which has previously been reported[108, 117, 135, 151]. It is important to decipher if and how co-existing *ESR1* mutant-cells interact, and if such interaction provides the tumor with an evolutionary advantage compared to

single *ESR1* mutant tumors. It is likely that tumors that represent genetic heterogeneity at the *ESR1* locus may differentially respond to antiestrogen treatments compared to the tumors with a single mutation in the *ESR1* gene. The ultimate goal of the research on *ESR1* mutations is to identify treatments that show efficacy in *ESR1* mutant-tumors, and we should expect that such treatment might depend on the specific mutation(s).

These data suggest that there are significant mutation-specific effects that need to be accounted for when determining the effect of mutation on progression in the clinical setting, and potentially in drug development. We also observed cell-line-dependent effects, for example, ligand-independent growth was more obvious in MCF7 compared to T47D cells. Cell-line-dependent effects have previously been described for the study of other mutations [152], and future studies need to address if and how this relates to inter-tumor heterogeneity with respect to the effects of *ESR1* mutation.

In the second section of this Chapter, I focused on investigating the mechanisms underlying endocrine resistance in *ESR1* mutant cell models based on the transcriptomic characterization from the first section. Recent data indicate that mutations in the ligand binding domain of *ESR1* confer constitutive ligand-independent gene expression, and that this may lead to endocrine resistance and poor survival[153, 154]. Using two breast cancer cell lines that are genome-edited with two of the most common *ESR1* mutations, I have characterized augmented IGF1 response in these cell lines, which is partially attributed to the upregulation of IRS1. Elevated levels of IRS1 in *ESR1* mutant-cells increased IGF1 potency through the PI3K-Akt axis, and triggered enhanced IGF1-induced growth. Of note, employment of an IGF1R inhibitor (OSI-906) in combination with a selective estrogen receptor degrader (fulvestrant) showed synergistic growth inhibitory effects in

*ESR1*-mutant cells yet showed antagonism in wild-type MCF7 cells. This study sheds light on a novel putative therapeutic strategy for *ESR1* mutant breast cancers (Fig. 10).

Initial reports from Gelsomino et al. identified enhanced IGF1R-ER cross-talk in breast cancer cells transfected with mutant *ESR1*, with a focus on the higher basal IGF1R activation levels in *ESR1* mutant cells[130]. I find a similar increase in IGF1R-ER crosstalk in genome-edited cells, corroborating the study of Gelsomino et al., however, there are several important differences. Gelsomino et al. found that IGF1R and IRS1 levels were unaltered in *ESR1* transfected cells, despite these being estrogen-regulated genes. In contrast, I found elevated IGF1R and IRS1. One potential reason for these discrepancies is the difference in cell model construction, perhaps mainly the CRISPR versus transfection, with CRISPR heterozygous mutant clones more closely mimicking the genetic alteration in human breast cancers[155, 156]. In addition, our findings were shown in two cell lines with two distinct high-frequent mutations allowing us to identify different context-dependent effects e.g. greater IGF sensitization MCF7 mutant *ESR1* clones[112, 152]. This pattern matched the observation of IGF1R upregulation in MCF7 mutant cells according to RNA-seq, which in part increases ligand activation and downstream signaling.

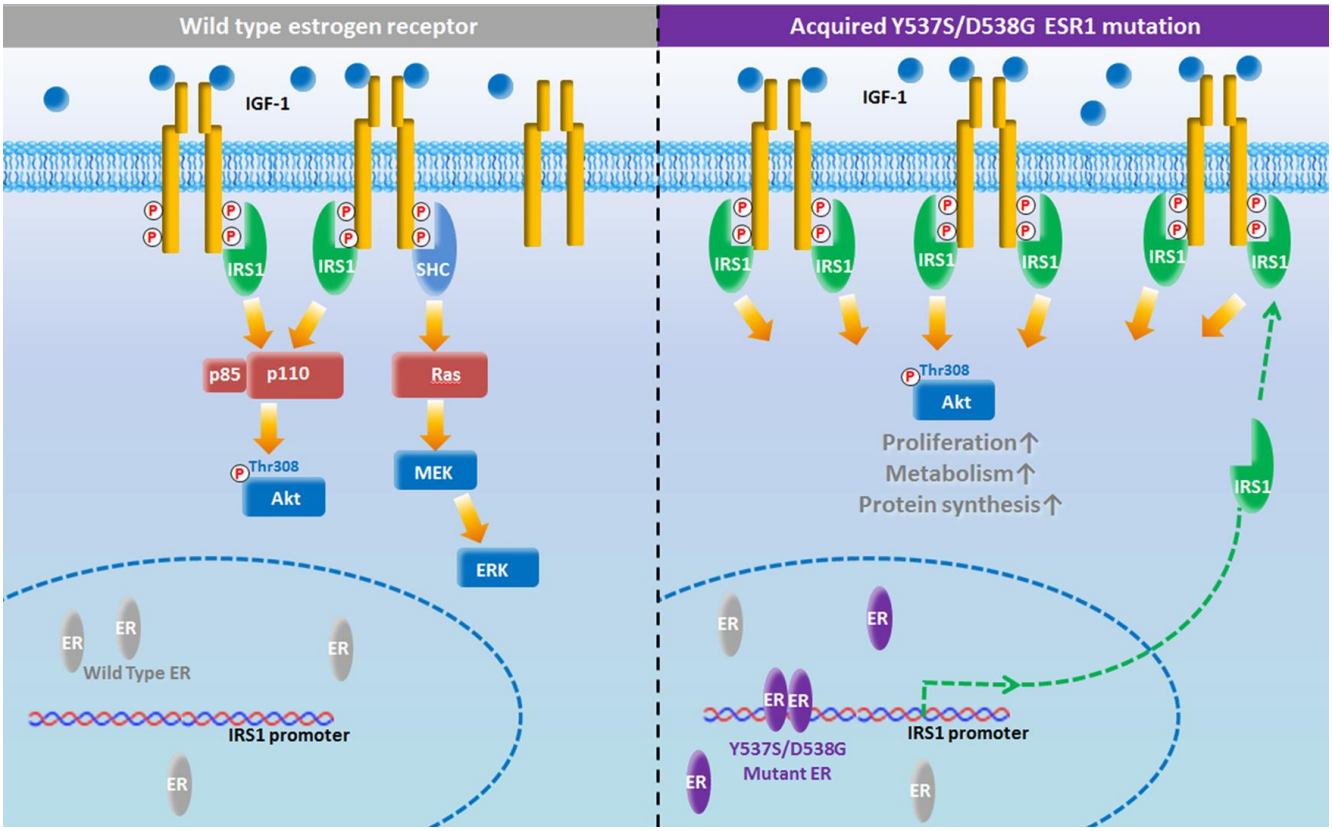



**Figure 12. Schematic view of enhanced IGF1 response in *ESR1* mutant cells.**

In *ESR1* WT cells, IGF1 activates IGF1 receptor, which recruits either IRS1 or Shc to increase PI3K-Akt and/or MAPK pathways. With the acquisition of Y537S or D538G mutations, mutant ER binds to IRS1 promoter and increases IRS1 expression levels. Upregulation of IRS1 further amplifies the IGF1 response via PI3K-Akt axis, and thus enhances cell proliferation.

According to our previous RNA-seq analysis, although estrogen-independent gene expression patterns were detected in all cell models, genes overlapping between Y537S and D538G mutant-cells were only 20.1% and 43.5% of all the constitutively expressed genes in T47D and MCF7 respectively, suggesting mutation-specific ligand-independent genes and their consequential phenotypes. For instance, it is interesting to observe the enhanced bFGF response in T47D-Y537S and EGF response in Y537S mutant exclusively of both cell lines (Fig. 8C). It could be presumably explained by either the upregulation of signaling downstream kinases

(*PRKCA*) or transcriptional factors (*STAT5A* and *MYC*) particularly under these signaling receptors. I also discerned higher basal phosphorylation levels in D538G but limited levels in Y537S mutants of MCF7 cell line. This may also be explained by either the exclusively downregulation of *TIMP3* (a MMPs inhibitor), or upregulation of *GAB2* (a signaling adaptor) in MCF7-D538G line, which potentially enhances the AKT basal levels through integrin or ER crosstalk.

Intriguingly, I found selective IGF-1 induced phosphorylation of S473 on AKT in *ESR1* mutant clones. Akt sites T308 and S473 sites are phosphorylated by PDK1 and mTORC2 respectively[157]. Previous studies have identified site-specific regulatory mechanisms of Akt phosphorylation and downstream signaling[157]. The Ser473 site phosphorylation is reported to be exclusively regulated by integrin-linked kinase (ILK), PKCα, and other RTK activation[158]. It is possible that *ESR1* mutant cells have these pathways activated, which increase basal pS473-Akt. Moreover, the mutant or cell line specific upregulations of kinase components genes (*DEPTOR*)[159] or signaling adaptor genes (*GAB2*)[160] will potentially enhances the basal pS473-Akt.

Additionally, the first section of this chapter[110] revealed that D538G mutation of T47D cell line exhibits relatively lower ligand-independent growth rate but stronger estradiol response after hormone deprivation, which indicates a weaker constitutive ligand-independent activation gene pattern in this particular line, including the group IGF-1 pathway genes shown in Fig. 7B. As a consequence, the IGF-1-induced proliferation was consistently weaker in T47D-D538G cell line, compared to other *ESR1* mutation cell lines (Fig. 8A).

Furthermore, as a ligand of insulin receptor (IR), IGF-1 has the capacity to activate IR[161, 162]. Interestingly, IR consistently decreased in all the mutant lines at both RNA (Appendix A Fig. 75A) and protein levels (Appendix A Fig. 75B), implicating that stronger growth response dominantly depends on IGF1R activation in *ESR1* mutant cells. Similarly, the inhibitory effects of OSI-906 also mainly rely on the block of IGF-1 activity rather than IR, even IR serves as one of the targets of OSI-906[163]. The equivalent IR isoform ratios (IR-A/IR-B) between different mutants of each cell line (Appendix A Fig. 75C) rules out the possibility that the altered IR isoform ratios leads to differentiated IGF-II response (Fig. 8C).

Finally, our data indicate that the enhanced IGF1 response is exclusive to *ESR1* mutant cells, and not found in TamR and LTED cells, indicating that it is not a general endocrine resistance mechanism. Consistent with this, Fagan et al. reported the loss of IGF1R expression and IGF1 response in tamoxifen-resistant cell models[148]. These phenotypic distinctions further stress the need to dissect out the specific mechanisms of resistance and suggest that precise treatment of endocrine resistance may require the monitoring of *ESR1* mutations, where IGF1R inhibitors may have effect. To this end, further analysis examining gene profiles and identify novel pathways, together with clinical archived materials, is warranted.

# 3.0 Hotspot *ESR1* Mutations Are Novel Drivers of Breast Cancer Metastasis

## 3.1 Introduction

More than 70% of breast cancers express the estrogen receptor-α (ER). Antiestrogen therapy, including either depletion of estradiol (E2) by aromatase inhibitors (AIs) or antagonizing ER activity by Selective Estrogen Receptor Modulators (SERMs) or Degraders (SERDs), is the mainstay for ER+ breast cancer treatment. Development of endocrine resistance, however, remains a large clinical and socioeconomic challenge[164, 165].

Somatic base pair missense mutations in *ESR1* are enriched in 20-30% of endocrine-resistant metastatic breast cancer[108, 166, 167], and can be detected in the blood of patients with advanced disease[131, 168]. Recent work from us and others has uncovered a crucial role for *ESR1* hotspot mutations in driving constitutive ER activity and decreased sensitivity towards ER antagonists[109-111]. Moreover, structural investigation of the two most frequent mutations, Y537S and D538G, has demonstrated that *ESR1* mutations stabilize helix 12 (H12) in an agonist conformation, thereby providing a mechanistic explanation for the constitutive activity[169].

The identification of *ESR1* mutations in endocrine resistant metastatic breast cancer suggests that mutant ER could play a role in conferring metastatic capacities in addition to its known function in mediating endocrine resistance. Recent *in vivo* studies showed that overexpressed mutant ER can induce metastasis of MCF7 and T47D cells[117, 129], and some

limited *in vitro* studies showed a gain of cell motility[126, 170] and growth in 3D[130]. Clinically, ligand binding domain *ESR1* mutations correlate with poor outcomes in patients with advanced disease[150, 168, 171]. Thus, there is an urgent need to decipher the mechanistic underpinnings of the potential role(s) of mutant ER in the metastasis process, in order to allow better identification of therapeutic vulnerabilities in *ESR1* mutant tumors.

Previous transcriptomic profiling performed by us and others has identified context-dependence of *ESR1* mutations, as well as significant differences between the two most frequent mutations, Y537S and D538G[109, 110, 117]. Differentially expressed genes vary widely between allele-specific mutations in different cell lines and are largely distinct from the estradiol-dependent wild-type ER transcriptome. Similarly, comparison of the wild-type and mutant ER cistromes has also revealed similar context-dependence and allele-specific effects on ER recruitment[109, 117]. These findings imply, with the high degree of molecular diversity that exists in tumors and patients, that somatic *ESR1* mutations might also have the potential to trigger different metastatic phenotypes. However, this phenomenon has not yet been systematically investigated.

In the results presented in this section, we explored metastatic ER gain-of-function phenotypes in genome-edited cell line models, guided by transcriptomic changes in clinical samples harboring *ESR1* mutations. We identified mechanisms underlying context and allele-specific metastatic phenotypes, and subsequently confirmed the alterations of a number of potential therapeutic targets in metastatic biopsies. We are optimistic that our systematic bedside-to-bench-to-bedside approach presented here will ultimately result in improved metastasis-free outcomes for patients with ER+ tumors.

## 3.2 Materials and Mthods

### 3.2.1 Human Tissue Studies from Women's Cancer Research Center (WCRC) and Charite cohort*

All the patients enrolled were approved within IRB protocols (PRO15050502) from University of Pittsburgh and Charite Universitaetsmedizin Berlin. Informed consent was obtained from all participating patients. Biopsies were obtained and divided into distant metastatic and local recurrent tumors. Genomic DNA was isolated from FFPE samples using Qiagen's All-prep Kit (#80234). *ESR1* mutations status was detected with droplet digital PCR towards Y537S/C/N and D538G mutations in *ESR1* LBD region pre-amplified products as previously reported[131]. cDNA samples synthesized from RNA were used for ddPCR screen for three local recurrent samples due to poor DNA quality. For the recurrence free survival comparison, patient with RFS=0 were excluded.

For the 54 ER+ metastatic tumor samples, genomic profiles were determined based on tumor RNA sequencing from other colleagues' projects. Specifically, the sample processing and data set generation of the bone metastasis cohort were conducted by Nolan Priedigkeit, MD, PhD and Rebecca Watters, PhD. Data set of brain metastasis cohort was from Nolan Priedigkeit, MD, PhD. RNA extraction and sequencing for the GI and ovarian metastases was performed as previously described for our brain and bone metastases cohorts and was completed by Michelle Boisen, PhD, Ahmend Basudan, PhD and Esther Elishaev, MD [172, 173]. Briefly, 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. Peter Lucas MD, PhD served as the pathologist for the brain, bone, and local recurrences cohorts. RNA was extracted from FFPE tissue using Qiagen's All-Prep Kit, and library preparation performed using Illumina's TruSeq RNA Access Library Preparation protocol. Transcript counts from all samples were quantified with Salmon v.0.8.2[174] and converted to gene-level counts with tximport[175]. The gene-level counts from all studies were then normalized together using TMM with edgeR[176]. Log2 transformed TMM-normalized counts per million: log2 (TMM-CPM + 1) expression values were used for the analysis.

*ddPCR on part of the metastatic tumors were performed by Amir Bahreini, PhD and Peilu Wang, MD. The RNA-seq data sets from 54 ER+ breast cancer metastatic samples were generated and processed by Nolan Priedigkeit, MD, PhD, Rebecca Watters, PhD, Ahmed Basudan, PhD, Michelle Boisen, PhD, Ahmend Basudan, PhD, Esther Elishaev, MD and Peter Lucas MD, PhD.

### 3.2.2 Droplet Digital PCR (ddPCR)

Detailed method of ddPCR has been described in Chapter 2 (section 2.3.5).

### 3.2.3 Cell Culture*

Maintenance of genome-edited MCF7 and T47D *ESR1* mutant cell models were described in the material and methods section in Chapter 2 section 2.2.1. Primary human breast cancer associated fibroblasts (CAFs) were purchased from Asterand Bioscience (PCD-10-0110) and

cultured in DMEM + 10% FBS. Ex vivo CTC lines, BRx07 and BRx68 were cultured as previously described[177]. Other *ESR1* mutant MCF7 cell models were kind gifts from Jason Gertz, PhD and Simak Ali, PhD.

*Experiments with ex vivo CTC cell lines were conducted by Yonatan Amazleg from Dr. Min Yu's group.

### 3.2.4 *In Vivo* Study

4-weeks old female *nu/nu* athymic mice were ordered from The Jackson Laboratory (002019 NU/J) according to University of Pittsburgh IACUC approved protocol #19095822. MCF7 and T47D *ESR1* mutant cells were hormone deprived and resuspended in PBS with $10^7$ cells/ml. 100 µl cell suspension was then injected into nude mice with 7 mice per group via tail vein. The tail vein injection was operated by Mr. Weizhou Hou. Mice were under observation weekly. According to the IACUC protocol, upon above 50% mice in any of the group showed the following pre-defined signs of euthanasia including above 20% weight loss, difficulty ambulating, anorexia, piloerection, hunched posture, rough and ungroomed hair coat and ungroomed appearance, excessive scratching and licking, mutilation of painful area, pallor due to severe anemia, tumor>$2cm^3$, respiratory distress, not responsive to external stimuli, the entire cohorts were then euthanized 13 weeks for MCF7 cells-injected mice and 22 weeks for T47D injected mice). Macro-mets as well as other potential organs (lung, liver and UG tracts) for metastatic spread were harvested. All the harvested tissues were processed for formalin fixed paraffin embedded (FFPE) preparation and hematoxylin and eosin (H&E) staining by the Histology Core

at Magee Women's Research Institute. Macro-metastatic tumor FFPE sections were further evaluated by pathologist Peter Lucas, MD, PhD. Micro-metastatic lesions in the lung were further examined and quantified by immunofluorescence staining as described below.

### 3.2.5 Compounds and Reagents

17β-estradiol (E2, #E8875) was obtained from Sigma, and Fulvestrant (#1047), carbenoxolone disodium (#3096) and EDTA (#2811) were purchased from Tocris. LGK974 (#14072) and T5224 (#22904) were purchased from Cayman. Marimastat (S7156) was obtained from SelleckChem. For the knockdown experiments, siRNA against *FOXA*1 (#M-010319), *DSC1* (#L-011995) and *GJA1* (#L-011042) were obtained from Dharmacon.

### 3.2.6 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

Detailed methodological description of qRT-PCR was presented in the materials and methods section of Chapter 2. All primer sequences are shown in Appendix D Table 10.

For ECM molecules $RT^2$ Profiler PCR array, experiments were done following the provided manufacturing protocols from Qiagen (PAHS-013Z) using three biological replicates. Cq values were normalized against the GEO mean of five house-keeping genes, and the expression FC was calculated comparing *ESR1* mutant vs WT. To select significantly altered genes, students' t test was performed to each pair of WT and mutant cells, p-values were adjusted by multiple

comparison. Genes filtered with Cq<35 in a least one group of each cell line and q-value below 0.1 and was considered as significantly differential expressed genes.

### 3.2.7 Immunoblotting

Detailed methodological description of immunoblotting was present in the materials and methods section of Chapter 2 section 2.2.6. All antibody information are shown in Appendix D Table 11.

For cell fractionation separation, the cell fractionation kit (Cell Signaling #9038) was used and all procedures were following the manufactory instruction.

### 3.2.8 Immunofluorescence (IF)

MCF7 cells were hormone deprived and seeded on coverslips. After attached, cells were fixed with 4% paraformaldehyde and blocked with 3% BSA solution plus 0.1% tritonX-100. Primary antibody against desmoglein 2 (Santa Cruz, sc-80663) was applied to stain the cells followed by secondary FITC-conjugated antibody (Thermo Scientific, A16079) and Hoechst (Thermo Scientific, #62249). Coverslips were mounted and images were taken using fluorescence microscope (Olympus, CZX16) under objective of 20X. Quantification was performed by dividing the integrated intensity of FITC signals by the number of nuclei in each image.

For the immunofluorescence staining of mice lung sections for evaluating micro-metastasis, 5 micron FFPE slides of each sample were baked in 60 °C for two hours.

Deparaffinization and rehydration were then performed follows 3x5' xylene, 3x5' 100% EtOH, 1x5' 95% EtOH, 1x3' 80% EtOH, 1x3' 70% EtOH and 1x3' PBS. The slides were then subjected to antigen retrieval step by boil in citrate buffer (16 mM sodium citrate, 4 mM citrate acid monohydrate, pH=6.0) in high pressure cooker for 20 minutes. Upon cooling down the lung sections were cycled using PAP pen and rinsed in PBST for 5 minutes. The slides were additionally incubated in 100 mM glycine solution to reduce background staining and incubate in blocking buffer (0.3% Triton X-100, 5% goat serum in PBS) for one hour. Primary antibodies were mixed in blocking buffer and the applied on all the slides in 4 °C overnight (Human CK19: Thermo Fisher 190-P1; CK9+18: abcam#53280). Slides were further washed with PBST for 3 times and incubated for secondary antibody for 1 hour in room temperature (Alexa Fluor 488: Thermo Fisher A32723; Alexa Fluor 568: Thermo Fisher A11011). Hoechst was further staining for nuclei visualization, and all the slides were mounted and imaged using fluorescence microscope (Olympus, CZX16). Two representative areas of each slides were selected and quantification was performed by Ms. Kara Burlbaugh blindly using ImageJ.

### 3.2.9 Immunohistochemistry (IHC)

For IHC staining for connexin 43 on mouse-derived tumor sections, all steps were identical as described in section 3.2.8 second part until primary antibody incubation. Additional incubation step in 3% $H_2O_2$ was added after antigen retrieval step to block endogenous peroxidases. Connexin 43 antibody (Sigma-Aldrich C6219, 1:100) was used for overnight incubation at 4 °C. DAKO secondary reagent (Agilent E0432) was then applied to all the slides following DAKO DAB

substrate visualization for 3 minutes. Slides were then conterstained hematoxylin for nuclei and mounted. Nikon 90i microscope was used for the entire image scanning.

### 3.2.10 Cell Growth Assay

3,000 (MCF7) or 4,000 (T47D) cells were seeded into either flat bottom 96-well ultra-low attachment plates (for 3D growth) (Corning, #3474) or regular 96-well plate (2D growth) (Corning, #353072). Cell numbers were quantified after desired days of culture with either Celltiter Glo luminescent cell viability kit (Promega, G7573) or FluoReporter Blue Fluorometric dsDNA quantification kit (Invitrogen, F2962). Fluorescent readouts were corrected by background measurement.

### 3.2.11 IncuCyte Living Imaging System

#### 3.2.11.1 Wound Scratching Assay

150,000 MCF7 or T47D cells were seeded into Imagelock 96-well plates (Essen Bioscience, #4379) coated with Matrigel (Corning, #356237). Wounds were scratched in the middle of each well using the Wound Maker (Essen Bioscience, #4493). Desired treatments were loaded after twice of PBS washing with the addition of 5 μg/ml Mitomycin C (Sigma-Aldrich, #10107409001) to block cell proliferation. The IncuCyte system was used to record the wound images every 4 hours and calculated the wound closure density was calculated using the manufacturer's wound scratch assay module.

#### 3.2.11.2 Aggregation Rate Assay

MCF7 (3,000) or T47D (4,000) cells were seeded into 96-well round bottom ultra-low attachment plates with 100 µl media of each well. Cell aggregation process was monitored by IncuCyte living imaging system every 1 hour. Spheroid areas were normalized to time 0 and plotted in the time course.

#### 3.2.11.3 Cell Growth Assay

MCF7 (3,000) or T47D (4,000) cells were seeded into 96-well plate with desired drug treatment. Cell density was captured and quantified by IncuCyte living image system every four hours. At least six biological replicates were included each time.

### 3.2.12 Calcein-labelled Cell-cell Interaction Assay

MCF7 and T47D cells were seeded into black-walled 96 well plate at 15, 000 cells per well for 24 hours to form a fully confluent monolayer. Separate cultures of cells were digested and labelled with 1 µM calcein AM (BD Pharmingen, #564061) for 30 minutes in room temperature. 40, 000 labelled cells were then loaded on the top of the monolayers and incubated for 1 hour at 37 °C. Cells were washed for 3 times after incubation by manually pouring out all the PBS. The plates were read using Victor X4 plate reader (PerkinElmer) under the excitation and emission wavelength of 485/535 nm. Cell-cell adhesion ratios were calculated by dividing the post-washed readouts to the pre-wash readouts after each time of wash. For the vacuum aspiration method, we used a standard laboratory vacuum pump with the modified speed around 100 ml/minutes.

### 3.2.13 Cell-ECM Adhesion Assay

30,000 cells/well were seeded into Collagen I (Thermo Fisher Scientific, A1142803) or uncoated 96-well plates. For the ECM array assay, cells were re-suspended and loaded into the ECM array plate (EMD Millipore, ECM540). After incubation at 37°C for 2 hours, the plates were washed with PBS for three times, and attached cells were quantified using the FluoReporter kit (Thermo Fisher Scientific, F2962). Adhesion ratios were calculated by dividing the remaining cell amounts in the washed wells to the initial cell amounts in pre-washed plates. All the experiments were repeated for at least three times independently, with at least six biological replicates each time. For *TIMP3* overexpression, PRK5M-*TIMP3* plasmid (Addgene, #31715) was transfected into targeted cells, which was subjected to adhesion assay after 24 hours.

### 3.2.14 Pan-MMP Activity Assay

FRET-based MMP activity assay were performed using MMP activity assay kit (abcam, ab112146) and following the manufacturing protocols. In brief, 30 μg protein from whole cell lysis were pre-mixed with APMA and incubate for 3 hours to mainly activate collagenases. MMP substrates were then loaded and fluorescence intensities were monitored under the excitation and emission wavelength of 490/525 nm after 0, 5, 10, 15, 20, 25, 30, 40, 50 and 60 minutes of incubation. Background emissions were removed and fold changes were calculated by normalizing to the initial readout at time 0.

### 3.2.15 Spheroid Invasion and Collective Migration Assay

Spheroid invasion assay was performed as previously described[178]. 3, 000 (MCF7) or 4, 000 (T47D) cells were seeded into 96 well round bottom ULA plate for spheroid formation for 2 days. For invasion assay, 1.3 mg/ml Type I collagen (Corning, #354236) supplemented with 1% NEAA (Thermo Fisher Scientific, #11140050) were directed added into each well. Images were taken at day 0 and day 6 (T47D)/ day4 (MCF7). Invasion areas were quantified by subtract day0 areas from day 6/4 areas using Adobe Photoshop. For collective migration assay, spheroids were gently transferred into a collagen coated 96-well plate in the presence of 5 μg/ml Mitomycin C to block cell proliferation. Spheroid migration images were taken at day0 and day4, migratory distances were calculated using the mean values of spheroid weights and heights normalized to day 0.

### 3.2.16 Boyden Chamber Chemotaxis Assay

For chemotaxis assay, QCM chemotaxis cell migration kit (Millipore Sigma, ECM508) was used. All procedures were following the manufactory instruction. Briefly, T47D cells were hormone deprived and starved in serum free medium for 24 hours. Cells were digested and diluted into $10^6$ cells/ml. 300 μl of cell suspensions were loaded in the inner chamber of the 8 μm cell inserts with biological duplicates. 10% CSS or serum free medium was loaded in the outer chamber. After 72 hours, cells stained with 0.1% crystal violet and cells in the inner chamber were wiped using cotton sticks. In the inserts were imaged and remained cells were then dissolved in extraction buffer and subjected to colorimetric measurement under OD450.

### 3.2.17 Quantitative Microfluidic Fluorescence Microscope (qMFM)*

qMFM assays were performed as previously described [179]. Cells labelled with calcein were seeded onto coverslips for 24 hours. Then the same type of cells labelled with DiD cell staining solution (Thermo Fisher Scientific, V22887) were loaded into the microfluidic system under the shear stress of 6 dynes/cm to allow for attachments. Videos were recorded for 200 frames with 100ms/frame. Nikon Element Software was used for quantification, with a module calculating the cells moving slower than 3 μm/s in each frame. Adhered cells from T0 were subtracted, and cell-cell interaction events were normalized to the total calcein signals of each video.

*This experiment was in collaboration with Dr. Maritza Montanez and Dr. Prithu Sundd for the operation of the qMFM device, Mr. Callen Wallace and Dr. Simon Watkins for the development of automatic quantification program.

### 3.2.18 Ibidi Microfluidic System*

MCF7 and T47D *ESR1* mutant cells were hormone deprived for 3 days and diluted to 1,000,000 cells in 14 ml medium and loaded into the ibidi pump system (ibidi, #10902). Cells were constantly flowing with 15 dynes/cm shear stress for two hours and were imaged immediately after seeded back to flat bottom ULA plate. For each group, six wells were images and two images were recorded for each well. T0 cells were also imaged as the control of initial time point. Cell numbers in cluster or non-cluster versions were manually counted. Cell cluster ratios were calculated by dividing the cell numbers in clusters to the total cell numbers, cell clustering grade were divided

by the cell numbers in each cluster. For CBX treatment, cells were pre-treated with 100 μM CBX for two days before flowing.

*These experiments were in collaboration with Dr. Yu Jiang. All these experiments were performed by Yang Wu.

### 3.2.19 Calcein Dye Transfer Assay

MCF7 WT and mutant cells were trypsinized and labelled with 1 μM calcein for 30 minutes. 10, 000 labelled and unlabeled cells were mixed 1:1 and incubated in Corning round bottom tube (#352054) for 12 hours. Cells were then subjected to flow cytometry analysis gating on single cell population. GFP positive cells were analyzed by setting the threshold with positive and negative controls. Exchanged dye ratios were calculated by doubling the GFP+ cell percentages minus 50%. This experiment was done with three biological replicates each time.

### 3.2.20 Top-Flash Luciferase Assay

125,000 T47D cells were seeded into 12-well plate. M50 Super 8X Top-Flash luciferase (Addgene, #12456) and Renilla plasmids were co-transfected into all the groups as previous described[180]. Cells were lysed after 24 hours of transfection and luciferase values were measured with Dual-luciferase Reporter Assay system (Promega, E1910). Relative Wnt activity was calculated by as the ratio of firefly luciferase activity over Renilla luciferase activity. This experiment was repeated for nine times independently, with three biological replicates each time. The results shown were combined with all experiments.

### 3.2.21 Hanging-drop Aggregation Assay

Hanging-drop assay was performed as previously described[181]. In brief, MCF7 WT or *ESR1* mutant cells were trypsinized, centrifuged, and resuspended as single-cell suspensions at 5 × $10^5$ cells/ml. 30-µl drops of cell suspension were pipetted onto the inside surface of 35-mm culture dish lids, and dishes were filled with 2 ml media to prevent evaporation. At each time point, the lid was inverted for imaging, and drops were triturated 10 times through a 20-µl pipet. 4 µl 16% PFA was added, and each drop was spread onto a glass slide. Cells numbers were then counted in each cluster at each time points. This experiment was performed with at least three biological replicates each time.

### 3.2.22 Chromatin Immunoprecipitation (ChIP)

ChIP were performed as previously described [182]. Briefly, hormone-deprived WT and mutant cells were treated with vehicle or 1nM E2 for 45 minutes. Chromatin DNA was then extracted from each sample. The immunoprecipitation was performed using ERα (sc543) and rabbit IgG (sc2027) antibodies (Santa Cruz Biotechnologies). FOXA1 (ab23738), Histone 3 acetylation at K27 site (ab4739) and Histone 3 monomethylation at K4 site (ab8895) antibodies were obtained from Abcam.

For ER ChIP-seq, DNA samples were pooled from each individual clone with the same genotype and at least 10 ng DNA of each sample was sent to McGill University Sequencing Core for library preparation using TruSeq ChIP Library Preparation Kit (Illumina IP-202) and

sequenced using Illumina Hiseq 2500 Platform (ER ChIP-seq). Over 16M single end 50bp reads were allocated for each sample.

For FOXA1 ChIP-seq, DNA samples were originally prepared from pooled cells in biological triplicates, and then the pool of replicate DNA were sent to the Health Sciences Sequencing Core at Children's Hospital of Pittsburgh for library preparation using TruSeq ChIP Library Preparation Kit (Illumina IP-202) and sequenced using Illumina NextSeq platform. Over 80M single end 150bp reads were allocated for each sample.

For ChIP-qPCR, DNA was diluted and subjected to qRT-PCR as described above. Fold enrichment method was used to quantify the binding enrichment at the selected sites. IgG-IP'ed samples were used as negative control. The primer sequences for *DSC1* and *DSG1* region ChIP-qPCR are shown below: Peak1-Forward: AATGGACCCAACGAGTTCTC; Peak1-Reverse: AGACATACCAAAGACACAGCC.; Peak2-Forward: TGTGTACTGAGGAACCTACTCTC; Peak2-Reverse: CTGTTCATTTGTGGGGCTAAC. All sequences are from 5' to 3'.

### 3.2.23 ChIP-sequencing Analysis*

ChIP-seq reads were aligned to reference genome assembly using Bowtie 2.0 [183], and peaks were called using MACS2.0 with p value below $10^{-5}$ (ER ChIP-seq) or q value below 0.05 (FOXA1 ChIP-seq) [184]. We used Diffbind package [185] to perform principle component analysis, identify differentially expressed binding sites and analyze intersection ratios with other data sets. Briefly, all the BED files for each cell line were merged and binding intensity was estimated at each site based on the normalized read counts in the BAM files. The pairwise

comparison between WT and mutant samples were performed to calculate the fold change (FC) and the binding sites were sub-classified into three categories: gained sites (FC>2), lost sites (FC<-2), and not-changed sites (-0.2<FC<0.2). Heatmaps and intensity plots for binding peaks were visualized by EaSeq[186]. Genomic feature distribution and proximal gene annotation were called using ChIPseeker [187], taking the promoter region as +/- 3000 bp of the transcriptional start site (TSS) and 100kb or 200kb as the peak mediated region for ER and FOXA1 respectively. For gene annotation from FOXA1 binding sites, FOXA1 peaks were prefiltered with FDR value below 1E-10 to select those high confident peaks. For motif analysis, BED files were converted into fasta versions using bedtools[188], and uploaded into MEME Suit[189] for Analysis of Motif Enrichment (AME) using JASPAR CORE and UniPROBE Mouse data base. Enriched motifs were filtered with E value below $10^{-5}$. For estrogen response element (ERE) scanning, EREFinder[190] was used in default setting. Kd values were extracted from the final output for plotting.

* ER ChIP-seq data generation and processing were conducted by Amir Bahreini, PhD.

### 3.2.24 Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq)*

ATAC-seq was performed as previously described[191]. Briefly, T47D *ESR1* mutant cells were hormone deprived in the individual clone manner with or without 10 nM E2 treatment for 1 hour. Tn5 transposase was loaded on cells and integrates its adaptor payload into regions of accessible chromatin. DNA fragments were then amplified and subjected to high-throughput sequencing via HiSeq 2500 platform to detect open chromatin locations. For the data processing, qualified reads will be aligned to hg19 reference genome using Bowtie 2. Sam files were converted

into BAM files and sorted with Samtools. ATAC-seq peaks were called using MACS2 normalized to the Input control with the cutoff of $p<10^{-5}$. Diffbind R package was used to determine the differential binding sites between WT and mutant cells and call PCA plot. ChIPseeker was used to annotate the genes around the predefined range of those called peaks using 200 kb as the ATAC sites-mediated regions.

*ATAC-seq and data process steps were done by Spencer Arnesen from Jason Gertz, PhD's group at University of Utah.

### 3.2.25 RNA-sequencing Analysis

RNA-seq data processing was described in Chapter 1 section 2.2.11 using Salmon package [110]. Differential Expression Analysis was done using DEseq2 package[192]. In brief, genes were prefiltered with the maximum log2CPM>1 across all samples. DE genes with q value below 0.1 were used for Ingenuity Pathway Analysis[193]. GSEA analysis was performed using the Broad GSEA Application[194]. Significant enriched pathways were filtered using FDR<0.25. Gene set variation analysis were performed using GSVA package[195]. All gene sets used in this study are shown in the corresponding table in Appendix D. Data visualizations were performed using "ggpubr[196]" and "VennDiagram[197]".

### 3.2.26 Drug Synergy Analysis

Drug synergy was calculated based on the Bliss independence model using the SynergyFinder R package[198]. Bliss synergy scores were used to determine the synergistic effects.

### 3.2.27 Statistical analysis

GraphPad Prism software version 7, and R version 3.6.1 were used for statistical analysis. All experimental data were shown as mean ± standard deviations from biological replicates, unless otherwise stated. Specific tests were indicated in the corresponding legends. All tests were two-tailed, with $p<0.05$ considered statistically significant.

### 3.2.28 Data Availability

The ChIP-seq data have been deposited in the Gene Expression Omnibus database. The accession numbers for the data reported in this study is GSE125117 (not publicly available yet). RNA-seq data from MET500 cohorts were requested from the original resource[118]. Briefly, the MET500 FASTQ files for hybrid-capture RNA-Seq breast metastatic samples were downloaded via dbGaP (phs000673.v2.p1). Transcript quantification was performed with Salmon v0.8.2 (quasi-mapping mode, 31-kmer index using GRCh38 Ensembl v82 transcript annotations, seqBias and gcBias corrections), followed by mapping to gene counts via tximport [199], and normalization via log2 transformed TMM-normalized counts-per million (log2 CPM) with the

edgeR package. ER-positivity for MET500 was defined as RNA expression of *ESR1* greater than or equal to the median. RNA-Seq data from the WCRC cohort are available at https://github.com/leeoesterreich. DFCI cohort RNA-sequencing data were originally published by Jeselsohn et al.[117], and further expanded by Nikhil Wagle, MD. All the histone modification marker ChIP-seq data were obtained from GSE63109. All the T47D FOXA1 data were downloaded and reprocessed from GSE23893, GSE32465, GSE25710 and GSE72249. All the raw data are available upon request from the authors.

## 3.3 Results

### 3.3.1 Hotspot *ESR1* Mutations Are Enriched in Distant Recurrence and Alter Multiple Molecular Functions

#### 3.3.1.1 Significant Enrichment of *ESR1* Mutations in Distant Metastases Versus Local Recurrences

To seek clinical evidence for the potential metastasis-conferring roles of *ESR1* LBD mutations, we compared the *ESR1* mutation frequencies between distant metastatic and local recurrent tumor samples. Of note, such a comparison has not been previously possible, given that until recently, most prior studies included little to no local recurrences. Analysis of the MSK-IMPACT panel sequencing data[116] showed that while 137/837 distant metastases (16.4%) harbor an *ESR1* mutation, none are found in the 14 local recurrence samples in the data set [116]. Another recent study from Zundelevich et al. identified 15/41 (36%) local recurrences harbor *ESR1*

mutations [200], To expand this observation, we additionally screened 75 ER+ recurrent tumors, including 27 local recurrences and 48 distant metastatic tumors from the Women's Cancer Research Center (WCRC) and Charite cohorts for *ESR1* hotspot (Y537S/C/N and D538G) mutations using the highly sensitive droplet digital PCR (ddPCR) method. we identified 12 mutation-positive cases among the distant metastases (25%), whereas none of the local recurrences harbored *ESR1* mutations (Table 1). Consistent with previous studies, the most frequent mutations were D538G (71.4%) and Y537S (21.4%) (Appendix B Fig. 82A and Appendix D Table 12), and two metastases harbored two concurrent hotspot mutations. There was no significant difference in recurrence-free survival times for patients with distant vs local recurrence (Appendix B Fig. 82B), excluding the possibility that the observed differences could have simply been caused by duration of time to recurrence. The different findings between ours and Zundelevich et al. [200] was likely because 26/27 local recurrences from our cohort were first recurrences, whereas 6/15 of patients from the latter cohort also developed distant metastasis. These results provide clinical evidence supporting the idea that *ESR1* mutations may confer a neomorphic metastatic capacity in addition to endocrine resistance.

**Table 1. Significant enrichment of *ESR1* mutations in ER+ distant compared to local recurrences.**

| Site of Recurrence | Total Number (N=75) | *ESR1* WT (N=63) | *ESR1* Mutant (N=12) | P value |
|---|---|---|---|---|
| Distant Metastasis | 48 | 36 (75%) | 12 (25%) | 0.0031 |
| Local Recurrence | 27 | 27 (100%) | 0 (0%) | |

48 distant metastatic and 27 local ER positive recurrence samples were harvested from the WCRC and Charite cohorts. Genomic DNA was isolated from either FFPE or frozen tumor tissues, and subjected to droplet digital PCR (ddPCR) detection with specific probes for Y537S, Y537C, Y537N and D538G point mutations (cDNA samples were used for 3 local recurrent samples). Hotspot *ESR1* mutation incidences between distant metastatic and local recurrent samples were compared using Fisher's exact test (** $p<0.01$).

*Part of ddPCR of distant metastasis was operated by Amir Bahreini, PhD and Peilu Wang, MD.

#### 3.3.1.2 Comprehensive Transcriptional Analysis Revealed Multiple Altered Molecular Functions in *ESR1* Mutant Tumors

To identify candidate functional pathways mediating the metastatic properties of *ESR1* mutant cells, I compared WT and *ESR1* mutant tumor transcriptomes from i) our local WCRC cohort of ER+ tumors[173, 201], and ii) the recently reported MET500 cohort[118] and iii) the Metastatic Breast Cancer Project from Dana-Faber Cancer Institute, which contain 46 *ESR1* WT

and 8 mutant, 45 *ESR1* WT and 11 mutant, and 32 *ESR1* WT and 98 mutant breast tumors respectively (Appendix B Fig. 83).

I performed both Gene Set Enrichment Analysis (GSEA) and Ingenuity Pathway Analysis (IPA) following differential gene expression analysis (Figure 13).

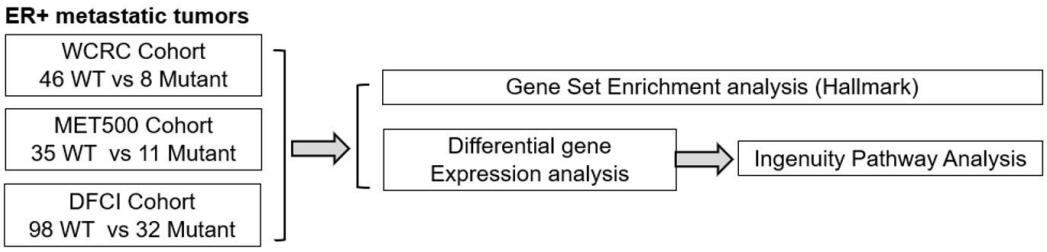



**Figure 13. Schematic view of comprehensive transcriptomic analysis on three ER+ metastatic breast cancer cohorts.**

Gene Set Enrichment Analysis identified "Estrogen Response Early" and "Estrogen Response Late" pathways as the two most enriched pathways in *ESR1* mutant tumors, which is consistent with the ligand-independent ER activation in various *in vitro* studies (Fig. 14A and 14B). In addition, I also found "Glycolysis" and "Adipogenesis" among the top 5 consistent pathways enriched in *ESR1* mutant tumors, indicating the potential reprogramming of metabolic pathways in *ESR1* mutant tumors (Fig. 14A).

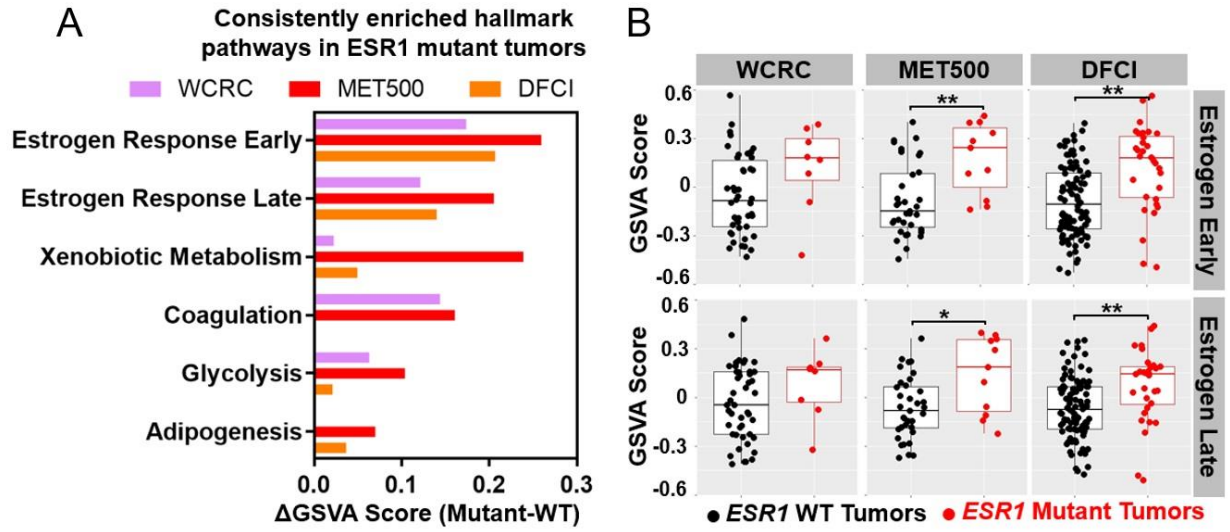



**Figure 14. Gene set enrichment analysis confirms the hyperactivsating ER signaling in *ESR1* mutant tumors**

Bar graph showing the top 5 consistent pathways among all three metastatic breast cancer cohorts. Differences of enrichment scores between WT and mutant tumors were calculated based on gene set variation analysis of each tumors with the 50 MSigDB curated Hallmark pathways. Pathways were first filtered by trend consistencies (co-increased or co-decreased in three cohorts) and ranked by the average delta score of the three cohorts. B. Box plots representing the enrichment levels of "Estrogen Response Early" and "Estrogen Response Late" signatures in *ESR1* mutant versus *ESR1* WT metastatic tumors in each cohort. Mann Whitney U test was used. (* $p<0.05$; ** $p<0.01$)

Moreover, differential gene expression analysis identified 188/92/126 up-and 234/217/278 down-regulated genes in the WCRC/MET500 and DFCI cohorts respectively (q value <0.1, FC>1.5) (Fig.15A). Ingenuity Pathway Analysis (IPA) identified "Cell-to-cell signaling and interaction" and "Cellular Movement" among the top five altered pathways in *ESR1* mutant tumors in all three cohorts (Fig. 15B). Taken together, these results indicate that *ESR1* mutations might mediate metastatic phenotypes through effects on cell-to-cell signaling and interaction and cell movement, which I investigated further using previously established genome-edited MCF7 and T47D cell line models[110].

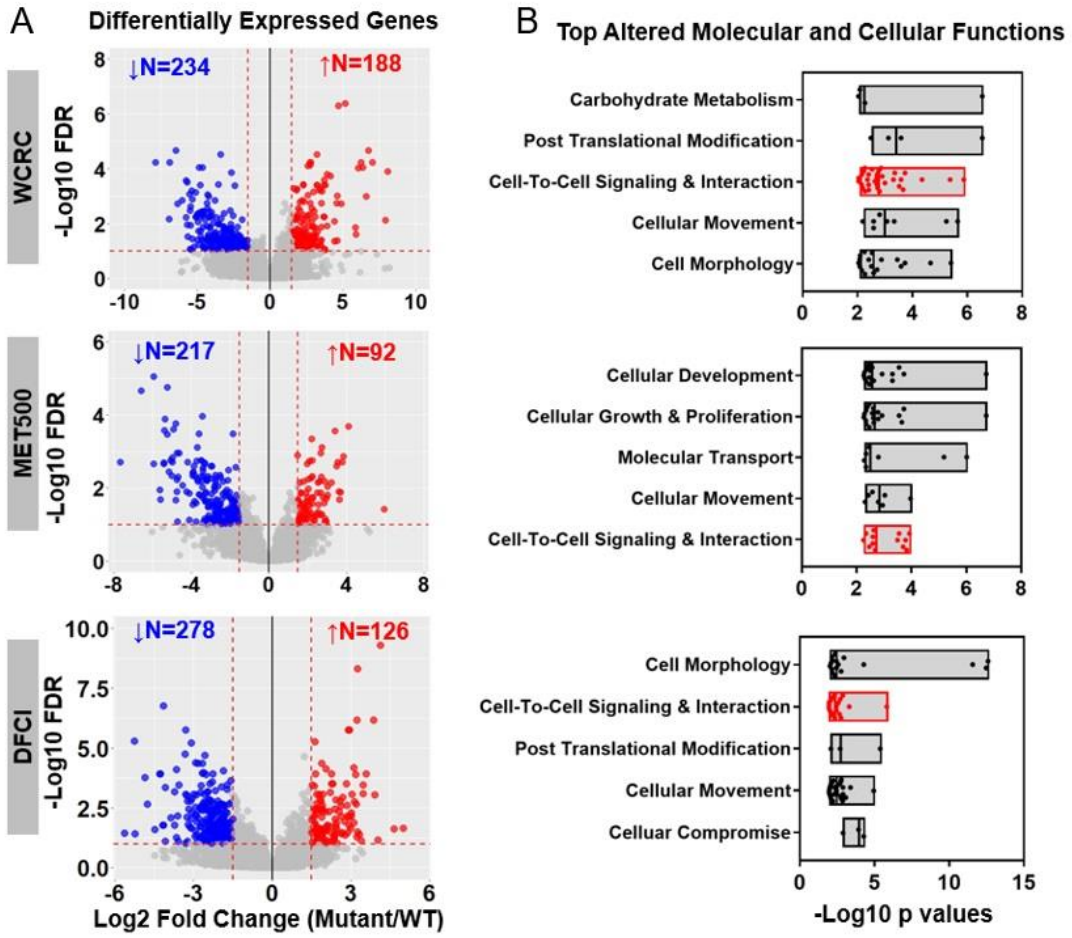



**Figure 15. Ingenuity pathway analysis with differentially expressing genes revealed enrichment of metastatic pathways in *ESR1* mutant tumors.**

A. Volcano plots representing the differentially expressing genes (DE genes) in *ESR1* mutant tumors versus WT tumors in the three metastatic breast cancer cohorts. DE genes were selected using the cutoff of q<0.1 and FC>1.5. Genes that were upregulated or downregulated were labelled in red and blue respectively. B. Top 5 altered cellular and molecular function based on all the DE genes generated from IPA software.

### 3.3.2 Hotspot *ESR1* Mutations Reprogram Cell-cell Adhesome to Facilitate Breast Cancer Metastasis

#### 3.3.2.1 *ESR1* Mutant Cells Exhibit Stronger Cell-cell Adhesion in Static Conditions

I first assessed the cell-cell adhesion properties by visual inspection of cell clusters formed in suspension culture (Fig. 16A). Following hormone deprivation and six days of culture in ultra-low attachment (ULA) plates, I observed more compact cell clusters in both MCF7 and T47D cell lines for both Y537S and D538G mutants as compared to WT cells. A time course study confirmed enhanced cluster formation as early as day 1, with stronger effects in MCF7 mutant cells and in T47D-Y537S cells than in T47D-D538G (Appendix B Fig. 84). Similar observations were made in MCF7 and T47D individual clones compared to the pool of mutant clones, eliminating the possibility of clonal effects (Appendix B Fig. 85A).

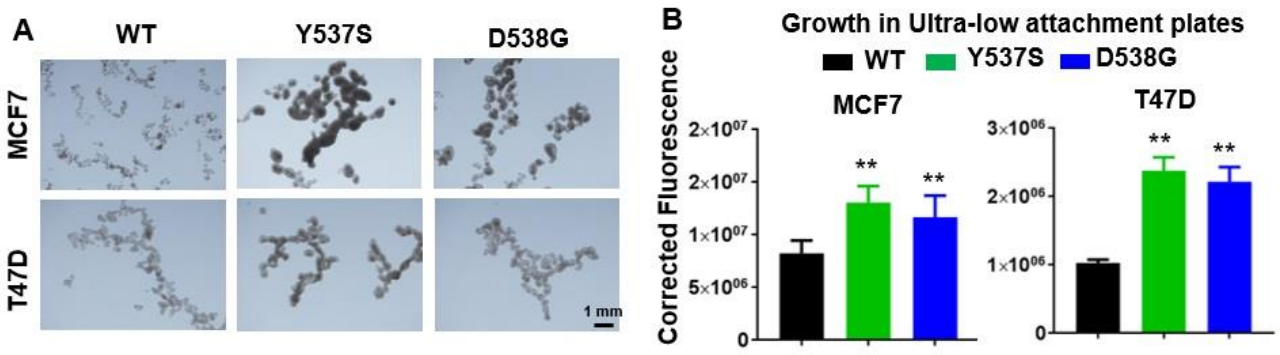



**Figure 16. *ESR1* mutant cells form larger multicelluar clusters and showed enahcned proliferation in suspension condition.**

A. Representative images of MCF7 and T47D spheroids after seeding into 6-well ultralow attachment plates for 6 days after hormone deprivation. Scale bar: 1mm. Data were from one representative experiment with three biological replicates of three independent experiments. B. MCF7 or T47D WT/*ESR1* mutant cells were seeded into flat bottom ULA plates. Cell numbers were quantified using Celltiter Glo. Fluorescence readouts were corrected by background measurements. Dunnett's test was used between WT and each mutant. Data were from one representative experiment with six biological replicates of six independent experiments. (* $p<0.05$; ** $p<0.01$)

*Fig. 16A is provided by Jian Chen.

Since *ESR1* mutant cells displayed significantly increased ligand-independent growth in ULA (Fig. 16B and Appendix B Fig. 85B), I wanted to rule out the possibility that the increased cluster formation was simply a result of increased cell number.

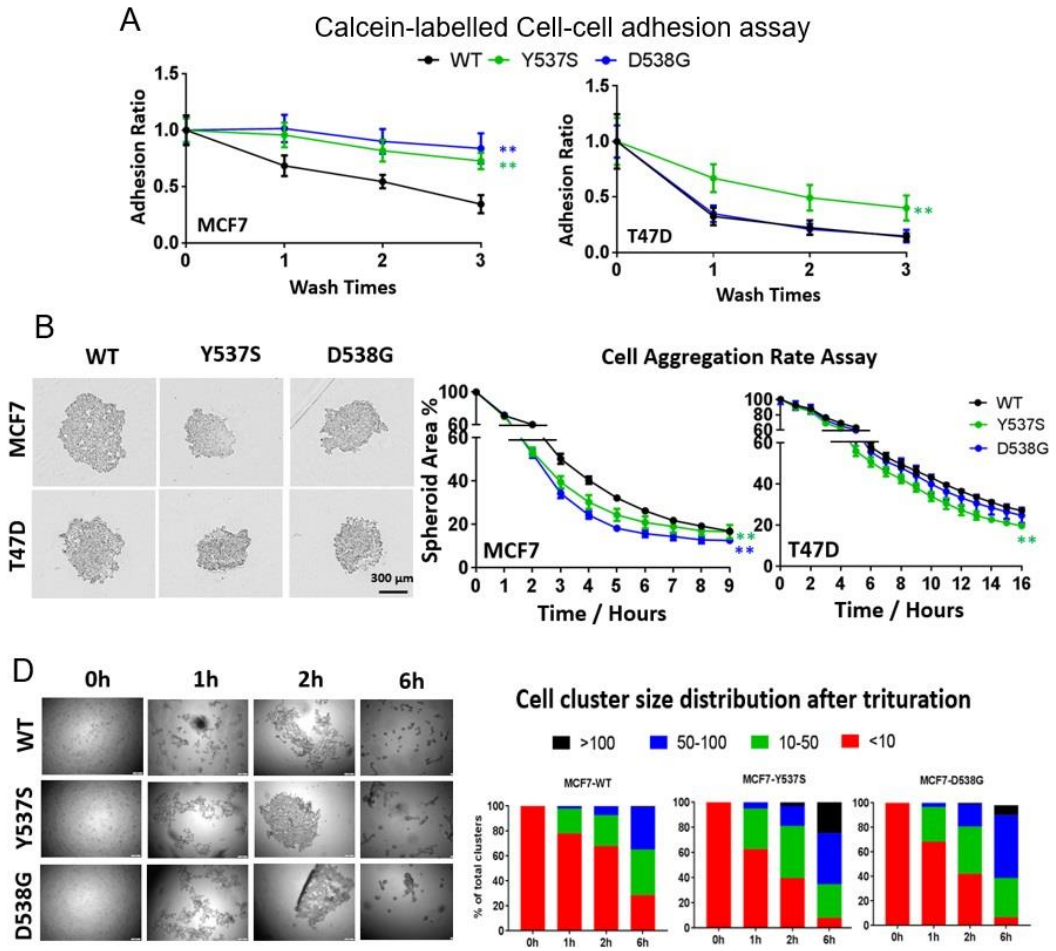



**Figure 17. *ESR1* mutant cells exhibited stronger cell-cell interaction under static condition.**

A. Calcein labelled cell-cell adhesion assay was performed in MCF7 and T47D WT/ mutant cells. Calcein-labelled cells were loaded on the top of monolayers of adhered unlabeled cells in a black-walled 96-well plate. Cells were washed 3 times after incubation. The plates were read under the excitation and emission wavelength of 485/535 nm. Adhesion ratios were calculated by dividing the remaining cells after each time of wash to the initial readout from unwashed wells. A pairwise Two-way ANOVA between WT and each mutant was utilized. Data shown were from a representative experiment among five (MCF7) or two (T47D) independent repeats. B and C. MCF7/T47D cells were seeded into round bottom ULA plates and the cell aggregation process followed by IncuCyte living imaging system every 1 hour. Spheroid areas were normalized to time 0. Representative images after 3 hours (MCF7) and 16 hours (T47D) of aggregation are shown in panel B. A pairwise Two-way ANOVA between WT and each mutant was utilized. Data shown were from a representative experiment among five (MCF7) or two (T47D) independent repeats. D. Representative images of MCF7 cells after hanged in droplet. 15, 000 cells in 30-µl drops of cell suspension were pipetted onto the inside surface of 35-mm culture dish lids. The lids were inverted and cell clusters were captured in the droplets for 0, 1, 2 and 6 hours under 4X objectives. Cells clusters were triturated for 10 times and cell numbers distribution in each cluster were

counted. Images shown here were from a representative experiment among three independent repeats.. (* $p<0.05$; ** $p<0.01$)

I therefore set out to directly quantify homotypic cell-cell interactions via measuring the adhesion of calcein labelled *ESR1* WT or mutant cells. These assays showed that both MCF7 mutant cell models exhibited significantly stronger cell-cell adhesions compared to the WT cells[178] (Fig. 17A). The same trend was also observed using an alternative method, in which we applied vacuum aspiration instead of shaking off the cells by manual force (Appendix B Fig. 86). In T47D cells, similar effect was observed, however, it was limited to the Y537S mutant (Fig. 17A). These assays were complemented by quantification of cell aggregation rates as a direct reflection of cell-cell adhesion[202]. Visual inspection (Left panel, Fig. 17B) showed more compact clusters, and quantitative measurement using IncuCyte confirmed faster aggregation rates in MCF7-Y537S/D538G and T47D-Y537S cells, as early as two hours after plating (Right panel, Fig. 17B and Appendix B Fig. 87). In addition, I measured aggregate formation in spheroids, through counting the number of cells in hanging droplets, as previously described[181] (Fig. 17C). The faster spontaneous aggregation was also reproduced in two other independent CIRPSR MCF7 *ESR1* mutant models (Appendix B Fig. 88). Consistent with earlier results, there were significantly more cells in the spheroids formed by the *ESR1* mutant MCF7 cells, as early as one to two hours after plating, reflective of increased cell-cell adhesion.

Finally, I tested whether *ESR1* mutations would affect heterotypic cell-cell interactions, and thus tested the interaction of MCF7 WT and mutant cells with human breast cancer associated fibroblasts (CAFs), given their increasingly recognized role in tumor cell interactions[203].

Similar to what was observed for homotypic interaction, I observed significantly increased adhesions between CAFs and MCF7 mutant cells compared to WT cells (Fig. 18).

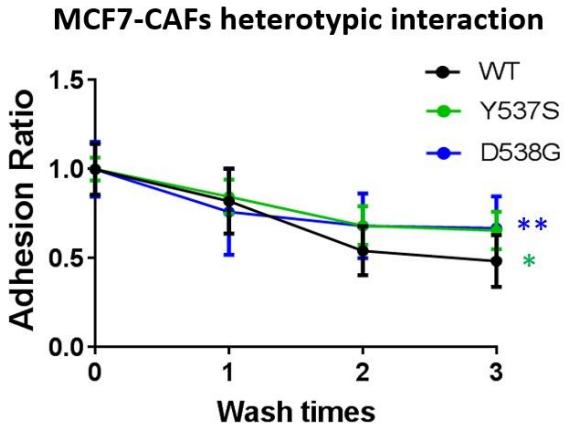



**Figure 18. *ESR1* mutant cells show enhanced heterotypic interaction with cancer-associated fibroblats.**

Heterotypic cell-cell adhesion assay was performed with MCF7 WT/mutant cells and calcein-labelled breast cancer-associated fibroblasts (CAFs). Calcein-labelled CAFs were loaded on the top of a monolayers of adhered unlabeled *ESR1* mutant cells. Same methods were used as described in figure 17A. This experiment was done once. (* $p<0.05$; ** $p<0.01$)

### 3.3.2.2 *ESR1* Mutant Cells Exhibit Stronger Cell-cell Adhesion in Microfluidic Conditions

Cell-cell interaction has been reported to affect several stages of metastasis, including collective invasion, intravasation, dissemination and circulation[41, 204, 205]. To test whether the mutations could affect cell-cell adhesion of tumor cells in circulation, we utilized a quantitative microfluidic fluorescence microscope (qMFM) imaging system simulating blood flow[179]. Quantification of dynamic adhesion events normalized to adhesion surfaces revealed an enhanced cell-cell adhesion capacity of *ESR1* mutant MCF7 cells, with more pronounced effects in D538G cells (Fig. 19A). In addition, we tested the spontaneous clustering balance after two hours of microfluidic condition with physiological shear stress using ibidi pump system[206]. MCF7-*ESR1*

mutant cells showed higher clustering probability and they formed or preserved significantly larger clusters than WT cell (Fig. 19B). Of note, the higher cluster ratios and larger cluster grades were also reproduced in two other independent CIRPSR MCF7 *ESR1* mutant models (Appendix B Fig. 89). Together, these studies show that mutations in *ESR1* cause increased cell-cell attachment under static and fluidic conditions, and that the effect size differs between the mutation sites and genetic backgrounds.

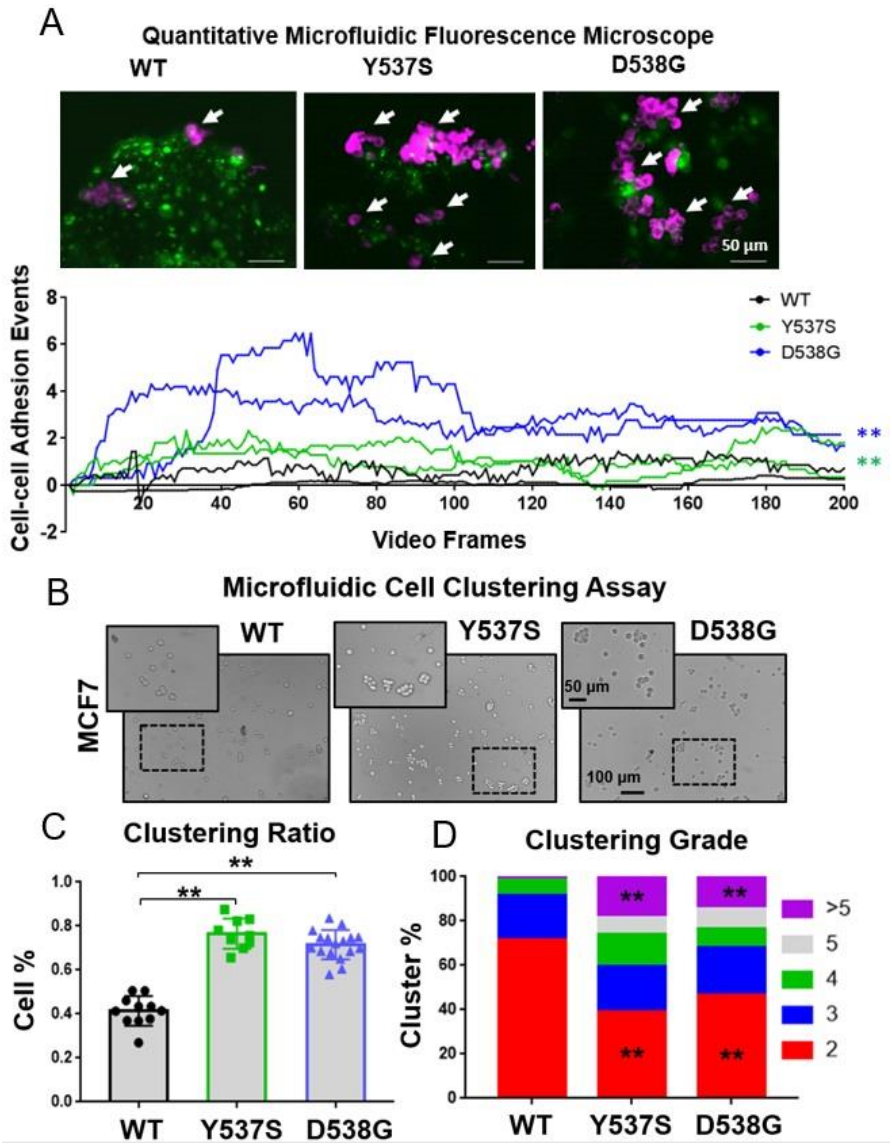



**Figure 19. *ESR1* mutant cells exhibited stronger cell-cell interaction under microfluidic condition.**

A. Representative dynamic adhesion curves of MCF7 *ESR1* mutant cells with qMFM. Cells labelled with calcein (green) were seeded on a coverslip for 24 hours. Then the same type of cells labelled with DiD (pink) were loaded into the microfluidic system. Videos were recorded for 200 frames with 100ms/frame. Representative images from each video of qMFM are shown in the top panel. Adhered cells are indicated with white arrows. Calculated cell-cell interaction events were normalized to the total calcein signals of attached cells in each video. At least two biological replicates were used in this experiment. Dunnett's test was used between WT and mutant cells. Scale bar: 50 µm. This experiment was done once. B. Representative images of cell cluster formation after two hours of flow with physiological shear stress with ibidi microfluidic system.

C. Bar graph representing the percentage of cells in cluster version after quantification of cluster and single cell numbers. Dunnett's test was used between WT and mutant cells. D. Stack bar chart showing the cluster grade distribution. Fisher's exact test was used between WT and each mutant cell type under the same cluster grade. Data shown were from a representative experiment among three independent repeats. (* $p<0.05$, ** $p<0.01$).

*Fig. 19A is in collaboration with Dr. Maritza Montanez and Dr. Prithu Sundd for the operation of the qMFM device, Mr. Callen Wallace and Dr. Simon Watkins for the development of automatic quantification program. Fig. 19B, 19C and 19D are provided by Yang Wu.

#### 3.3.2.3 *ESR1* Mutant Cells Induced Enhanced Metastasis *In Vivo*

To test whether *ESR1* mutant cells induce metastasis *in vivo*, particularly via the alteration of cell clustering capacity in bloodstream, we next conducted a tail vein injection of both MCF7 and T47D *ESR1* mutant models into nude mice. Upon above 50% mice if any of the group showed the IACUC pre-defined signs of euthanasia, the entire cohorts were then euthanized. Macro-metastasis was examined by H&E staining whereas micro-metastasis was tested specifically in lung by immunofluorescence staining using human specific CK19 antibody.

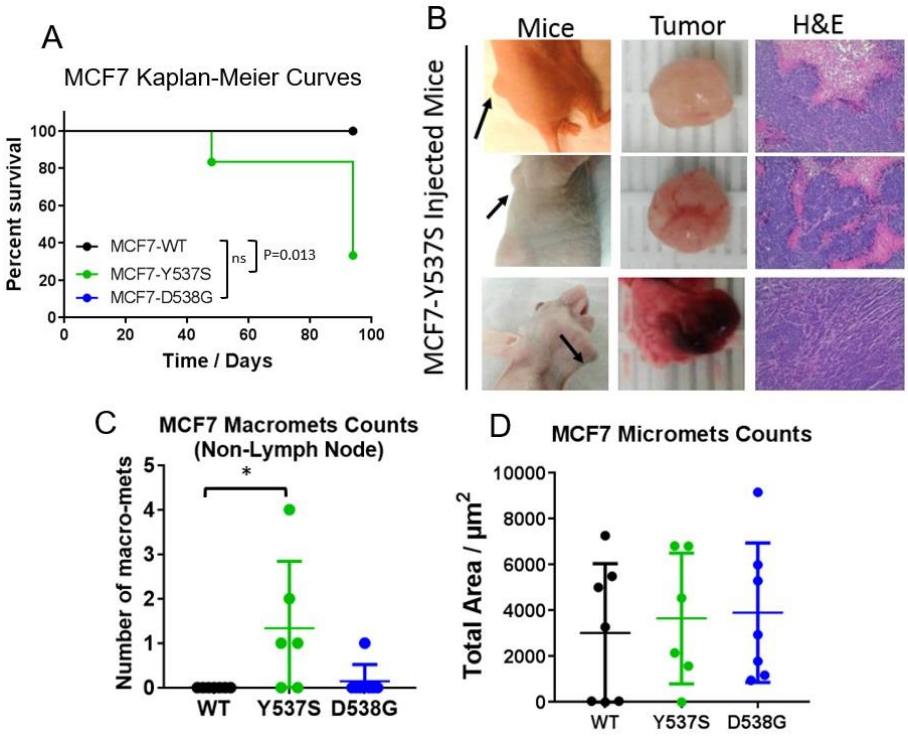



**Figure 20. MCF7-Y537S cells induces more macro-mets *in vivo*.**

A. Kaplan-meier curves of MCF7 *ESR1* mutant and WT cells-injected mice (WT n=7; Y537S n=6; D538G n=7) over 12 weeks. Pair-wised log rank test were utilized. B. Representative images of MCF7-Y537S-induecd macro-mets with the views of visible tumors on mice (left panel), dissected tumors from mice (middle panel) and the corresponding H&E staining of the tumorous portions from the entire tumor sections (right panel). The three tumors shown here were found from back and neck. C. Dot plots showing the macro-mets (non-lymph node mets) counts per mouse from MCF7-*ESR1* mutant-injected mice. Pair-wised Mann-Whitney U test was used to compare the macro-mets numbers in each group to WT cells-injected groups. D. Quantification of lung-micromets areas based on CK19 staining. Two representative areas of each slides were selected and quantified by Ms. Kara Burlbaugh blindly using ImageJ. Pair-wised Mann-Whitney U test was used. This experiment was done once. (* $p<0.05$)

Starting from 10 weeks after injection, we observed multiple distant macromets developed in 4/6 MCF7-Y537S mutant cells-injected mice. We euthanized all the MCF7-injected mice at week 12 and counted them as uncensored death. Overall, we were able to determine significantly poor survival of the MCF7-Y537S mouse group up to 12 weeks (Fig. 20A). 3/7 MCF7-WT cells

exclusively developed lymph node tumors around kidney, whereas MCF7-Y537S cells derived more distant metastatic tumors at other sites than WT cells. These metastatic sites included back, leg, abdominal and jaw (Fig. 20B). Only one macromets was observed in D538G group was discerned (Fig. 20B). The more pronounced metastatic capacity in MCF7-Y537S than D538G cells was possibly an outcome of the later latency of MCF7-D538G subtype recently reported by Jeselsohn et al. We further quantified and compared the micro-metastasis loci areas in the lung sections and detected no difference between WT and mutant cells-injected mice, this is likely due to the high basal lung colonization capacity of MCF7 cell models (Fig. 20D).

On the other hand, T47D cells did not induce visible macro-metastatic tumors, whereas we observed sever weight loss and water retention in multiple WT and mutant cell-injected mice of that cohort at week 23. We thus euthanized the cohort accordingly. No differential survival was observed between WT and mutant groups (Fig. 21A). After dissection, we only observed two macromets from each T47D mutant group (Fig. 21B). However, we identified both T47D-Y537S and D538G mutant cells resulted in more lung micro-metastatic loci, with a more pronounced effect by D538G cells (Fig. 21C and 21D).

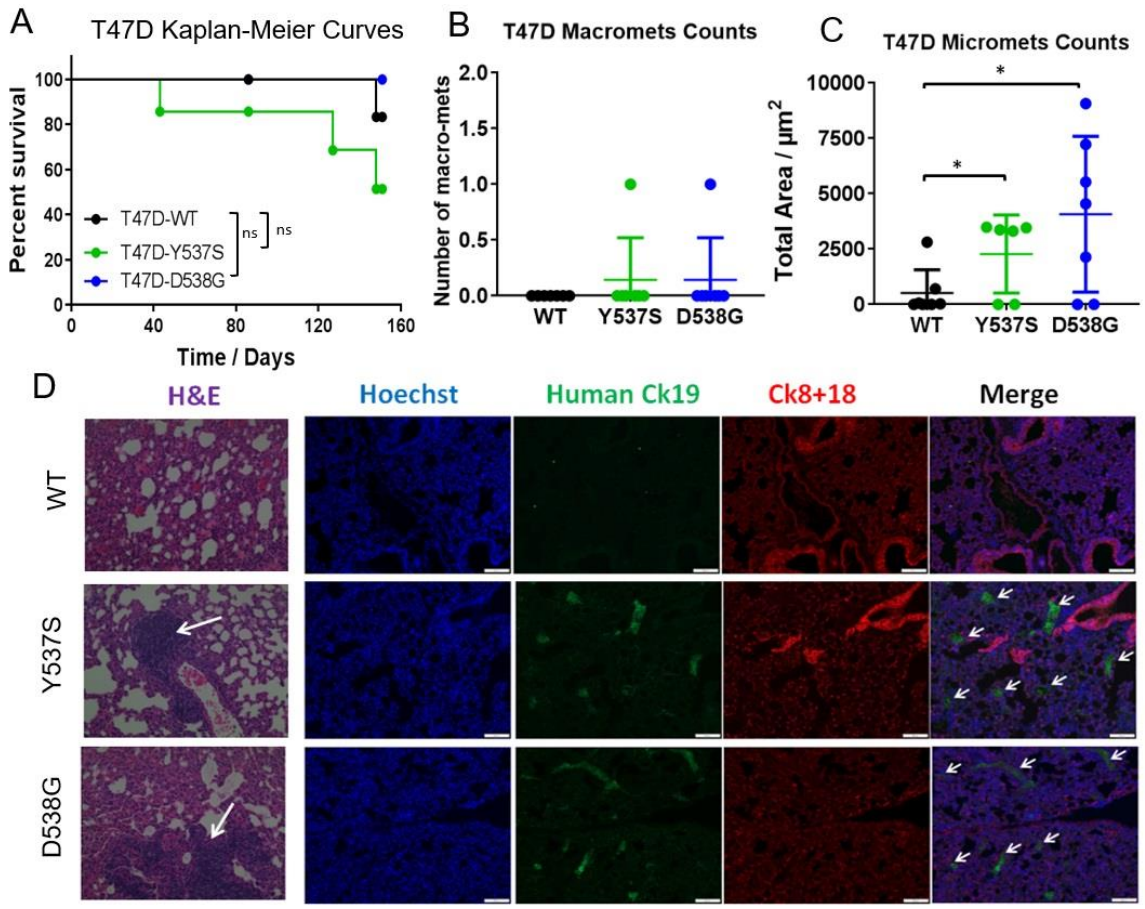



**Figure 21. T47D *ESR1* mutant cells induced more lung micro-mets *in vivo*.**

A. Kaplan-meier curves of T47D *ESR1* mutant and WT cells-injected mice (WT n=7; Y537S n=6; D538G n=7) over 23 weeks. Pair-wised log rank test were utilized. B. Dot plots showing the macro-mets counts per mouse from T47D-*ESR1* mutant-injected mice. Pair-wised Mann-Whitney U test was used to compare the macro-mets numbers in each group to WT cells-injected groups. C. Quantification of lung-micromets areas based on CK19 staining and blind quantification. Two representative areas of each slides were selected and quantified by Ms. Kara Burlbaugh blindly using ImageJ. Pair-wised Mann-Whitney U test was used. D. Representative images of micro-metastatic evaluations on T47D-*ESR1* mutant cells-injected mice, including H&E staining (first column) and immunofluorescence for mouse lung tissue nuclear (blue), CK8+18 (red) and human-specific CK19 (green) images. (* $p<0.05$) This experiment was done once.

Taken together, MCF7 *ESR1* mutant cell models are more invasive than T47D cells based on the tail vein injection results. Within each cell line background, MCF7-Y537S and both T47D mutant

cells mutant cells exhibited higher metastatic potential *in vivo* based on macro- or micro-metastatic evaluation respectively. This experiment again highlights the functional aspects of *ESR1* mutations are largely dependent on the genetic backgrounds and mutation subtypes.

#### 3.3.2.4 *ESR1* Mutant Cells Reprogram Desmosome and Gap Junction Adhesomes

Given the different phenotypes seen in cell adhesion, I more closely investigated the four major cell-cell junction subtypes – desmosomes, gap junctions, tight junctions and adherens junctions with gene set variation analysis (GSVA) of RNA-Seq data[207, 208] (Appendix D Table 13). Expression of desmosome genes were enriched in both MCF7-Y537S/D538G and T47D-Y537S cells, and gap junction genes were additionally enriched in MCF7-Y537S cells as compared to WT cells. (Fig. 22A). Tight junctions were enriched in WT cells, and there were no differences in AJ gene expression (Appendix B Fig. 90). Individual gene expression analysis ($FC>1.2$; $p<0.05$) identified 18 commonly upregulated desmosome and 4 gap junction genes in both MCF7 *ESR1* mutant cells (Fig. 22B).

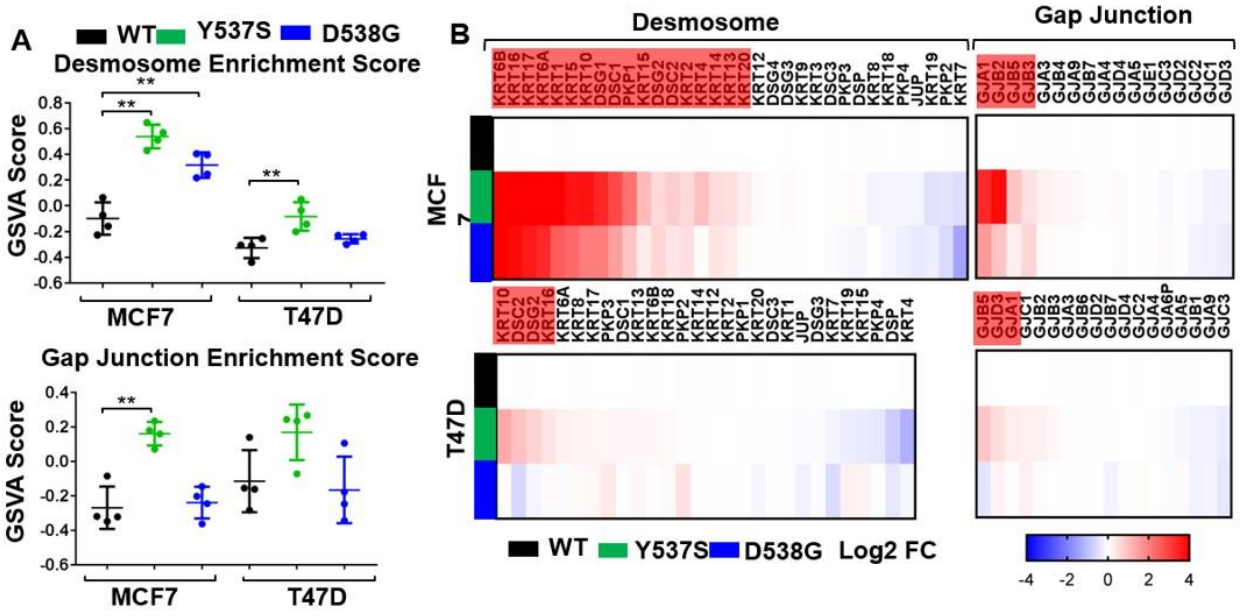



**Figure 22. RNA-seq reveals enrichment of desmosome and gap junction gene network in *ESR1* mutant cells.**

A. Gene Set Variation Analysis (GSVA) between MCF7 and T47D mutant and WT transcriptomes of desmosome and gap junction gene sets. Each cell type has four biological replicates. Dunnett's test was used to test the significance between cell lines. B. Heatmaps show all desmosome and gap junction component genes in MCF7 and T47D *ESR1* mutant cells. Data were extracted from RNA-sequencing results. The expression levels were normalized to WT cells. Genes with counts=0 in more than one replicate in all cell types were filtered out. Genes with log2FC>1.2 and p <0.05 are labelled in red. (** p<0.01)

In addition to keratins, induction of classical desmosome genes *DSC1/2*, *DSG1/2* and *PKP1*, and gap junction genes *GJA1*, *GJB2* and *GJB5* was observed and confirmed by qRT-PCR in MCF7 cells (Fig. 23A). Higher protein levels were also observed for *DSC1*, *DSG2*, *PKP1*, *GJA1* (Cx43), and *GJB2* (Cx26) (Fig. 23B). Immunofluorescence staining allowed for visualization of higher *DSG2* expression in MCF7 Y537S (D538G showed an increasing trend with p value equals to 0.32) at cell-cell contact areas (Fig. 23C and 23D). Furthermore, I validated the overexpression of the desmosome and gap junction genes *DSC2*, *DSG2* and *PKP1* in another recently described

genome edited MCF7-Y537S cell line[109] (Appendix B Fig. 91). In addition, the expression of connexin 43 was also validated in 5 out of 6 Y537S-derived macro-metastatic tumors by IHC staining. (Appendix B Fig. 92)

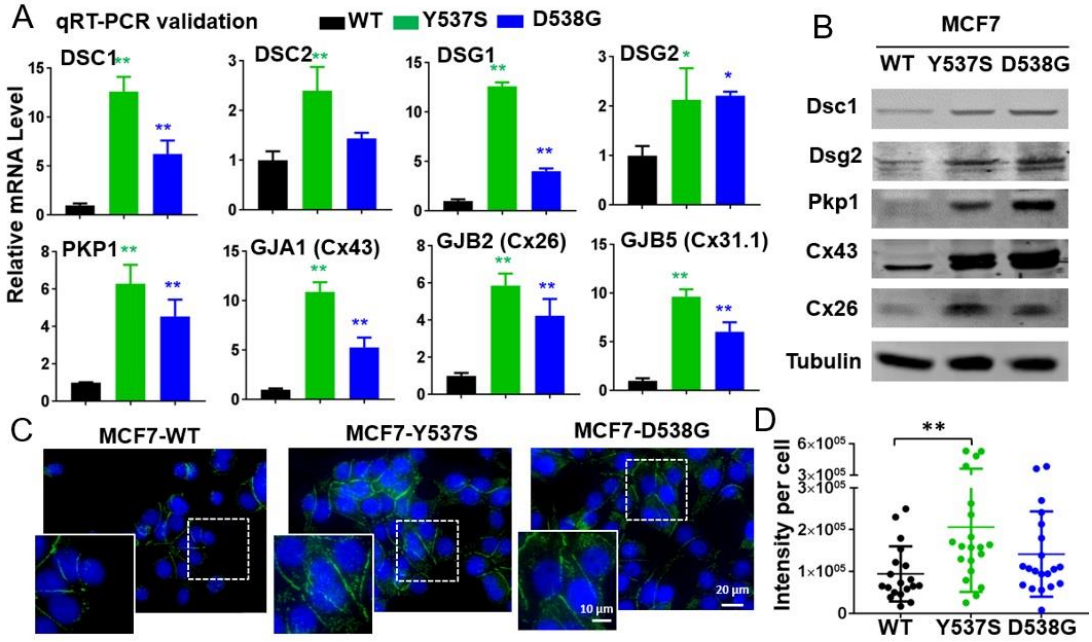



**Figure 23. Validation of essential desmosome and gap junction genes in MCF7 *ESR1* mutant cell models.**

A. qRT-PCR validation of selected altered candidate desmosome and 3 gap junction genes in MCF7 *ESR1* mutant cells. Dunnett's test was used to compare the gene expression between WT and each mutant. Data were from one representative experiment with three biological replicates of four independent experiments. B. Western blot validation of the expressional level of *DSG2*, *DSC1*, *PKP1*, Cx43 and Cx26 in MCF7 WT and *ESR1* mutant cells after hormone deprivation. Tubulin was detected as a loading control. Data were from one representative experiment of at least three independent experiments. C. Representative images of immunofluorescence staining showing the distribution of desmoglein 2 (*DSG2*) in MCF7 WT and *ESR1* mutant cells. D. *DSG2* signal intensities were quantified and normalized to cell numbers of each image. Magnificent images of the selective regions were shown in the left corner. Data from 20 regions of images were combined from four independent experiments and quantified in a blind manner. Dunnett's test was used to test the significance between WT and mutant cells. Scale bar: 20 and 10 μm. (* $p<0.05$, ** $p<0.01$)

*Fig. 23C is provided by Jian Chen.

Consistent with the weaker cell-cell adhesion phenotypes in T47D mutant cells (Fig. 17A and 17B), gene expression changes were less pronounced, with induction of three and four gap junction and desmosome genes, respectively, in Y537S and minor effects in D538G cells (Fig. 22B). The upregulation of *DSC2, GJA1, GJB2* and *GJB5* were further validated by qRT-PCR in T47D-Y537S mutant cells (Fig. 24).

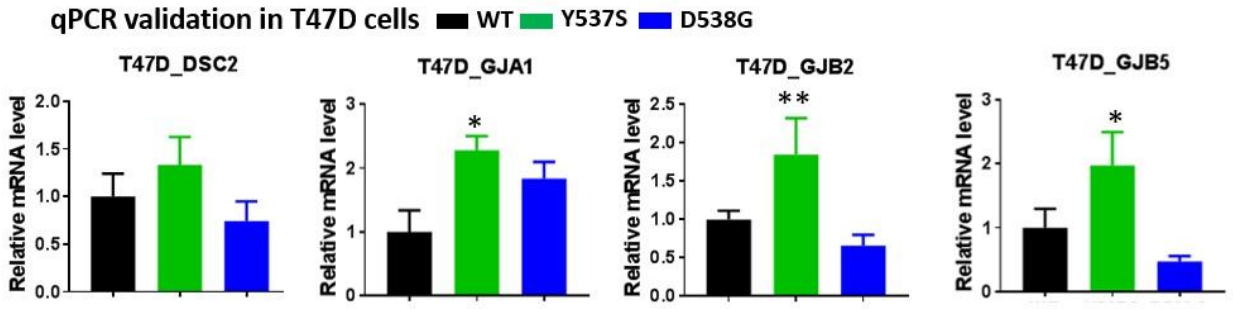



**Figure 24. Validation of essential desmosome and gap junction genes in T47D *ESR1* mutant cell models.**

qRT-PCR validation of top altered desmosome and gap junction genes in T47D *ESR1* mutant cells. ΔΔCt method was used to analyze relative mRNA fold changes normalized to WT cells and RPLP0 levels were measured as the internal control. Dunnett's test was used to compare the gene expression between WT and each mutant. Data were from one representative experiment with three biological replicates of two independent repeats. (* $p<0.05$; ** $p<0.01$)

### 3.3.2.5 Gap Junctions Are Essential for The Enhanced Cell-cell Adhesion in *ESR1* Mutant Cells

I next investigated the functional roles of the reprogrammed adhesome in the *ESR1* mutant MCF7 cells. Transient individual knock-down of *DSC1* and *GJA1* did not cause any changes in MCF7 Y537S and D538G cell-cell adhesion (Appendix B Fig. 93). However, treatment with the irreversible pan-gap junction inhibitor Carbenoxolone (CBX) caused significant inhibition of cell-

cell aggregation, similar to that of the calcium chelator EDTA, suggesting redundancy in the reprogrammed adhesome in the *ESR1* mutant cells (Fig. 25A). These findings were further validated under microfluidic condition, where treatment of CBX significantly blocked cell clustering trend (Fig. 25B) and the formation/preservation of larger cell clusters (Fig. 25C).

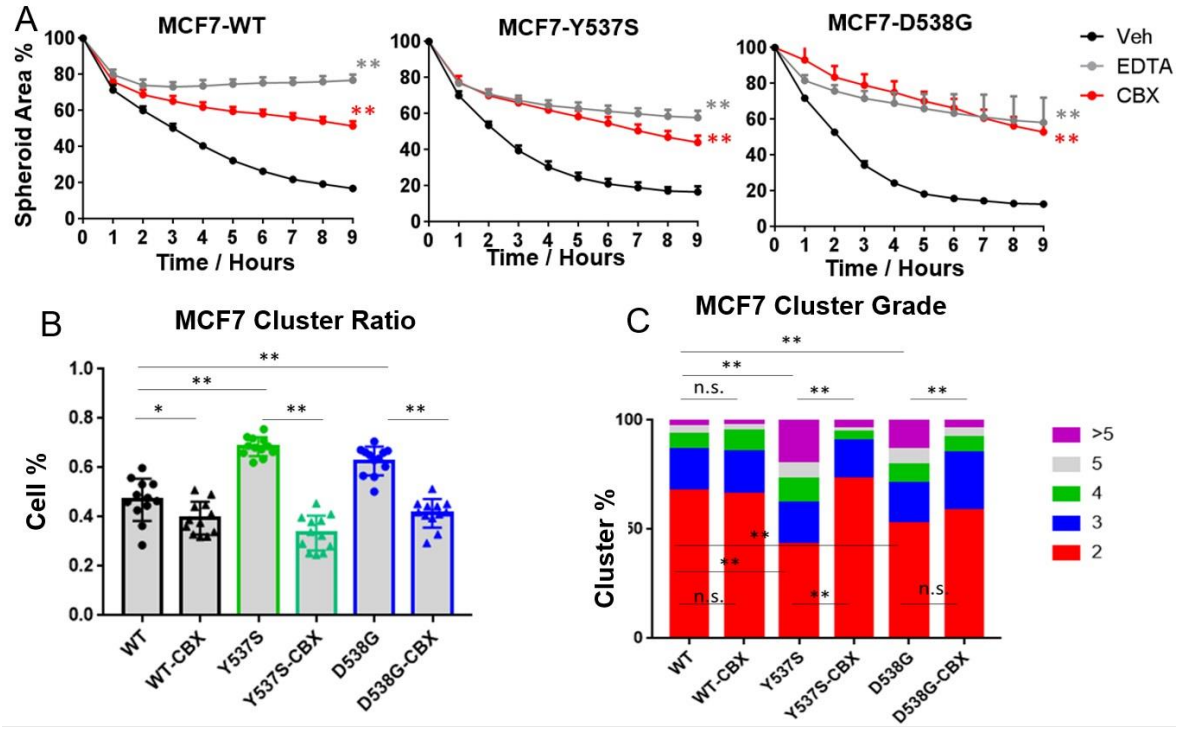



**Figure 25. Gap junctions are essential for the enhanced cell-cell adhesion in *ESR1* mutant cells.**

A. Cell aggregation rate assay in round bottom ULA plates was performed to MCF7 WT and mutant cells with the treatment of vehicle, 2mM EDTA or 100 μM CBX. Images of cell aggregation were captured every 1 hour and spheroid areas were normalized to time 0. Pairwise two-way ANOVA between vehicle and each condition in each cell type was used to test the drug effects. Data were from one representative experiment with six biological replicates of three independent experiments. B and C. Cell cluster ratio and cluster grade distribution of MCF7 WT and *ESR1* mutant cells after 2 hours of microfluidic shear stress with or without 100 μM CBX pretreatment for two days. Pair-wised t test was used to examine the effects of CBX between each group for cluster ratio. Pair-wised Fisher's exact test was used to test the CBX effects on cluster grade distribution. Data were from one representative experiment among three independent experiments. (* $p<0.05$; ** $p<0.01$)

*Figure 25B and 25C are provided by Yang Wu.

In addition, I utilized calcein-based dye transfer assays, which revealed significantly higher dye exchange ratios in MCF7 mutant compared to WT cells, likely reflecting increased number of active gap junctions on the cell surface of *ESR1* mutant cells (Fig. 26). In summary, MCF7-Y537S/D538G and T47D-Y537S showed increased expression of desmosome and gap junction component genes, with the overexpression of the gap junction genes contributing to the enhanced cell-cell adhesion phenotype.

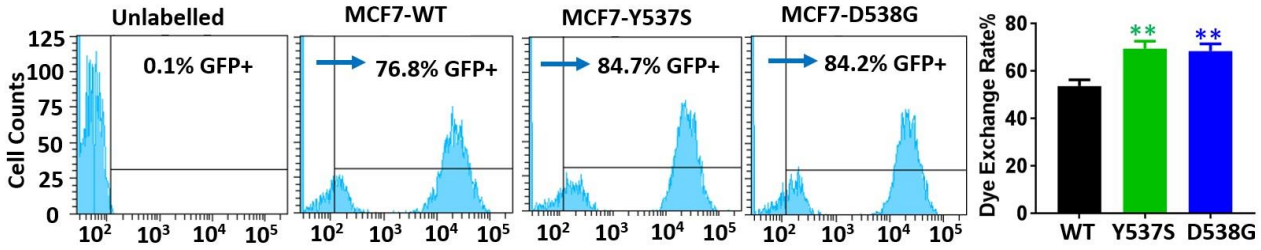



**Figure 26. MCF7 *ESR1* mutant cells show enhanced functional gap jucntions.**

MCF7 WT/mutant cells were labelled with calcein and mixed equally with unlabeled cells. Cells were incubated for 12 hours and GFP+ cells populations were tested by flow cytometry. Dye exchange ratio was calculated by 2x (GFP+ cell percentages - 50%). Dunnett's test was used to compare the dye exchange ratio. Data were from one representative experiment with three biological replicates of three independent experiments. (** $p<0.01$)

### 3.3.2.6 Desmosome and Gap Junction Signature Enrichment is Recapitulated in *ESR1* Mutant Metastatic Tumors

I next sought to extend our findings to clinical specimens. We first analyzed the RNA-seq data set composed of 51 pairs of ER+ primary-matched metastatic tumors (7 *ESR1* mutant and 44 *ESR1* WT pairs) merged from our in-house WCRC cohort[172, 173, 209] (36 pairs) and DFCI cohort (15 pairs)[210] (Fig. 27A and Appendix D Table 14). After subtracting each enrichment

score of primary tumors from the matched metastatic lesions (Δ GSVA Score), I found that metastatic tumors harboring *ESR1* mutations exhibited significantly higher enrichment levels of the sum of desmosome and gap junction signatures than WT ones (Fig. 27B). This finding further confirmed our *in vitro* results in clinical samples.

**A**

**ER+ primary-matched metastatic tumors**

| Cohorts | ESR1 WT (N=44) | ESR1 Mutated (N= 7) |
|---|---|---|
| WCRC | 32 | 4 |
| DFCI | 12 | 3 |

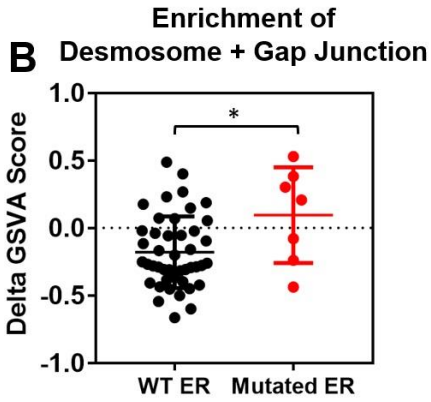



**Figure 27. Desmosome and gap junction signature enrichment is recapitulated in *ESR1* mutant metastatic tumors**

A. Paired ER+ metastatic tumor sample information in WCRC and DFCI cohorts. B. Gene Set Variation Analyses were performed on patient matched primary paired samples with gene sets of desmosome/gap junction. Delta GSVA score of each sample was calculated by subtracting the scores of primary tumors from the matched metastatic tumors to eliminate the individual backgrounds. Mann-Whitney U test was performed to compare the Delta GSVA scores between WT or *ESR1* mutation harboring tumors. (* $p<0.05$)

### 3.3.2.7 Enhanced Cell-cell Adhesion Is Recapitulated in Ex Vivo Circulating Tumor Cell Models with Hotspot *ESR1* Mutations

Stronger adhesion of *ESR1* mutant cells under fluidic conditions suggests that circulating tumor cells (CTCs) with mutant ER might form more clusters. Yu et al. recently generated six *ex vivo* CTC cell lines isolated from patients with advanced ER+ breast tumors, three of which harbor *ESR1* mutations[177]. Mining RNA-seq data from WT and *ESR1* mutant CTCs, we observed overexpression of three out of the top eight gap junction and desmosome genes (*DSC2, DSG2,*

*GJA1*) (Fig. 28A) in *ESR1* mutant CTC lines. In addition, suspension culture of the validated *ex vivo* CTC lines exhibited the formation of tighter cell clusters in *ESR1* mutation-harboring CTCs (Fig. 28B). Ingenuity pathway analysis of genes differentially expressed between WT and mutant CTCs (n=100 upregulated and 49 downregulated genes; q value <0.1) identified "Cell Death and Survival" as the top altered pathway in *ESR1* mutant CTC lines, in addition to "Cell movement" and Cell-to-Cell Interactions" (Fig. 28C), pathways also seen in *ESR1* mutant tumors. Altogether, these data indicate that the enhanced intercellular adhesion may confer a survival advantage to *ESR1* mutant CTCs in the blood stream.

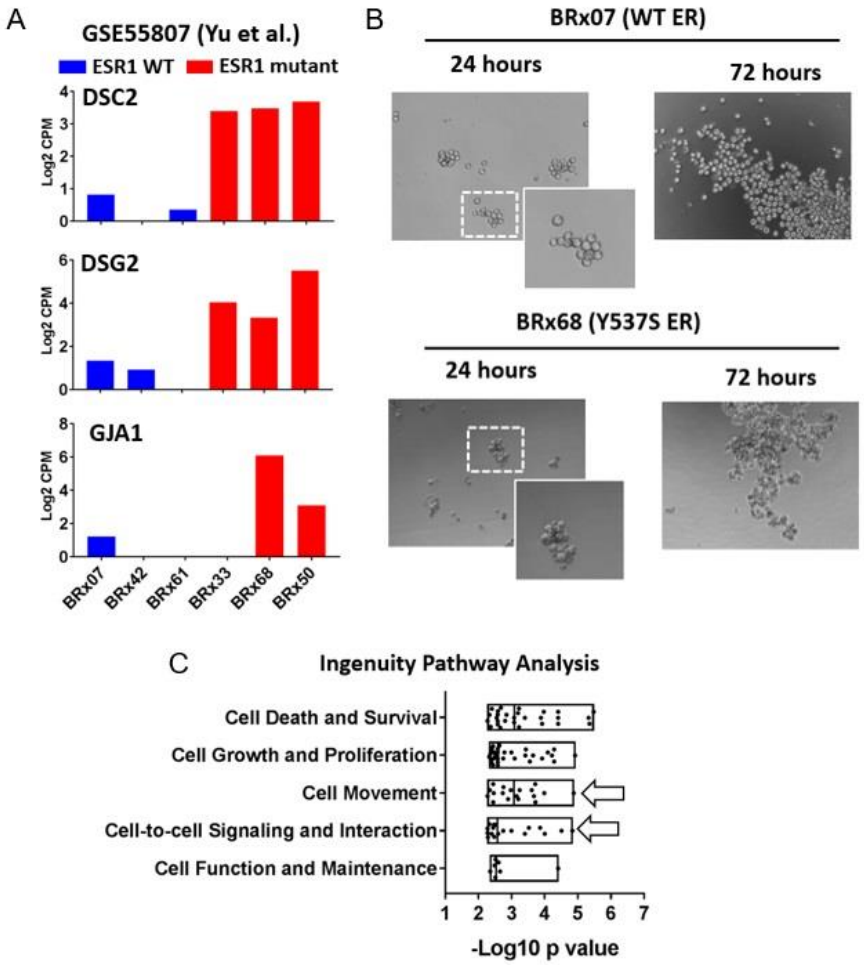



**Figure 28. Circulation tumor cells with *ESR1* mutations form more compact clusters.**

A. Comparison of *DSC2*, *DSG2* and *GJA1* gene expression levels between 3 *ESR1* mutant and 3 WT ex vivo circulation tumor cell lines. Data were extracted from public available RNA-sequencing data (GSE55807). B. Representative images of Brx68 and Brx07 CTC cell clusters suspended in ultra-low attachment plate for 24 and 72 hours. Objective of 10X was used. Images of selected clusters are highlighted with larger magnification. C. Top five altered cellular and molecular functions and enriched pathways in *ESR1* mutant CTCs using IPA.

* Fig. 28B was provided by Yonatan Amzaleg from Dr. Min Yu's group.

#### 3.3.2.8 *ESR1* Mutation Allele-specific Cistromes Reveal ER Independent Regulations

Our systematic evaluation of *ESR1* mutant-associated metastatic phenotypes identified a group of candidate genes driving cell-cell adhesion. We next set out to determine whether an altered mutant ER cistrome causes constitutive expression of the identified desmosome genes (*DSC1, DSC2, DSG1, DSG2, PKP1*) and gap junction genes (*GJA1, GJB2, GJB5*). We therefore mapped ER binding globally by performing ER chromatin immunoprecipitation sequencing (ChIP-seq) in WT and *ESR1* mutant cells in the absence or presence of estradiol.

As expected, hormone deprivation led to low amounts of ER binding sites in WT MCF7 (n=125) and T47D (n=615) cells, whereas E2 stimulation triggered substantial ER recruitment events (MCF7, n=12,472; T47D, n=1,724) (Appendix D Table. 15). Although we called overall fewer ER peaks than other publicly available ER ChIP-seq experiments [117, 128], Y537S and D538G ER were consistently recruited to binding sites in the absence of hormone, resulting in 657 binding sites in MCF7-Y537S, 1,016 in MCF7-D538G, 1,096 in T47D-Y537S and 1,468 in T47D-D538G cells (Appendix D Table 15). Co-occupancy analysis using ChIP-seq data from two additional genome-edited cell models[109, 117] revealed considerable overlap varying between 36% to 68% dependent on genetic background and conditions (Appendix B Fig. 94). Strikingly, none of the identified desmosome or gap junction genes showed functional ER binding sites in their promoter or enhancer regions (-/+ 100kb of transcriptional start sites). Together, these data indicate a lack of direct ER regulation of the mutant *ESR1*-target genes mediating metastatic properties. We therefore further analyzed the ChIP-seq data to potentially elucidate ER-independent mechanisms of transcriptional regulation.

As a first step, we performed a general characterization of the data set. We first compared all ER peak intensities in the absence of hormone at E2-stimulated ER binding sites in WT cells. In both MCF7 and T47D cell lines, all mutant cells displayed the distributions of stronger ER binding peaks compared to WT-vehicle groups at those ER-regulated sites (Appendix B Fig. 95A), which was also validated in terms of average binding intensities (Appendix B Fig. 95B). As expected, Principle Component Analysis (PCA) using all E2-induced ER binding sites segregated mutant-vehicle binding sites from both WT-vehicle and mutant-E2 (Fig. 29A). We also compared the intensities of ER mutant binding sites to WT binding sites in the absence of estrogen and visualized the peaks with differential binding intensities (Fig. 29B). The analysis showed that at least 60% of binding sites exhibited increased binding intensities in mutant cells (MCF7-Y537S:75%; MCF7-D538G: 81%; T47D-Y537S: 72%; T47D-D538G: 61%), suggesting a dominant ligand-independent redistribution of mutated ER.

Co-occupancy analyses between WT-E2 and mutant-vehicle sets illustrate that while MCF7-Y537S/D538G and T47D-Y537S cells shared at least 60% of their binding sites with WT, T47D-D538G only shared 33% with WT (Fig. 29C). Of note, E2 treatment caused 7-fold more induced peaks in MCF7 compared to T47D, consistent with the results in our recent transcriptomic study[110], and other previous cistromic studies[211]. Analysis of genomic feature distributions showed that the ER in T47D-D538G had an enrichment of binding in intergenic regions (Fig. 29D and Appendix B Fig. 95C). In total, shared peaks revealed 57% and 41% overlap between the two cell lines, representing classical ER-target genes, such as *GREB1* (Fig. 29C). Overall, our data are in line with previously described ligand-independent and neomorphic transcriptomics and

cistromics data[110, 117], while additionally showing strong context-dependence of the binding events.

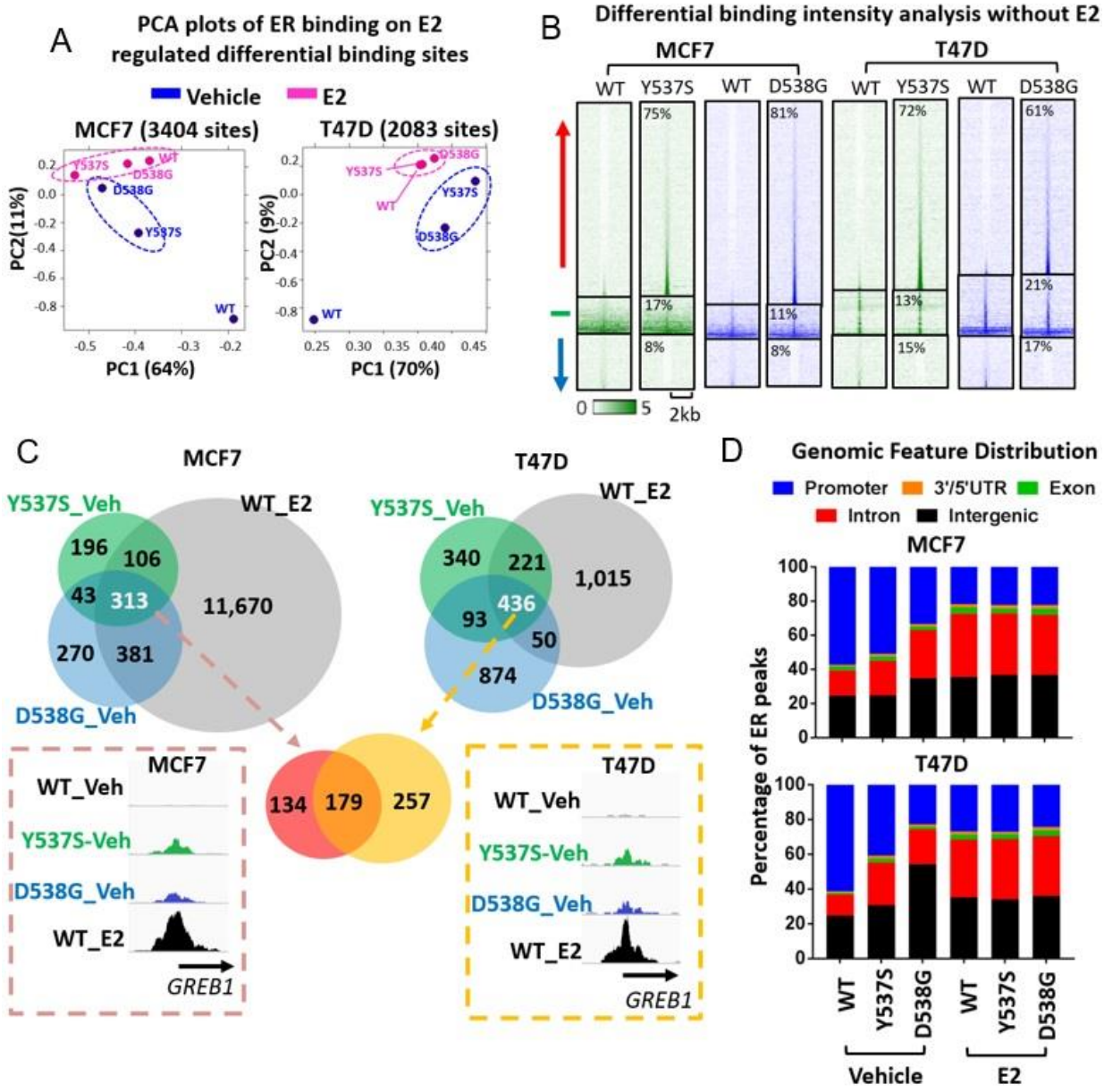



**Figure 29. Mutation allele-specific cistromes reveal ER independent regulation.**

A. Principle component analysis (PCA) of WT and mutant cells on basis of the differential ER binding sites derived from E2 treatment in WT and mutant cells. In total, 2,083 and 3,404 binding sites were identified to be differentially expressed in T47D and MCF7 cells, respectively ($P<0.01$). B. Heatmaps of differential ER binding intensities in Y537S, D538G mutants compared to WT ER in a pairwise manner, shown in a horizontal window of ± 2kb from the peak center. The pairwise comparison between WT and mutant samples were performed to calculate the fold change (FC) of intensities and the binding sites were sub-classified into sites with increased intensity

(FC>2, red arrow), decreased intensity (FC<-2, blue arrow), and non-changed intensity (-0.2<FC<0.2, green line). Percentage of each subgroups were labelled on the heatmaps respectively. C. Venn Diagrams showing the occupancy intersection between WT-E2, Y537S-vehicle and D538G-vehicle groups in MCF7 and T47D cell lines. Shared peaks were further intersected between the two cell lines. ER binding at the *GREB1* gene promoter region in both cell lines are shown as an example of E2-regulated and ligand-independent binding. D. Genomic feature distribution patterns of all ER binding sites in the presence or absence of E2 in all cell types. The promoter regions are defined as +/- 3 kb around the transcriptional start sites.

* ER ChIP-seq data was generated by Amir Bahreini, PhD. Fig. is from Amir Bahreini, PhD.

#### 3.3.2.9 *ESR1* mutations induce *GJA1* expression via cFos-cJun secondary regulation

Given no additional gained ER binding sites were detected in our ChIP-seq, I then alternatively hypothesized that mutated ER drives cell-cell adhesion genes via an indirect regulation. To examine this, I set out a transient knockdown of *ESR1* in MCF7 models. Among the top six increased desmosome and gap junction genes, knockdown of *ESR1* only diminished the expression of *GJA1* in *ESR1* mutant cells, whereas the expression of *DSC1*, *DSG1*, *PKP1*, *GJB2* and *GJB5* were increased (Fig. 30). In addition, E2 treatment did not increase *GJA1* expressional levels in MCF7-WT cells (Appendix B Fig. 96). These results suggest that these genes are regulated by distinct mechanisms, and *GJA1* is likely to be the only gene as an outcome of a secondary transcriptional effects of canonical ER ligand-independent activation or an ER-dependent epigenetic reprogramming.

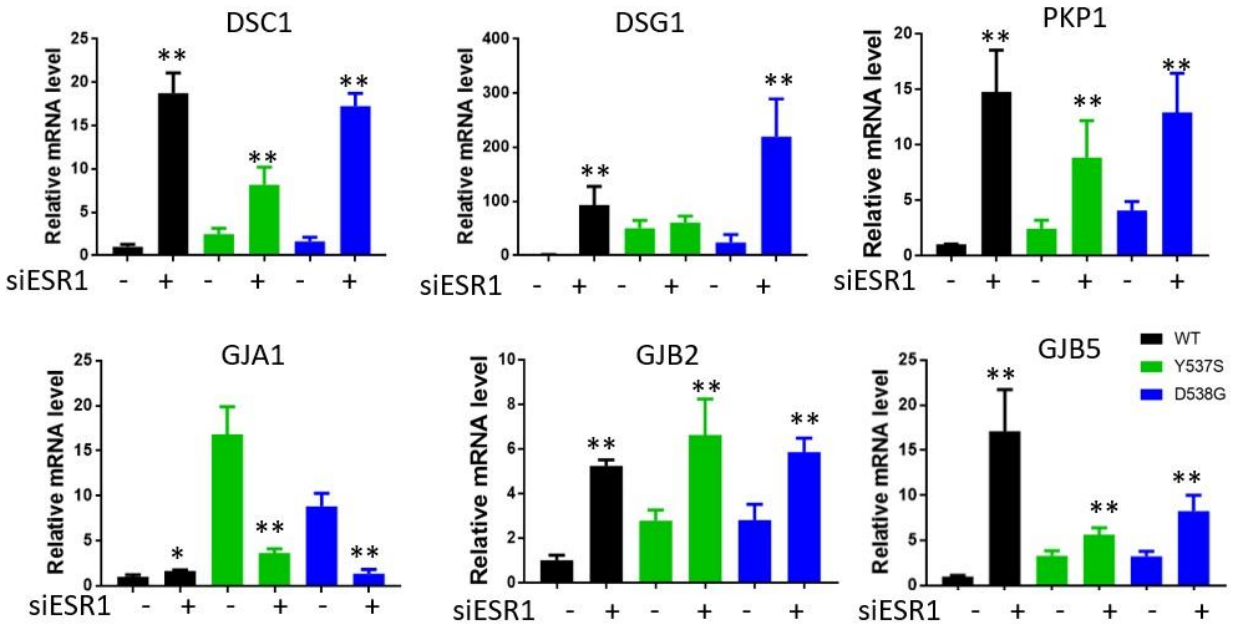



**Figure 30. Knockdown of *ESR1* diminished *GJA1* expression in *ESR1* mutant cells.**

Bar graphs representing qRT-PCR measurement of *DSC1, DSG1, PKP1, GJA1, GJB2 and GJB5* mRNA levels in MCF7 WT and *ESR1* mutant cells with siRNA knockdown of *ESR1* for 7 days. ΔΔCt method was used to analyze relative mRNA fold changes normalized to WT cells and RPLP0 levels were measured as the internal control. Each bar represents mean ± SD with three biological replicates. This experiment was replicated for three times independently, and representative results from one experiment were shown. Student's test was used to compare the gene expression between scramble and knockdown groups of each cell type. (* $p<0.05$; ** $p<0.01$)

I then further searched for potential transcriptional factors that related to *GJA1* transactivation and compared their levels in WT and *ESR1* mutant cell models. Previous publications[212] indicated six major transcriptional factor families that directly involve in *GJA1* transcription. Among which, the *FOS* and *LEF1* showed significant upregulation in both MCF7 *ESR1* mutant cells and E2-treated WT cells, suggesting that they are directly induced by the canonical ER ligand-independent program (Fig. 31A). Gene set enrichment analysis further confirmed the AP1 gene signature is significantly enriched in both Y537S and D538G *ESR1* mutant cells (Fig. 31B).

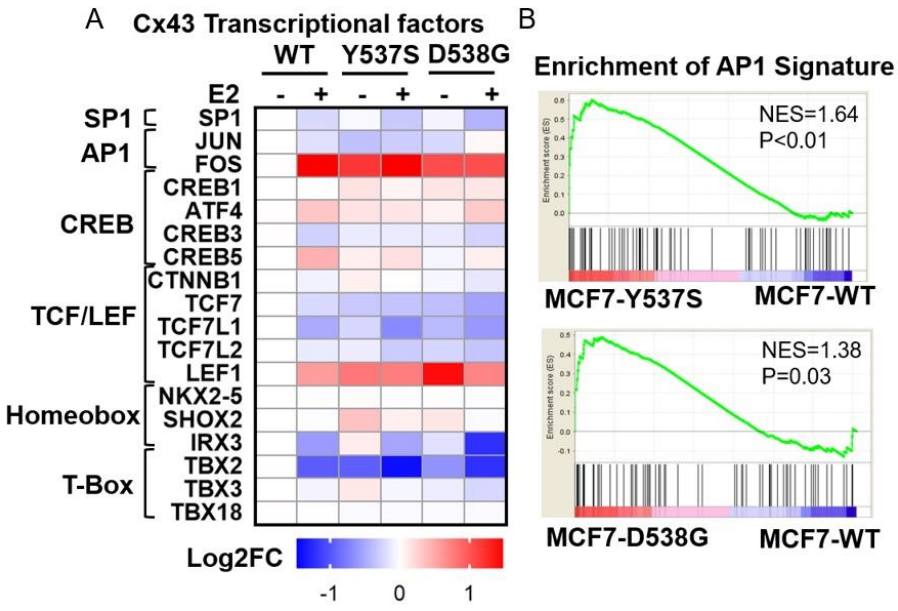



**Figure 31. AP1 transcriptional signature is enriched in MCF7 *ESR1* mutant cells.**

A. Heatmap representing the expression levels of the 18 transcriptional factors associated with *GJA1*. Data were extracted from RNA-seq with four biological replicates and normalized to WT-veh groups. B. Gene set enrichment plots of AP1 transcriptional signatures in the transcriptomes of *ESR1* mutant cells.

To further validate this at protein levels, I set out to a similar seven-days knockdown assay of *ESR1* in MCF7 models and tested the expression level of Cx43 *(GJA1*) using immunoblotting. Knockdown of *ESR1* significantly blocked the expression of both Cx43 and cFOS (Fig. 32A). To further examine whether Cx43 expression dependents on cFOS upregulation in *ESR1* mutant cells, I treated MCF7 *ESR1* mutant cell models with a compound, T5224, that specifically inhibits the interaction of cFOS with AP1 sites. Blockade of cFOS-AP1 interaction partially abolished the upregulation of Cx43 (Fig. 32B). Taken together, our results suggest that Cx43 is likely induced by the secondary upregulation of cFOS via a transactivation mechanism, which is directly triggered by ligand-independent activation of mutant ER.

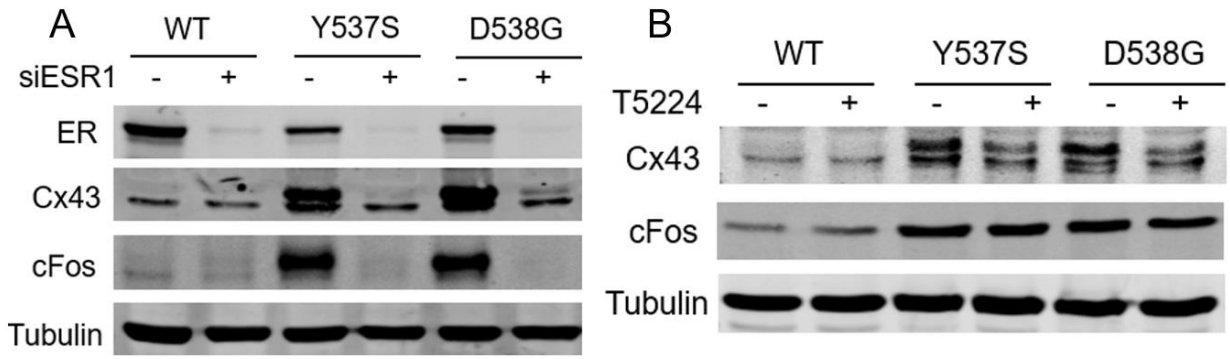



**Figure 32. Increased Cx43 is dependent on cFOS upregulation in *ESR1* mutant cells.**

Western blot validation of the expressional level of ER, Cx43, cFOS in MCF7 WT and *ESR1* mutant cells after seven days of *ESR1* knockdown (A) or three days of 20 μM T5224 treatment (B). Tubulin was detected as a loading control. Data were from one representative experiment of two independent repeats of each panel.

#### 3.3.2.10 *ESR1* Mutations Regulate Desmosome Gene TranscriptionVia Epigenetic Reprogramming

Given the identification of ER-independent regulation, alongside the lack of detected ER binding sites at the majority of our candidate gene loci, I thus hypothesize that mutant ER triggers long-term ER-independent epigenetic reprogramming to increase chromatin accessibility and transcription of those genes. Overlapping ChIP peaks, in the absence of E2, with histone modification markers showed increased active marks (H3K27ac, H3K4me1/me3 in MCF7; H3K27ac and H3K4me2 in T47D) and decreased inactive marks (H3K9me3 in both cell line, H3K27me3 in T47D) in mutants (Fig. 33A). However, T47D-D538G again showed opposite trends compared to other mutants. This analysis demonstrated the contextual epigenetic remodeling in ER mutant cells, with active alterations in MCF7-Y537S/D538G and T47D-Y537S cells, and decreased global chromatin accessibility in T47D-D538G cells, in the absence of E2. Higher intersection ratios were still detectable in MCF7 mutant cells towards H3K27ac, H3K4me3 and H3K9me3, whereas epigenetic activities were in general decreased in T47D mutant cells

versus WT-E2 group. Furthermore, GSVA using a epigenetic regulator gene set (n=652) from the EpiFactors database[213], indicate significant enrichment of epigenetic signature in MCF7 mutants vs. WT cell (Fig. 31B).

To further verify whether active histone modification is elevated on our target gene loci, I performed a H3K27ac and H3K4me1 ChIP-qPCR to compare the recruitment intensities between MCF7 WT and mutant cells on the nearest two histone modification sites around *DSC1* and *DSG1* loci, the two most-upregulated desmosome component genes in MCF7 mutant cells (Fig. 33C). The results showed enhanced H3K27ac and H3K4me1 recruitments in both Y537S and D538G cells at the *DSC1* peak, whereas only enhanced H3K27 acetylation were observed at the *DSG1* peak, suggesting the activation of desmosome genes via an ER-independent epigenetic mechanism in mutant cells (Fig. 33D).

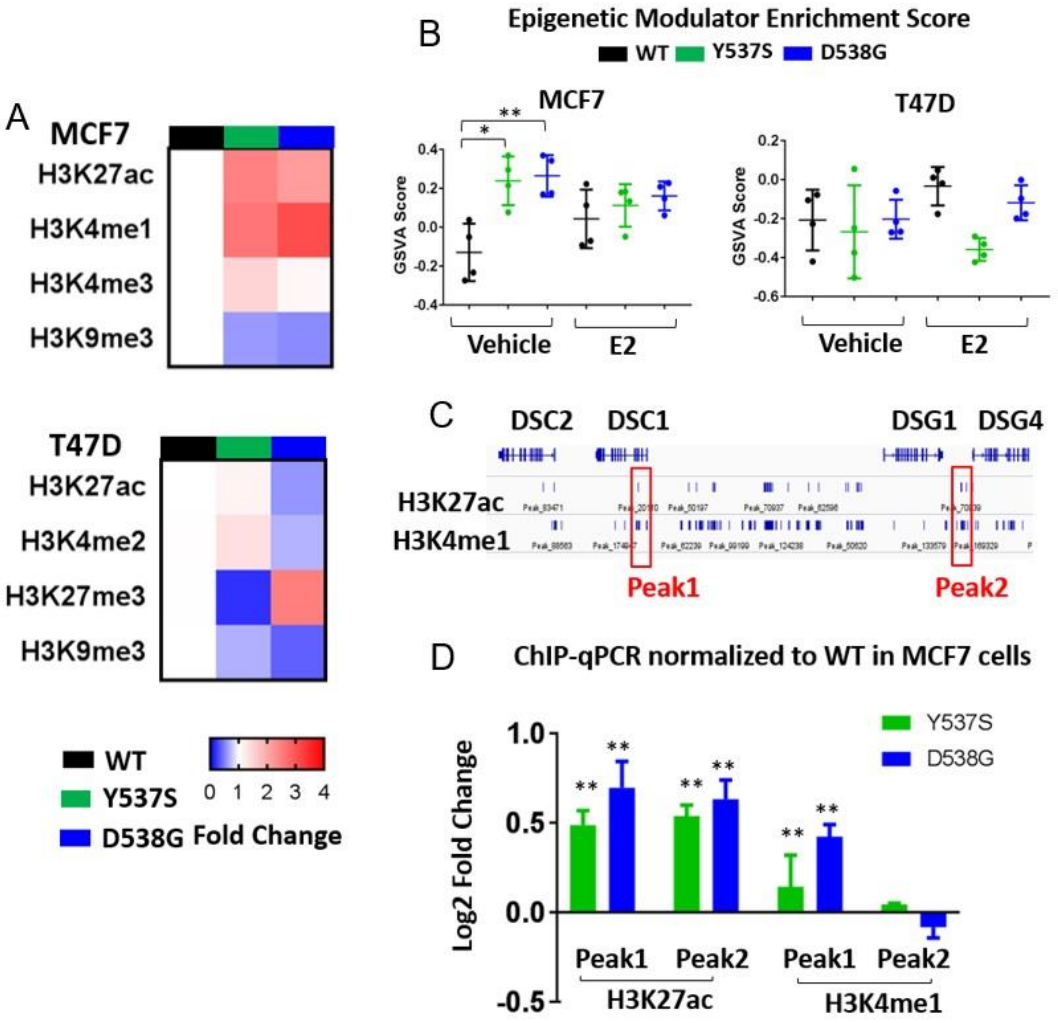



**Figure 33. *ESR1* mutations induce desmosome via epigenetic reprogramming**

A. Co-occupancy analysis of ER binding peaks to various histone modification binding sites from public available data sets within the same cell lines (MCF7: GSE96352, T47D:GSE63109). Intersection ratios of mutant cells were normalized to WT-Vehicle groups. B. GSVA enrichment score of epigenetic regulator gene sets in the transcriptomes of MCF7 and T47D WT/mutant cells. Log2TPM values were used from RNA-seq data set. Each cell type has four biological replicates. Dunnett's test was used to test the significance between cell line. C. Schematic view of histone modification binding sites around *DSC1* and *DSG1* gene bodies selected for ChIP-qPCR. D. Enrichment of H3K27ac and H3K4me1 bindings at the putative histone modification sites around desmosome gene locus. Percentage of input were calculated based on ChIP-qPCR, and then normalized to WT cells. Dunnett's test was used. (* $p<0.05$; ** $p<0.01$)

Collectively, our results provided stronger evidence that *ESR1* mutations gained enhanced expression of desmosome and gap junction genes, which potentially enhanced cell-cell adhesion and facilitates cell cluster formation/preservation especially when traveling in the bloodstream. Escalated CTC cluster formation then further enhance cell metastatic capacity in blood vessel and caused malignant metastasis in ER+ breast cancer patients. The mechanism underlying of this section is summarized in Figure 34.

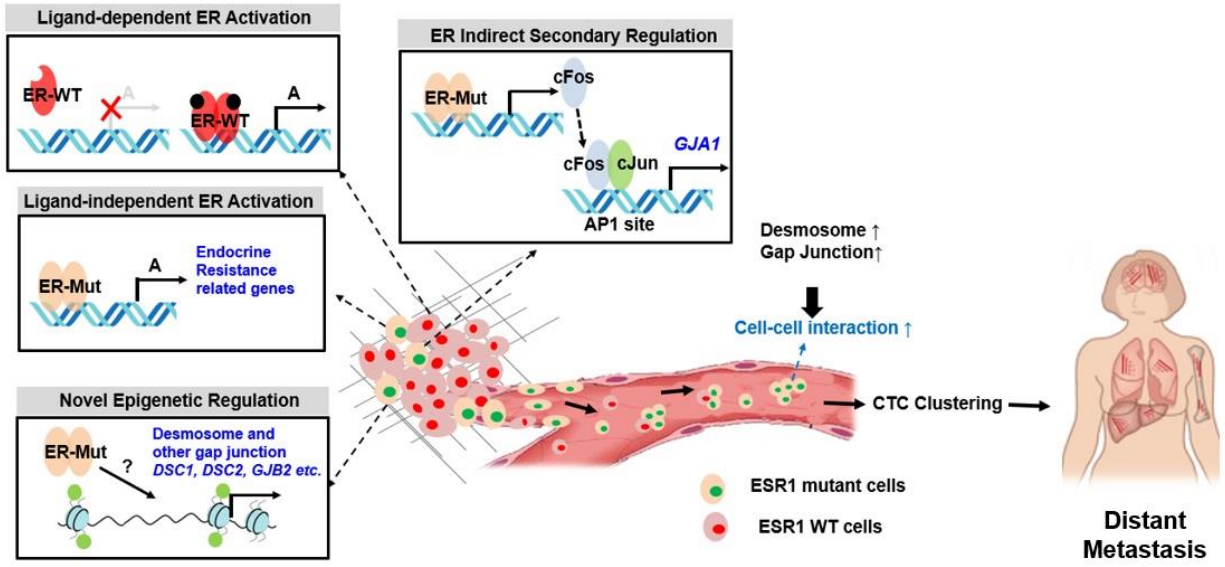



**Figure 34. Schematic model of *ESR1* mutation-driven metastasis via cell-cell adhesome reprogramming.**

Mutated ER triggers either ligand-independent transcriptional or novel epigenetic regulation. Ligand independent transcription includes either constitutive binding to canonical ER regulated sites or mutant-ER-selective binding events with recruitment of novel co-regulatory factors to induce novel target genes. Novel ER-indirect regulation includes increasing histone modification to increase chromatin accessibility of specific gene loci such as desmosome gens. In addition, *ESR1* mutations also triggers gene expression via a secondary indirect regulation, exemplified by *GJA1*. Consequently, Increased desmosome and gap junction expression, results in enhanced cell-cell adhesion, which ultimately facilitates the metastasis of *ESR1* mutant cells.

### 3.3.3 Hotspot *ESR1* Mutations Inhibit Adhesive and Enhance Invasive Properties Via *TIMP3*-MMP Axis

#### 3.3.3.1 *ESR1* Mutant Cells Show Diminished ECM Adhesion

Besides cell-cell adhesion, metastasis is additionally mediated by coordinated changes in cell-matrix interaction[214, 215], I therefore examined the correlated phenotypic changes by assessing the adhesive properties of *ESR1* mutant cell models on seven major ECM proteins (Fig. 35A). MCF7 *ESR1* mutant lines consistently showed diminishing trends of adhesion on collagen I, collagen IV, laminin, tenascin, and vitronectin, whereas alleviated adhesions on collagen I, collagen II, fibronectin, laminin and vitronectin were detected in T47D *ESR1* mutant cells. Of note, a strong cell line context-dependent adhesive alteration was observed on fibronectin. I next compared the expression of the genes or gene sets (Appendix D Table 16) encoding these individual ECM components in ER+ primary tumors from TCGA (n=808) and METABRIC (n=1,505) cohorts (Appendix B Fig. 97). I determined the collagen I genes (*COL1A1* and *COL1A2*) as the most abundant ECM component, indicating the potentially high clinical impacts resulted from the differential adhesion in primary tumors. We further repeated our adhesion assay on collagen I and confirmed the consistently attenuated adhesion in all mutant cells (Figure 35B and 35C). The same trend was also observed in each individual *ESR1* mutant clone of both cell lines (Appendix B Fig. 98), eliminating the false positive possibility from clonal specific effects. This was additionally visualized in coculture adhesion assays on collagen I using differentially labelled MCF7 and T47D *ESR1* WT and mutant cells, which recapitulated the mutant-specific adhesive deficiencies in a co-existing system that reflects *ESR1* mutation-harboring primary tumors

(Appendix B Fig. 99). Of note, the diminished adhesive properties in *ESR1* mutant cells were also reproducible in two other MCF7 CIRSPR cell models from Jason Gertz and Simak Ali's group (Appendix B Fig. 100).

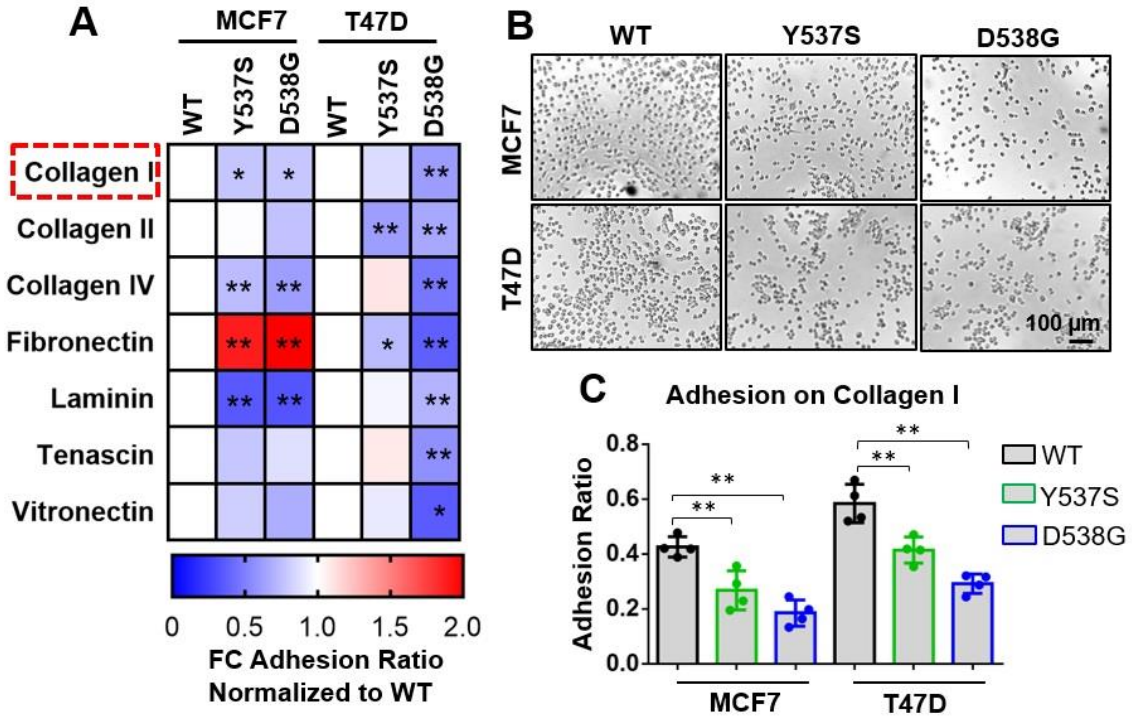



**Figure 35. *ESR1* mutant cells show diminished adhesion on collagen I.**

A. Heatmap representation of ECM adhesion array with MCF7 and T47D *ESR1* mutant cells. Adhesion ratio of each condition with biological quadruplicates was quantified by dividing the numbers of remining cells after wash to the original total cells. All data were further normalized to WT cells within each cell line. Dunnett's test was applied to each condition of each cell line. This array was done once. B. Representative images of collagen I adhesion assay. Cells remaining on collagen I after three times of PBS wash were imaged in bright field. C. Quantification of adhesion ratios on collagen I of each cell type. Bar graphs represent the mean ± SD with at least four biological replicates in each group. This experiment was reproduced for 10 (MCF7) and 6 (T47D) times independently, and representative results from one experiment were shown. Dunnett's test was utilized within each cell line. (* $p<0.05$, ** $p<0.01$)

#### 3.3.3.2 Decreased *TIMP3* Expression Drives the Loss of Collagen I Adhesion in *ESR1* Mutant Cells

I further sought to explore the molecular mechanisms causing the unique defect of collagen I adhesion in *ESR1* mutant cells. Given the integrin family proteins have been well-characterized as the direct collagen I adhesion receptors, I first compared the enrichment of the overall integrin gene sets (Appendix B Fig. 101) in our cell models, using the transcriptome data from our previous study[110]. No differences of integrin overall enrichment were observed in both MCF7 mutants and T47D-Y537S cells, whereas T47D-D538G showed a slightly decreasing trend compared to WT cells (Appendix B Fig. 101A). An additional comparison of individual integrin genes identified a few integrin genes exhibiting cell-line specific alterations, such as *ITGA2*, *ITGAM* and *ITGA9* in MCF7, and *ITGA1*, *ITGA10* in T47D *ESR1* mutants (Appendix B Fig. 101B). These inconsistent changes of integrin components suggest a possibly alternative and indirect ECM adhesion regulator that leads to the phenotype.

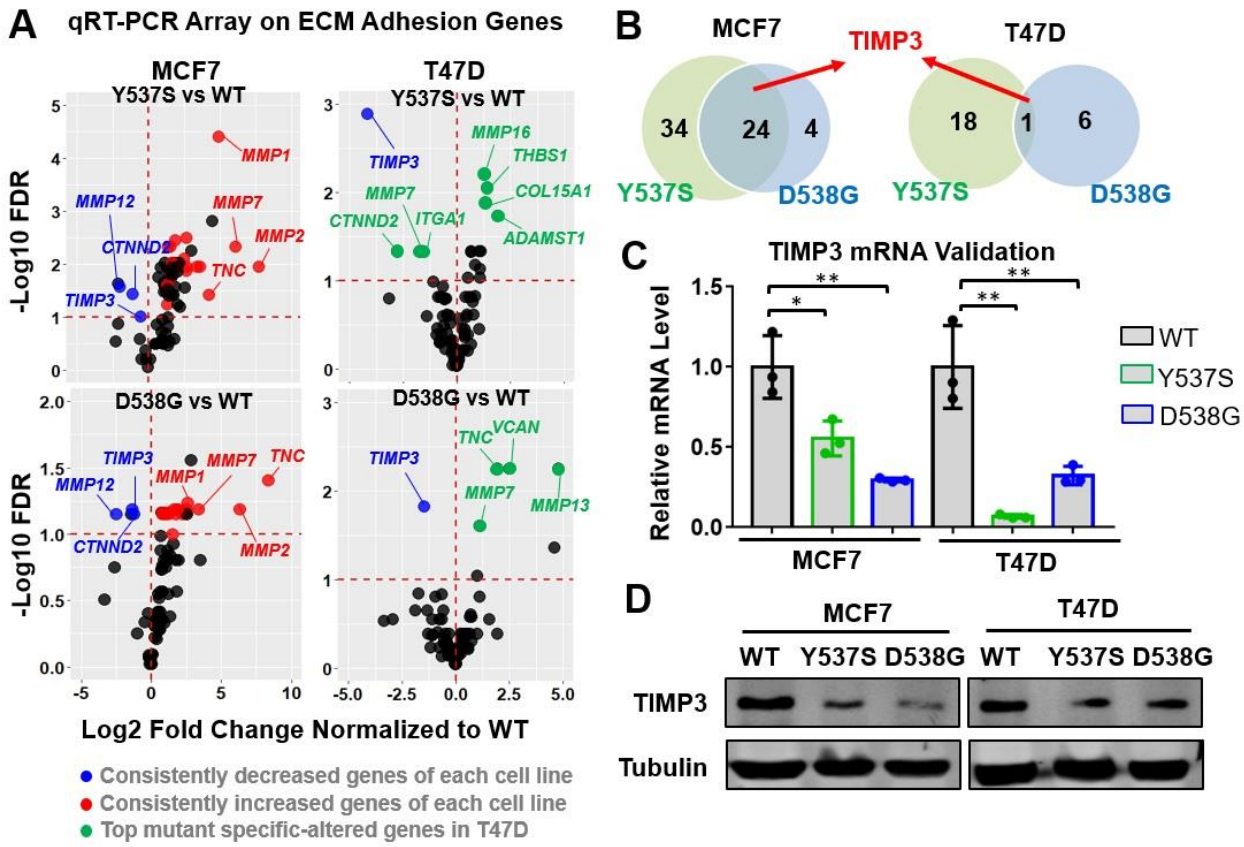



**Figure 36. qRT-PCR array revealed allevieated *TIMP3* expression in *ESR1* mutant cells.**

A. Volcano plots showing the alterations of 84 ECM adhesion genes in all the mutant cell types in a pairwise comparison to WT cells. Genes filtered with average Ct<35 in at least one group and FDR<0.1 were considered as significantly altered genes in *ESR1* mutant cells. Overlapped downregulated (blue) or upregulated (red) genes between the two mutants of each cell line were further highlighted, with gene name labelled for the top targets. B. Venn diagrams showing the consistently and differentially expression genes between the two mutant types within each cell line. *TIMP3* was pointed as the only intersected gene in all four cell types. This qPCR array was done once. C. qRT-PCR validation of *TIMP3* expression in WT and *ESR1* mutant cells. Ct values were normalized to RPLP0 and further normalized to WT cells. Bar graphs represent the mean ± SD with biological triplicates in each group. Five independent experiments of each cell line were reproduced and data were from one representative experiment. Dunnett's test was utilized within each cell line. D. Western blot validation of *TIMP3* from whole cell lysates after hormone deprivation. Tubulin was used as a loading control. This experiment was reproduced for six (MCF7) and three (T47D) times independently and representative images from one experiment were shown for each cell line. (* $p<0.05$, ** $p<0.01$)

To narrow down this target list, I repeated RNA analysis focusing on gene expression of 84 ECM adhesion-related genes using a qRT-PCR array (Appendix D Table 17). Pairwise comparisons between each mutant line and their corresponding WT cells revealed a strong context-dependent pattern of ECM regulation network reprogramming (Fig. 36A). MCF7 *ESR1* mutant cells displayed more differentially expressing ECM adhesion genes (58 for Y537S, 28 for D538G) than T47D models (19 for Y537S, 7 for D538G). Intersection between MCF7 Y537S and D538G mutants showed 20 (e.g. *MMP1*, *TNC*) and 4 (e.g. *TIMP3*, *CTNND2*) consistently up- and down-regulated genes respectively. In contrast, T47D cell line showed strong mutation subtype-specific patterns. The uniquely altered genes of each mutant include multiple MMPs (e.g. *MMP7, MMP13*) and integrins (e.g. *ITGA1*), whereas the two T47D mutants only shared one commonly downregulated gene-*TIMP3* (Fig. 36B). *TIMP3*, encoding tissue metallopeptidase inhibitor 3, was turned out to be the only gene shared between all four mutant cell types (Fig. 36B), and this was confirmed at mRNA level by qRT-PCR in the both pooled and individual clones (Fig. 36C and Appendix B Fig. 102), as well as at protein levels by immunoblot (Fig. 36D).

*TIMP3* expression was also significantly lower in additional genome-edited and overexpression *ESR1* mutant models previously described by Harrod et al.[109] and Jeselsohn et al[117] (Appendix B Fig. 103). Overexpression of *TIMP3* rescued the collagen I adhesive defects in *ESR1* mutant cells (Fig. 37A and 37B), implying a causative role for *TIMP3* downregulation in the decreased cell-matrix adhesion phenotype of the *ESR1* mutant cells.

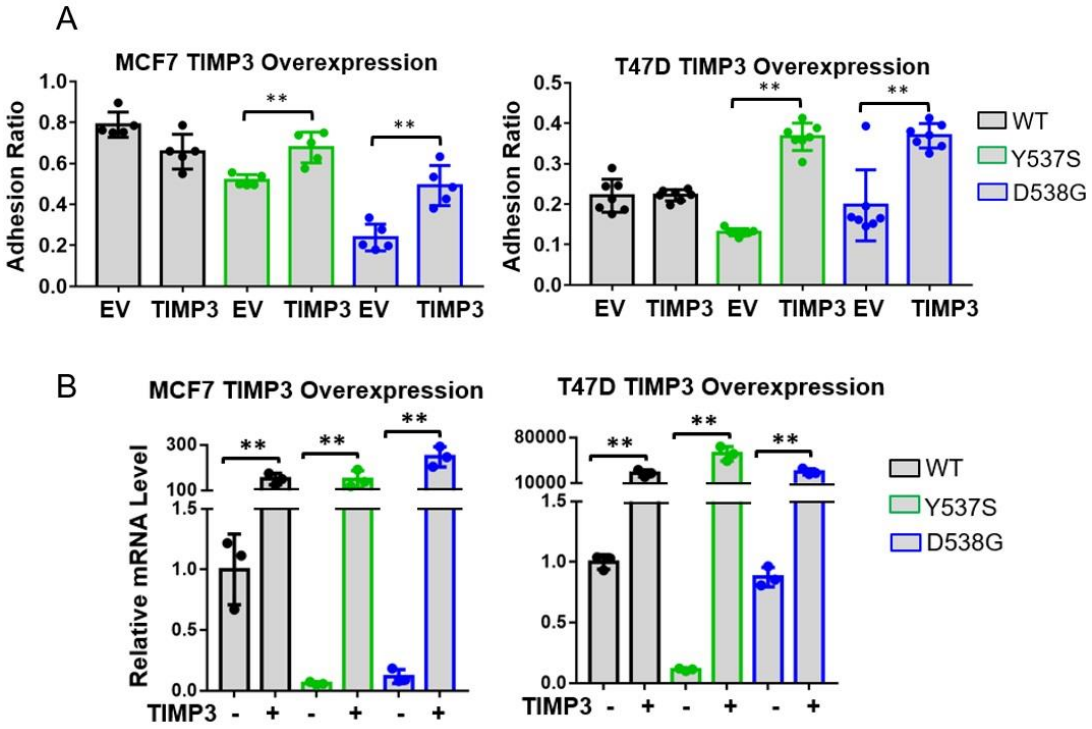



**Figure 37. *TIMP3* overexpression rescued the diminished ECM adhesion of *ESR1* mutant cells.**

A. Quantification of adhesion ratios on collagen I of each cell type with transfection of pcDNA empty vector or *TIMP3* plasmids. Bar graphs represent the mean ± SD with at least four biological replicates in each group. This experiment was replicated for four times independently, and representative results from one experiment were shown. Student's t test was used to compare empty vector and *TIMP3* overexpression groups. B. qRT-PCR validation of *TIMP3* overexpression in MCF7 and T47D cells. Ct values were normalized to RPLP0 and further normalized to WT empty vector group. Bar graphs represent the mean ± SD with biological triplicates in each group. Pair-wised student's t test was used to compare the effects to *TIMP3* overexpression. The validation was done only once (* $p<0.05$, ** $p<0.01$)

### 3.3.3.3 *ESR1* Mutant Cells Display Enhanced MMP Activities and Invasive Properties in Collagen I

Given *TIMP3* is well characterized as a MMP inhibitor, I then hypothesized that dampened *TIMP3* expression contributes to the elevated MMP activity. To examine this, I compared MMP

activities between *ESR1* WT and mutant cells. A FRET-based MMP enzymatic activity assay revealed significantly increased MMP activation in all mutant cells with a more pronounced effects MCF7-Y537S cells (Fig. 37A), indicating that the *ESR1* mutant cells have increased capacity for matrix digestion. This was further validated by a spheroid invasion assay in collagen I. *ESR1* mutant cell-derived spheroids showed significantly lager invasive areas than WT spheroids when surrounded by collagen I (Fig. 37B and 37C), whereas no differences in cell growth were observed in collagen I within the same period (Appendix B Fig. 104), suggesting the effect was exclusively from escalated invasion. The enhanced invasive properties in *ESR1* mutant cells were further supported by a cocultured spheroid invasion assay with differentially labelled WT and mutant cells. More dominant mutant cells-derived fluorescent signals were visualized in the invasive edge of each spheroid (Appendix B Fig. 105). In summary, these results suggest that *ESR1* mutant cells gained enhanced invasive capacities via increased MMP activities resulted from *TIMP3* down-regulation.

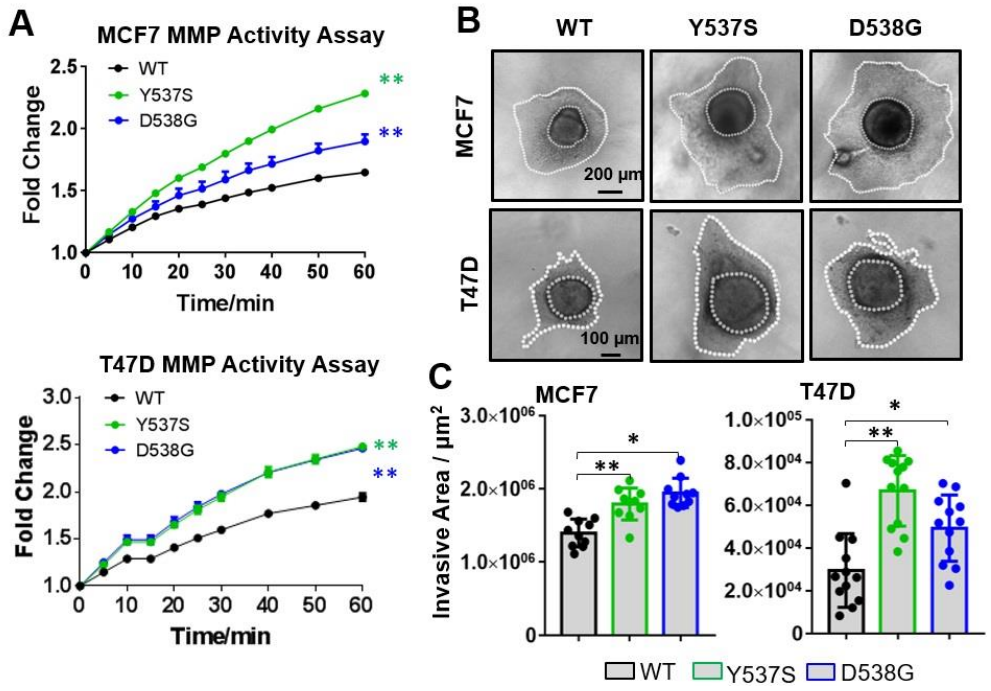



**Figure 38. *ESR1* mutant cells display enhanced MMP activities and invasive properties in collagen I**

A. Graphic view of pan-MMPs FRET kinetic assay. MMPs in cell lysates were pre-activated and mixed with MMP substrates. Fluorescence were measured in a time course and normalized to the readouts of WT cells. This experiment was reproduced for three times independently of each cell line. Representative results from one experiment were shown. Pair-wised two-way ANOVA between WT and each mutant cell type were performed. B. Representative images of spheroid-based collagen invasion assay of *ESR1* WT and mutant cell models. MCF7 and T47D spheroids were mixed in collagen I for 4 and 6 days respectively. Bright field images were taken accordingly. C. Quantification of invasive areas from b. Invasive areas were calculated by subtracting each original spheroid area from the corresponding total area at endpoint. Three independent experiments were performed with at least five biological replicates each time. Representative data from one experiment were shown. Dunnett's test was used to compare between WT and each mutant cells. (* $p<0.05$; ** $p<0.01$)

### 3.3.3.4 Enhanced Invasion of *ESR1* Mutant Cells Are Sensitive to MMP Inhibition

Last but not least, I sought to identify clinically actionable vulnerabilities of *ESR1* mutations-driven metastasis. Marimastat, a pan-MMP inhibitor, has been reported in multiple

phase II/III clinical trials towards advanced metastatic breast cancer[216, 217]. Although no improvement to progression-free survival has been reached so far, I tent to repurpose it and evaluate the efficiency in targeting *ESR1* mutant breast cancer.

First, Marimastat pre-treated MCF7 cells all showed significantly increased adhesion on collagen, with more pronounced effects in *ESR1* mutant cells (Appendix B Fig. 106). This was in agreement with the *TIMP3* rescued-adhesion results (Fig. 37) and insinuated that the adhesive defect of *ESR1* mutant cells was driven by the altered *TIMP3*-MMP axis. I next assessed the effects on cell invasion using the dose range as previously described[218, 219]. Marimastat treatment substantially reduced the escalated invasion of *ESR1* mutant cells in a dose dependent manner (Fig. 39A-39D). Together, these data demonstrate that targeting MMP is an attractive way to block the gained invasive properties in *ESR1* mutant cells. The overall mechanism of the *TIMP3*-MMP axis-mediated alteration in adhesive and invasive properties of *ESR1* mutant cells are summarized in Figure 40.

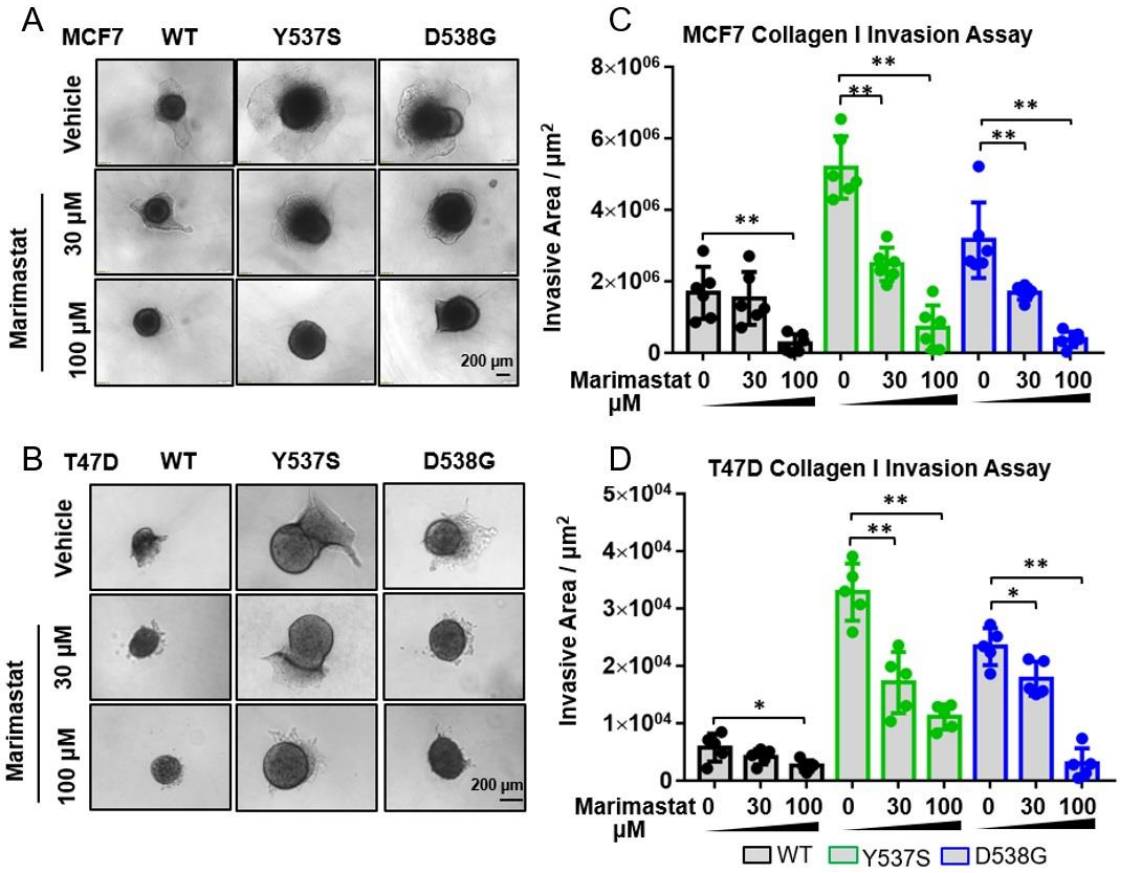



**Figure 39. Enhanced invasion of *ESR1* mutant cells are sensitive to MMP inhibition.**

A & B Representative images spheroid-based collagen invasion assay with different doses of Marimastat treatments in MCF7 (A) and T47D (B) cell models for 4 and 6 days respectively. C & D Quantification of corresponding invasive areas from d and f. Three independent experiments were performed with at least five biological replicates each time. Representative data from one experiment were shown. Student's t test was used to compare the effects of Marimastat treatments to vehicle controls. (* $p<0.05$, ** $p<0.01$)

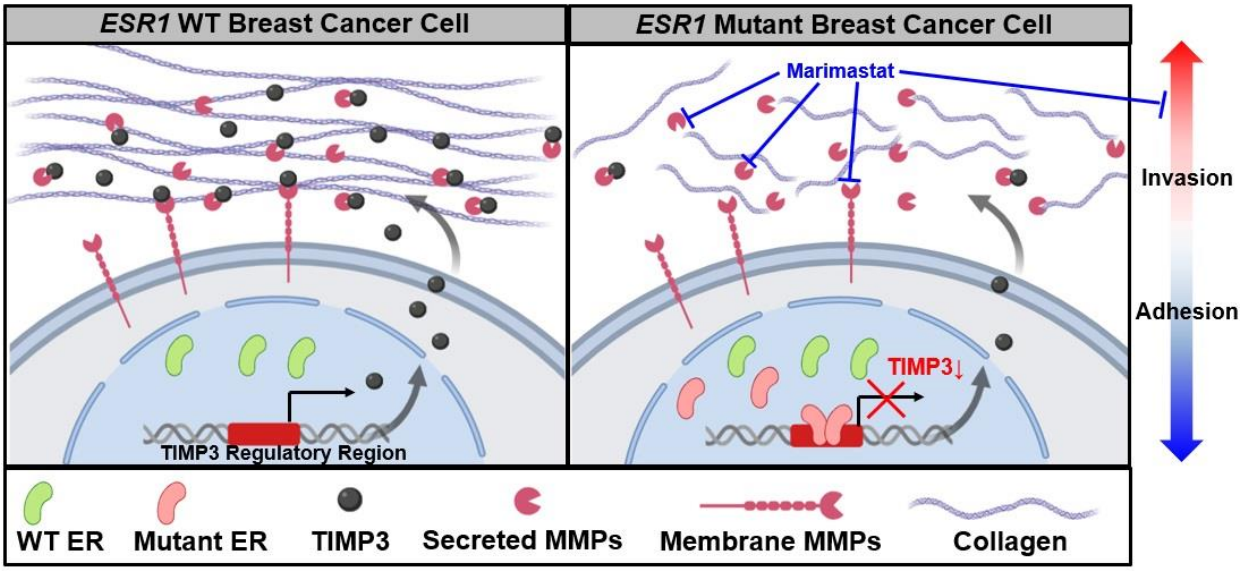



**Figure 40. Schematic model of altered adhesive and invasive properties in *ESR1* mutant breast cancer cells via *TIMP3*-MMP axis.**

Left panel: In breast cancer cells expressing WT ER, *TIMP3* expression is sustained at basal level which blocks the activities of both secreted and membrane-anchored MMPs. Those cells thus display limited capacities to digest collagens and invade. Right panel: Hotspot *ESR1* mutations suppresses *TIMP3* expression via ligand-independent bindings to *TIMP3* regulatory regions. Decreased *TIMP3* expression leads to 1) diminished adhesion to collagens and 2) enhanced MMP activities and invasive properties in collagen. Together, these alterations facilitate *ESR1* mutant cells to escape from the primary site and contribute to metastatic spread.

### 3.3.4 D538G *ESR1* Mutation Subtype Uniquely Enhances Contextual Cell Migration Via Canonical Wnt Pathway

#### 3.3.4.1 T47D-D538G Mutant Cells Uniquely Exhibit Enhanced Migration *In vitro*

Since pathway analysis of genes differentially expressed between *ESR1* WT and mutant tumors had also uncovered "Cell movement", we next asked whether mutant ER could also provide the cells with an increased migratory ability. Wound scratch assays identified significantly increased cell motility in T47D-D538G model (Fig. 41A and 41B), but not in T47D-Y537S (Fig. 41B) or MCF7 mutant cells (Appendix B Fig. 107). Of note, the migratory leading edge of T47D-D538G cells exhibited a much shaper morphology, implicating a more invasive property (Fig. 39C). This enhanced motility was shared between the three individual T47D-D538G clones excluding clonal artifacts (Appendix B Fig. 108). This striking effect was not a result of increased proliferation in hormone-depleted medium since motility assays were performed in the presence of mitomycin, which abolished the proliferation advantages of mutant cells (Appendix B Fig. 109). In addition, the D538G-specific migration was also discerned in an independent T47D overexpression cell model described in Chapter 1 (Appendix B Fig. 110).

To mimic the situation of collective migration from a cluster of cells, I utilized a spheroid-based collective migration assay on Type I collagen (Left panel, Fig. 41D). The distance to the leading edges of T47D-D538G (but not Y537S) mutant cells was significantly larger compared to WT spheroids (Right panel, Fig. 41D). Intriguingly, MCF7 *ESR1* mutant cells showed a slight decrease in motility in this assay (Appendix B Fig. 111). A co-culture assay mixing labelled WT and D538G mutant cells recapitulated enhanced migratory capacities of T47D-D538G over WT

cells (Appendix B Fig. 112). The enhanced migratory capacity in T47D-D538G was also reproducible in a transwell chemotaxis assay (Appendix B Fig. 113). In summary, these data demonstrated unique motility of T47D-D538G cells, re-confirming the allele and context-dependence of the metastatic properties gained through *ESR1* mutation.

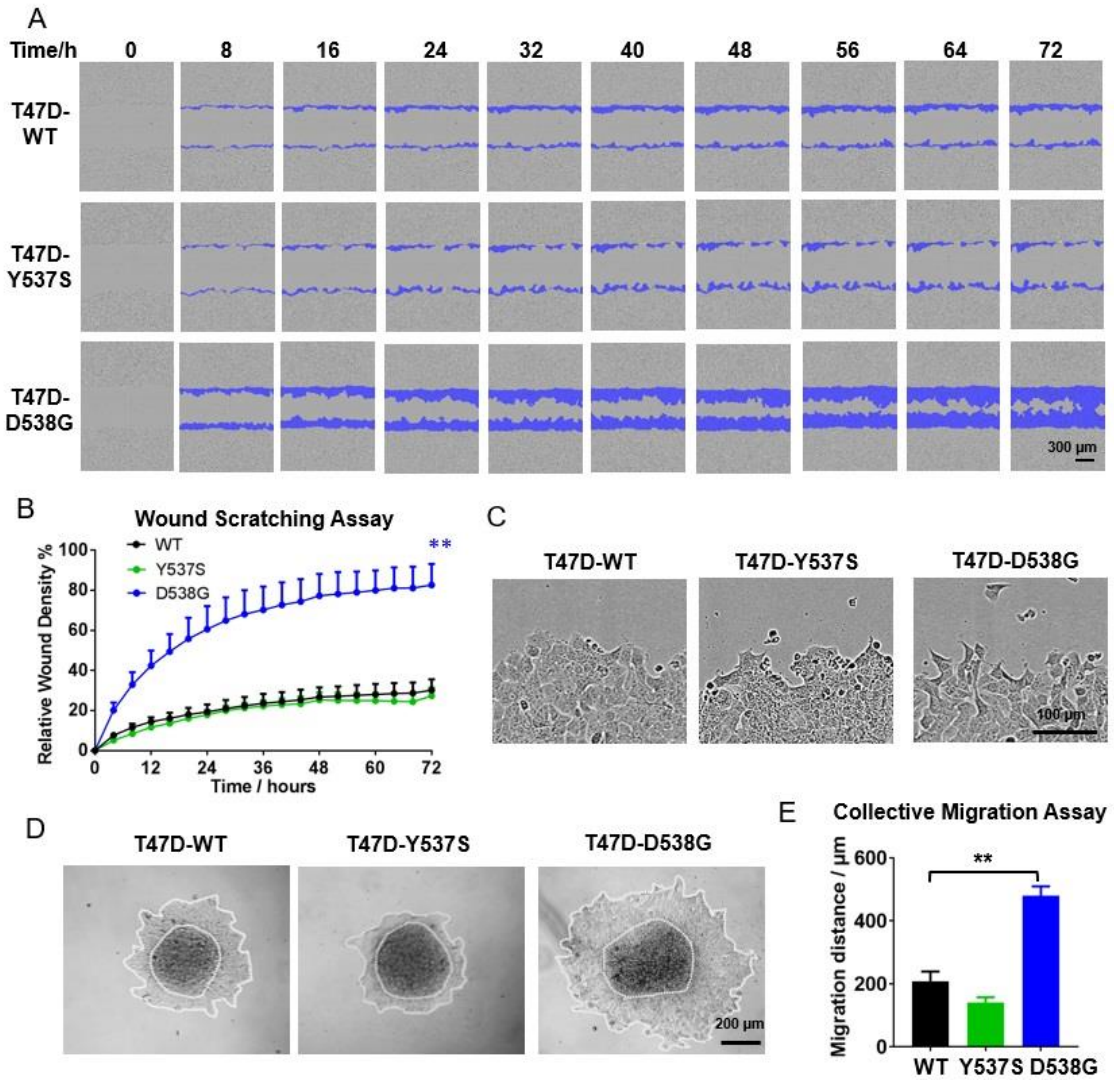



**Figure 41.T47D-D538G mutant cells uniquely exhibit enhanced migration *in vitro***

A-B) Representative images (A) and quantification (B) of wound scratch assay performed using IncuCyte living imaging system over 72 hours. The migratory edges were labelled in blue. Cell migration rates were quantified based on wound closure densities. Pairwise two-way ANOVA between WT and each mutant was performed. Scale bar: 300 μm. Data were from one representative experiment with eight biological replicates among nine independent repeats. D. Representative images and quantification of spheroid collective migration of T47D mutant cells. T47D cells were initially seeded into round bottom ULA plates to form spheroids, which were then transferred onto Collagen I coated plates. E. Collective migration was measured after 4 days. Migratory distances of four representative directions are indicated with white arrows. Migratory distance was calculated based on the mean radius of each spheroid normalized to the original areas. Three independent experiments were reproduced. Representative data from one experiment were shown. Dunnett's test was used. Scale bar: 200 μm. (** $P<0.01$)

### 3.3.4.2 Canonical Wnt Pathways Are Exclusively Enriched in D538G *ESR1* Mutant Cell and Tumor Transcriptome

In an effort to understand the mechanisms underlying the migratory phenotype of T47D-D538G cells, I next sought to identify pathways uniquely enriched in these cells. GSEA identified E2F targets, MYC targets and estrogen responses in both T47D mutants (Fig. 42 A), but angiogenesis and Wnt-β-catenin signaling as pathways uniquely enriched in T47D-D538G (Fig. 42A and Appendix D Table 18). Further enrichment analysis of the "Reactome Wnt Pathway"-a gene set depicting Wnt pathway components confirmed the higher enrichment in T47D-D538G cells and an increasing trend in tumors harboring D538G mutation from DFCI cohort (Fig. 42B).

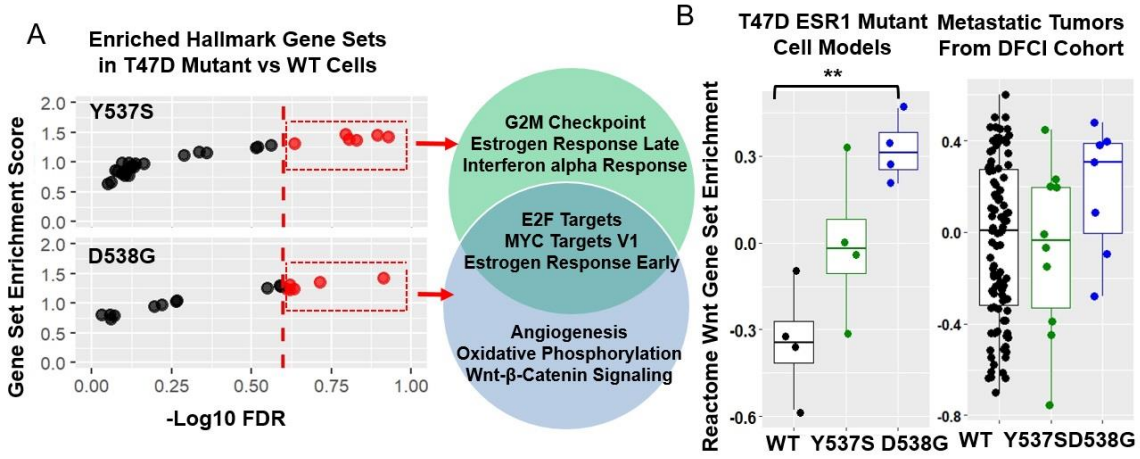



**Figure 42. Canonical Wnt pathways are exclusively enriched in D538G *ESR1* mutant cell and tumor transcriptome.**

A. Dot plots representing the enrichment distribution of the 50 MSigDB curated Hallmark gene sets in T47D-Y537S and T47D-D538G models respectively. Significant enriched gene sets (FDR<0.05) were highlighted in red, whose names were specified in the venn diagram in the right panel. Gene sets enriched in Y537S and D538G cell models were in green and blue circles respectively. B. Box plot showed the enrichment levels of the "Reactome Wnt Component Gene Sets" in T47D-*ESR1* mutant cells and in 130 DFCI ER+ metastatic tumor samples with Y537S and D538G mutant tumors separated. Dunnett's test was used. (** p<0.01)

### 3.3.4.3 Hyperactivation of Canonical Wnt Pathway Prompts D538G-driven Migration

Hyperactivation of the canonical Wnt-β-catenin pathway was further confirmed by increased phosphorylation of GSK3β at Ser9 and GSK3α at Ser21 and increased protein levels of β-catenin in T47D-D538G (Fig. 43A), as well as using a Top-Flash luciferase assay (Fig. 43B). Of note, the increased accumulation of β-catenin was also validated in a separate cell fractionation immunoblotting, which showed the higher β-catenin levels in cell nuclei portions but not membrane portions (Appendix B Fig. 114). Moreover, T47D-D538G *ESR1* mutant cells also displayed enhanced response towards canonical Wnt ligand Wnt3A, implicating the potential increased level of Wnt signaling component that potentiate the downstream response (Fig. 43C). To examine whether the hyperactivation of Wnt pathway is associated with the observed migratory phenotype, we next overexpressed a dominant negative TCF4 plasmid, which repressed canonical Wnt signaling transactivation, into WT and D538G cells, and were able to detect the complete abolishment of T47D-D538G migration (Fig. 43D and 43E), suggesting the Wnt hyperactivation is required for the migration phenotype of T47D-D538G.

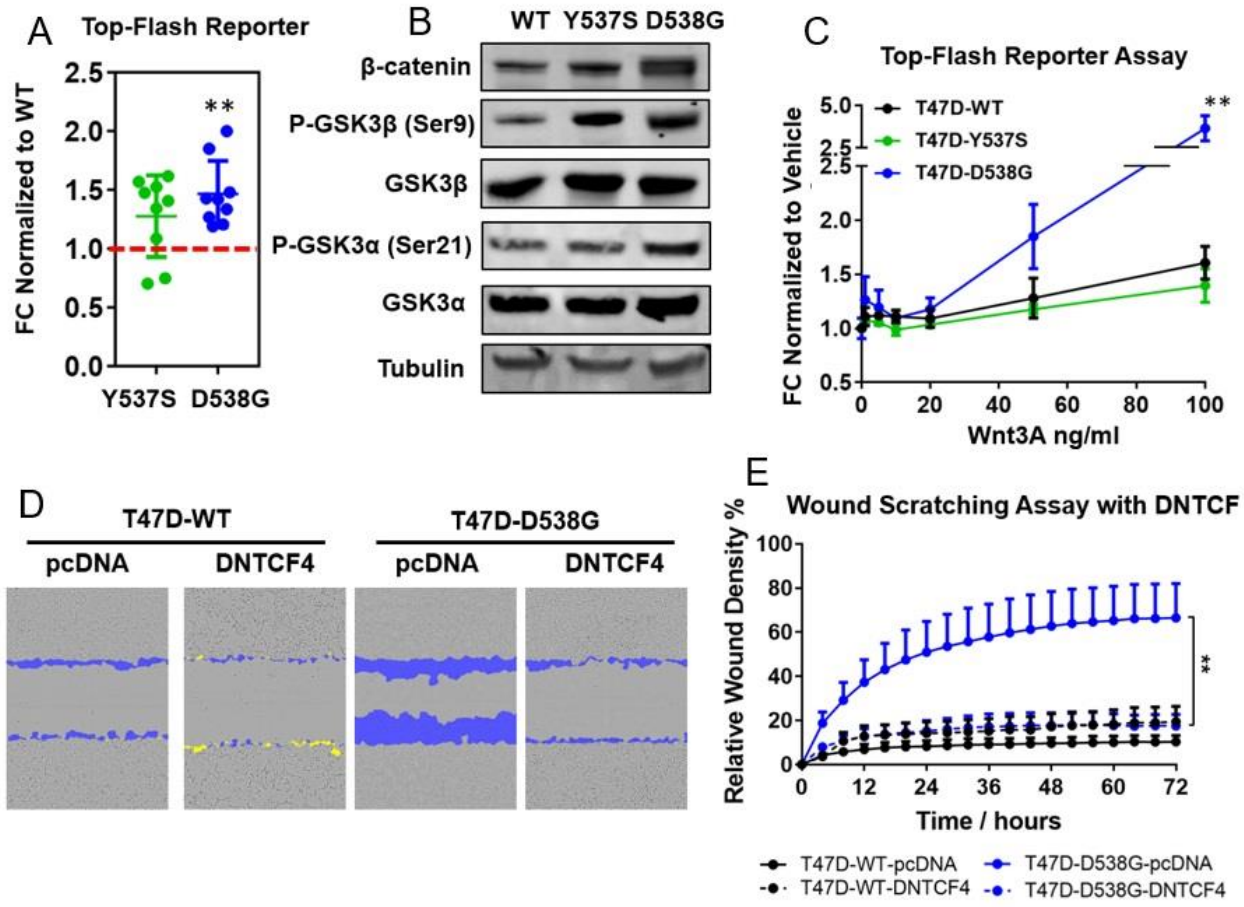



**Figure 43. Hyperactivation of canonical Wnt pathway prompts D538G-driven migration**

A. Top-flash luciferase assay on T47D-WT and mutant cells at basal levels. Luminescence readouts of top-flash were normalized to Renilla, fold changes were further calculated by normalizing to WT cells. The results were combined from nine independent experiments. B. Immunoblot detection of β-catenin, phospho-GSK3β (Ser9), phospho-GSK3α (Ser21) and total GSK3β, GSK3α levels in T47D WT and mutant cells after hormone deprivation. Tubulin was detected as a loading control. Data were from one representative experiment three independent repeats. C. Top-flash luciferase assay with dose gradient treatment of Wnt ligand Wnt3A. Luminescence readouts of top-flash were normalized to Renilla, fold changes were further calculated by normalizing to vehicle controls. Pair-wised Two-way ANOVA was utilized. This response curve was measure once, but the response at 100 ng/ml was reproduced for 3 times. D and E. Wound scratch assay in T47D-WT and D538G cells with or without transfection od dominant negative TCF4 plasmid for 72 hours. Representative images of 0 and 72 hours were shown in the left panel. Pairwise two-way ANOVA between vehicle and treatment conditions were performed. Data were from one representative experiment with six biological replicates of two independent experiments. (* $p<0.05$; ** $p<0.01$)

#### 3.3.4.4 Combination Treatment of Porcupine Inhibitor and Fulvestrant Exhibit Synergistic Inhibitory Effects towards T47D-D538G Migration

To address the potential clinical relevance of this finding, I next utilized the porcupine inhibitor LGK974, which prevents the secretion of Wnt ligands and is currently in a clinical trial for patients with advanced solid tumors including breast cancer (NCT01351103)[220, 221]. Treatment with LGK974 resulted in 20% and 40% inhibition of migration of T47D *ESR1* WT and D538G mutant cells, respectively (Fig. 44A and 44B) whereas no inhibitory effects were observed in cell growth with the identical condition (Appendix B Fig. 115), suggesting that Wnt signaling partially contribute to the migratory phenotype but not endocrine resistance. I next studied the combination of LGK974 and the selective ER degrader (SERD) Fulvestrant in migration assays and detected significant synergy (Fig. 44C and 44D), suggesting that combination therapy co-targeting the Wnt pathway and ER signaling might overcome the metastatic phenotypes of *ESR1* mutant tumors with Wnt activation.

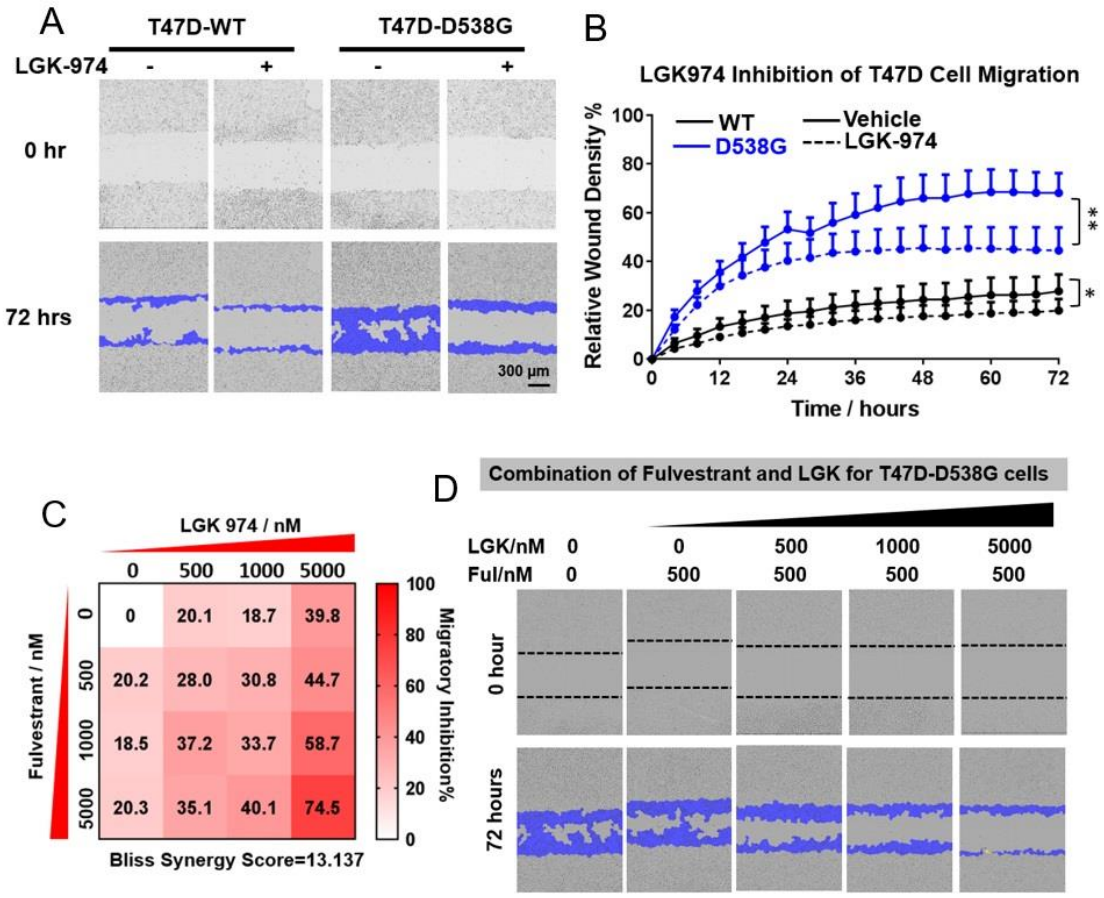



**Figure 44. Combination treatment of porcupine inhibitor and fulvestrant exhibit synergistic inhibitory effects towards T47D-D538G migration.**

A. Representing images of T47D-WT and D538G cells with or without 5µM LGK974 treatment at 0 or 72 hours. B. Wound scratch assay in T47D-WT and D538G cells with or without 5µM LGK974 treatment for 72 hours. Pairwise two-way ANOVA between vehicle and treatment conditions were performed. Data were from one representative experiment with six biological replicates of three independent experiments. C and D. IncuCyte migration assay with combination treatment of four doses of LGK974 and Fulvestrant in T47D-D538G cells was performed. Inhibition rates were calculated using the wound density at 48 hours normalized to vehicle control and labelled in the heatmap. SynergyFinder package was used to determine the synergy score. Positive scores are considered as synergistic combination. Data were from one representative experiment with six biological replicates of four independent experiments. (* $p<0.05$; ** $p<0.01$)

### 3.3.4.5 Multiple Wnt Component Genes Are Upregulated in T47D-D538G Cells

To further elucidate the specific altered Wnt component genes that potentially lead to the gain of migratory phenotypes, I further compared the expression levels of Wnt pathway component genes under five categories. While the majority of upregulated Wnt component genes were not consistent between both mutants and are not E2-inducible, I found five genes that are uniquely increased in D538G (*TCF4*, *LRP5*, *TCF3*, *CTNNB1*, *FZD4*) and two other genes (*FZD6*, *WNT6A*) that are increased in both mutants with a larger magnitude in D538G. These genes covered a diverse range of Wnt signaling pathway, including receptors (*LRP5*, *FZD4*, *FZD6*), ligands (*WNT6A*) and nuclear factors (*CTNNB1*, *TCF3*, *TCF4*). Although I have identified a few Wnt component genes that potentially contribute to the Wnt hyperactivation, further functional validations of these target genes are still required to complete this mechanistic investigation.

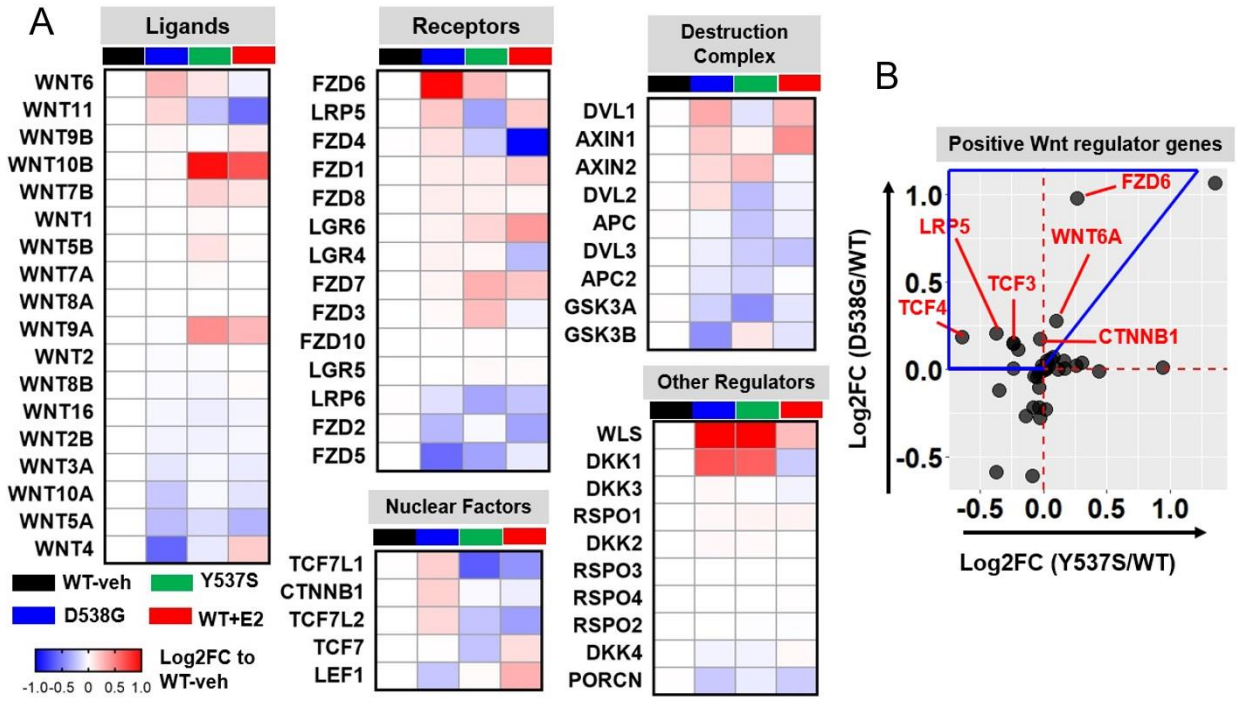



**Figure 45. Multiple Wnt component genes are upregulated in T47D-D538G cells.**

A. Heatmaps showing the gene expression patterns of all Wnt signaling component genes in T47D *ESR1* WT and mutant cell models. All data were extracted from RNA-seq with four biological replicates in Log2(TPM+1) values. Log2 fold changes were calculated by normalizing to WT-veh groups. B. Dot plot representing the fold changes of all Wnt signaling component genes in both T47D *ESR1* mutant cell models normalized to WT ones. The blue frame highlights the unique D538G increased genes and genes that are increased in both mutants with a larger magnitude in D538G.

### 3.3.5 FOXA1 Reprogramming Contributes to T47D-D538G Specific Migration

#### 3.3.5.1 D538G ER Tends to Bind to Non-canonical ER Binding Sites

Given the extensive upregulation of these Wnt component genes in T47D-D538G cell model, I next sought to investigate how D538G mutant ER induces these candidate genes. I first hypothesized that D538G ER induced these genes via the canonical ligand-independent manner. However, mining the isogenic T47D ER ChIP-seq data revealed no gained ER binding sites around these gene loci in T47D-D538G cell types, indicating potential ER indirect regulation. To decipher these indirect regulatory mechanisms, we therefore further analyzed the binding sites of WT and mutant ER in T47D cells.

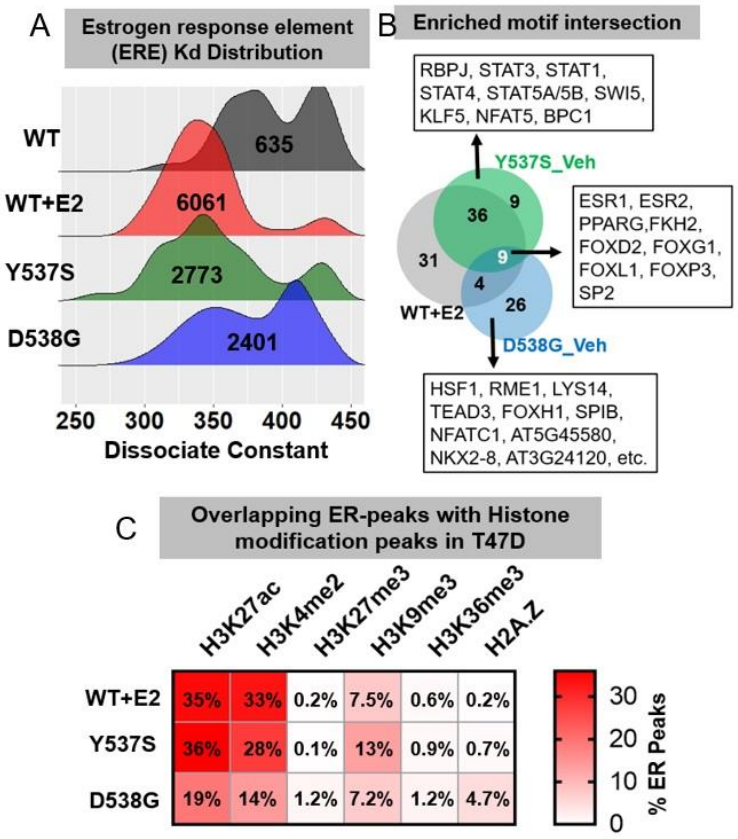



**Figure 46. D538G ER tends to bind to non-canonical ER binding sites at more repressive chromatin.**

A. Distribution of the dissociation constants of all putative estrogen response element on ER peaks in T47D *ESR1* mutant cell models. ERE numbers of each group were labelled in the plot. B. Venn diagram showing the intersection of significantly enriched motifs ($E<10^{-5}$) in T47D-WT+E2/Y537S/D538G ER cistromes. Motif names were pointed out in the frames besides each portion. (Only top10 enriched motifs were specified in T47D-D538G). C. Heatmap showing the percentage of ER intersecting with various of histone modification markers.

We first scanned the entire ER binding site sequences for estrogen response element with the algorithm EREFinder [190], and were able to identify 6,061, 2,773 and 2,401 EREs from WT+E2, Y537S and D538G groups respectively (Fig. 46A), suggesting that the ER binding sites in both mutants are less likely to be canonical ER binding sites with robust ERE motifs (Fig. 46A). In

addition, we found the EREs derived from D538G ER binding sites showed larger dissociation constant, indicating the D538G ER bindings are less stable than WT+E2 and Y537S groups (Fig. 46A). Motif enrichment analysis further conformed that D538G ER binding sites derives extensively more unique noncanonical motifs than Y537S, whereas the canonical *ESR1* motifs still remain significant enriched in all three groups (Fig. 46B). Furthermore, D538G ER showed lower possibility to bind to active open chromatin regions that harbor H3K4 di-methylation or H3K27 acetylation, suggesting that D538G ER binding are partially exclusive from some active histone modification markers (Fig. 46C).

#### 3.3.5.2 T47D Y537S and D538G Mutations Reshape Accessible Genomic Landscapes

Given the tremendous ER redistribution in *ESR1* mutant cells that are distinct from WT+E2 groups, we then sought to examine whether genomic remodeling shapes the accessible genome in *ESR1* mutant cells and leads to the unique mutant ER cistromes. We therefore further performed an ATAC-seq on our T47D *ESR1* mutant cell models to elucidate the accessible genomic landscapes.

ATAC-seq peak calling showed that E2 conferred very limited, if any, effects to the open chromatin sites, whereas all three clones of D538G cells showed fewer ATAC-peaks, indicating that T47D-D538G harbors overall more closed chromatin landscape (Fig. 47A and Appendix D Table 19). PCA plot of ATAC peaks illustrated that both Y537S and D538G cells exhibit unique pattern of accessible genomic sites that segregated from WT cells, whereas WT-E2 groups still clustered with vehicle groups (Fig. 47B). We further performed a differentially accessible site analysis against WT-vehicle groups after merging all the peaks together and counts normalization.

Consistent to PCA plot, E2 only induced less than 20 differential accessible sites whereas *ESR1* mutant cells caused 5,981/2,285 increased and 8,478/3,647 decreased accessible sites in Y537S and D538G cells respectively (Fig. 47C). Interestingly, the closed sites are more prevalent than open sites in *ESR1* mutant cells. Upon intersecting ER binding peaks with ATAC sites, we found approximately half of ER binding sites were located in open chromatin in WT and Y537S cells, whereas only 25% D538G ER binding peaks were intersected with accessible sites (Fig. 47D). These findings are consistent with the fewer overlapped ER peaks with active histone modification markers in D538G mutant cells described above (Fig. 47C).

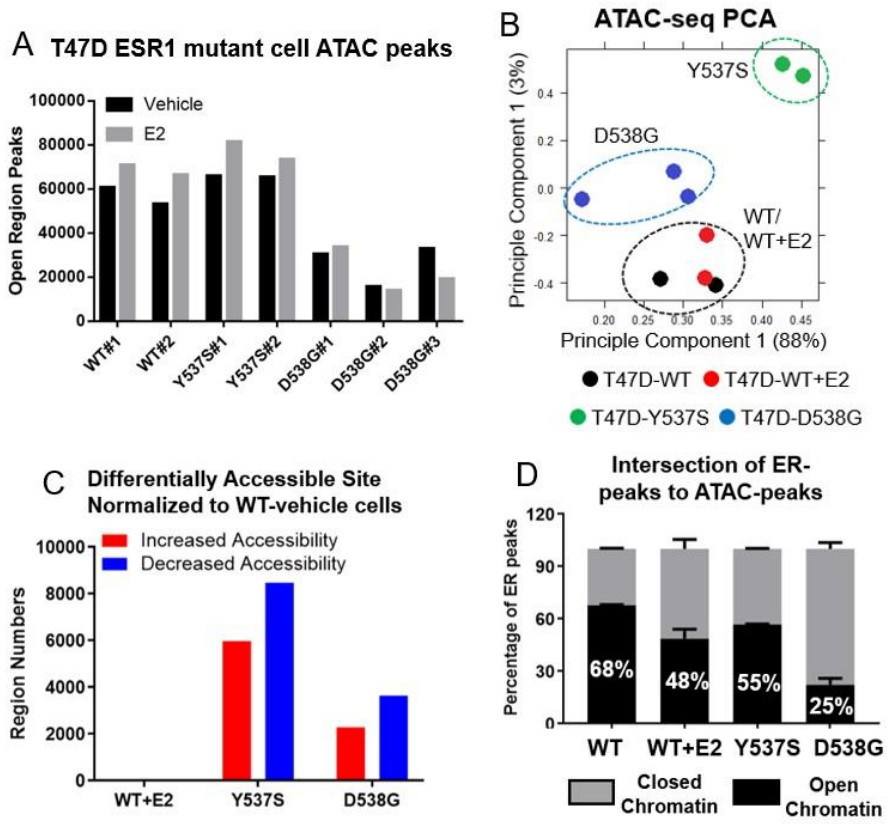



**Figure 47. T47D Y537S and D538G mutant cells reshape accessible genomic landscape.**

A. Bar graph showing ATAC peak numbers called in T47D *ESR1* mutant individual clones in the presence or absence of 10 nM E2 for 1 hour. Peaks were called with MACS2 using $p<10^{-5}$ as the cutoff. B. PCA plot showing the ATAC peak distribution of T47D-*ESR1* mutant cells. C. Bar graph showing the differentially accessible sites in T47D *ESR1* mutant cells, all significant sites were called versus WT-veh groups and separated into increased or decreased sites. D. Bar chart showing the percentage of ER locates in open (ATAC peaks overlapped) or closed (non-ATAC peaks overlapped) chromatin.

*ATAC-seq data was generated and processed by Spencer Arnesen from Dr. Jason Gertz's group.

### 3.3.5.3 D538G ER Binding Sites Are Less Dependent on FOXA1

Integrating ER ChIP-seq and ATAC-seq data sets strongly suggest that T47D-D538G ER displays a distinct binding pattern enriched in closed chromatin. Arnesen et al. recently reported the enrichment of FOXA1 motifs in T47D-D538G mutant specific gained open chromatin regions

[222]. In order to understand this non-canonical ER binding mechanism, we further hypothesized that the well-characterized ER upstream pioneer factor, FOXA1, shows a reprogrammed distribution, which minimizes its role in guiding canonical ER binding and causes many noncanonical ER distribution, especially in the T47D-D538G model.

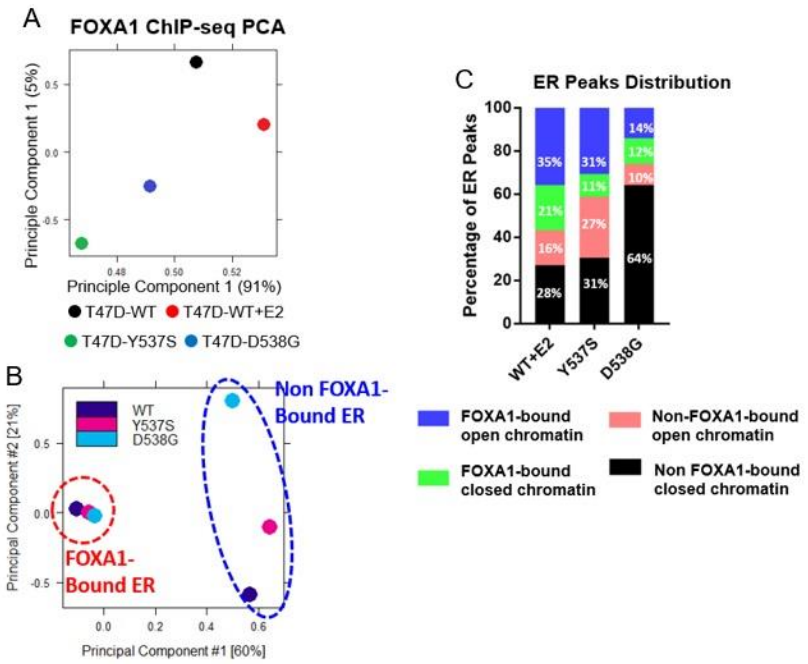



**Figure 48. D538G ER binding sites are less dependent on FOXA1in T47D cell line.**

A. PCA plot showing the FOXA1 peak distribution of T47D WT, WT+E2, Y537S and D538G cells. B. PCA plot showing the distribution of FOXA1-bound and non-FOXA1-bound ER peaks in in WT+E2, Y537S and D538G groups. C. Stacked plot showing the distribution of ER in terms of intersection of ATAC and FOXA1 peaks in T47D WT+E2, Y537S and D538G groups.

To elucidate this, we conducted a FOXA1 ChIP-seq in T47D *ESR1* mutant cell models with WT, WT+E2, Y537S and D538G groups respectively. First, FOXA1 peaks in T47D-WT cells showed 44%-82% overlap with other public available FOXA1 binding sites in T47D cells under either FBS or CSS conditions (Appendix B Fig. 116) [68, 223-225]. Peak calling results suggest

both Y537S and D538G *ESR1* mutant cells gained more FOXA1 binding sites, whereas E2 treatment slightly decreased FOXA1 overall binding sites (Appendix D Table 20). FOXA1 binding pattern was segregated in all four groups (Fig. 48A), suggesting both E2 treatment and *ESR1* mutations drive FOXA1 redistribution at different levels. In addition, general characterization demonstrated no differences of genomic feature distribution were induced by either E2 treatment or *ESR1* mutations (Appendix B Fig. 117).

Given FOXA1 serves as an ER pioneer, we further investigated whether the pioneering effects of FOXA1 towards ER was altered. We separated ER peaks into FOXA1-bound and FOXA1-unbound ER and found that FOXA1-bound ER showed high similarities in WT+E2, Y537S and D538G groups, whereas non-FOXA1-bound ER was tremendously different (Fig. 48B). A comprehensive intersection ratio comparison demonstrated that D538G FOXA1-unbound but not bound ER tends to bind to intergenic regions with poor chromatin accessibility (Appendix B Fig. 118). A further triple integrating analysis with ATAC-seq data revealed only 14% T47D-D538G ER was putatively located in FOXA1-driven open chromatin, whereas WT+E2 and Y537S groups showed 35% and 31% ER peaks following such loop (Fig. 48C). Taken together, D538G ER binding in T47D cells is less dependent on FOXA1.

#### 3.3.5.4 FOXA1 Redistribution Is Associated with Novel Target Genes in *ESR1* Mutant Via Enhanced Chromatin Accessibility

Given both Y537S and D538G *ESR1* mutant cells harbor more FOXA1 peaks than WT cells (Appendix D Table 20) albeit they showed minimized role in guiding ER, particularly in T47D-D538G cells, we next questioned whether FOXA1 gains novel functional binding sites that

potentially drive novel accessible chromatin regions and contribute to *ESR1* mutant-specific novel target genes and metastatic phenotypes.

We first intersected the entire FOXA1 binding sites to ATAC sites and found that the open chromatin regions in T47D-D538G cells are more dependent on its FOXA1 binding compared to WT cells, whereas no differences were observed in Y537S mutants (Fig. 49A). In line with this, binding intensities of FOXA1 on the open chromatin were stronger in T47D-D538G clones (Fig. 49B), suggesting that the altered FOXA1 distribution in *ESR1* mutant cells is highly associated with open chromatin. To further investigate the potential novel target genes that induced by this FOXA1-driven mechanism, we conducted integrating analysis of the three gene collections of each mutants: 1) novel target genes derived from RNA-seq; 2) mutant-specific gained ATAC peaks annotated genes; 3) mutant-specific gained FOXA1 peaks annotated genes ((Fig. 49C). Intersection analysis revealed 61 and 26 novel target genes that were potentially mediated by this mechanism in Y537S and D538G respectively (Appendix D Table 21), with only two genes overlapped (*CLSTN2, FEM1C*). This novel mechanism was exemplified by genes such as *GFRA*, whose expression showed no E2 regulation but specifically increased in T47D-D538G cells, with gained FOXA1 binding and open chromatin regions at its proximate loci (Fig. 49D). Of note, I have identified an altered Wnt signaling component gene, TCF4, among the list of D538G FOXA1-driven novel genes. Potential connection of this epigenetic mechanism to Wnt hyperactivation is currently under examination.

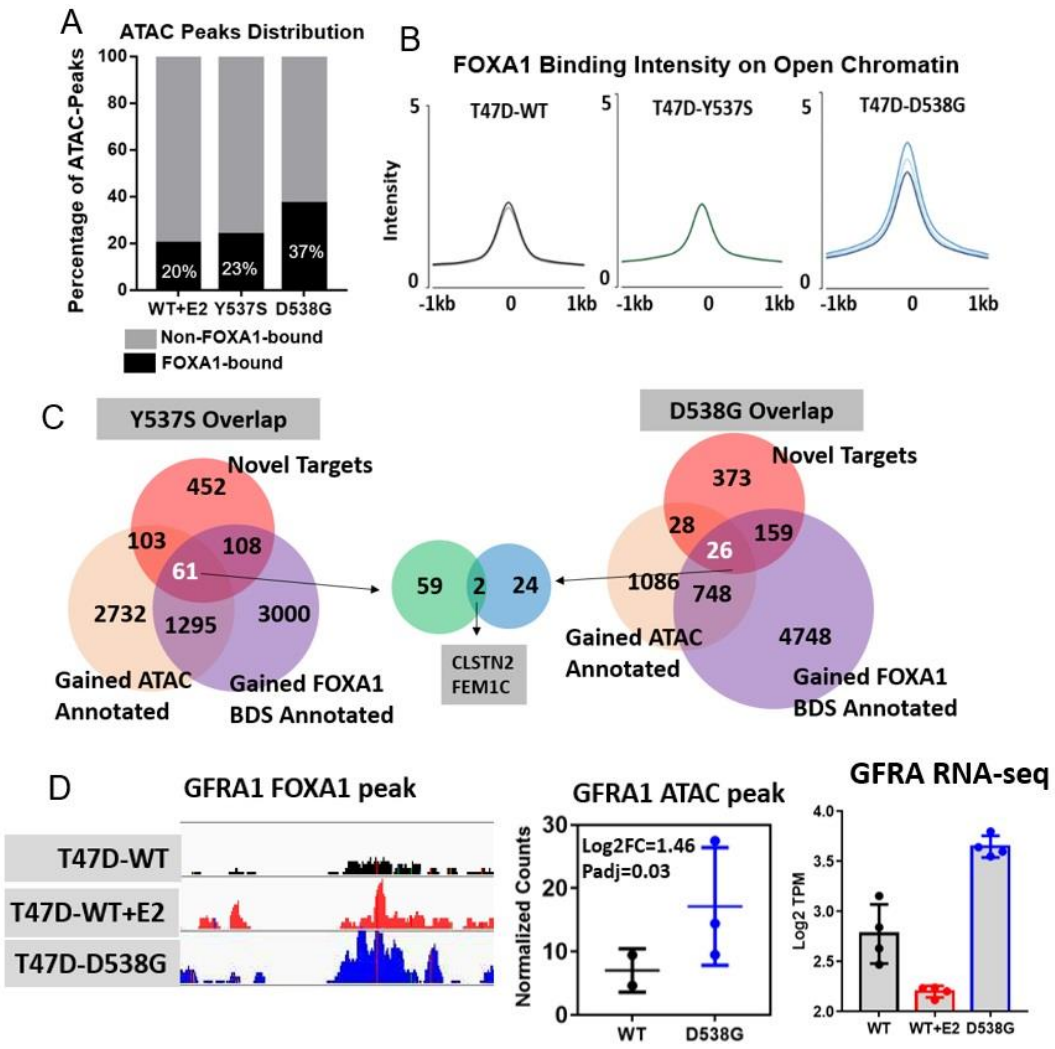



**Figure 49. FOXA1 redistribution is associated with novel target genes in *ESR1* mutant via enhanced chromatin accessibility.**

A. Bar charts showing the percentage of ATAC peaks bound or unbound to FOXA1. B. Intensity plot representing the average binding signals of FOXA1 peaks on open chromatin within a window of 2 kb. C. Venn diagram showing the intersection of gene annotated from either gained ATAC and FOXA1 peaks (+/- 3kb of TSS with 200 kb of the peak flank) and differentially expression novel target genes from RNA-seq. Intersected genes from each mutant cell were further overlapped. D. Left panel: FOXA1 binding signal in T47D-WT, W+E2 and D538G groups in GFRA1 gene loci. Middle panel: dot plot comparing the normalized ATAC-seq signals of one differential ATAC peak between WT and D538G cells. Right panel: Bar chart showing the expression of GFRA1 in T47D WT, WT+E2 and D538G cells extracting from RNA-seq data set with four biological replicates.

### 3.3.5.5 T47D-D538G Migration Depends on FOXA1 Rather Than ER

To examine whether FOXA1 contributes to the novel migratory phenotype in T47D-D538G and evaluate how the effective size compared to ER, we tested the migratory properties of T47D WT/D538G cells following either FOXA1 knockdown or Fulvestrant treatment (Fig. 50).FOXA1 knockdown fully rescued the enhanced migration in T47D-D538G cells (Fig. 50A and 50B), whereas Fulvestrant treatment only dampened 15% of this phenotype (Fig. 50C and 50D), indicating that the T47D-D538G migration depended more on FOXA1 rather than ER. Given I found the FOXA1 redistribution in T47D-D538G cells is associated with a few novel target genes via chromatin accessibility alterations, further functional investigations are necessary and critical to determine the specific gene (s) mediating this phenotype.

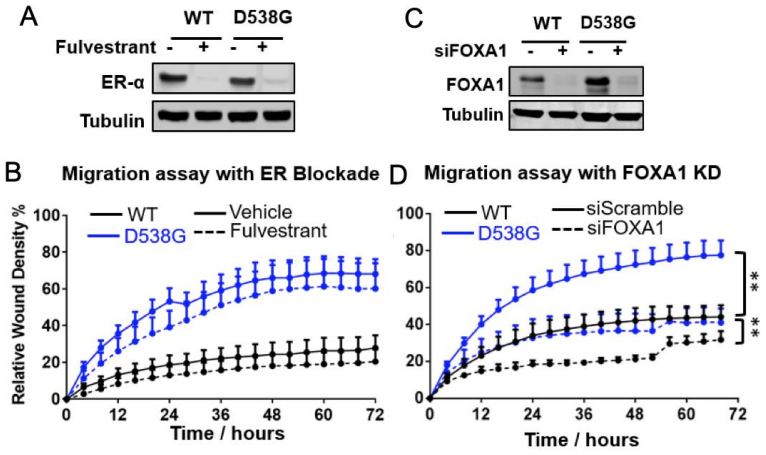



**Figure 50. T47D-D538G migration depends on FOXA1 rather than ER.**

A. Immunoblot validation of Fulvestrant-induced ER degradation and FOXA1 knockdown. Cell lysis were subjected to ER and FOXA1 detection. Tubulin was detected as a loading control. B. Wound scratch assay in T47D-D538G and WT cells with 1 μM of Fulvestrant treatment (Left panel) or knockdown of FOXA1 (Right panel) for 72 hours. Cell migration rates were quantified based on wound closure density. Pairwise two-way ANOVA between siScramble/siFOXA1 or vehicle/Fulvestrant conditions in each cell type was performed. Data were from one representative experiment with six biological replicates of three independent experiments. (* $p<0.05$; ** $p<0.01$)

Collectively, in this section, I identified the unique enhancement of cell migration in T47D-D538G cell models. Two potential mechanisms are further elucidated. Firstly, transcriptomic analysis revealed the hyperactivation of the canonical Wnt signaling with increase of various Wnt components in T47D-D538G cell, which is required to induce T47D-D538G specific migration. Secondly, ER binding pattern in T47D-D538G cells is distinct from WT+E2 groups, in terms of locating at distant intergenic regions with more repressive chromatin accessibility. This is likely due to the reprogramming of ER pioneer factor, FOXA1. FOXA1 reprogramming also exhibits novel binding sites that potentially open non-canonical chromatin regions and induce novel target genes that might mediate migratory phenotypes. Potential connection of these two mechanisms is currently under examination. The detailed mechanism of this process is summarized in Figure 51.

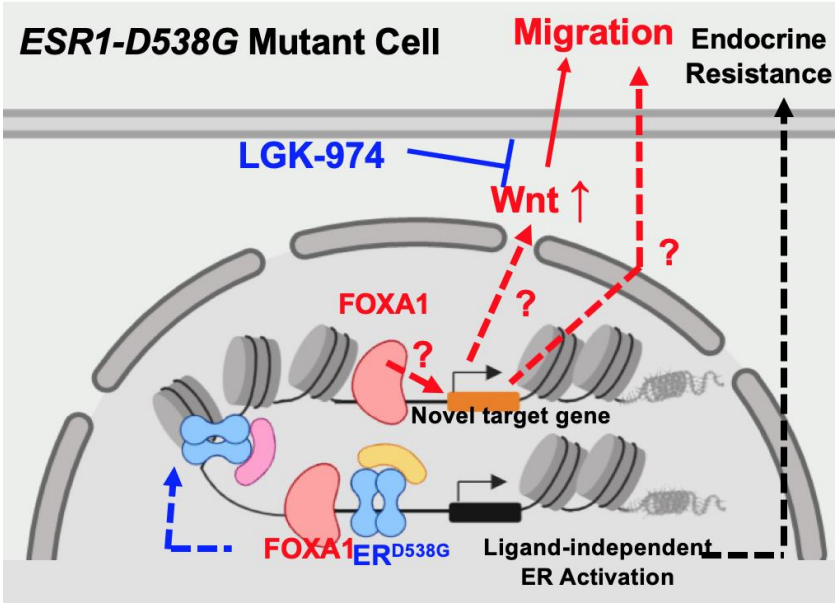



**Figure 51. Schematic view of mechanisms underlying the T47D-D538G specific migratory phenotype.**

In T47D-D538G cells, part of ER shows ligand-independent binding at FOXA1-pionnered locus and thus convert to endocrine resistance, whereas FOXA1 is also redistributed and open non-canonical chromatin regions that induce novel target genes which might confer the migratory pohenotype. On the other hand, canonical Wnt pathway is hyperactivated in T47D-D538G and this is also linked to the migration phenotype. However, a few points are still not well understoond, 1) how does FOXA1 contribute to the novel gene regulation; 2) whether the FOXA1-driven novel genes are related to the Wnt hyperactivation; 3) What other FOXA1-driven novel genes might involve in this phenotypic regulation.

## 3.4 Discussion

Hotspot somatic mutations clustering in the LBD of ER make up a prevalent molecular mechanism that drives antiestrogen resistance in approximately 30% of patients with advanced ER+ breast cancer. There is an urgent need for a deeper understanding of these resistance mechanisms in order to develop novel and personalized therapeutic approaches. Utilizing unique clinical samples, *in silico* analysis of large data sets, and robust studies using two genome-edited

cell line models with two distinct hotspot mutations, our study provides the first clinical evidence of *ESR1* mutations conferring gain-of-function metastatic properties and uncovers complex and context-dependent mechanisms. We identify *ESR1* mutations as a novel multimodal metastatic driver hijacking adhesive and migratory networks, likely affecting multiple steps during metastatic spread. Mechanistically, we uncovered novel ER-indirect regulation of expression of key metastatic candidate genes, distinct from previously described[109, 110, 226] canonical ligand-independent gene induction.

We discovered enhanced cell-cell adhesion via upregulated desmosome and gap junction networks in cell lines and clinical samples with *ESR1* mutations. We propose that this key alteration could support increased metastases in ER mutant tumors through facilitating the formation of homo- or heterotypic CTC clusters, thereby providing a favorable environment for CTC dissemination, as previously described [41]. Notably, an association between *ESR1* mutations and increased CTC cluster formation was recently observed in clinical samples[227]. This idea is further supported by previous data showing that upregulation of the desmosome gene plakoglobin (*JUP*) as part of a signature for CTC cluster formation[41]. We observed increased expression of plakophilin, desmocollin, and desmoglein in *ESR1* mutant cells, suggesting the importance of the broad desmosome network reprogramming in functional cell clustering. Moreover, gap junction are not commonly reported to mediate physical cell-cell adhesion, whereas our data suggested that gap junction blockade diminished that effects. It is possible that strong gap junction expression could additionally regulates cell-cell physical adhesion at some specific environments such as microfluidic conditions. Enhanced gap junction genes might also potentiate intercellular calcium or proton signaling, facilitating the elongated survival of various types of cells attached to *ESR1*

mutant cells in the metastasis[228]. Furthermore, our western blot result in Fig. 23B reveals an increased upper band for connexin 43 in *ESR1* mutant cells, implicating the potential enhanced Cx43 phosphorylation, which is an important modification to mediate Cx43 activity, assembly, and its life cycle on cell membrane [229-231], which is consistent with the increased gap junction function determined in Fig. 26. More experiments are warrant to decipher the mechanisms involved including the specific phosphorylation sites and the their impacts on gap junction kinetics. Previous studies have validated the anti-tumor effects of FDA-approved gap junction blockers carbenoxolone and mefloquine *in vivo*[232, 233]. Our results call for additional preclinical studies using drugs which target cell-cell interactions, with the ultimate goal to test them in CTC-targeted clinical trials for patients with *ESR1* mutant advanced disease.

In addition to increased cell-cell adhesion, we also identified decreased cell-ECM adhesion and enhanced invasion driven by an altered *TIMP3*-MMP axis. These finding suggest that *ESR1* mutant cells might gain invasive potential even before the intravasation stage, and as such, this finding warrants further testing of MMP inhibitors such as Marimastat[216] for efficacy in *ESR1* mutant tumors. Previous studies using similar *ESR1* mutation cell models described enhanced migratory properties in *ESR1* mutant cells[126, 170], but no mechanisms were uncovered. In addition, mutation-subtype and cell line context-dependent effects are still pronounced, as previously described in our cell model transcriptomic analysis[110]. In our ECM adhesion array, we observed markedly more pronounced adhesive defects in D538G than Y537S mutant of T47D cell line (Fig. 35A), which is reflected by the more evident integrin downregulation in this line (Appendix B Fig. 101). Additionally, MCF7 mutant cells showed uniquely enhanced adhesion on fibronectin, which is likely due to the cell line contextual upregulation of relevant adhesion

molecules such as ITGAV[234] and CD44[235]. According to our qPCR array, multiple other context-dependent molecular mechanisms might involve in the decreased collagen adhesion (Fig. 36A). For instance, the unique gain of MMP1 expression in both MCF7 *ESR1* mutant cells and the decreased ITGA1 expression in T47D-Y537S line might directly contribute to the cell line specific adhesive and invasive phenotypes. Taken together, these additional contextual effects are setting the stage for a more personalized therapeutic targeting strategy for *ESR1* mutant breast cancer.

Here we also identify a critical role for Wnt-β-catenin signaling in migratory properties and show that co-targeting of Wnt and ER resulted in synergistic inhibition of cell migration. Intriguingly, the strong effect we observed on migration was unique to T47D-D538G, a discovery that was made possible through our use of multiple genome edited mutation models. This finding might help to explain the high frequency of D538G mutations in metastatic samples, while endocrine resistance are stronger in Y537S mutant cells[108, 110, 117, 127, 236]. Since higher Wnt activity and β-catenin accumulation were also observed in T47D-Y537S cells (Fig. 43A and 43B), we conclude that Wnt activation plays a necessary but insufficient role in inducing migratory properties in T47D-D538G cells. It is also possible that some genes uniquely regulated by Y537S in T47D cells might inhibit the migration. For instance, the gap junction component, connexin 43 (*GJA1*), which is exclusively upregulated in T47D-Y537S cells, has been reported to play an inhibitory role in epithelial cell migration[237]. These data support strong allele and context dependent effects of the *ESR1* mutation on metastatic phenotypes. Such context-dependent alterations have been shown in previous work from us and others in the transcriptome and cistromes of *ESR1* mutant cells[109, 110, 117]. Similarly, a recent study has shown that MCF7-

Y537S mutant cells have a shorter latency compared to D538G mutant cells to develop distant metastasis[117].

Our global cistrome analysis points towards as unique ER binding site redistribution in each mutant cell types, with T47D-D538G showing increased distal intergenic ER binding events. In addition, different mutation subtypes exhibit distinct clinical features. Recent BOLERO2 trial showed significant differences in overall survival and everolimus response between Y537S and D538G mutations[150]. A more recent PALOMA3 trial suggested a potential palbociclib resistance uniquely gained in patients bearing the Y537S mutation[151]. Taken together, these proof-of-concept studies are setting the stage for a more contextual and personalized therapeutic targeting strategy for *ESR1* mutant breast cancer.

In the present study, we sought to address the mechanisms of induction of the candidate metastatic driver genes through ER ChIP-seq. Interestingly, however, none of the metastatic candidate genes, in mutant cells gained proximal ER binding site. On one hand, this may have been due to a technical reason, given the possibility that our strict hormone deprivation protocol and selection of CSS possibly may have depleted those binding signals with intermediate intensities, thus, resulting in less sensitive readouts[139]. Notably, the ChIP-seq data set from Harrod et al. showed additional gained ER binding intensities around *DSC2*, *DSG2* and *TIMP3* gene loci in MCF7-Y537S cells alongside the consistent altered mRNA levels of these three genes compared to WT-vehicle sets, indicating the potential ligand-independent regulation on these genes. [128]. On the other hand, as we investigated in this study, there may be novel ER-indirect regulation is involved in the gene regulation network. The observed “lack of ER dependence” on the gene induction and metastatic phenotypes indicates a long-term non-canonical ER effects on

chromatin structure remodeling and novel transcriptional regulator recruitment. For instance, D538G mutations in the T47D background might induce some unique FOXA1 redistribution events to specific enhancers controlling the key migration driver gene(s). This is supported by a recent study which uncovered that FOXA1 occupancy depends on steroid hormone receptor actions and mutual regulation [223]. In addition, we also found increasing H3K4me1 and H3K27ac modifications at the desmosome gene loci in Fig. 33. This could be induced by enhanced enzymatic activities (MLL3/4, UTX, TRX and CBP) [238, 239] or enhanced chromatin accessibility in *ESR1* mutant cells. Thus, delineating the epigenomic landscapes of *ESR1* mutant cells remains a major area of ongoing research.

Our study partially addresses the metastatic phenotypes and mechanisms associated with hotspot *ESR1* mutations, albeit with some remaining limitations, such as the lack of *in vivo* validation with specific drug treatment of our robust *in vitro* findings for a full preclinical study. Furthermore, no detailed genetic or epigenetic mechanisms were investigated in our study to explain how mutant ER indirectly triggers the induction of metastatic candidate genes. Finally, a relative small-scale patient cohort, especially for those paired primary-metastatic *ESR1* samples, was used in our study, which warrants validation in future studies with larger clinical cohorts. Nevertheless, our study serves as a timely and important pre-clinical report uncovering mechanistic insights into *ESR1* mutations that pave the way towards personalized treatment of patients with advanced metastatic breast cancer.

# 4.0 Upregulated Basal Cytokeratins in *ESR1* Mutant Breast Cancer Are Associated with Enhanced Immune Response

## 4.1 Introduction

Breast cancer is a highly heterogeneous disease that cause a tremendous barrier for the development of treatments in the clinic[9, 10]. The identification of different biological subtypes of this disease occurs primarily through the use of immunohistochemistry and gene expression profiling[9, 10]. Broadly, these intrinsic subtypes include luminal ER positive (luminal A and luminal B), HER2 enriched and basal-like based on the status of essential hormone receptors (ER, PR and HER2) and Ki67 staining[12]. Tumors with different molecular subtypes showed distinct gene expressional pattern[240, 241], and this guides the therapeutic decision-making for individual patients in the modern age of precision medicine. On the basis of this, the advent of novel technologies to aid in the identification of new markers will also be critical.

Breast cancer intrinsic subtypes are associated with distinct patterns of metastatic spread and immune profiles[242, 243]. The basal subtype, which represents 15-25% of cases, is characterized by an expression profile similar to that of myoepithelial normal mammary cells, and highly overlaps with triple negative breast cancers[17]. Basal-like breast cancers have been reported to be featured by higher aggressiveness than luminal subtypes and have poor metastatic-free survival [17, 244]. Numerous studies have uncovered some essential molecular mechanisms for the high invasiveness, including but not limited to the alterations of CCL5/CCR5 axis[245],

amplified EGFR[246] kinase signaling and activation TGF-β signaling[247]. Recently, various groups have provided the evidence supporting the application of immune therapy in treating patients with basal breast cancer. Soliman et al. have showed the greater levels of PD-L1 constitutively and with IFN γ signaling in basal breast cancer, suggesting the therapeutic strategies with anti-PD-L1 immune check point inhibitor[248]. Molecular signature studies have uncovered the higher immune- infiltrations in basal-like breast cancer[242]. Moreover, based on the recent IMpassion130 clinical trial (NCT02425891), the FDA has granted an accelerated approval for atezolizumab, a monoclonal antibody drug targeting PD-L1, plus chemotherapy for the treatment of patients with PD-L1-positive, unresectable, locally advanced or metastatic triple negative breast cancer [249]. However, the potential application of immune therapy on luminal subtype breast cancer still largely remains unknown and additional research is warranted in the future.

Among the four intrinsic subtypes, basal and luminal are the two molecular subtypes with opposite histochemical features and notable differences in prognosis[250, 251]. However, the dynamic and continuous interval between cancer "luminal-ness" and "basal-ness" allows inter-subtype shift, which confers novel molecular features and delivers insights into precision therapeutic designs[201, 252]. Estrogen receptor has been well characterized as a luminal lineage marker[253]. Hotspot mutations at its ligand-binding domain were recently documented in 20%-40% endocrine resistance breast cancer patients, which promotes ligand-independent ER activation and metastasis[109, 110, 117]. In the face of various recently discovered targets and pathways, the nature of their intrinsic subtype balance is less well understood. In this study, I aim to examine the potential alterations of basal marker enrichment in *ESR1* mutant breast cancers and investigate the mechanisms of induction of the leading genes-basal cytokeratins by ESR1

mutations. In addition, using clinical sample transcriptomic data sets, I also tend to identify the clinical impact by the gain of basal cytokeratins, with a focus on the immune pathway alterations.

## 4.2 Materials and Methods

### 4.2.1 Human Tissue and Blood Studies*

Access of RNA-seq data from paired metastatic ER+ breast cancers merged from WCRC and DFCI cohort was described in Chapter 3 section 3.2.28.

For patient blood study, all patients provided written informed consents for research usage and all procedures were under the University of Pittsburgh Institutional Review Broad regulation (PRO15050502). 18 patients diagnosed with late stage metastatic ER+ breast cancer were recruited. Hotspot *ESR1* mutation identification procedure was described in our previous study[135]. In brief, blood samples were harvested in EDTA tubes (BD, #367856) and cfDNA was isolated from plasma samples using Qiagen circulating nucleic acid kit (#55114). *ESR1* ligand binding domain was pre-amplified in cfDNA and the products were subjected to droplet digital PCR detection with Y537S/C/N and D538G probes.

*The blood screen was in collaboration with Rekha Gyanchadani, PhD. Beth Knapick and Jian Chen provided significant technical assist in the blood processing and ddPCR operation.

### 4.2.2 Cell Culture

Maintenance of genome-edited MCF7 and T47D *ESR1* mutant cell models were described in the material and methods section in Chapter 2 section 2.2.1. For all the experiments, hormone deprivation was performed for all the experiments unless stated otherwise, cells were maintained in phenol-red-free IMEM (Gibco, A10488) with 10% and 5% charcoal-stripped serum (CSS, Gemini, #100-119) for T47D and MCF7 respectively, twice a day for three days. Clones with the same genotypes were equally pooled for downstream analysis. For other ER+ cell lines, ZR75-1 (CRL-1500), MDA-MB-134-VI (HTB-23) and MDA-MB-330 (HTB-127) were obtained from the ATCC. BCK4 cells were developed as reported previously[254]. Cell lines were maintained in the following media (Life Technologies) with 10% FBS: MDAMB-134 and MDA-MB-330 in 1:1 DMEM: L-15, ZR75-1 in RPMI, and BCK4 in MEM with nonessential amino acids (Thermo Fisher, #11140050) and insulin (Sigma-Aldrich, #91077C).

### 4.2.3 Compounds and Reagents

Progesterone (P4, P0130) and RU486 (m8046) were obtained from Sigma-Aldrich.

### 4.2.4 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

MCF7 were seeded into 6-well plate with 120, 000 cells per well respectively with biological triplicates. After desired treatments, RNAs were extracted, and cDNAs were synthesized using PrimeScript RT Master Mix (Takara Bio, #RR036). qRT-PCR reactions were

performed with SybrGreen Supermix (BioRad, #1726275), and the ΔΔCt method was used to analyze relative mRNA fold changes and RPLP0 levels were measured as the internal control. All primer sequences are shown in Appendix D Table 22. siRNA against *ESR1*(L-003401), *PGR* (L-003433) and non-targeting scrambled control (D-001820-01) were purchased from Dharmacon.

### 4.2.5 Immunoblotting

Detailed methodological description of immunoblotting was present in the materials and methods section of Chapter 2 section 2.2.6. Antibodies against ER (#8644), HA-tag (#3724) and PR (#3176) were purchased from Cell Signaling. Tubulin antibody was obtained from Sigma (T6557).

### 4.2.6 S100A8/S100A9 Heterodimer ELISA

Human S100A8/S100A9 heterodimer amounts in human plasma samples were quantified using S100A8/S100A9 heterodimer Quantikine ELISA kit (R&D System, DS8900) following the manufacture protocol. All plasma samples were first diluted in calibration buffer with 1:50 ratio and directly loaded into antibody-coated plate with three replicates.

### 4.2.7 Chromatin Immunoprecipitation (ChIP)

Detailed method of ChIP was described in Chapter 2, section 2.2.10. For CTCF ChIP, CTCF antibody was purchased from Millipore Sigma (#07-729). qPCR primers for the selected regions were shown in Appendix D Table 23.

### 4.2.8 Generation of Sorlie and TCGA Basal/Luminal Gene Sets

For Sorlie gene sets, original panel of intrinsic genes were downloaded from Stanford Genomics Breast Cancer Consortium (http://genome-www.stanford.edu/breast_cancer/). 453 genes were annotated from 553 probes. Expression of these 453 genes was matched to 33 luminal and 39 basal breast cancer cell lines. Significantly higher (FDR<0.01) intrinsic genes in basal or luminal cells were called as basal (n=75) or luminal (n=68) markers in Sorlie gene sets. For TCGA gene sets, differentially expressed genes were called between basal and luminal A or basal and luminal B ER+ tumors using raw counts. Top 200 increased genes of these two comparisons were further intersected. Overlapped DE genes in basal (n=164) and luminal (n=139) tumors were called as TCGA gene sets. Detailed method illustrations are shown in Fig. 52.

### 4.2.9 RNA Sequencing Analysis*

RNA sequencing data sets were analyzed using R version 3.6.1. Log2 (TPM+1) values were used for the RNA-seq of our *ESR1* mutant cell models and TMM normalized Log2(CPM+1)

values were used for other cell line RNA-seq data. Transcript per million (TPM) expression data from The Cancer Genome Atlas (TCGA) was downloaded from the Gene expression Omnibus database (GSE62944). Raw microarray data from Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) was downloaded from Synapse software platform (syn1688369; Sage Bionetworks, Seattle, WA, USA) respectively. TCGA reads were reprocessed using Salmon v0.14.1[174] and Log2 (TPM+1) values were used. For METABRIC data set, probes with the highest IQR were selected to represent genes. For pan-breast cancer cell line transcriptomic clustering, 97 breast cancer cell line RNA-seq data were reprocessed using Salmon and merged from three studies[255-257], batch effects were removed using "limma[258]" package. Gene set variation analysis were performed using "GSVA" package[195]. Survival comparisons were processed using "survival" and "survminer" packages[259] using Cox Proportional-Hazards model. Data visualizations were performed using "ggpubr[196]", "VennDiagram[197]" and "plot3D[260]".

*TCGA and METABRIC data set reprocessing were conducted by Kevin Levine, MD, PhD. 97 breast cancer cell line data set integration and processing were performed by Nolan Priedigkeit, MD, PhD.

### 4.2.10 Tumor Mutation Burden Analysis

Tumor mutation burden (TMB) calculation was performed as previous described[261]. Briefly, TCGA mutation annotation files from 982 patients were downloaded from FireBrowse (firebrowse.org/) data version 2016_01_28 and mutation subtypes were summarized using

"maftool" package[262]. Mutations subtypes were classified into truncated (nonsense, frame-shift deletion, frame-shift insertion, splice-site) and non-truncated mutations (missense, in-frame deletion, in-frame insertion, nonstop). TMB was calculated as 2X Truncating mutation numbers + non-truncating mutation numbers.

### 4.2.11 Data Availability

MSigDB curated gene sets were downloaded from GSEA website (www.gsea-msigdb.org/). TCGA biospecimen immune profile data were downloaded from Saltz et al[263]. TCGA mutation annotation format (MAF) files and methylation data were downloaded from FireBrowse website. CTCF and PR ChIP-seq data were obtained from GSE85106 and GSE40724 respectively and were visualized using Integrative Genomics Viewer (IGV). ChIA-PET data were downloaded from GSE72816 and were visualized using 3D Genome Browser. DFCI cohort RNA-sequencing data were originally published by Jeselsohn et al.[117], and further expanded by Nikhil Wagle, MD. RNA-Seq data from the WCRC cohorts are available at Lee-Oesterreich Lab Github repository. RNA-seq data from Jason Gertz cell model will be published separately. RNA-seq/microarray data for multiple ER+ breast cancer cell lines with or without E2 treatment were obtained from GSE89888 (MCF7_1, T47D_1), GSE51403 (MCF7_2), GSE108304 (T47D_2), GSE3834 (BT474), GSE61368 (ZR75-1), GSE50695 (MM134 and SUM44). All the raw data and scripts are available upon request from the author.

## 4.3 Results

### 4.3.1 Basal Subtype Marker Are Enriched in *ESR1* Mutant Cell Models and Clinical Samples

#### 4.3.1.1 Generation of Four Pairs of Luminal/Basal Gene Sets

To comprehensively test whether *ESR1* mutations confer "luminal-ness" and "basal-ness" imbalance, I adapted four pairs of luminal-basal gene sets (Fig. 52 and Fig. 53A). In brief, gene sets of Charafe-Jauffret et al.[264] and Huper et al.[265] were derived from microarray-based differentially expressed genes between luminal/basal breast cancer cell lines or luminal/basal cells isolated from normal mammary tissues respectively (Fig. 52), both of which were curated in MSigDB[266]. I also generated two other gene set pairs from i) differentially expressed (FDR<0.01) intrinsic subtype genes[267] between luminal (n=33) and basal (n=39) breast cancer cell lines based on public available RNA-seq data sets (Fig. 52)[255-257] and ii) differentially expressed genes between luminal and basal primary tumors from TCGA data set (Fig. 52).

Inter-gene-set overlaps were limited in both basal and luminal series despite the repeated presence of some classic basal (*KRT16*, *KRT6A*, etc.) and luminal (*ESR1*, *FOXA1*, etc.) marker genes (Fig. 53B), reflecting the differences in methodology and sources. As a quality control, all four basal gene sets showed significantly enrichment in basal versus luminal breast cancer cell lines and primary tumors from both TCGA and METABRIC (Appendix C Fig. 118A and 118B), and *vice versa* for luminal genes sets except for Huper luminal markers, possibly due to its origin of normal mammary tissue[265] (Appendix C Fig. 118C and 118D).

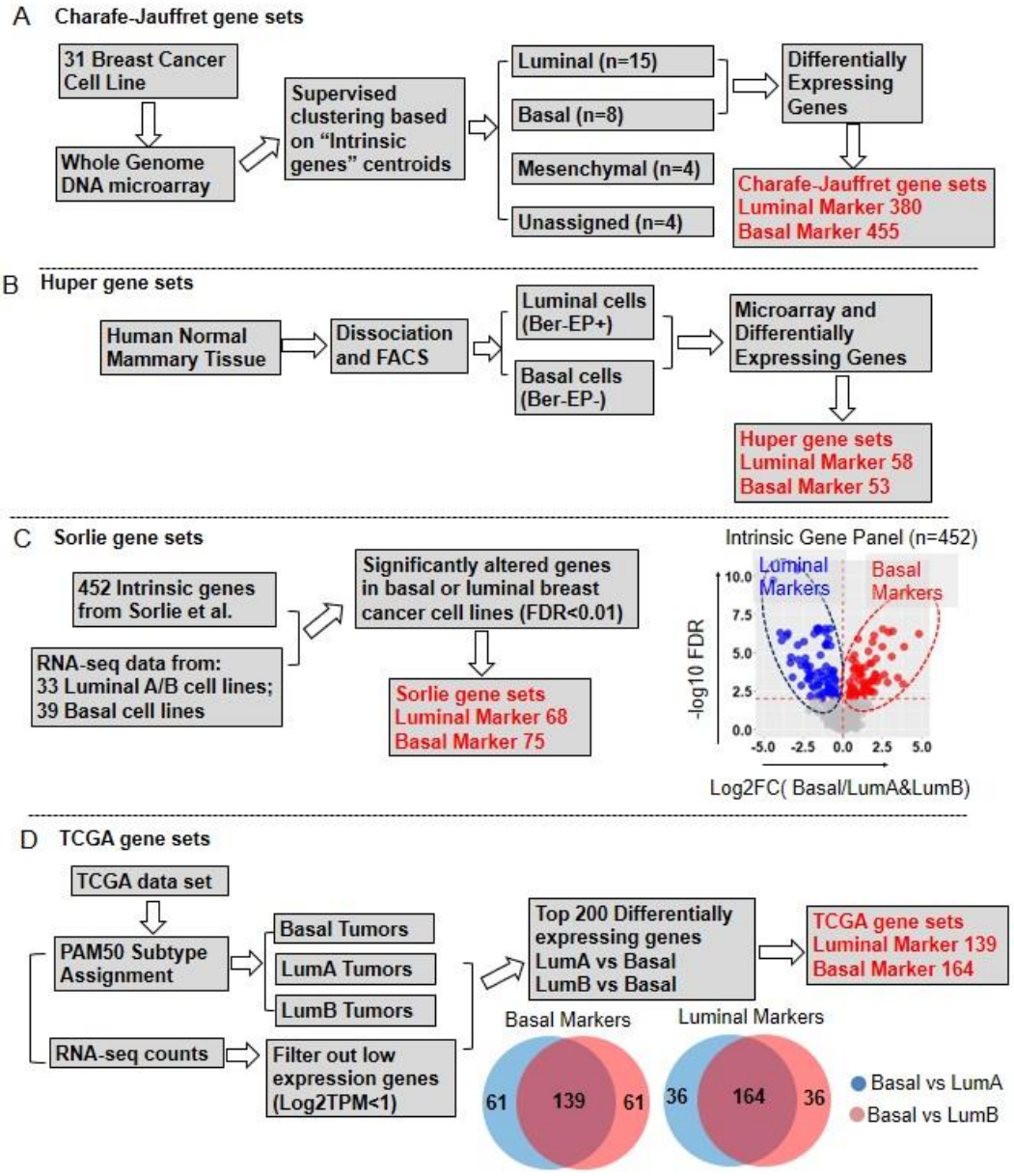



**Figure 52. Schematic flow charts of the generation process of the four pairs of luminal/basal gene sets.**

### 4.3.1.2 Basal Gene Sets Are Enriched in *ESR1* Mutant Cell Models and Metastatic Lesions

While applying them to *ESR1* mutant cell transcriptomes[110] from our previous study, I found consistently increased enrichment of basal markers in Y537S and D538G MCF7 *ESR1*

genome-edited mutant cells, whereas no differences were observed under estrogen treatment (Fig. 53C). Furthermore, no consistent alterations were observed in all four luminal marker enrichments (Appendix C Fig. 119A). The similar enrichment patterns were recapitulated in a secondary CRISPR-engineered MCF7 *ESR1* mutant cells from Dr. Jason Gertz, PhD. (Appendix C Fig. 119B) and T47D *ESR1* mutant cell models[110] (Appendix C Fig. 119C).

I next sought to extend our findings to clinical specimens. I first analyzed the RNA-seq data set composed of 51 pairs of ER+ primary-matched metastatic tumors (7 *ESR1* mutant and 44 *ESR1* WT pairs) merged from our in-house WCRC cohort[172, 173, 209] (36 pairs) and DFCI cohort (15 pairs)[210] (Appendix D Table 14). After subtracting each enrichment score of primary tumors from the matched metastatic lesions (Δ GSVA Score), I found that metastatic tumors harboring *ESR1* mutations exhibited significantly higher enrichment levels of basal but not luminal marker genes than WT ones (Fig. 53D and Appendix C Fig. 119D).

A Basal /Luminal Gene Set Pairs Used in This Study

| Pair Number | Pair Name | Luminal markers | Basal markers |
|---|---|---|---|
| 1 | Charafe-Jauffret | 380 | 455 |
| 2 | Huper | 58 | 53 |
| 3 | Sorlie | 68 | 75 |
| 4 | TCGA | 139 | 164 |

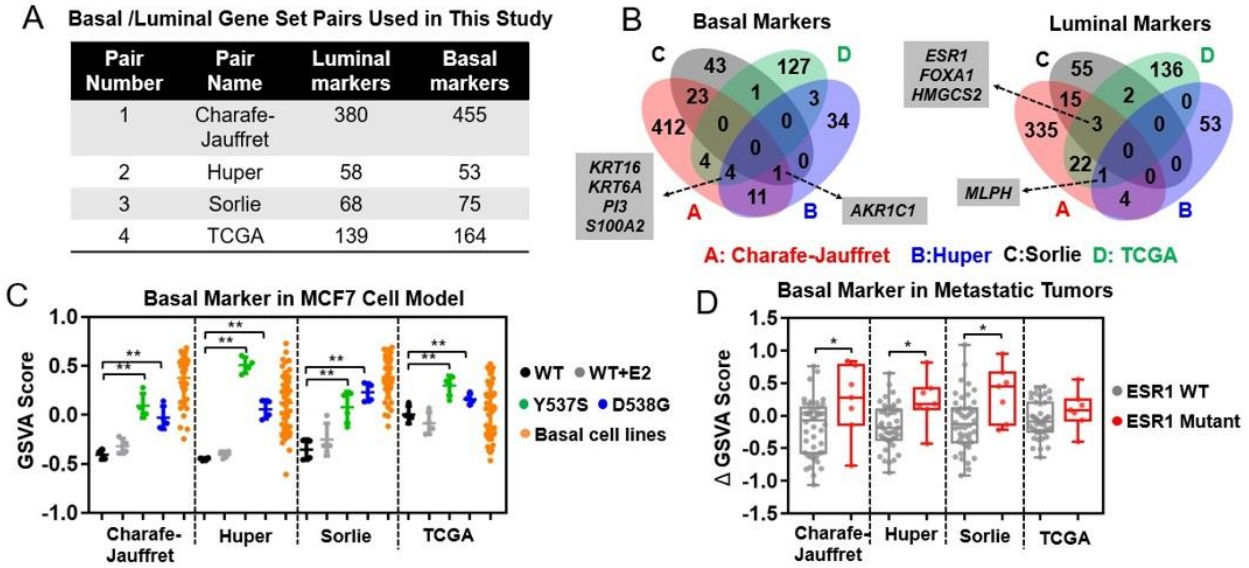



**Figure 53. Basal gene sets are enriched in *ESR1* mutant cell models and metastatic lesions**

A. Summary of all four pairs of luminal/basal gene sets applied in this study with gene numbers specified in each set. B. Venn diagram representing the overlap of basal (left) or luminal (right) gene sets among all four pairs. Intersected genes among three of the gene sets were labelled in grey frames. C. Dot plots showing GSVA score of the four pairs of basal marker gene sets enrichment in MCF7 genome-edited cell models. Scores from luminal and basal breast cancer cell lines were used as positive controls. Dunnett's test was used within each group. D. Box plots representing basal marker enrichments in primary-matched paired metastatic samples. Delta GSVA score of each sample was calculated by subtracting the scores of primary tumors from the matched metastatic tumors. Mann-Whitney U test was performed to compare the Delta GSVA scores between WT (N=44) or *ESR1* mutation-harboring (N=7) paired tumors. (* $p<0.05$; ** $p<0.01$)

### 4.3.2 Basal Cytokeratins Are Upregulated in *ESR1* Mutant Cells and Tumors Via Dual Mechanisms

#### 4.3.2.1 Basal Cytokeratins Are Consistently Upregulated in *ESR1* Mutant Cells and Tumors

I further investigated the individual basal marker genes among the Charafe-Jauffret gene set (n=455) to identify the leading genes of enrichment, as it has the largest amount of basal genes and overlapped genes with other three sets. Integrating the RNA-seq results from MCF7 cell models[110] and clinical samples identified a group of basal cytokeratins (BCKs) as the top consistently increased basal markers (Fig. 54, Appendix C Fig. 121and Appendix D Table 24).

Among all six BCKs, increased *KRT5, 16* and *17* were further recapitulated in independent genome-edited (Gertz) (Fig. 55A and 55B) and overexpression MCF7 *ESR1* mutant cell models (Fig 55D-55F). However, BCKs were not consistently increased in T47D *ESR1* mutant cell models, despite the agreement of increased overall basal marker enrichment (Fig. 55C). Other non-canonical basal genes (*WLS, HTRA1*, etc) alternatively lead to the enrichment (Appendix D Table 25), representing the inter-patient heterogeneity.

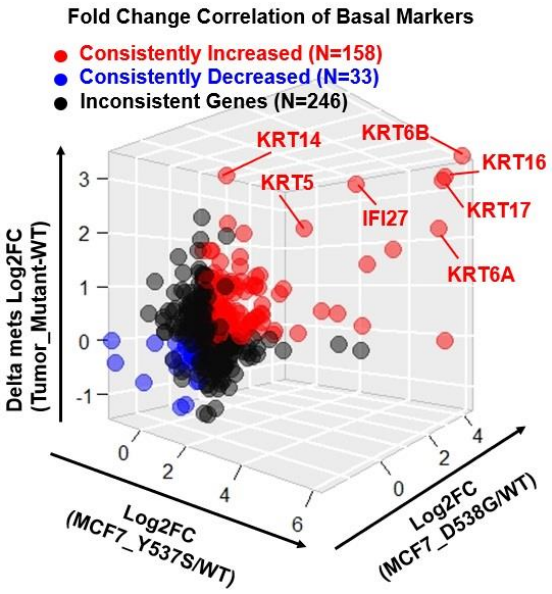



**Figure 54. Basal cytokeratins are consistently upregulated in *ESR1* mutant cells and tumors.**

Three-dimensional plot showing the correlation of basal marker gene fold changes in Y537S/D538G cells (normalized to WT vehicle) and paired mutant tumors (normalized to WT tumors). Consistently increased genes were highlighted in red. Top 7 increased genes were labelled with gene names.

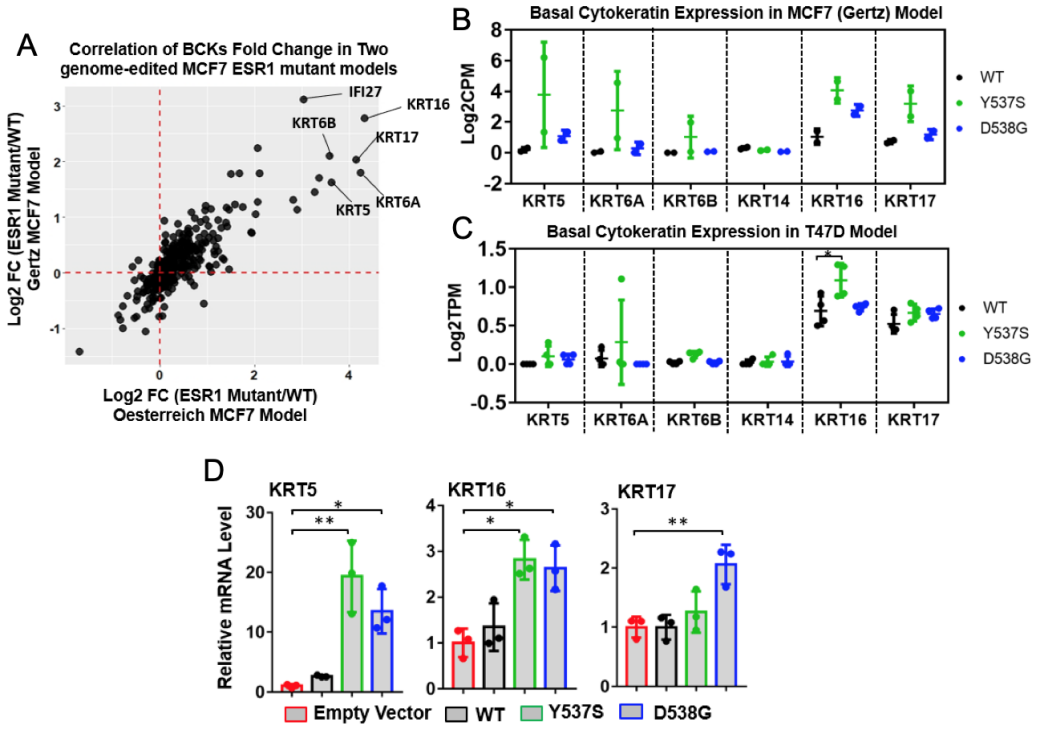



**Figure 55. Basal cytokeratins are increased in other *ESR1* mutant breast cancers cell models.**

A. Two-dimensional plot showing the correlation of basal marker gene fold changes in SO and JG MCF7 genome-edited Y537S/D538G cells (normalized to WT vehicle). Top 6 increased genes were labelled with gene names. B and C, Dot plot represents all six basal cytokeratins expression in the Gertz MCF7 (B) and SO T47D (C) *ESR1* mutant cell models. Each dots represent the expressional level from a single clone under each genotype. D. Bar graphs representing qRT-PCR measurement of *KRT5, 16* and *17* mRNA levels in MCF7 overexpression *ESR1* mutant cell models. ΔΔCt method was used to analyze relative mRNA fold changes normalized to empty vector cells and RPLP0 levels were measured as the internal control. Each bar represents mean ± SD with three biological replicates. This experiment was reproduced twice. Dunnett's test was used to compare the gene expression of each *ESR1* mutant group to WT cells. (* $p<0.05$, ** $p<0.01$).

### 4.3.2.2 Basal Cytokeratins Are Negatively Regulated by ER

As mutated ER is known to trigger gene expression in the absence of ligand[110, 117], I next asked whether induction of BCKs are partial outcomes of ligand-independent transcriptional program. Analysis of eight public available RNA-seq/microarray data sets[110, 117, 129, 268-

270] of six ER+ luminal breast cancer cell lines illustrated that none of the six BCKs were inducible by estradiol, whereas slight contextual E2 effects on luminal cytokeratins were observed (Fig. 56 A).

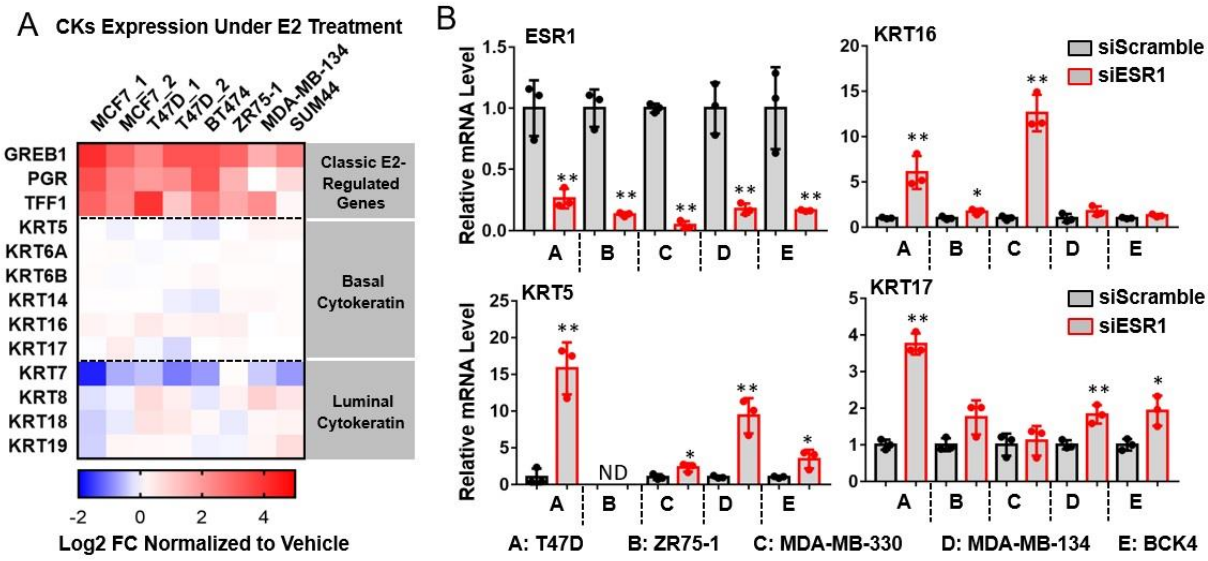



**Figure 56. Basal cytokeratins not E2-regulated and negatively associated with estrogen receptor.**

A. Heatmap representing the expression fold changes of the six basal cytokeratins and four luminal cytokeratins under E2 treatment in 6 ER+ breast cancer lines from public available data base. Classic E2-regulated genes were set as positive controls here. RNA-seq/microarray data for multiple ER+ breast cancer cell lines with or without E2 treatment were obtained from GSE89888 (MCF7_1, T47D_1), GSE51403 (MCF7_2), GSE108304 (T47D_2), GSE3834 (BT474), GSE61368 (ZR75-1), GSE50695 (MM134 and SUM44). B. Bar graphs representing qRT-PCR measurement of *ESR1*, *KRT5*, *16* and *17* mRNA levels in five ER+ breast cancer cells with siRNA knockdown of *ESR1* for 7 days. ΔΔCt method was used to analyze relative mRNA fold changes normalized to WT cells and RPLP0 levels were measured as the internal control. Each bar represents mean ± SD with three biological replicates. This experiment was done once. Student's test was used to compare the gene expression between scramble and knockdown groups (* $p<0.05$, ** $p<0.01$).

I further examined their correlation with ER expression in ER+ primary tumors. Surprisingly, strong negative correlations were observed in both TCGA and METABRIC restricting to the six BCKs but not luminal cytokeratins (Fig. 57A). I next tested the ER-dependence of *KRT5/16/17* induction in MCF7 *ESR1* mutant cells as they were the only three consistently increased BCKs in both Bahreini et al.[110] and Jason Gertz's MCF7 genome-edited cell models. Transient knockdown of ER significantly increased *KRT5/16/17* expression in MCF7 cells (Fig. 57B) and five other ER+ breast cancer cell lines (Fig. 56B). Together, these results suggest that BCKs induction in *ESR1* mutant cells were not resulted from short-term canonical ER transactivation, and I thus next tested the hypothesis of epigenetic reprogramming.

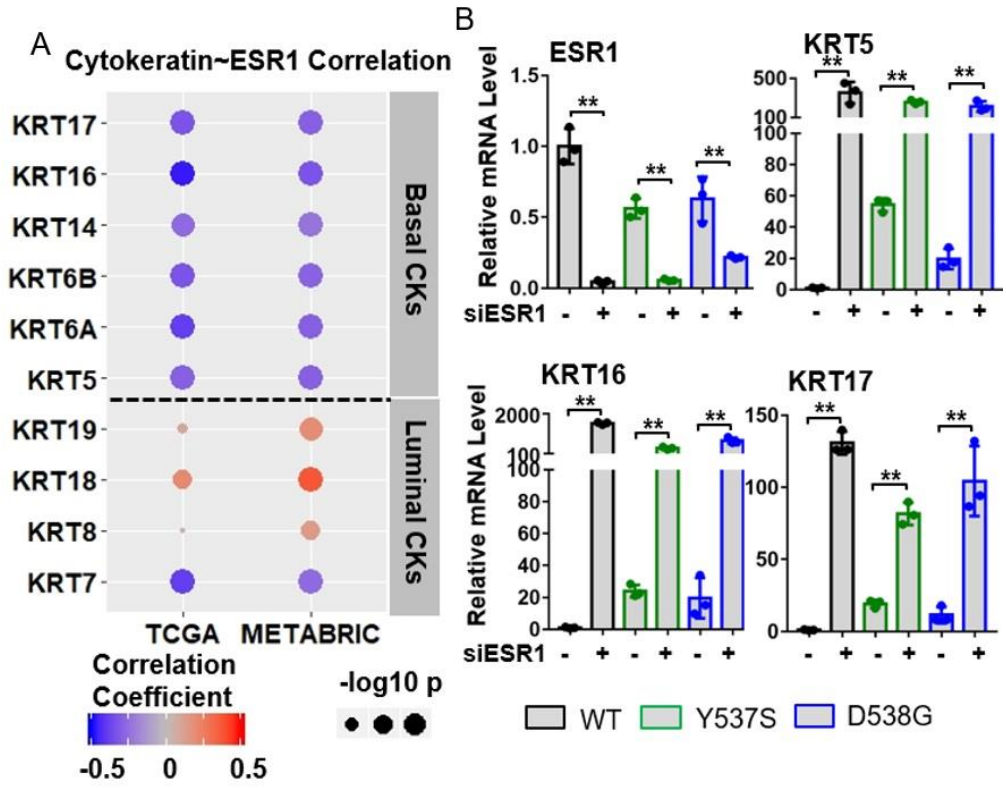



**Figure 57. Basal cytokeratins are negatively regulated by both WT and mutant ER.**

A. Graphic view of pearson correlation between *ESR1* and each basal or luminal cytokeratin in ER+ breast cancer patients in TCGA and METABRIC cohorts. Color scale represent correlation coefficient and dot size shows significance. B. Bar graphs representing qRT-PCR measurement of *ESR1*, *KRT5*, *16* and *17* mRNA levels in MCF7 WT and *ESR1* mutant cells with siRNA knockdown of *ESR1* for 7 days. ΔΔCt method was used to analyze relative mRNA fold changes normalized to WT cells and RPLP0 levels were measured as the internal control. Each bar represents mean ± SD with three biological replicates. This experiment was replicated for three times independently, and representative results from one experiment were shown. Student's test was used to compare the gene expression between scramble and knockdown groups. (* $p<0.05$; ** $p<0.01$)

#### 4.3.2.3 *ESR1* Mutations-Induced *KRT*14/16/17 Are Associated with A Putative Chromatin Loop

A recent study by Arnesen et al. suggested CTCF genomic binding regions are more accessible in *ESR1* mutant cells. In line with this, I found that the enrichment of CTCF gene

signature derived from MCF7 cells[271] was significantly higher in our MCF7 *ESR1* mutant cells (Fig. 58A and Appendix D Table 26), suggesting the active CTCF-driven epigenetic reprogramming. Previous CTCF ChIA-PET [272] and ChIP-seq [271] in MCF7 cells showed a putative CTCF-driven chromatin loop around the *KRT14/16/17* but not *KRT5/6A/6B* loci (Fig. 58B), also known as a topological associated domain (TAD) that provides enhancer access to BCKs. CTCF recruitment at those putative sites was not E2-regulated but sharply increased in *ESR1* mutant cells (Fig. 58C), suggesting increased TAD formation at this locus in *ESR1* mutant cells.

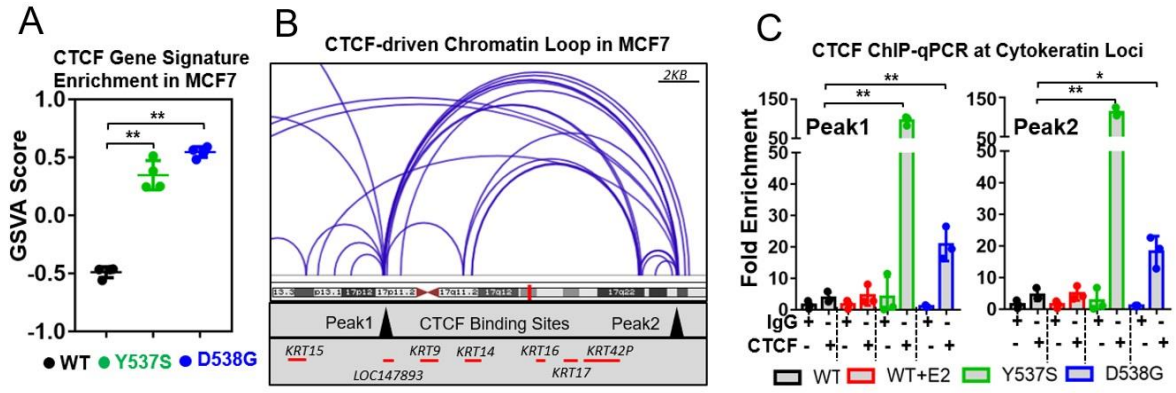



**Figure 58. *ESR1* mutations induce KRT14/16/17 via a putative chromatin loop.**

A. Dot plot showing enrichment levels of CTCF-gene signature enrichment levels in MCF7 WT and *ESR1* mutant cells. Dunnett's test was utilized to compare the difference. B. Illustration of CTCF-binding sites and CTCF-driven chromatin loops within *KRT14/16/17* genomic region in MCF7 cells. C. Bar graphs displaying CTCF binding events measured by ChIP-qPCR at putative binding sites illustrated in I. CTCF binding fold enrichments were normalized to the average of IgG binding. Each bar represents mean ± SD with three biological replicates. This experiment was repeated twice independently, and representative results from one experiment were shown. Dunnett's test was performed to compare CTCF binding in MCF7 WT and *ESR1* mutant cells. (* $p<0.05$; ** $p<0.01$)

#### 4.3.2.4 Basal Cytokeratins Are Partially Regulated by Progesterone Receptor

Additionally, progesterone receptor (PR) has been identified as a positive regulator to BCKs such as *KRT5*[273, 274]. PR ChIP-seq[275] showed PR binding sites around proximal regions of both and *KRT5/6A/6B* genomic loci and *KRT14/16/17* putative looping region (Fig. 59B). Given PR is upregulated in *ESR1* mutant cells (Fig. 59C), I then investigated the possibility of PR serving as an activator. Knockdown of PR partially rescued the increased expression of *KRT5/16/17*(Fig. 59A). Furthermore, both PR agonist P4 and PR antagonist RU486 induced their expression in Y537S mutant cells, whereas RU486 also trigged KRT5 and 16 in D538G cells. The effects of RU486 were likely due to the previous reported partial agonistic activity towards

progesterone receptor[276] and glucocorticoid receptor[277] (Fig. 59D). Taken together, our data demonstrated the dual-mechanistic induction of basal cytokeratins in *ESR1* mutant cells: *KRT14/16/17* are possibly induced by a TAD complimentary with PR transcriptional regulation, while *KRT5* might be merely mediated by PR.

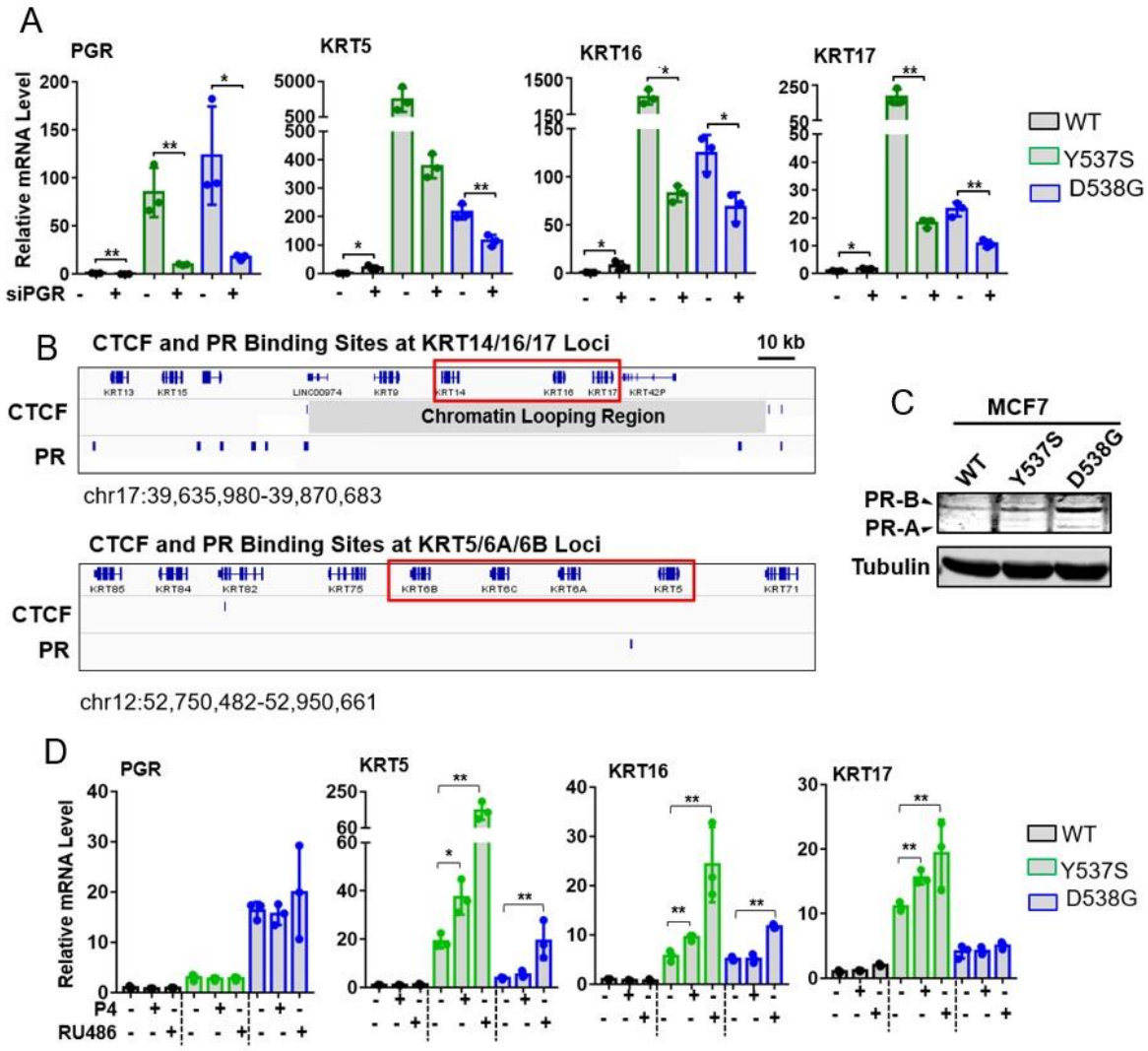



**Figure 59. Basal cytokeratins are partially regulated by progesterone receptors.**

A. Bar graphs showing qRT-PCR measurement of *PGR*, *KRT5*, *16* and *17* mRNA levels in MCF7 *ESR1* WT and mutant cells with siRNA knockdown of *PGR* for 7 days. This experiment was repeated three times independently, and representative results from one experiment were shown. B. Schematic illustration of CTCF and PR binding sites distribution at *KRT14/16/17* and *KRT5/6A/6B* locus. C. Immunoblot detection of PR expression in *ESR1* mutant cells after hormone deprivation and tubulin was used as loading control. This experiment was one once. D. Bar graphs showing qRT-PCR measurement of *PGR*, *KRT5*, *16* and *17* mRNA levels in MCF7 *ESR1* WT and mutant cells with either 100 nM P4 or 1 uM RU486 treatment for 3 days. ΔΔCt method was used to analyze relative mRNA fold changes normalized to WT cells and RPLP0 levels were measured as the internal control. Each bar represents mean ± SD with three biological replicates. This experiment was repeated twice independently, and representative results from one experiment were shown. Student's test was used to compare the gene expression between scramble and knockdown groups (* $p<0.05$, ** $p<0.01$).

### 4.3.3 *ESR1* Mutant Metastatic Samples Show Enhanced Immune Infiltration

#### 4.3.3.1 *ESR1* Mutant Tumors Share Enhanced Immune Response with Basal Cytokeratin Higher ER+ Tumors

BCKs are used as biomarkers to define basal subtype and predict clinical outcomes and are highly expressed in triple negative breast cancers[278-280]. Yet the functional aspects behind gain of BCKs in *ESR1* mutant tumors are unclear. To distinguish specific BCKs-associated pathways from other potential *ESR1* mutant-driven effects, I intersected all the significantly enriched hallmark pathways in *ESR1* mutant metastatic tumors with those enriched pathways in a subset of BCKs-high (top quantile GSVA score) ER+ LumA tumors from TCGA and METABRIC cohorts. Markedly, I identified four immune-related pathways (57.1%) out of seven shared in total (Fig. 60A, Appendix D Table 27 and Appendix C Fig. 122 A).

Bioinformatic evaluation first confirmed the higher overall immune fractions in both BCKs-high LumA tumors in TCGA (and a trend in METABIRC), albeit lower than actual basal tumors (Fig. 60B). BCKs-high tumors in TCGA exhibited higher lymphocyte score and leukocyte fractions according to recent biospecimens report[263] (Fig. 60C) concurrently presenting significantly higher PDCD1 expression (Appendix C Fig. 122 B). BCKs-high ER+ LumA patients also consistently showed better prognostic outcomes in both TCGA and METABRIC (Fig. 60D), likely as a result of the stronger immune infiltration. Of note, such effect was not observed in ER+ LumB patients (Appendix C Fig. 122 C).

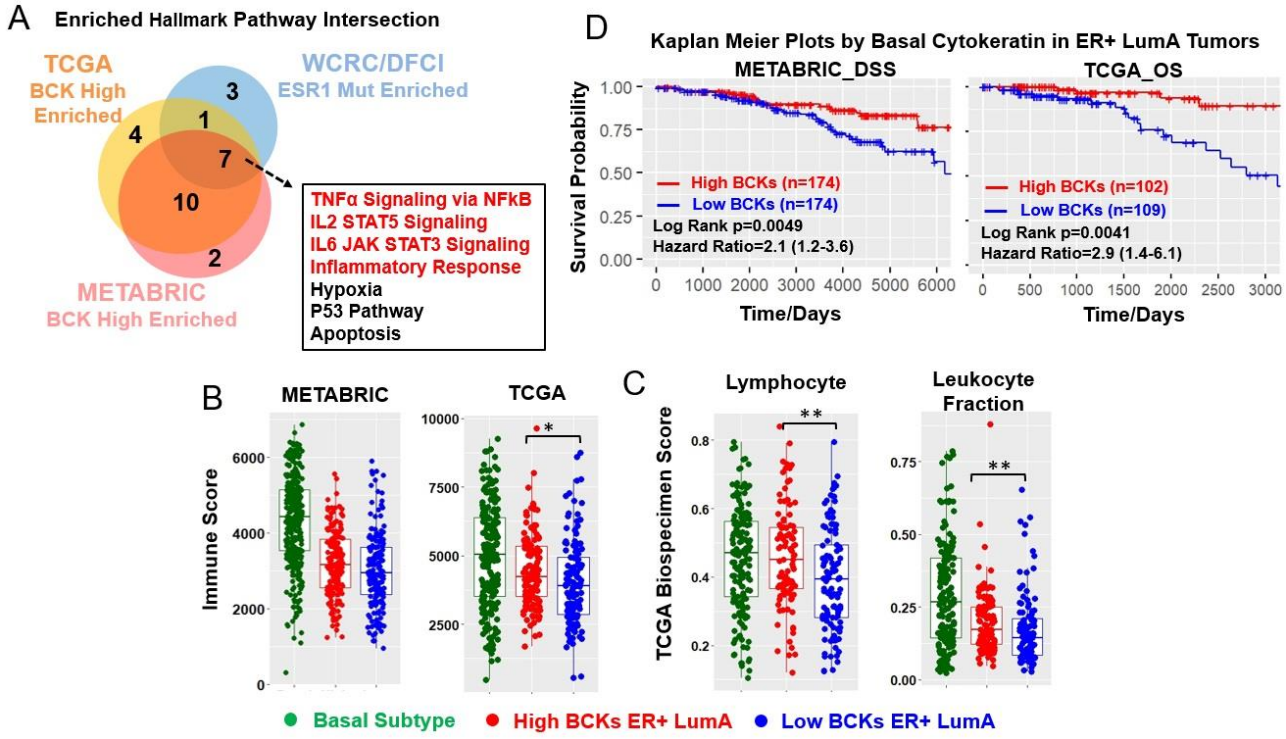



**Figure 60. BCK Higher ER+ lumA tumors show enhanced immune response.**

A, Venn diagrams showing the interaction of significantly enriched hallmark pathways in three sets of comparisons: BCKs-high vs low in 1) TCGA and 2) METABRIC and 3) *ESR1* Mutant vs WT metastatic tumors. Overlapped pathways were labelled in frame, immune related pathways were further highlighted in red. B, Box plots showing the immune fractions scores across basal tumors and BCK-high and low subsets in ER+ LumA tumors based on ESTIMATE evaluations. Mann Whitney U test was used for comparison. C, Box plots representing lymphocytes and leukocyte fractions comparison between TCGA BCK-high and low ER+ LumA subsets based on biospecimen data. Mann Whitney U test was applied. D. Kaplan-Meier plots showing the disease-specific survival and overall survival between BCKs high and low subsets of ER+ LumA patients from METABRIC (DSS) and TCGA (OS) cohorts. BCKs high and low were defined by the upper and bottom quartiles of each subset. Censored patients were labelled in cross symbols. Log rank test was used and hazard ratio with 95% CI were labelled. (* $p<0.05$, ** $p<0.01$)

On the other hand, *ESR1* mutant metastatic tumors also exhibited significantly higher immune portions than WT tumors (Fig. 61A). Particularly, enrichment analysis with individual

immune cell signatures[281, 282] revealed significantly higher CD8+ T cell, NK cell and macrophages enrichments in *ESR1* mutant tumors (Fig. 61B).

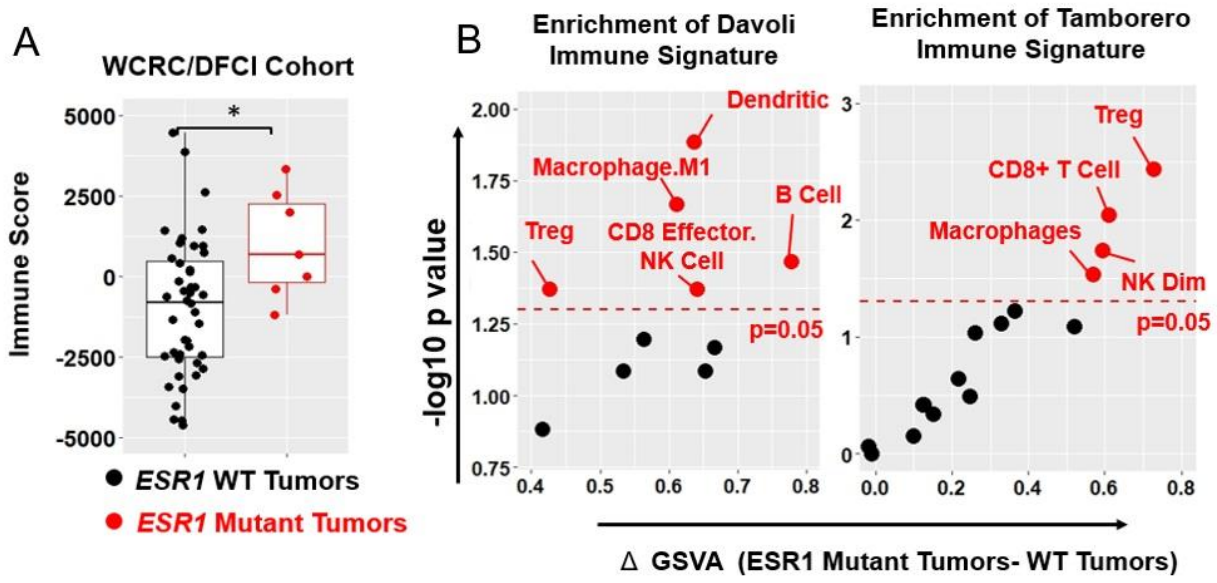



**Figure 61.*ESR1* mutant tumors show enhanced immune response.**

A. Box plots showing the immune scores between *ESR1* mutant and WT metastatic lesions. Immue scores of each samples were calculated using ESTIMATE and represents the immue cell fractions of each sample. B. Dot plot showing the enrichment alterations in terms of immune cell subtypes in *ESR1* mutant metastatic lesions using two sets well-defined immune cell signatures. (* $p<0.05$)

#### 4.3.3.2 S100A8-S100A9/TLR4 Signaling Is Consistently Enhanced in *ESR1* Mutant Tumors and Basal Cytokeratin High Tumors

Basal breast cancers harbor high immune infiltrations due to their nature of high tumor mutation burdens (TMBs). However, TMBs are not different in BCKs-high and low ER+ LumA tumors in TCGA (Appendix C Fig. 122 D), indicating another immune activation mechanism. A further correlation based on the fold changes of 141 immune genes referred by ESTIMATE[283] across TCGA/METABRIC BCKs-high ER+ LumA and *ESR1* mutant tumors identified S100A8

and S100A9 as the top consistently increased immune-related targets (Fig. 62A and Appendix D Table 28), which correlated with enhanced immune infiltrations in BCKs-high ER+ LumA tumors (Appendix C Fig. 122 E). S100A8 and S100A9 typically form heterodimers in tumor microenvironment and triggers inflammatory response[284, 285]. In agreement with metastatic tissues, I found significantly higher S100A8/9 heterodimer concentrations in plasma from patients with *ESR1* mutant progressive disease (Fig. 62B). High S100A8/9 expression was also observed in MCF7 *ESR1* mutant cells (Appendix C Fig. 122 F), and its corresponding signature of downstream Toll-like receptor 4 (TLR4) pathway was also enriched in MCF7 *ESR1* mutant cell models (Appendix C Fig. 122 G) and metastatic lesions (Fig. 62C).

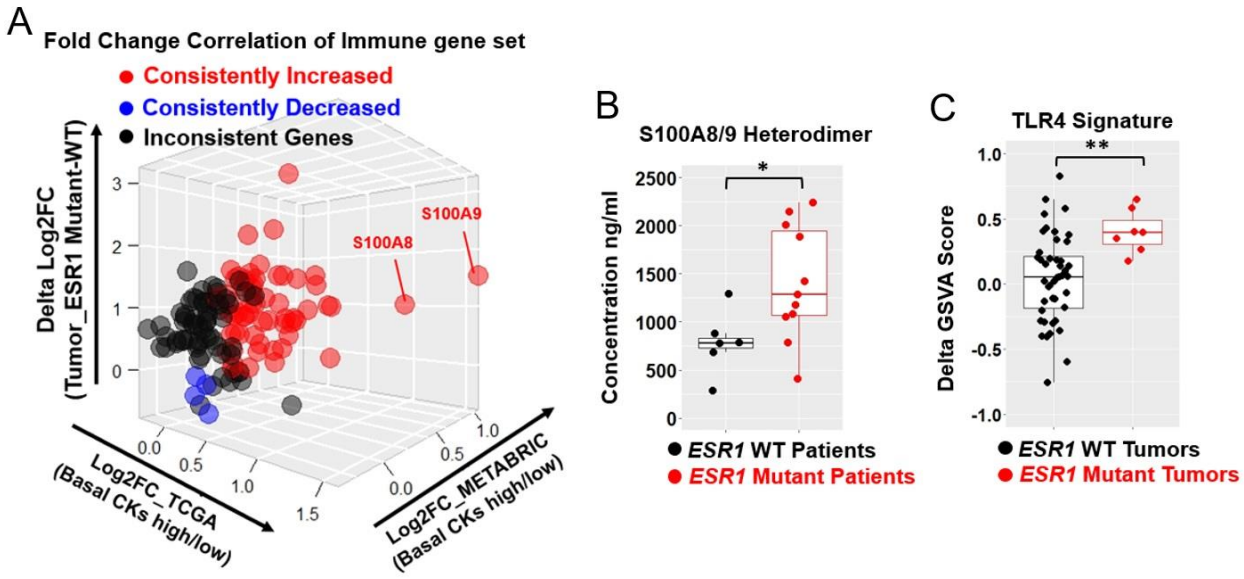



**Figure 62. S100A8-S100A9/TLR4 signaling are consistently enhanced in *ESR1* mutant tumors and basal cytokeratin high ER+ LumA tumors.**

A. Three-dimensional plot showing the correlation of immune gene sets (n=141) fold changes in TCGA and METABRIC BCKs-high tumors (normalized to BCKs-subsets) and paired mutant tumors (normalized to WT tumors). Consistently increased genes were highlighted in red. Top 2 increased genes were labelled with gene names. B. Box plot showing S100A8/9 heterodimer concentrations in *ESR1* WT and mutant patient plasma samples. Mann Whitney U test was utilized. This experiment was done once with three technical replicates. C. Box plot showing the comparison of TLR4 signaling signature enrichments in primary-matched paired metastatic samples. Delta GSVA score of each sample was calculated by subtracting the scores of primary tumors from the matched metastatic tumors. Mann-Whitney U test was performed between WT and mutant tumors. (* $p<0.05$, ** $p<0.01$)

## 4.4 Discussion

In this chapter, we utilized published and self-generated molecular gene signatures representing luminal and basal intrinsic subtypes from different resources and found the consistent enrichment of basal subtype marker genes in *ESR1* mutant cell models and clinical samples. We further identified the gain of basal cytokeratins among the top gained basal marker genes.

Increased basal cytokeratins expression is not a direct outcome of the acquired ER ligand-independent activation program, but instead associated with novel chromatin loops as well as progesterone receptor regulation. Clinically, high BCKs tumors share multiple enriched immune pathways with *ESR1* mutant tumors mainly correlated to enhanced S100A8/9-TLR4 signaling. Collectively, these observations show that both activating ER mutations gained basal cytokeratin expression, which is associated with epigenetic regulations and further implies immune therapeutic vulnerabilities of *ESR1* mutant tumors.

Although we identified basal cytokeratins in *ESR1* mutant cells among the top increased basal marker genes, we were also able to identify other few highly consistently increased basal marker genes (Fig. 54). For instance, the interferon-alpha inducible protein 27 (IFI27) is also consistently upregulated in both ours and Dr. Gertz's *ESR1* mutant cell models and clinical tumors, and previous studies have reported its role in regulating breast cancer innate immunity[286] and cisplatin resistance in gastric cancer[287]. In addition, the consistent upregulation of peptidase inhibitor 3 (PI3) has been reported to be highly associated altered TP53 pathways[288]. These additional findings imply that gain of basal-like features of *ESR1* mutant cells might confer other potential functional alterations. Further wet-bench works are still required to explore the functional aspects of those basal markers in the multiple acquired phenotypes I mentioned in Chapter 2 and 3 of this dissertation, such as endocrine resistance, cell-cell adhesion and cell-ECM adhesion.

In the previous chapter, we have identified a collection of genes that are ligand-independent activated/repressive in *ESR1* mutant cells but non-E2 regulated in WT cells through the transcriptomic characterization. Here, all six basal cytokeratins belong to such novel target gene category. Surprisingly, though lack of E2 regulation in WT cells, transient knockdown of ER

increased the expressional level of those basal cytokeratins. One possible explanation is that ER might serve as a negative epigenetic regulator that naturally suppresses the expression of those basal markers to maintain the luminal subtype feature. Similar findings have been reported before by Ariazi et al.[289]. In their studies, they proposed that ER silences basal marker, EMT and stem cell related genes via DNA methylation by recruiting essential methyl transferases like EZH2 and DNMTs[289]. To further validate this, bisulfide sequencing should be performed as the next bench step around the basal cytokeratin loci with *ESR1* knockdown to prove the causative relationship.

In the chapter, we have provided strong evidence towards the unique upregulation of S100A8/S100A9 signaling in *ESR1* mutant cells, tumors and blood samples, which is potentially associated with the enhanced overall immune fractions in *ESR1* mutant metastatic samples. S100A8/S100A9 typically exhibit pro- and anti-inflammatory properties in the context of breast cancer by forming heterodimers of S100A8/A9, alternatively known as calprotectin[290, 291], and associated with poor prognosis in multiple cancer types[292]. As blockade of S100A8/A9 activity using small-molecule inhibitors or antibodies improves pathological conditions in murine models[293], the heterodimer has potential as a therapeutic target in breast cancer harboring hotspot *ESR1* mutations and more pre-clinical studies are required.

Our study has discovered enrichment of basal markers in *ESR1* mutant cells and address the potential mechanisms as well as its clinical relevance, albeit with some remaining limitations, such as the lack of *in vivo* validation of basal cytokeratin overexpression. Furthermore, the enhanced immune infiltration requires additional validation by TIL counting on *ESR1* mutant tumor sections, and our clinical cohorts are still with limited numbers of *ESR1* mutant tumors. Additionally, the formation of CTCF-driven chromatin loop at the basal cytokeratin gene locus

requires additional confirmation by chromosome conformation capture. Nevertheless, our study serves as a robust pre-clinical report uncovering mechanistic insights into *ESR1* mutations and its role in conferring basal-like feature to ER+ breast cancer and implicates the immune therapeutic vulnerabilities to this subset of patients.

## 5.0 Summary of The Role of *ESR1* Mutations in Breast Cancer Endocrine Resistance and Metastasis from This Study

Recent studies have identified *ESR1* hotspot mutations in endocrine resistant metastatic breast cancer, with a the rare occurrence of in primary tumors (<5%) but high mutation frequencies (20-40%) in metastatic lesions and circulating-free DNA (cfDNA), associated with significantly worse clinical outcomes. However, the role of mutant ER plays in conferring endocrine resistance is still not fully characterized, and its additional function in mediating metastatic propensity remains largely unknown. In this thesis, we utilized robust genome-edited Y537S and D538G *ESR1* mutant cell models to characterize multiple endocrine resistant and metastatic phenotypic alterations, and further employed multi-omic data sets from cell lines and clinical samples to decipher this potential molecular mechanisms involved, which provided promising novel therapeutic targets to treat those patients harboring *ESR1* mutations. A detailed tabular summary of major findings are shown in Table 2 below.

**Table 2. Summary of the major findings of this thesis.**

| Chapter | Major features identified | Models | Proposed Mechanisms | Proposed Therapy |
|---|---|---|---|---|
| 2 | Ligand-independent growth and endocrine resistance | MCF7-Y537S/D538G<br>T47D-Y537S/D538G | Ligand-independent ER activation | Higher dose of SERDs/SERMs |
| | Enhanced IGF1 response | MCF7-Y537S/D538G<br>T47D-Y537S/D538G | IRS upregulation | IGF1R/ER co-target |
| 3 | Enhanced cell-cell adhesion | MCF7-Y537S/D538G<br>T47D-Y537S | Desmosme/gap junction gene increase | Gap junction blockade |
| | Diminished cell-ECM adhesion and enhanced invasion | MCF7-Y537S/D538G<br>T47D-Y537S/D538G | TIMP3 decreased and MMP hyperactivation | MMP blockade |
| | Enhanced migration | T47D-D538G | Wnt hyperactivation | Wnt/ER co-target |
| 4 | Increased basal marker enrichment and predicted immune infiltration | MCF7-Y537S/D538G | Basal cytokeratin upregulation via CTCF and PR | Immune Therapy |

In Chapter 2, we mainly focused on the mechanisms of endocrine resistance conferred by *ESR1* mutations. In the first part, we identified the ligand-independent ER activation which confers the anti-ER and anti-E2 therapeutic resistance to breast cancer cells. Transcriptomic analysis revealed highly mutation site- and context-dependent gene expression profiles. Despite the less sensitive response of *ESR1* mutant cells, our data support that higher doses of SERMs/SERDs treatment could still efficiently blocked the ligand-independent growth of those cells. In addition, the novel oral SERD AZD9496 shows the highest efficacy when comparing the drugs, which is proposed to be one of the therapeutic methods towards those patients with *ESR1* mutations. Future

directions include further assessing the efficacy of other novel SERDs towards our *ESR1* mutant cell models and investigating the mechanisms regarding to the differential drug response *in vitro and in vivo*. In addition, comprehensive integrative analysis with our RNA-seq data sets of cell models with other public available *ESR1* mutant cell model transcriptomic data sets are warrant in order to better elucidate how different *ESR1* mutations reprograms the transcriptome.

In the second part of Chapter 2, we furthered the observations of increased expression of several IGF1R signaling components in *ESR1* mutant cells according to RNA-seq analysis, and found that elevated levels of IRS1 in *ESR1* mutant-cells increased IGF1 potency through the PI3K-Akt axis, and triggered enhanced IGF1-induced growth. Combination treatment co-targeting ER and IGF1R induced synergistic inhibitory effects specific to *ESR1* mutant cells, implicating another therapeutic strategy to overcome endocrine resistance of *ESR1* mutant breast cancer in clinic. Future efforts regarding to this part may focus on better characterizing this therapeutic vulnerability in other *ESR1* mutant cell and mouse models to examine whether this strategy is efficient to the vast majority of *ESR1* mutant patients or just a subset of them. We will also assess whether the enhanced IGF1 response impacts metastasis properties of *ESR1* mutant cells or is merely specific to endocrine resistance features.

In addition to endocrine resistance, *ESR1* mutations also play essential roles in prompting breast cancer metastasis, according to the clinical evidence that the presence of *ESR1* mutations exclusively in distant but not in locally-recurrent tumors. As described in Chapter 3, our findings from *in vitro, in vivo* models and clinical specimens indicate ER mutant as a multimodal player potentially affecting different functional aspects at different stages of metastasis, including enhanced cell-cell adhesion, diminished cell-ECM adhesion and increased invasion, and unique

context-dependent increased migration. Of note, different identified mechanisms behind those gained metastatic features implicate potential novel therapeutic strategies in clinic.

Given the gap junction reprogramming attributes to the cell-cell adhesive phenotype, our data support that blockade of gap junction signaling abolishes enhanced cell-cell adhesion in *ESR1* mutant cells, thus provided a promising potential strategy to target *ESR1* mutant-CTC-clusters directly in bloodstream. Future efforts will be made to assess the correlation of *ESR1* mutations and CTC cluster prevalence in liquid biopsies from patients. We will also examine the efficacy of more FDA-approved gap junction blockers in affecting CTC clusters in *ESR1* mutant cell and mouse models to test the potential drug reposition to breast cancer. Furthermore, our data suggested that the increased expression of cell-cell adhesion genes is associated with differential active histone modification rather than transcriptional ligand-independent activation. Thus, additional experiments will be performed to further elucidate how these genes are induced via epigenetic regulations.

Secondly, the alteration of TIMP3/MMP axis causes the diminished cell-ECM adhesion and enhanced invasive properties. MMP inhibition using pan-inhibitor Marimastat efficiently dampened the mutant-specific gained cell invasive properties in collagen. Future directions may include deciphering the molecular mechanisms of TIMP3 repression by mutant ER in cell models, and assessing the efficacy of Marimastat in inhibiting *ESR1* mutation-driven metastasis in a spontaneous metastatic *in vivo* model.

Moreover, we found hyperactivation of canonical Wnt pathway is partially involving in the enhanced cell migratory phenotype in T47D-D538G *ESR1* mutant cell model, and this is potentially associated with novel FOXA1-driven reprogramming. Thus, we proposed combination

treatment of Porcupine inhibitor-LGK974 and fulvestrant to co-target Wnt pathway and ER to a specific subset of *ESR1* mutant patients with Wnt hyperactivation. Given we only observe this unique phenotype in on model, further validation from other *ESR1* mutant models are required to exclude the possibility of model-specific artifact. In addition, further efforts will be made to address the mechanism underlying FOXA1-redistribution in regulating Wnt component genes to unravel this unique signaling axis. Moreover, future clinical validations are needed to examine the high canonical Wnt pathway activation in D538G mutation subtypes with pathological chemical methodologies.

Last but not least, in Chapter 4, we also identified the unique enrichment of basal markers in *ESR1* mutant cells, which is partially caused by the gain of basal cytokeratins via potential CTCF-driven chromatin interaction or PR-induced transactivation. Clinically, tumors with high basal cytokeratins expression share multiple enriched immune pathways with *ESR1* mutant tumors mainly correlated to enhanced S100A8/9-TLR4 signaling. which provides another potential critical therapeutic possibility to those *ESR1* mutant patients regarding to immune therapies. Future directions may include further characterization of immune infiltrations in *ESR1* mutant and WT patient metastatic specimens using pathological methods and single-cell resolution sequencing to deepen the understanding of differential immune response in *ESR1* mutant patients. Moreover, the regulations from CTCF-driven loop and PR transactivation will need further experimental confirmation by regulatory region editing using CRISPR/Cas9 to assess their impacts on basal cytokeratin induction.

Markedly, our data suggest that the ways hotspot *ESR1* mutations impacting breast cancer endocrine resistance and metastasis follow a highly context- and mutation site-dependent manner.

First, despite the ligand-independent ER activation pattern is observed in all the cell models, the increased SERMs/SERDs resistance was more pronounced in Y537S rather than D538G cells, suggesting that those two mutation subtypes employ distinct conformational mechanisms to affect binding to those drugs. In addition, our RNA-seq analysis uncovers many differentially altered ligand-independent and novel target genes based on different mutant cell models, suggesting that the acquired phenotypes may depend on different activating transcripts in different patient genomic backgrounds and mutation subtypes. Furthermore, the identified metastatic phenotypes are also strongly context-dependent. For instance, the enhanced cell-cell adhesion was observed in MCF7-Y537S/D538G and T47D-Y537S cell models, whereas the migratory phenotypes was exclusively identified in T47D-D538G cell model. Overall, the contextual manner of *ESR1* mutants in mediating breast cancer endocrine resistance and metastasis implies that the future therapeutic decision making towards those patients might be additionally guided by mutation subtype classification and the personalized therapeutic target molecular validation following biopsies. Finally, to elucidate these contextual effects caused by *ESR1* mutations in clinic, continuous efforts in banking and sequencing more *ESR1* mutation metastatic tumor samples are warranty in the future.

Overall, our study addresses the endocrine resistance and metastatic phenotypes as well as the mechanisms associated with hotspot *ESR1* mutations, albeit with some remaining limitations, such as the lack of *in vivo* validation with specific drug treatment of our robust *in vitro* findings for a full preclinical study. Furthermore, no detailed genetic or epigenetic mechanisms were investigated in our study to explain how mutant ER indirectly triggers the induction of metastatic candidate genes. Finally, a relative small-scale patient cohort, especially for those paired primary-

metastatic *ESR1* samples, was used in our study, which warrants validation in future studies with larger clinical cohorts. Nevertheless, our study serves as a timely and important pre-clinical report uncovering mechanistic insights into *ESR1* mutations that pave the way towards personalized treatment of patients with advanced metastatic breast cancer.

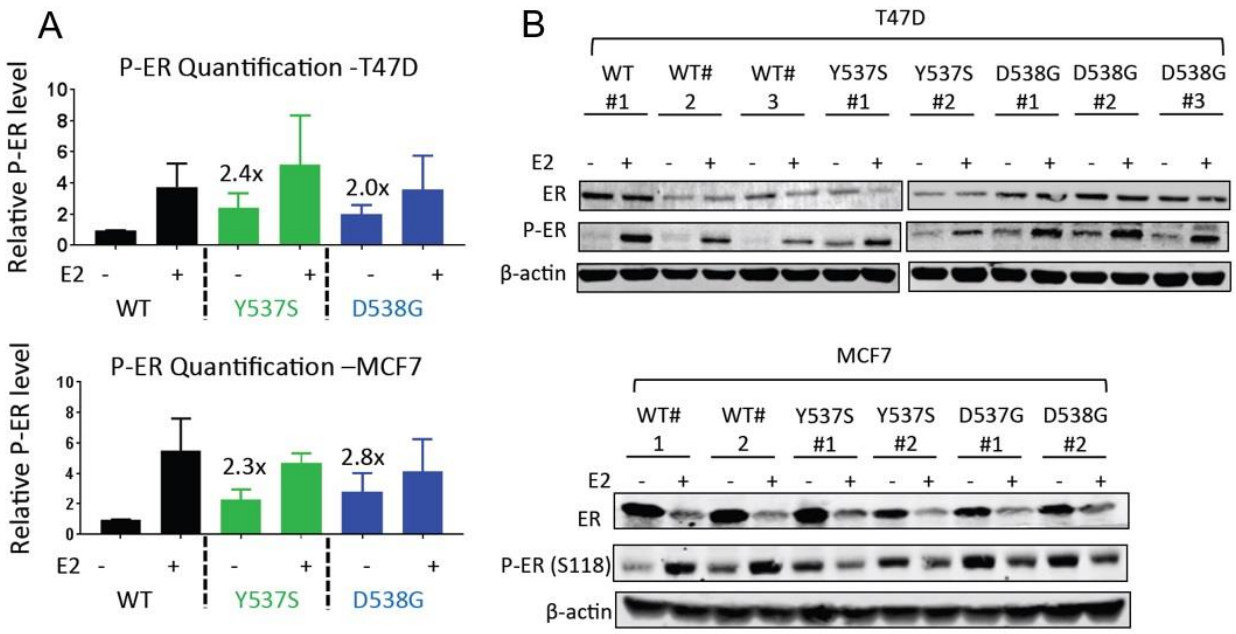



**Figure 63. Total ER and phospho ER bloDng in all clones of T47D and MCF7 cell lines.**

A. Quantification of P-ER(S118) bands from three independent experiments. Bands' densities were calculated by ImageJ. P-ER values were firstly divided to total ER level. The phosphor portions of each cell line were then normalized to vehicle-treated WT groups. B. Both T47D and MCF7 WT or mutant individual clones were hormone deprived and treated with or without 1 nM of E2 for 24 hours. The cells were lysed and subjected to western bot detec2on for ER and p-ER at Ser118 site. B-actin was used as a loading control. This experiment was done once.

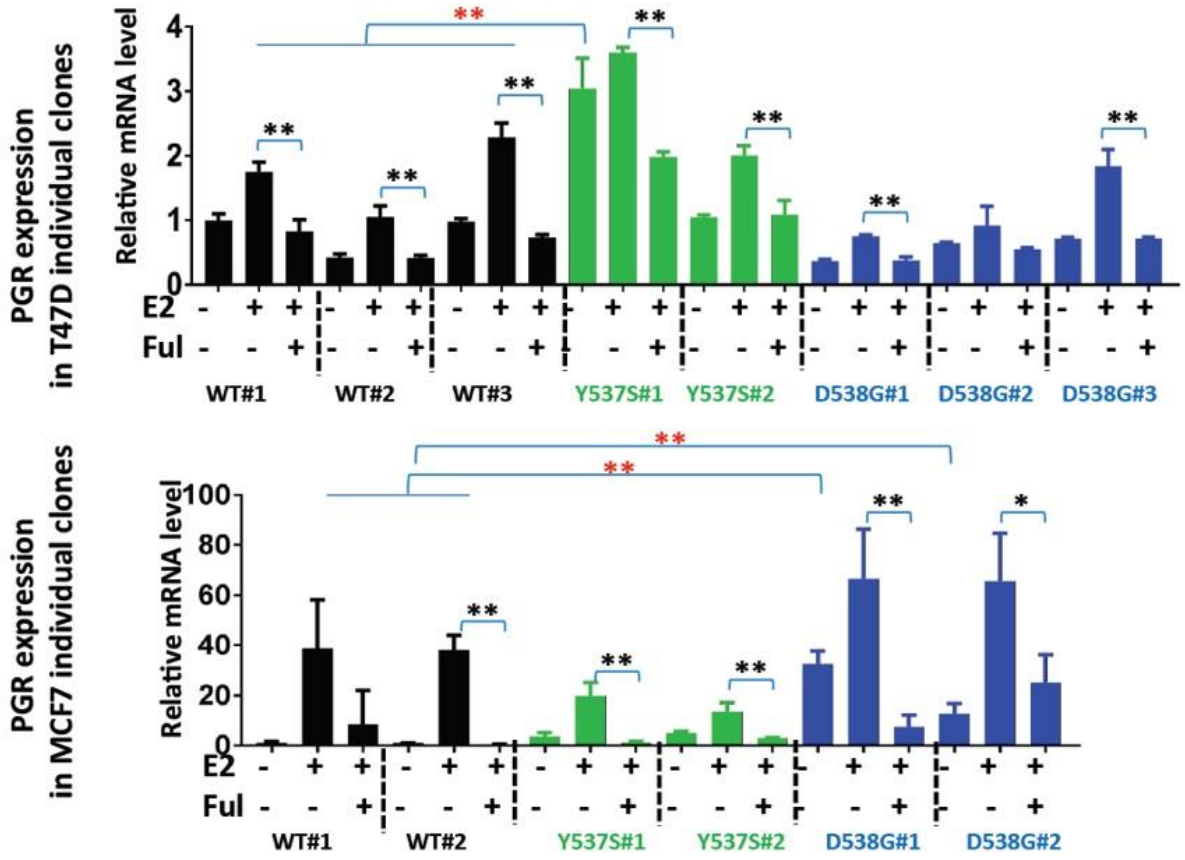
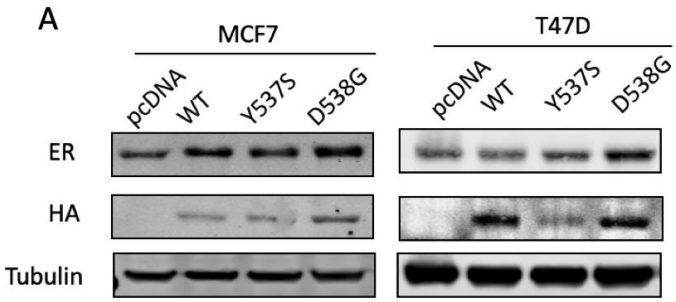



**Figure 64. *PGR* mRNA levels in T47D and MCF7 *ESR1* mutant cell models under E2 and fulvestrant treatment.**

The post-hormone-deprived MCF7 or T47D individual clones were treated with 1 nM of E2 combined with or without 1 µM of Ful for 24 hours. RT-qPCR was done using *PGR* primers. Dunnett's test was performed between the basal expression of *PGR* in each mutant clone and the average expression of *PGR* in the WT clones (* $p<0.05$, ** $p<0.01$, red) and student t-test was used to compare the response before and acer fulvestrant treatment. This experiment was done once. (* $p<0.05$, ** $p<0.01$, black).

A




B

| Method | pcDNA overexpression models | |
|---|---|---|
| Allele Freq. % | Y537S | D538G |
| MCF7_gDNA | 99.7 | 63.6 |
| MCF7_cDNA | 24 | 27.2 |
| T47D_gDNA | 86.9 | 83.1 |
| T47D_cDNA | 11.8 | 16.4 |

**Figure 65. Quality controls for the pcDNA overexpression Y537S and D538G *ESR1* mutant MCF7 and T47D cells.**

A. Immunoblot validation of MCF7 and T47D *ESR1* mutant overexpression cells. Total ER and overexpressed ER (HA-tagged) were detected and tubulin was used as loading control. B. Tabular view of mutation allele frequency in Y537S and D538G overexpression cell model at both DNA (gDNA) and RNA (cDNA) levels. This experiment was done once.

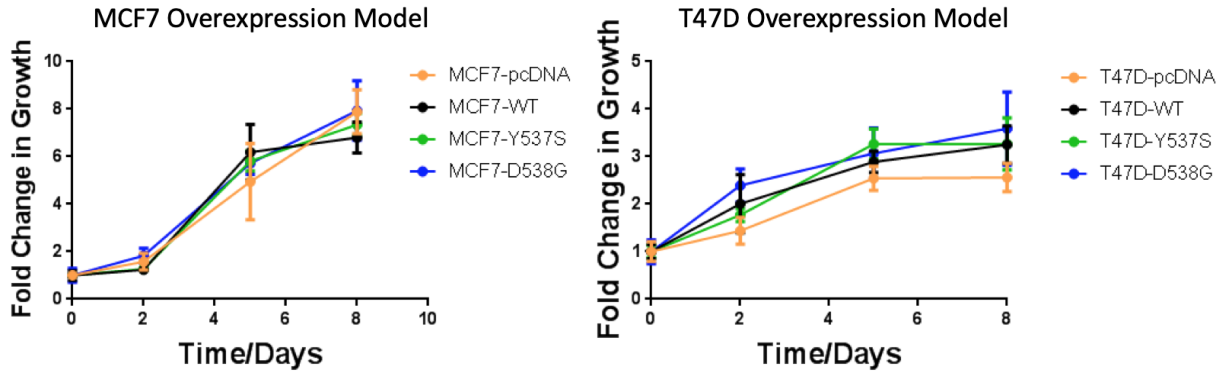



**Figure 66. Overexpression *ESR1* mutant cells do not exhbit ligand-independent growth in CSS.**

MCF7 and T47D empty vector, WT-ER, Y537S-ER and D538G-ER stably overexpressed cells were hormone deprived for 3 days and used for growth assay for 8 days starting from 4,000 cells per well. Cell numbers were quantified using FluoReporter kit. Pair-wised two-way ANOVA was performed to compare the growth ratio between WT and each mutant cell. This experiment was done once.

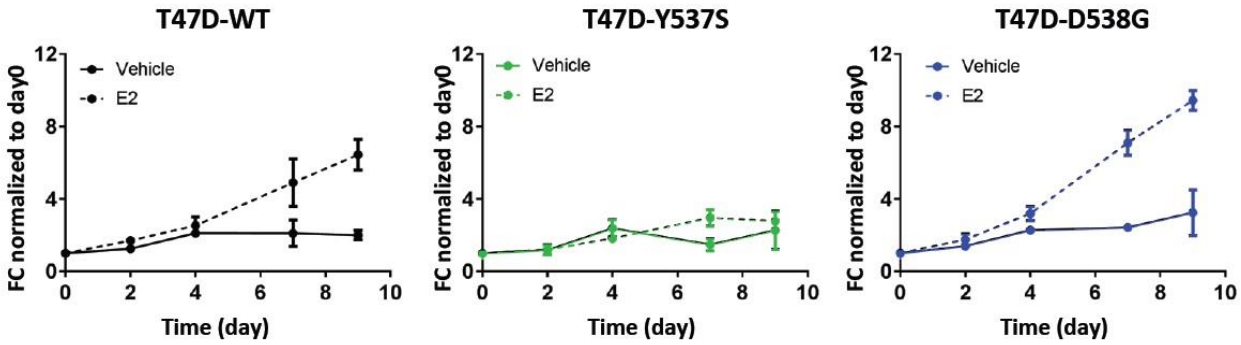



**Figure 67. The ligand-independent growth of T47D-Y537S clones depends on charcoal stripped serum.**

WT or mutant clones were hormone deprived for 3 days, pooled, and treated with veh or 1 nM E2 for up to 9 days. Cell numbers were quantified by FluorReporter kit. This experiment was repeated twice. (Gibco#12676 serum was used here)

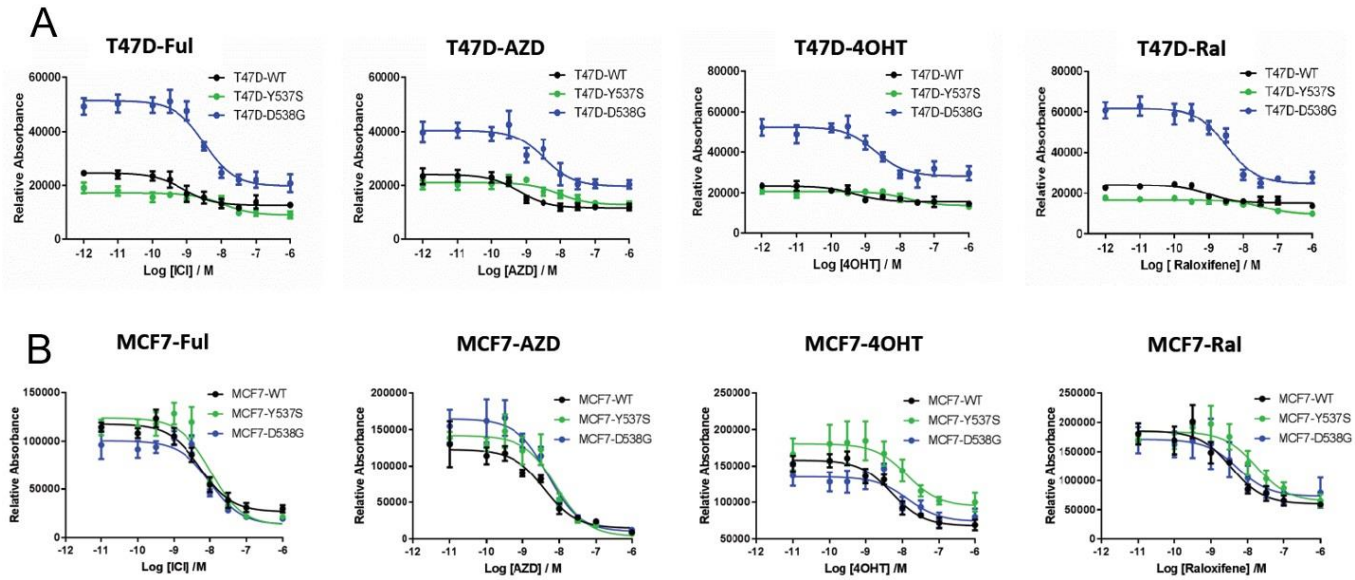



**Figure 68. Representatiuve dose response cures of T47D and MCF7 *ESR1* mutant cells towards anti-ER drug treatment.**

Dose response curves for 2D growth were plotted for Y537S and D538G mutants of T47D **(A)** and MCF7 **(B)** cells acer hormone deprivation for 3 days. The cells were treated with 20 pM E2 plus varying doses of Ful, AZD9496, 4OHT and Raloxifene. The dose response curves were fitted with nonlinear regression model in GraphPad Prism. This figure is a representative of one individual experiment that was repeated 6 2mes with consistent results. All experiments were performed in six biological replicates.

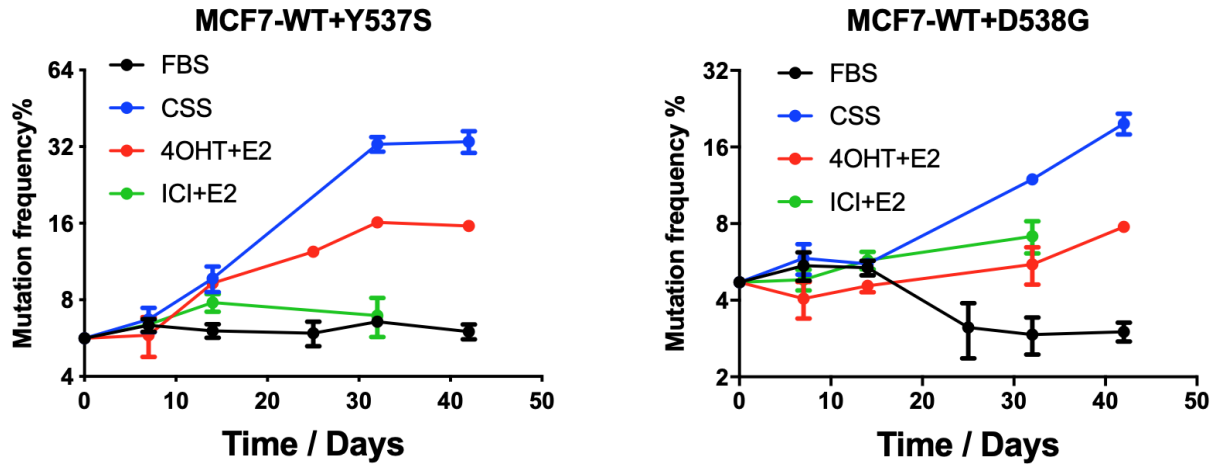



**Figure 69. MCF7 *ESR1* mutant cells show growth advaatges in CSS and anti-ER compounds.**

Pooled MCF7-WT/Y537S (left) and WT/D538G (right) cells were mixed at a ratio of 9:1 and grown in 10% FBS, 5% CSS, 5% CSS + 1 nM E2 + 100 nM 4OHT, or 5% CSS + 1 nM E2 + 30 nM fulvestrant over 45 days. The mutation allele frequency was analyzed at each indicated time points using digital droplet PCR. This experiment was done once.

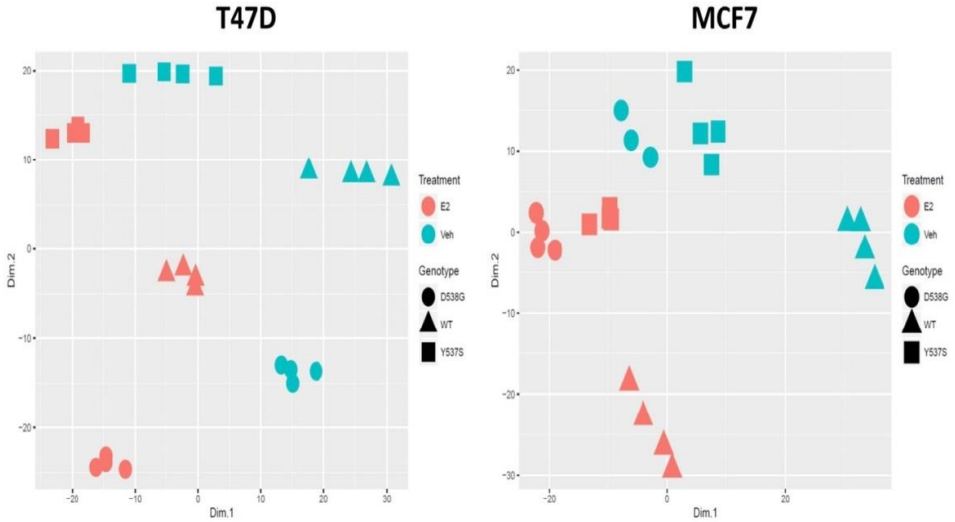



**Figure 70. PCA analysis of 1000 top variable genes between WT and mutants.**

The top 1000 most variable genes were selected based on interquartile range. The PCA analysis was performed and plotted using PCA function in R.
*This data was provided by Amir Bahreini, PhD.

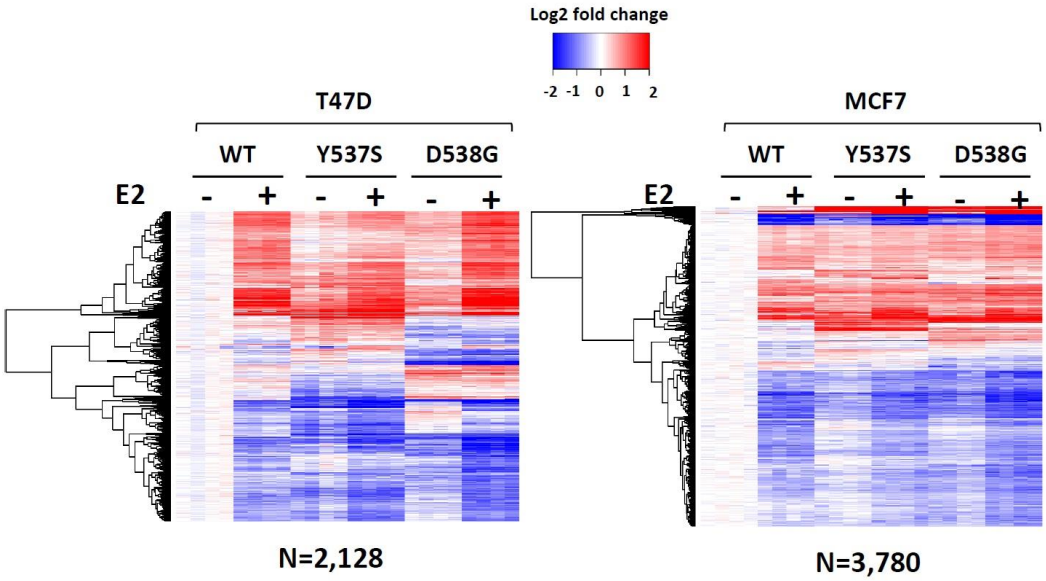



**Figure 71. Heatmap of variable genes in MCF7 and T47D *ESR1* mutant cells.**

Heatmap of variable genes (ANOVA test=p<0.0005, max FC>2) in mutants and WT cells. Gene expression TPM was estimated by Salmon package. ANOVA test was then used to identify differentially expressed genes between the samples. Genes with a p<0.0005 and fold change>2 that were differentially regulated in at least one mutant vs WT-veh were selected for this heatmap. *This data was provided by Amir Bahreini, PhD.

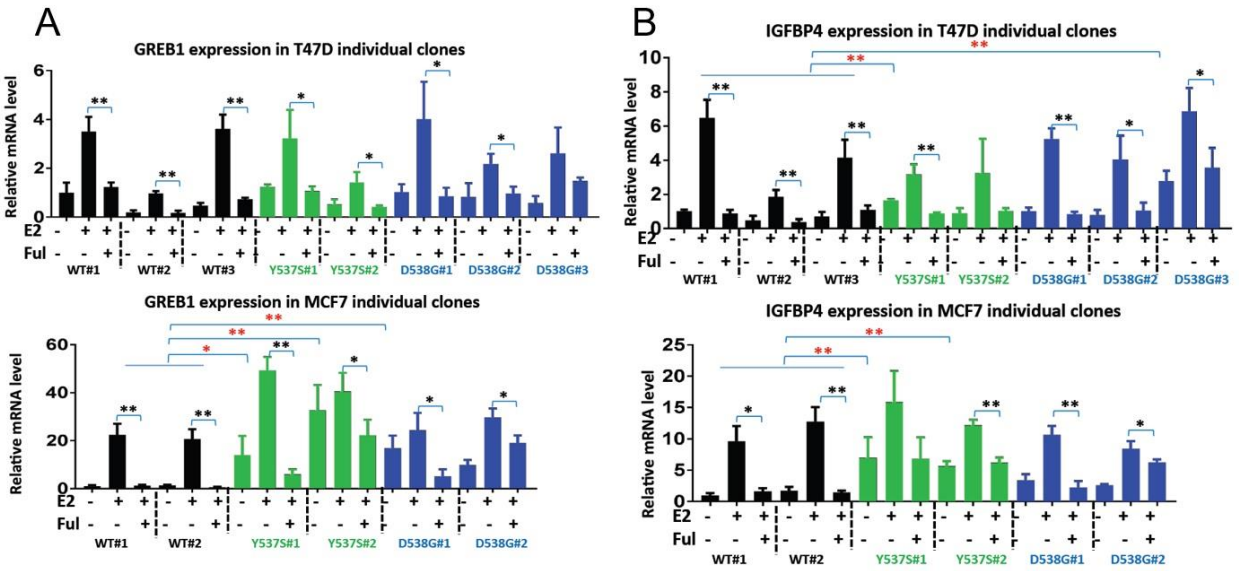



**Figure 72. mRNA expression of GREB1 and IGFBP4 in individual cloneso of T47D and MCF7 *ESR1* mutant cells.**

The post-hormone-deprived MCF7 or T47D cells (pooled or individual clones) were treated with 1 nM of E2 combined with or without 1 µM of Ful for 24 hours. RT-qPCR was done using GREB1 (A) or IGFBP4 (B) primers. All experiments were performed in three biological replicates. Dunnett's test was performed between the basal expressional levels in each mutant clones and the average expression of GREB1 and IGFBP4 in the WT clones (* $p<0.05$, **$p<0.01$, red) and student t-test was used to compare the response before and acer fulvestrant treatment. This experiment was repeated twice and data from representative experiment is shown here. ( * $p<0.05$, **$p<0.01$, black).

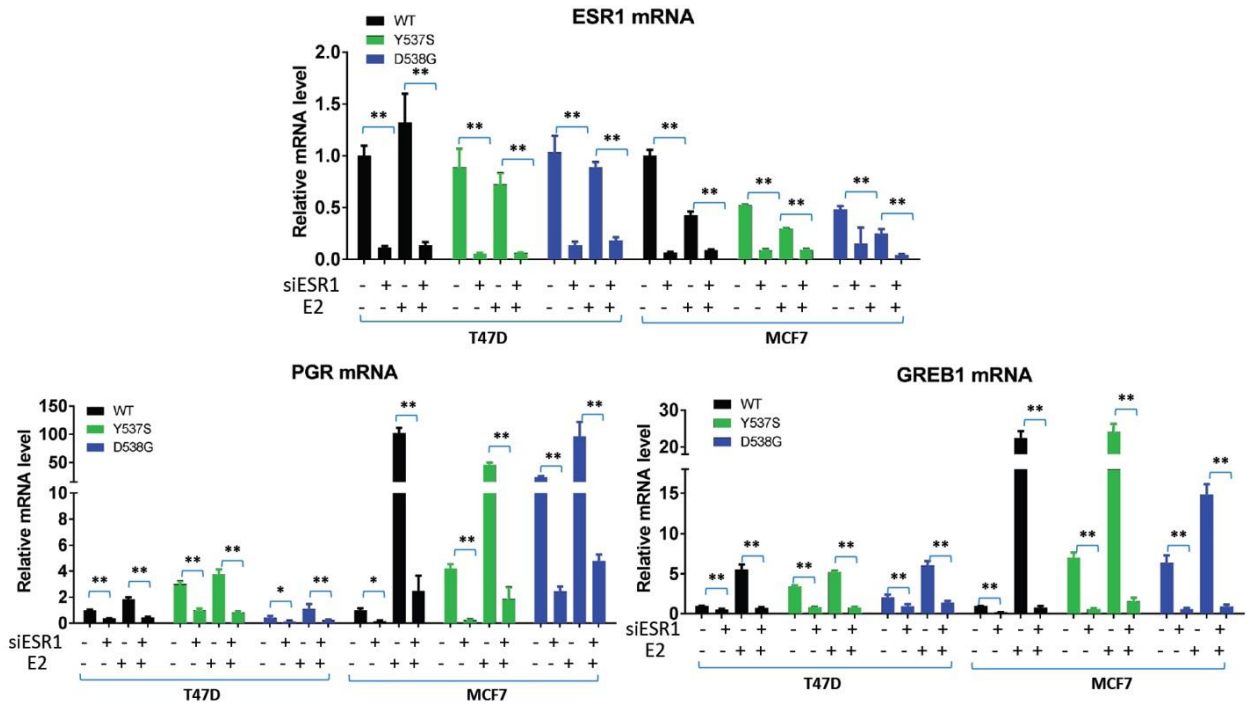



**Figure 73. mRNA expression of *ESR1*, *PGR* and GREB1 in *ESR1* mutant cells with or without *ESR1* transient knockdown.**

The post-hormone-deprived MCF7 or T47D cells (pooled or individual clones) were transfected with scramble siRNA or *ESR1* siRNA for 24 hours, and then treated with or without 1 nM of E2 for 24 hours. RT-qPCR was done using *ESR1*, *PGR* or IGFBP4 primers. All experiments were performed in three biological replicates. This experiment was done once. (Dunnett's test, * $p<0.05$; ** $p<0.01$)

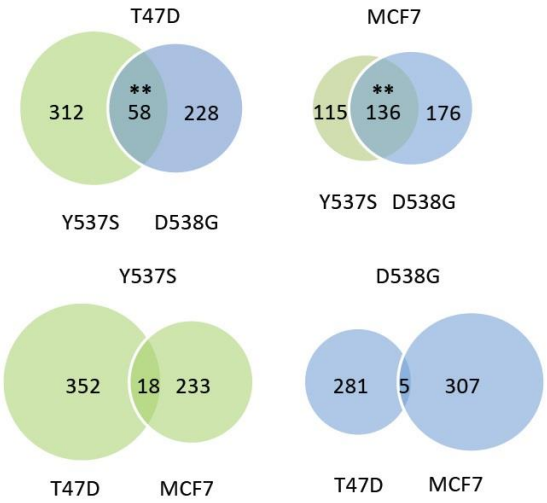



**Figure 74. Overlap of novel target genes between cell lines and mutation subtypes.**

Venn diagram overlap of novel ligand independent regulated genes of the *ESR1* mutations within one cell line (top panel) and between the cell lines (bottom panel) (Chi-square test, **p<0.01). *This data was provided by Amir Bahreini, PhD.

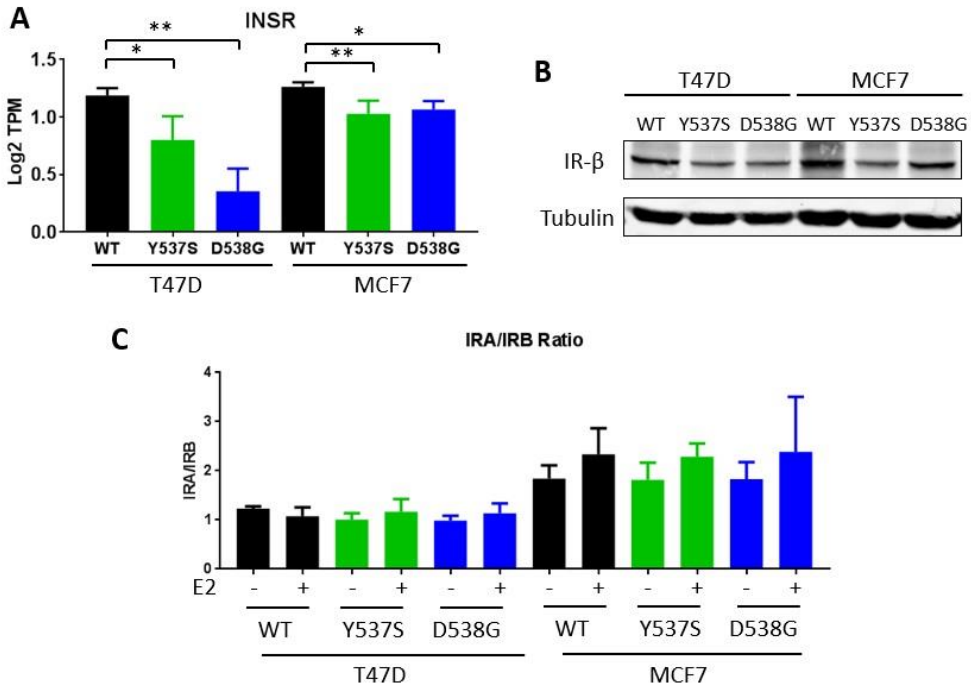



**Figure 75. Insulin receptor expression is decreased in T47D and MCF7 *ESR1* mutant cells.**

A and B. Insulin receptor levels are shown by RNA-seq (A) and western blot (B). MCF7 and T47D cells were hormone deprived. Protein were isolated and immune-blots were performed with the corresponding antibodies, with tubulin as the internal control. C. The ratio of IR-A/IR-B were calculated based on the TPM values from transcript reads of RNA-seq results. Dunnett's test was performed to compare the RNA levels of INSR between different mutants of each cell line. This experiment was done once. (* $p<0.05$; ** $p<0.01$)

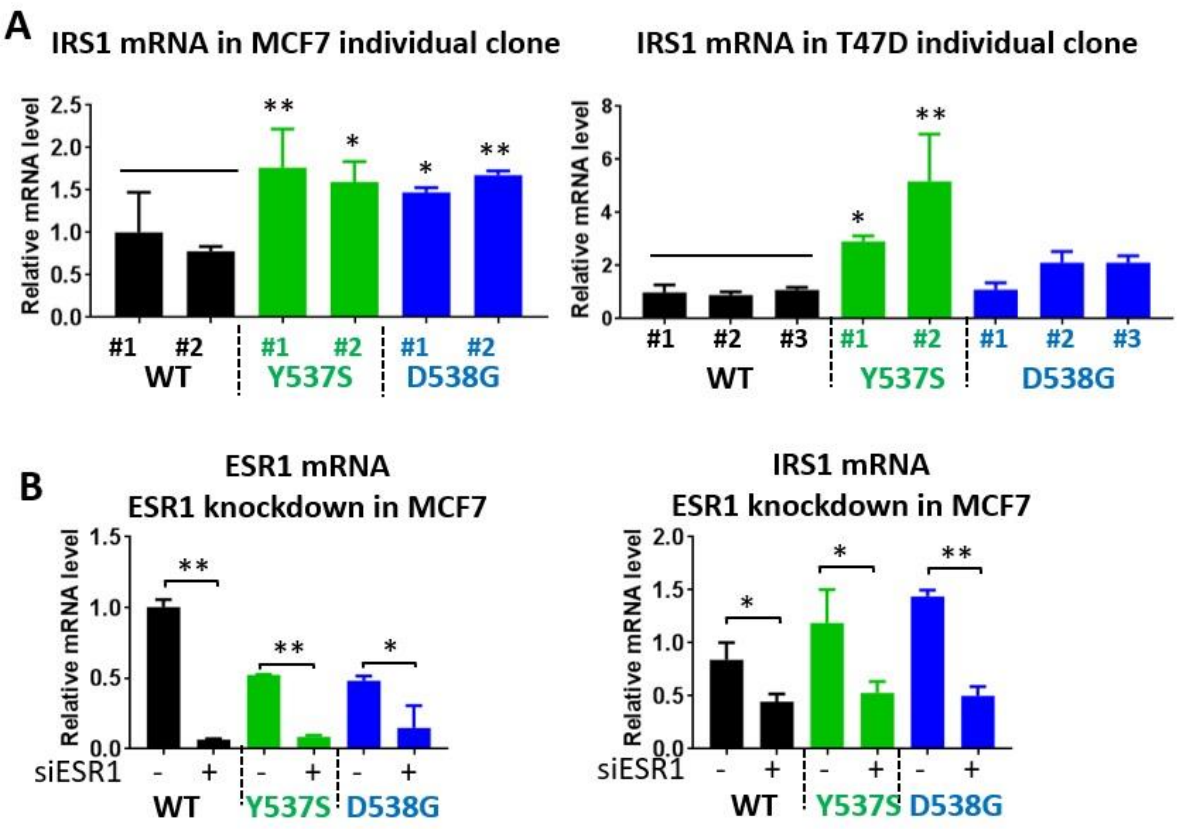
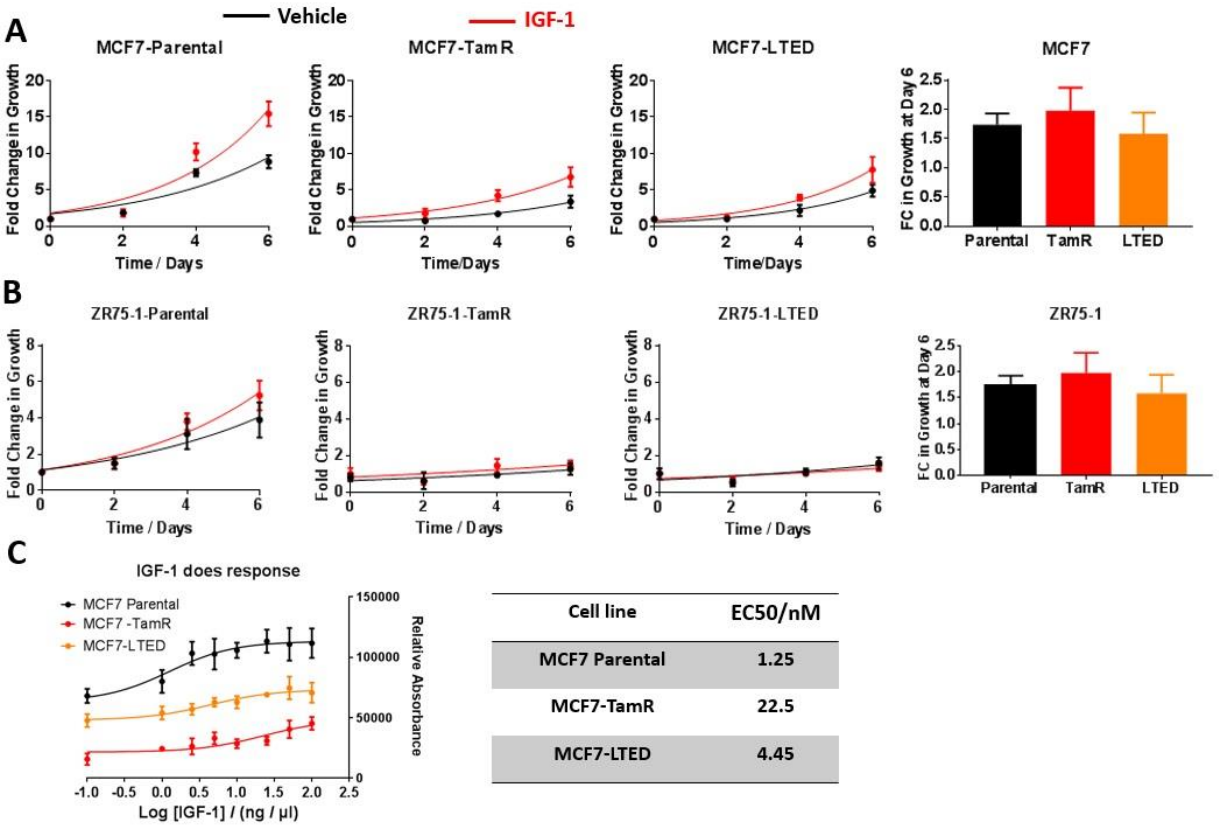



**Figure 76. IRS is upregulated in *ESR1* mutant cells and dependent on ER.**

A. RNA samples from hormone-deprived MCF7 and T47D clones were isolated individually, and qRT-PCR was performed (Dunnett's test for comparison of basal level * $p<0.05$, ** $p<0.01$). B. *ESR1* specific- or scramble siRNAs were reversely transfected into hormone-deprived MCF7 cells for 24 hours. RNA was isolated and qRT-PCR was performed with specific primers towards *ESR1* and IRS1. This experiment was done once. (Student's t-test for comparison mRNA levels in scramble and knockdown groups * $p<0.05$, ** $p<0.01$)




| Cell line | EC50/nM |
|---|---|
| MCF7 Parental | 1.25 |
| MCF7-TamR | 22.5 |
| MCF7-LTED | 4.45 |

**Figure 77. Growth response to IGF1 stimulation in tamoxifen resistant (TamR) and long-term estorgen deprivation (LTED) cell models.**

A & B. MCF7 and ZR75-1 parental, TamR or LTED cells were starved in serum-free medium for 24 hours. Cells were then seeded into 96-well plates and treated with or without 100 ng/ml of IGF-1 for up to 6 days. Cell numbers were quantified by FluoResporter kit. Growth fold change was normalized to day 0, and the fold changes at day6 were plotted on the right. C. EC50 values were determined by IGF-1 dose response growth curves following the identical procedure in A with various doses of IGF-1 and 20 pM of E2. This experiment was done once.

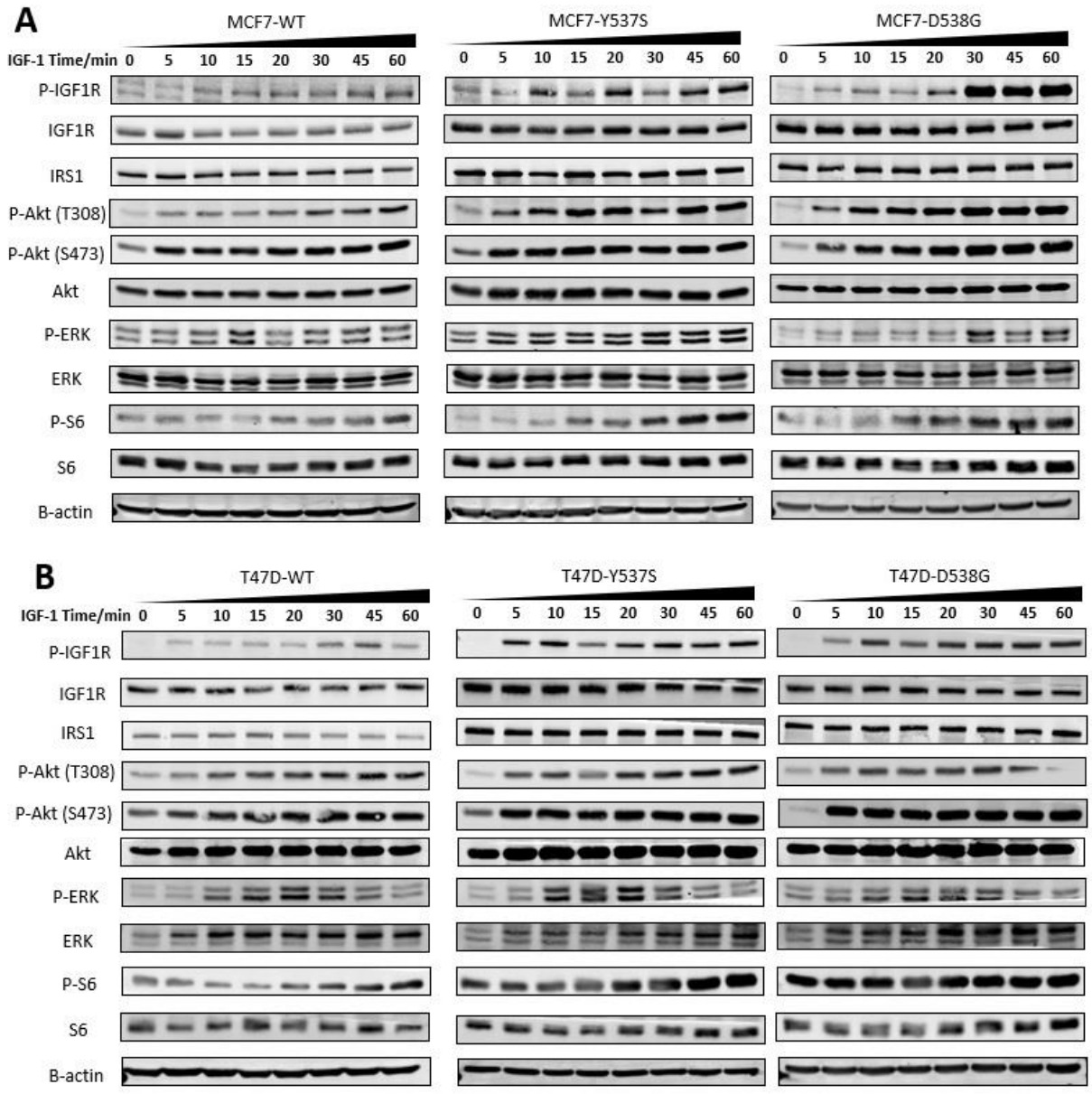



**Figure 78. Time course response of IGF-1R downstream signals in T47D and MCF7 *ESR1* mutant cells.**

A and B. MCF7 and T47D cells were hormone deprived and starved in serum-free medium for 24 hours. Cells were then stimulated with 100 ng/ml of IGF-1 for the corresponding time points. Protein samples were isolated and immune-blots were performed with the corresponding antibodies, with β-actin as the internal control. This experiment was done once.

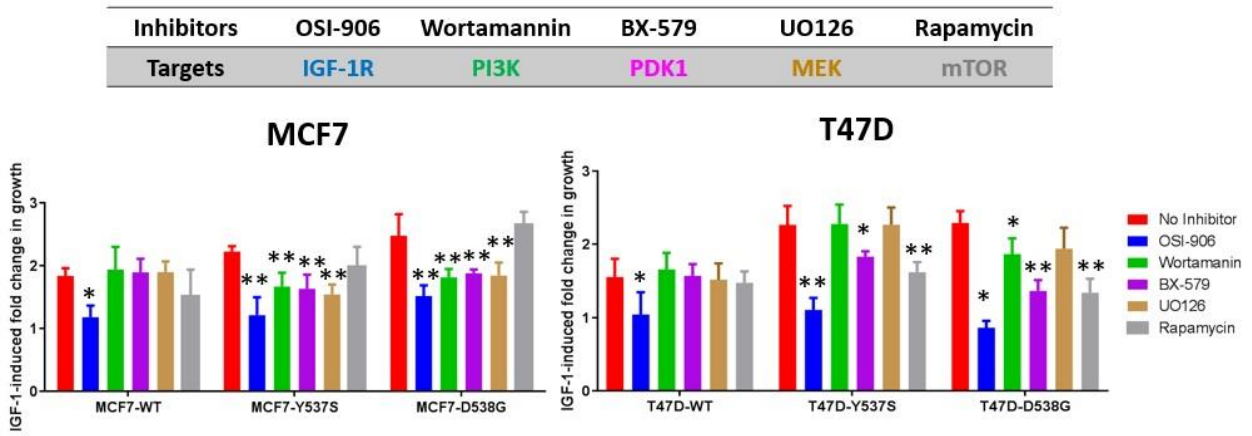



**Figure 79. IGF-1-induced cell growth with samll molecular inhibitors to various essential IGF-1R signaling targets.**

MCF7 cells and T47D cells were hormone deprived and starved in serum-free medium for 24 hours. Cells were then seeded into 96-well plates and treated with 100 ng/ml of IGF-1 as well as 100 nM of each inhibitor for 6 days. Growth fold changes at day6 were normalized to day 0. Dunnett's test ANOVA was performed to compare the growth rates between each inhibitor. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

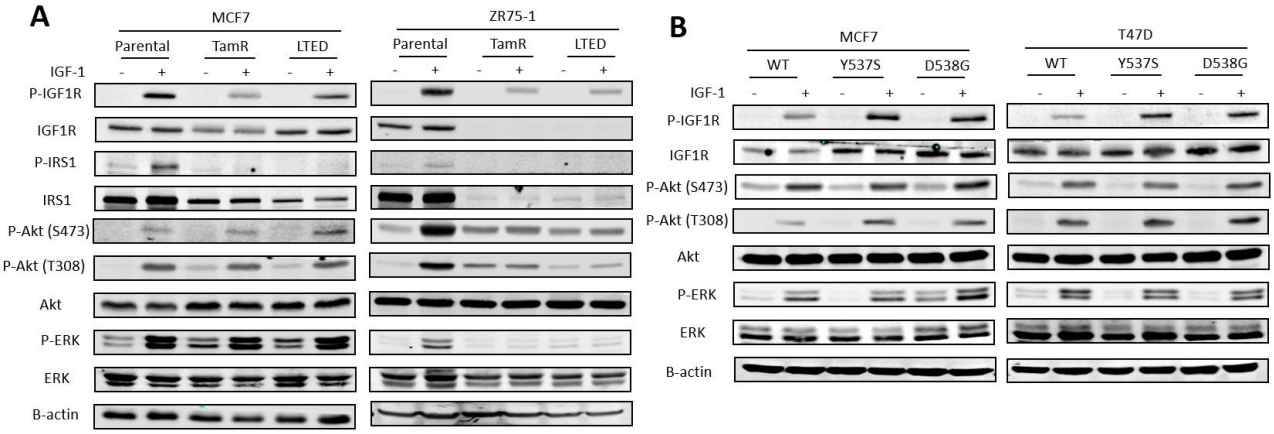
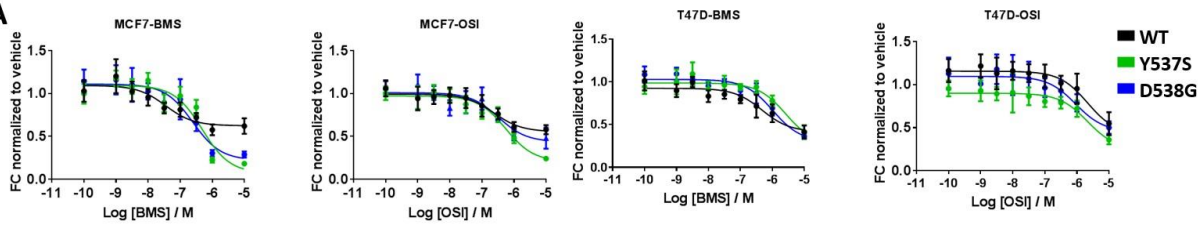



**Figure 80. IGF-1R signaling response in different endocrine resistance cell models.**

A and B. Cells with (*ESR1* mutant cells) or without (parental, TamR and LTED) hormone deprived and starved in serum-free medium for 24 hours. Cells were then stimulated with 100 ng/ml of IGF-1 for the 15 minutes. Protein samples were isolated and immune-blots were performed with the corresponding antibodies, with β-actin as the internal control. This experiment was done once.

A




B

| IC50/nM | BMS | OSI |
|---|---|---|
| MCF7-WT | 14.74 | 73.15 |
| MCF7-Y537S | 234.9 | 184.5 |
| MCF7-D538G | 91.9 | 117.8 |

| IC50/µM | BMS | OSI |
|---|---|---|
| T47D-WT | 0.4 | 2.79 |
| T47D-Y537S | 2.61 | 1.78 |
| T47D-D538G | 1.08 | 1.04 |

**Figure 81. Dose response curves and IC50 of IGF1R inhibitors in *ESR1* mutant cell models.**

Dose response growth assays were performed on hormone-deprived MCF7 and T47D cells with various doses of BMS and OSI-906 for 6 days under 20 pM of E2. IC50 values were calculated in PRSIM. This experiment was done once.

## Appendix B Supplementary Figures for Chapter Three

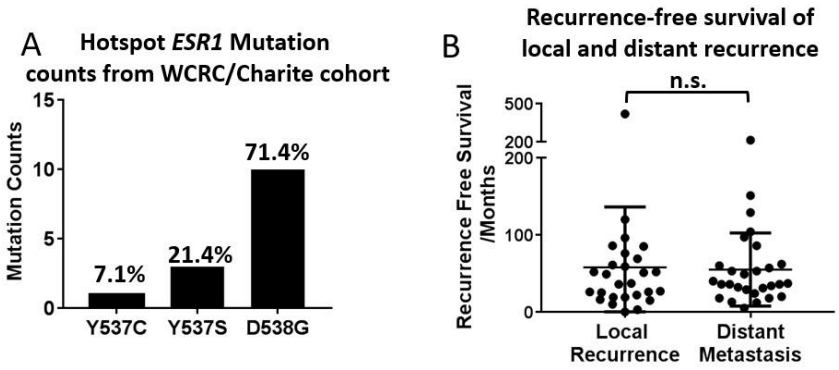



**Figure 82. Recurrence-free survival and mutation subtype distribution from WCRC/Charite cohort.**

A. Mutation subtype distribution for all detected mutations from WCRC/ Charite cohort. B. Comparison of recurrence-free survival lengths between part of distant metastatic (n=28) and local recurrent samples (n=27). Patients with RFS=0 were excluded in this analysis. Mann-Whitney U test was performed to compare the RFS between local recurrent and distant metastatic samples.

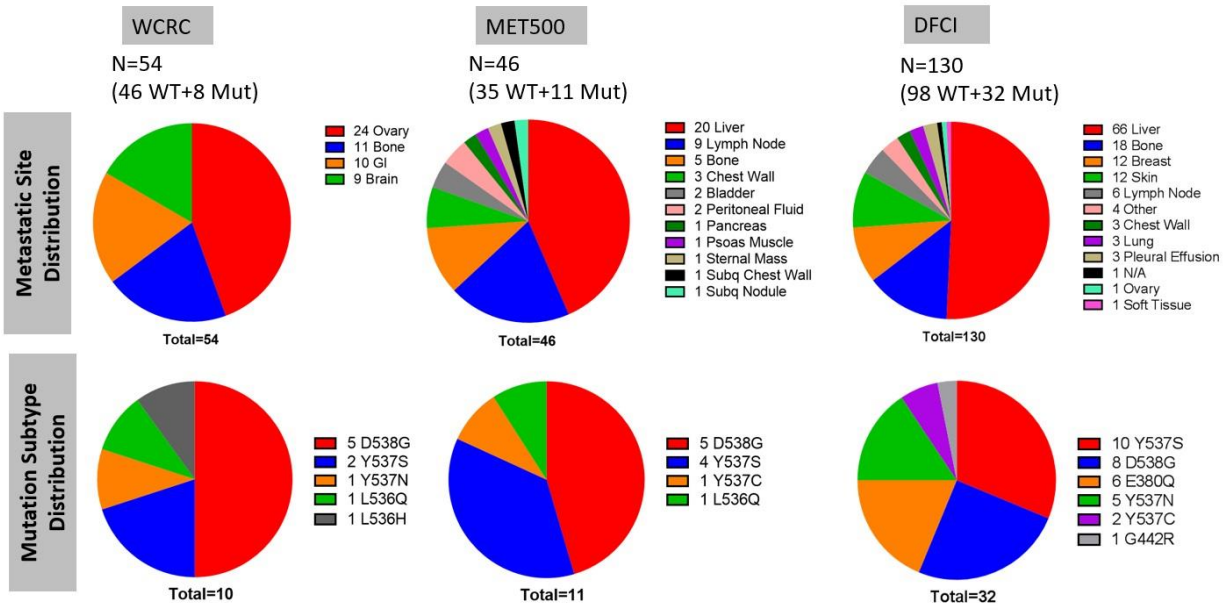



**Figure 83. Metastatic site and mutation subtype distribution of the three ER+ metastatic breast cancer cohorts.**

Pie charts represent the metastatic sites and mutation subtype distribution of each cohort. Metastatic sites were obtained from clinical records. Hotspot *ESR1* mutations were determined by ddPCR (WCRC), RNA-seq variants calling (WCRC) and whole exon sequencing (MET500/DFCI). For the latter two cohorts, only *ESR1* mutations that have been reported with experimental validations were selected here.

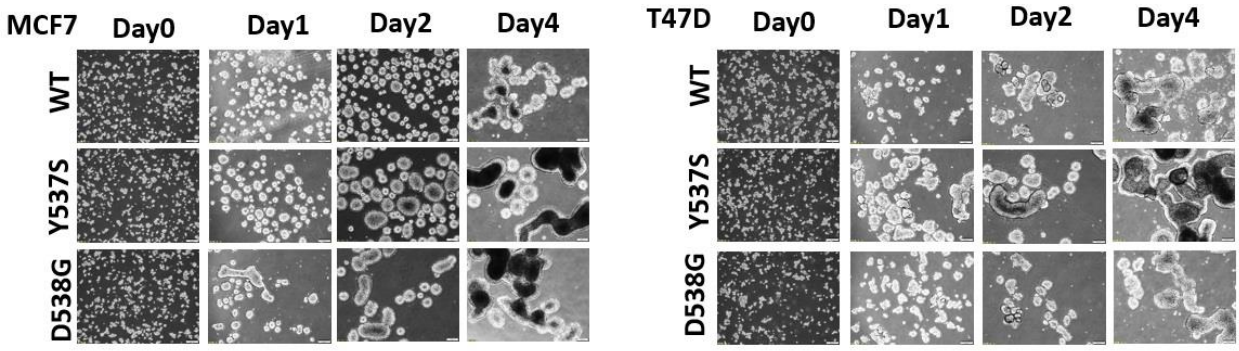



**Figure 84. *ESR1* mutant-cells form tighter clusters under suspension condition.**

Representative images of MCF7 and T47D spheroids after seeded into 6-well round bottom ULA plates at day 0, 1, 2 and 4. The objective of 4X were used. This experiment was repeated twice, data from representative experiment is shown here.

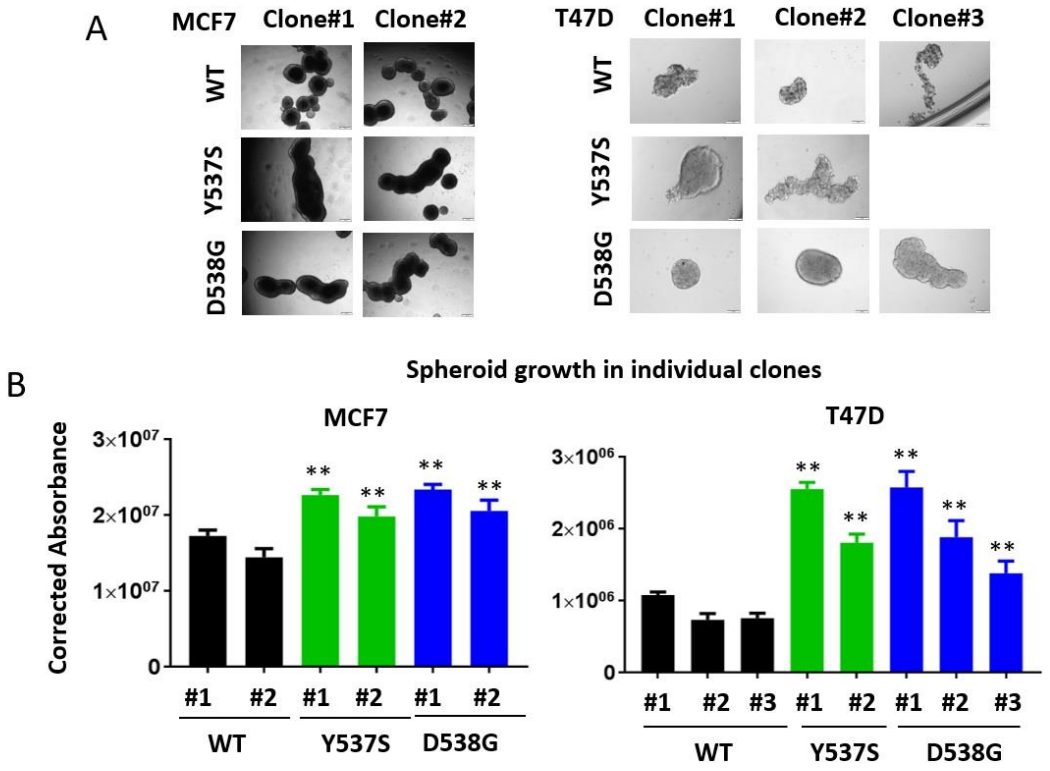



**Figure 85. *ESR1* mutant-cells exibit ligand-independent growth under suspension condition.**

A. Representative images of MCF7 (2 clones for each cell type) and T47D (3 clones for WT and D538G, 2 clones for Y537S) individual clone spheroids after seeded 3000 (MCF7) or 4000 (T47D) cells into 96-well ULA plates initially and let them grow for 6 days. The images were captured under 10X objectives. This experiment was done once. B. Individual clones of MCF7 or T47D WT/*ESR1* mutant cells were seeded into 96-well ULA plate with 3000 or 4000 initial cell densities respectively. Cell numbers were quantified at Day7 using Celltiter Glo. Fluorescence readouts were corrected by background measurements. Dunnett's test was used to between each mutant clone and the mean of WT clones. This experiment was done once. (* $p<0.05$; ** $p<0.01$)

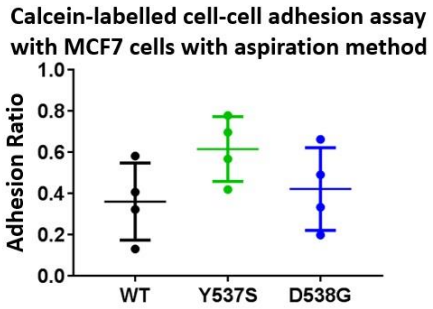



**Figure 86. MCF7 *ESR1* mutant cells show stronger static cell-cell interaction via aspiration method in the calcein-labelled cell-cell adhesion assay.**

Calcein labelled cell-cell adhesion assay on MCF7 cells was repeated using vacuum aspiration method for the washing steps. Adhesion ratios after three times of wash were extracted from all four independent experiments, and Dunnett's test was used to compare between WT and each mutant.

*This experiment was repeated by Dr. Jennifer Xavier.

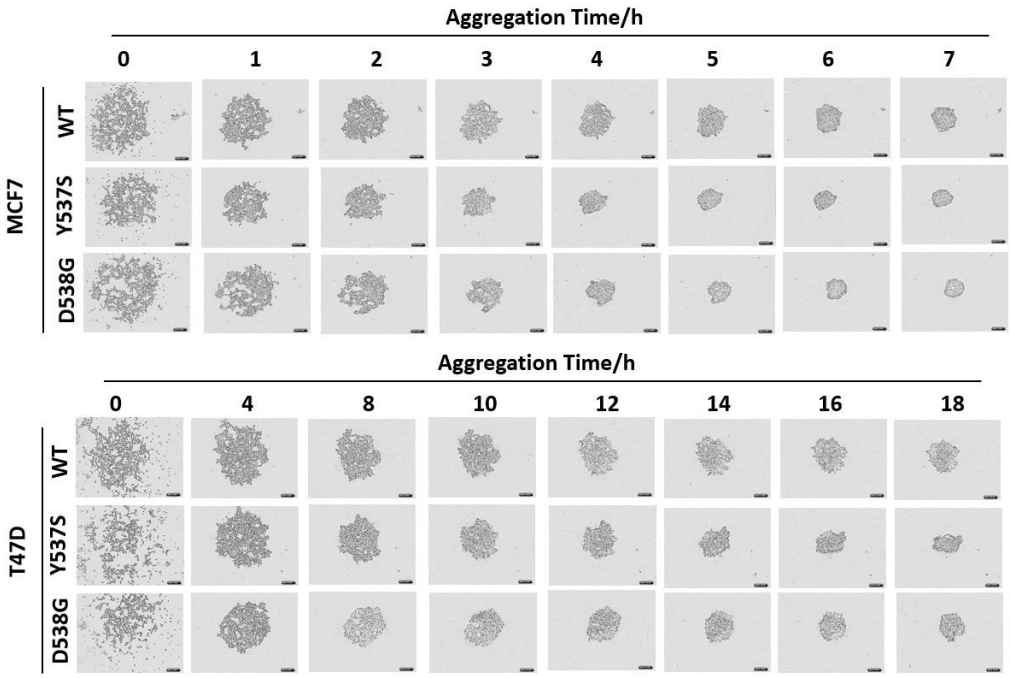



**Figure 87. Images of spontaneous cell aggregation assay with *ESR1* mutant cell models.**

Representative images of IncuCyte cell aggregation assay on MCF7 (Top panel, 0-7 hours) and T47D (Lower panel, 0-18 hours) WT/mutant cells in a time course. Images were taken with objectives of 10X.

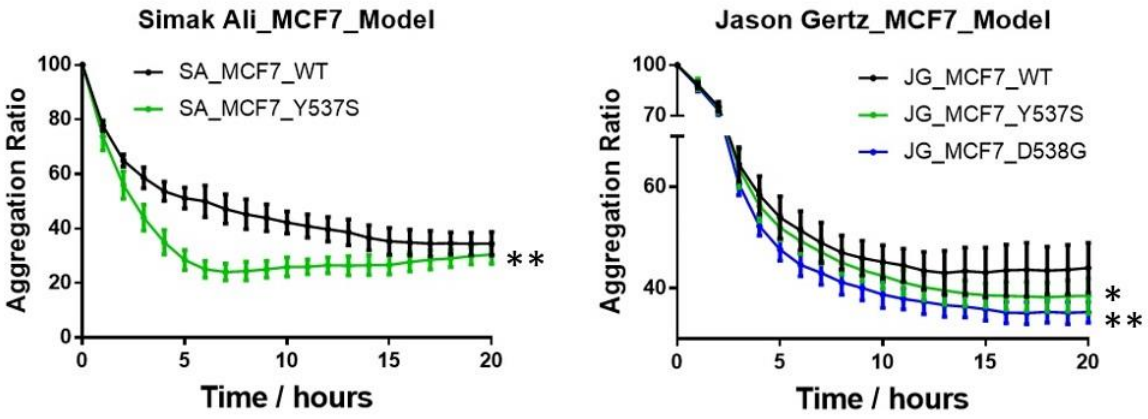



**Figure 88. Two additional MCF7 *ESR1* mutant cell models show faster spontaneous aggregation in suspention condition.**

MCF7 CRISPR cells derived from Simak Ali and Jason Gertz's groups were seeded into round bottom ULA plates and the cell aggregation process followed by IncuCyte living imaging system every 1 hour. Spheroid areas were normalized to time 0. Representative images after 20 hours of aggregation are shown in panel A pairwise Two-way ANOVA between WT and each mutant was utilized. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

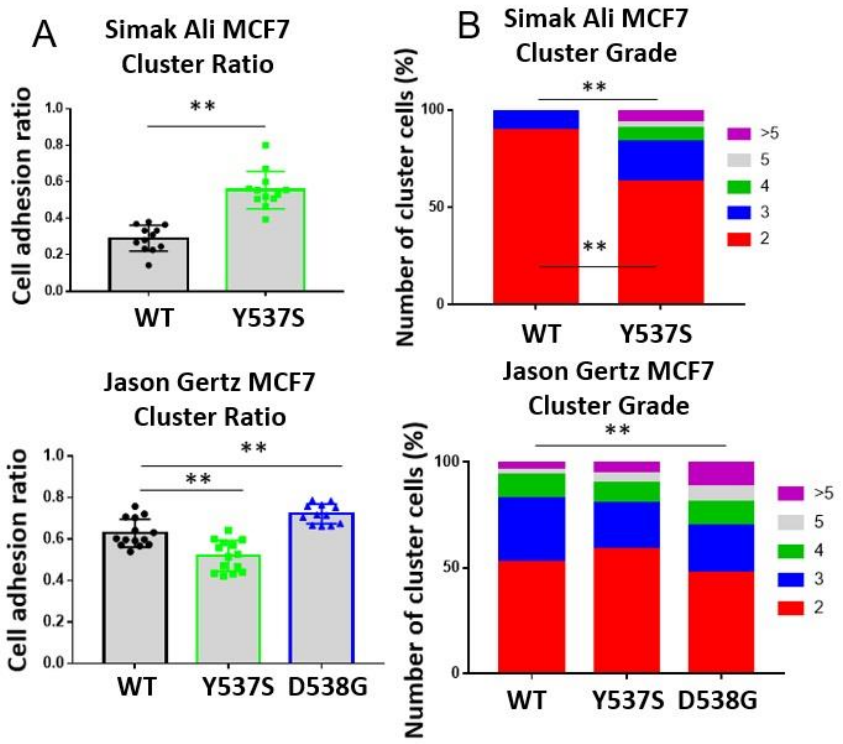



**Figure 89. Two additional MCF7 *ESR1* mutant cell models show enhanced cell-cell adhesion in microfludic conditions.**

A. Bar graph representing the percentage of cells in cluster version dividing by the total cell numbers of cluster and single cells in Jason Gertz and Simak Ali MCF7 *ESR1* mutant cell models.. Dunnett's test was used between WT and mutant cells. D. Stack bar chart showing the cluster grade distribution. Fisher's exact test was used between WT and each mutant cell type under the same cluster grade. This experiment was done once. (* $p<0.05$, ** $p<0.01$).

*These data were provided by Yang Wu.

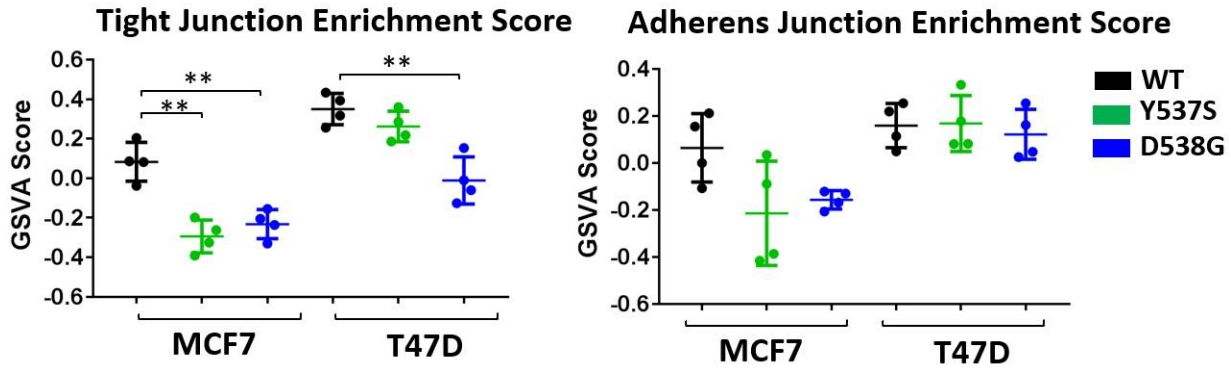



**Figure 90. Enrichment of tight junction and adhenrens junction gene sets in MCF7 and T47D *ESR1* mutant cells.**

Gene Set Variation Analysis (GSVA) between MCF7 and T47D mutant and WT transcriptomes from RNA-seq on tight junction and adherens junction gene sets. Each cell type has four biological replicates. Dunnett's test was used to test the significance between cell line. (* $p<0.05$, ** $p<0.01$)

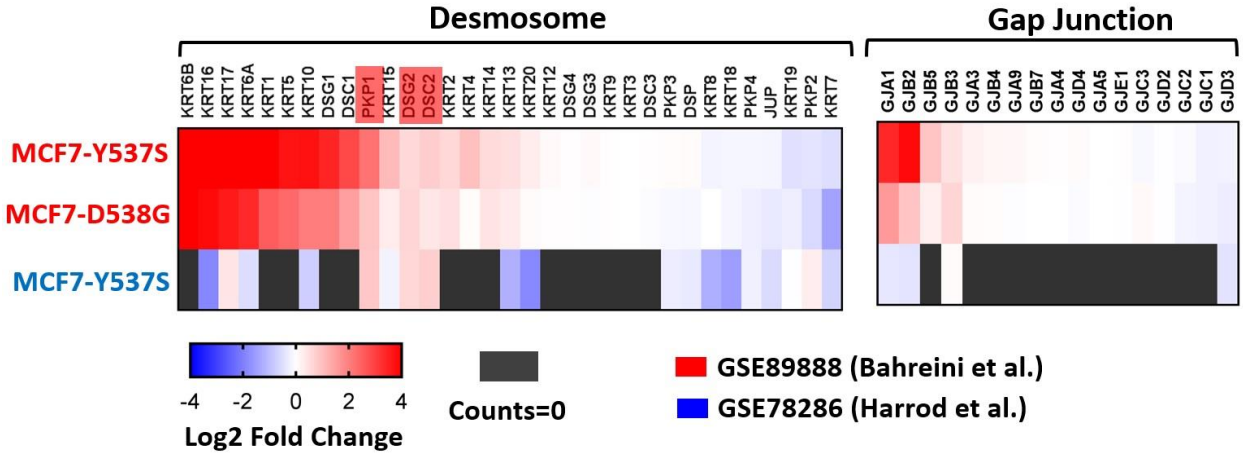



**Figure 91. Expression of desmosme and gap junction genes in another CRISPR MCF7 cell model.**

Comparison of desmosome and gap junction gene expression between another genome-edited *ESR1* mutant cell RNA-seq data sets. Genes with counts=0 in more than one replicate in all cell types were filtered out. Consistent upregulated genes across different cell models were labelled in red.

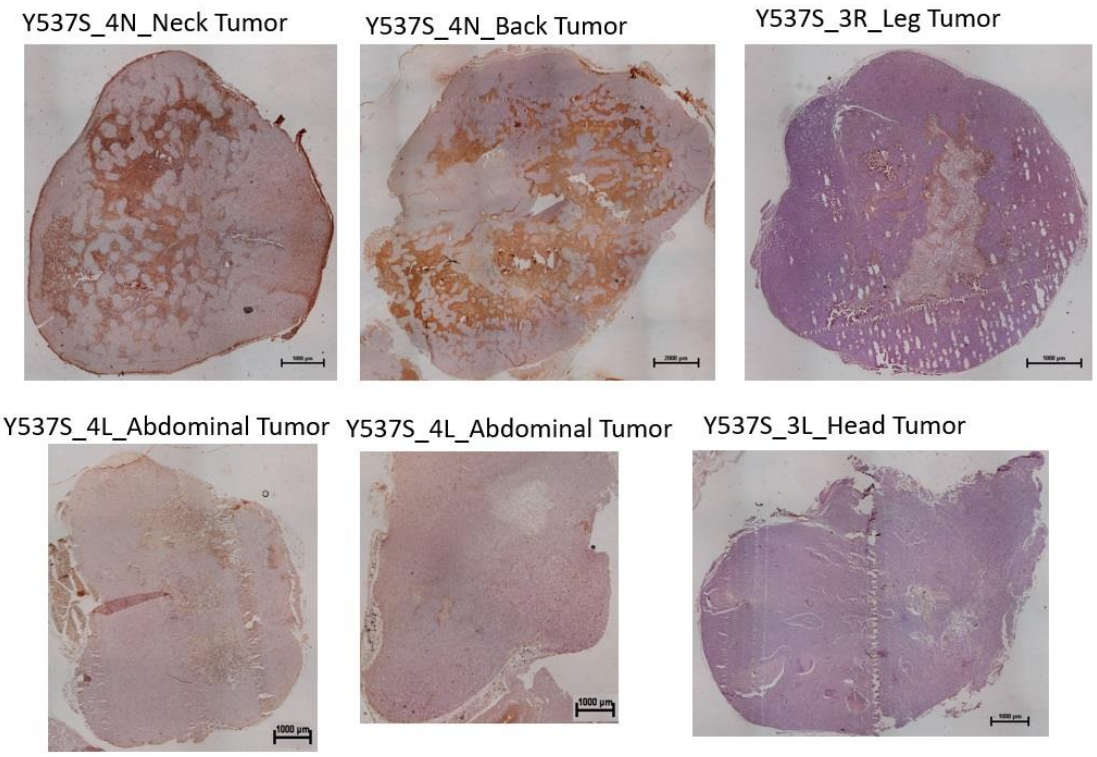



**Figure 92. IHC staining for Connexin 43 in MCF7 Y537S *ESR1* mutant cells-derived metastatic tumors.**

Whole section scanning images of Connexin 43 IHC staining in six MCF7-Y537S cell derived macro-metastatic tumors. This experiment was done once.

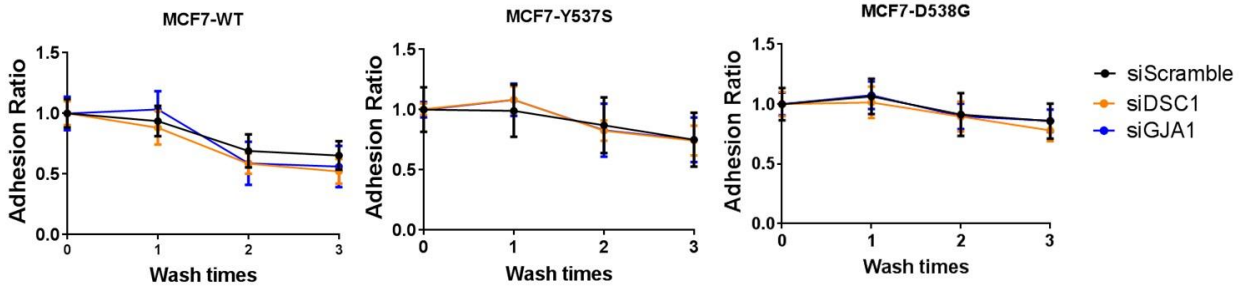



**Figure 93. Kncokdown of *DSC1* or *GJA1* does not rescue the stronger cell-cell adhesion in MCF7 *ESR1* mutant cells.**

siScramble or siRNA towards *DSC1* or *GJA1* were transfected into MCF7 WT/mutant cells. Calcein-labelled cell-cell adhesion assay was performed. A pairwise Two-way ANOVA between scramble control and knockdown groups were utilized in each cell type. This experiment was done once.

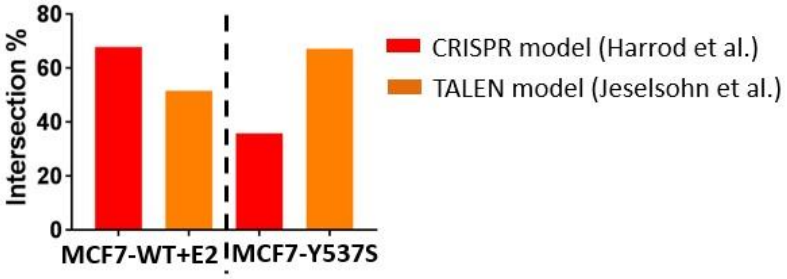



**Figure 94. Intersection of ChIP-seq peaks to other genome-edited cell models**

Intersection ratios of ER binding peaks from the present study with public available ChIP-seq data sets from other genome-edited *ESR1* mutant cell models.

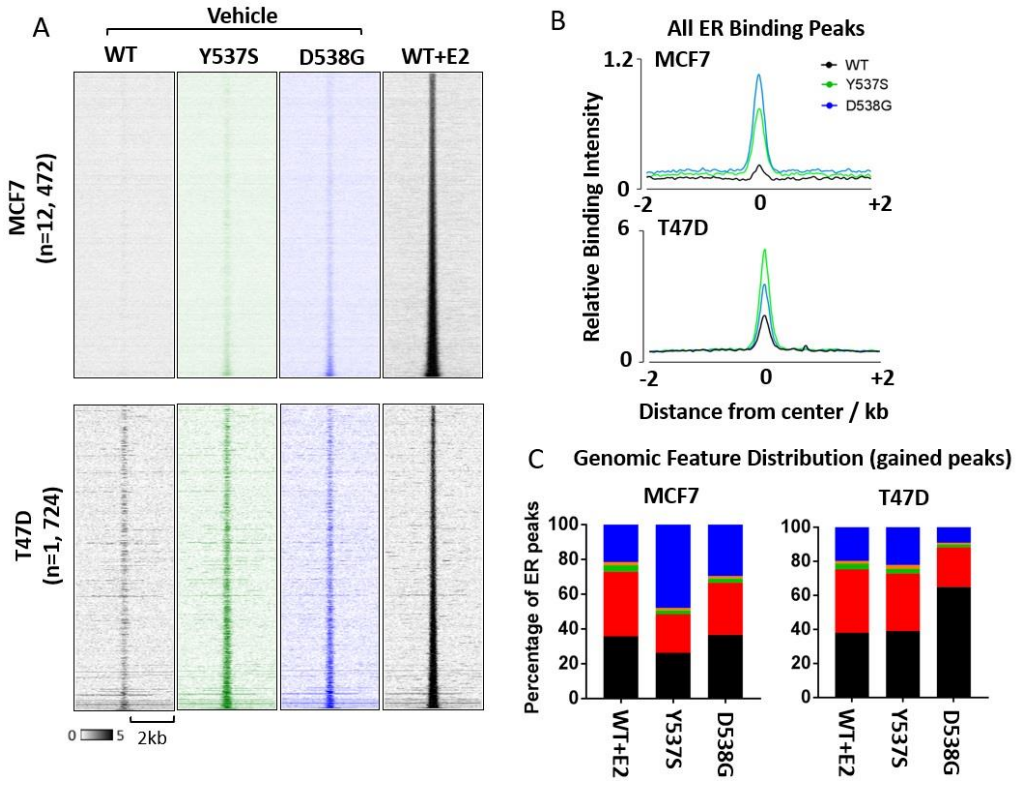



**Figure 95. Mutated ER cistromes depend on genomic contexts.**

A. Heatmaps showed the global ER binding distributions and intensities (normalized read counts) of WT, Y537S, D538G vehicle and WT-E2 groups on the regions of WT-E2 regulated sites, shown in a horizontal window of ± 2kb from the peak center. B. Average binding intensities towards all the binding sites from WT, Y537S and D538G in the absence of E2 of each cell lines in a window of ± 2kb from the peak center. Binding intensities are normalized to the WT-E2 region sets. C. Genomic feature distribution patterns of gained ER binding sites corrected to WT-vehicle groups in in all cell types. The promoter regions are defined as +/- 3 kb around the transcriptional start sites (TSS).

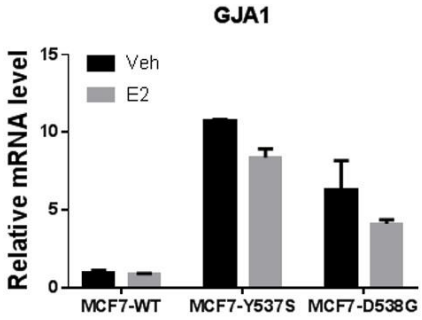



**Figure 96. Estrogen treament does not altere *GJA1* exporession in MCF7 WT cells.**

Hormone deprived MCF7 cells were treated with 1 nM E2 for 24 hours. RNA was isolated and subjected for reverse transcription and qRT-PCR validation with the primers for *GJA1*. Students t-test was used to compare the effects of E2 stimulation on each cell type. This experiment was repeated twice, data from one representative experiment is shown here.

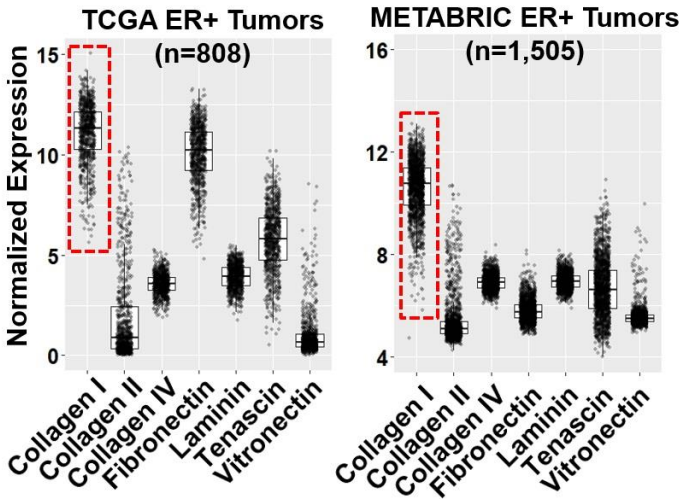



**Figure 97. Collagen I is the most enriched ECM component in ER+ primary tumors.**

Boxplots with individual values of the abundance of each individual ECM component from ER+ tumors from TCGA and METABRIC cohorts. The abundance calculations were based on the normalized gene (s) expression with log2TPM (TCGA) or probe intensities (METABRIC). Collagen I was highlighted as the most abundant component.

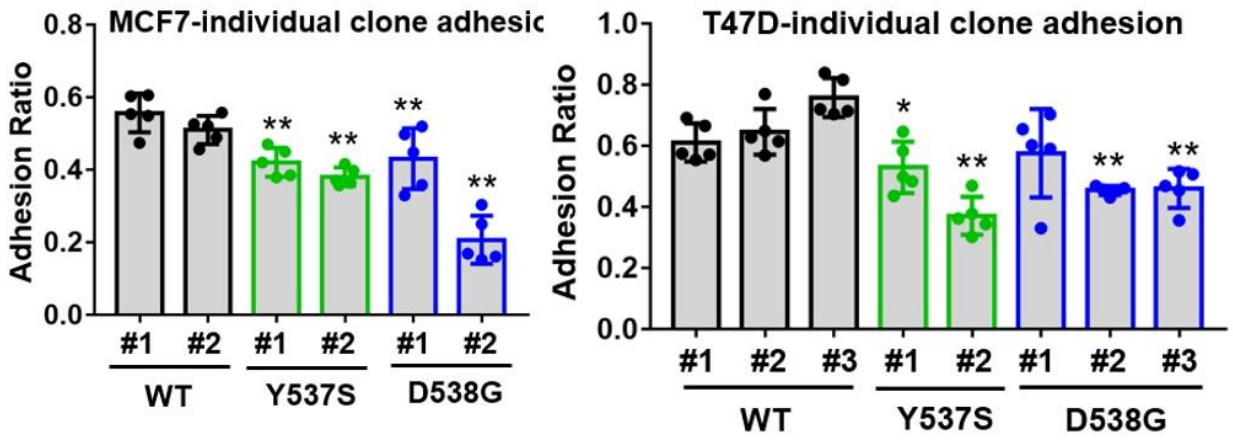



**Figure 98. *ESR1* mutant cells show diminished adhesion on collagen I in each individual clone.**

Quantification of adhesion ratios on collagen I of each cell type in the individual clone manner of MCF7 (Left) and T47D (Right) cell lines. Bar graphs represent the mean ± SD with five biological replicates in each group. Dunnett's test was utilized within each cell line to compare between the average of WT clones to each single mutant clone. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

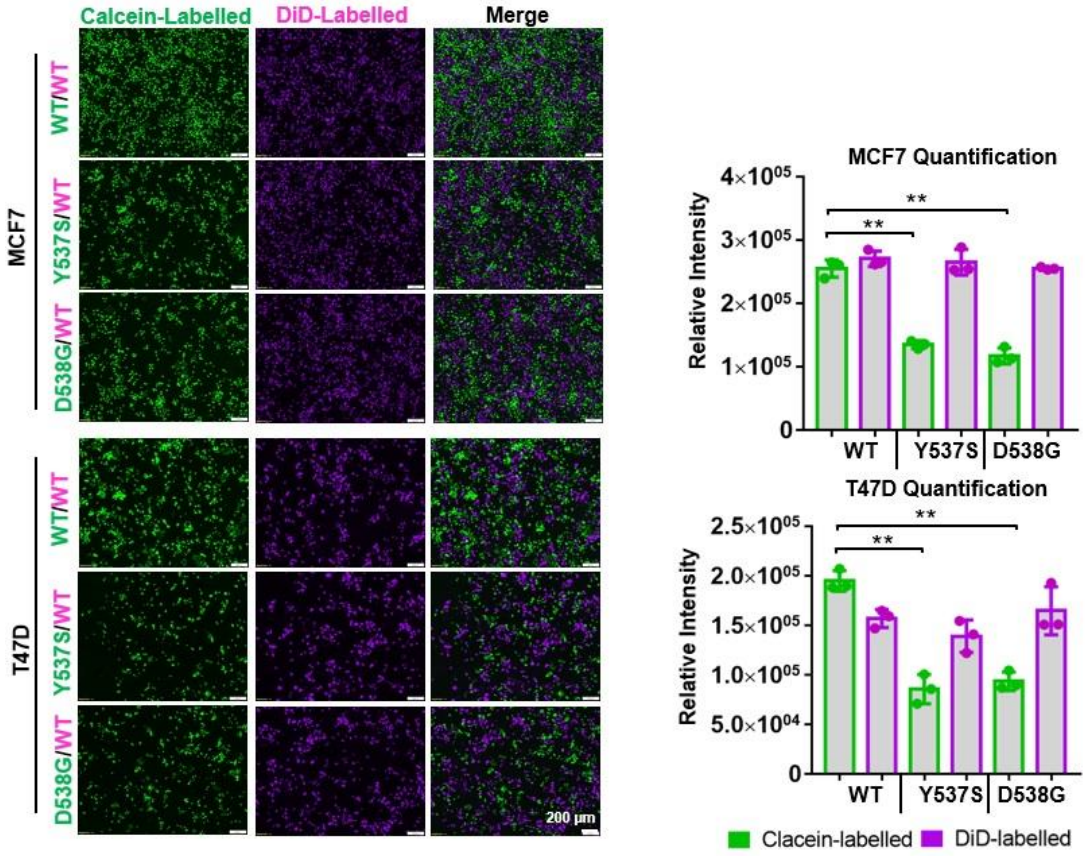



**Figure 99. *ESR1* mutant cells show attenuated adhesive properties on collagen I in a co-culture system.**

Representative images of co-culture adhesion assay in collagen I of MCF7 and T47D cell lines. WT cells with DiD (pink) labelling were equally mixed with calcein (green) labelled WT/Y537S/D538G cells. The mixed cells were subjected for adhesion assay, and images of ether single GFP/Cy5 channel or merged were shown. Right panel: Quantification of GFP or Cy5 single channel signal intensities from the left panel. Bar graphs represent the mean ± SD with three biological replicates in each group. Dunnett's test was used to compare the signal intensities between WT and each mutant cell within each individual channel. This experiment was done once for T47D and twice for MCF7. (* $p<0.05$, ** $p<0.01$)

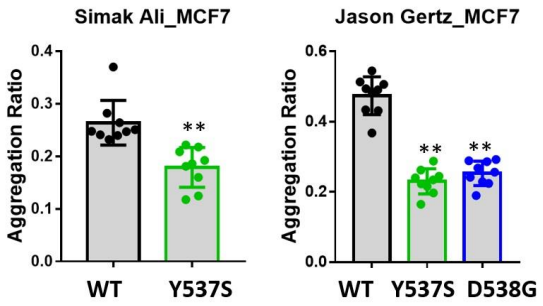



**Figure 100. Two additional MCF7 *ESR1* mutant cell models show diminished adhesive proterties on collagen I.**

Quantification of adhesion ratios on collagen I of each cell type. Bar graphs represent the mean ± SD with at least four biological replicates in each group. Dunnett's test was utilized within each cell line. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

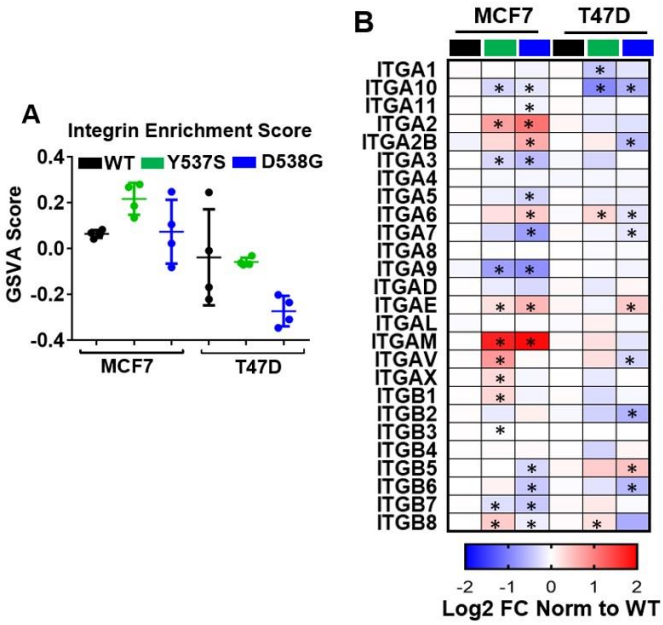



**Figure 101. Integrin gene signature is not enriched in *ESR1* mutant cell models.**

A. Graphic view of GSVA enrichment of Integrin gene sets in the transcriptome of *ESR1* WT and mutant cell models. Dunnett's test was used within each cell line. B. Heatmap view of integrin family genes in RNA-seq results of MCF7 and T47D *ESR1* mutant cells. Log2TPM of each gene in mutant cells was normalized to WT cells within each cell line. Pair-wised student's t test was performed following false discovery ratio correction. Significantly altered genes were defined as FDR<0.1 and labelled with asterisk signs.

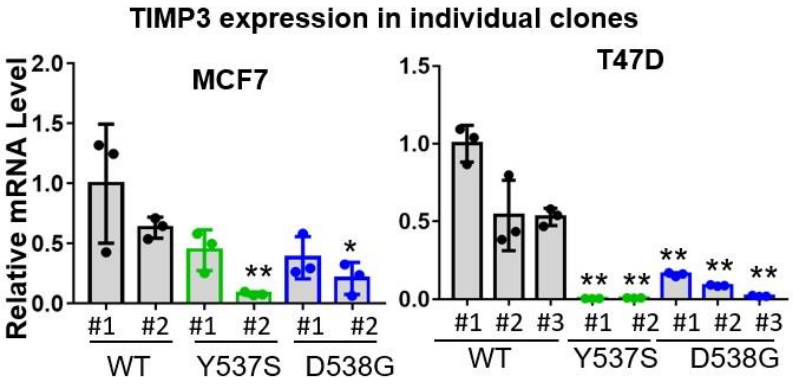



**Figure 102. *TIMP3* expression is decreased consistently in MCF7 and T47D *ESR1* mutant cells in the individual clone manner.**

qRT-PCR validation of *TIMP3* expression in WT and *ESR1* mutant cells in the individual clone manner. Ct values were normalized to RPLP0 and further normalized to WT#1 clone. Bar graphs represent the mean ± SD with biological triplicates in each group. Dunnett's test was utilized within each cell line to compare between the average of WT clones to each single mutant clone. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

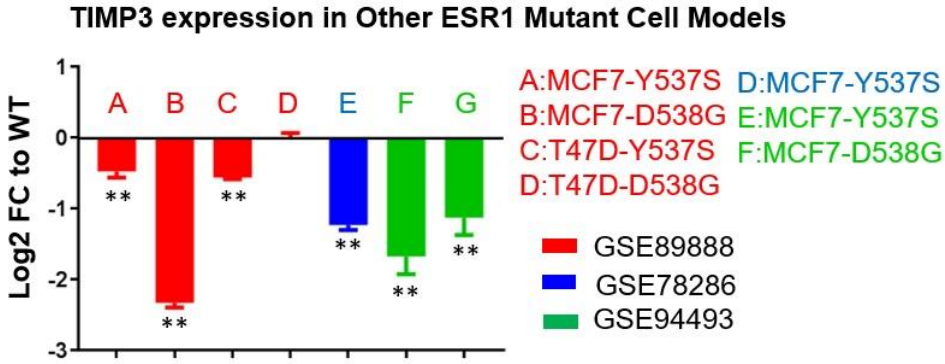



**Figure 103. *TIMP3* is downregulated in other *ESR1* mutant cell models.**

Expression fold change of *TIMP3* in *ESR1* mutant cells normalized to their WT controls from three publicly available RNA-seq data sets of distinct *ESR1* mutant cell models. Bar graphs represent the mean ± SD with corresponding biological replicates. (** $p<0.01$)

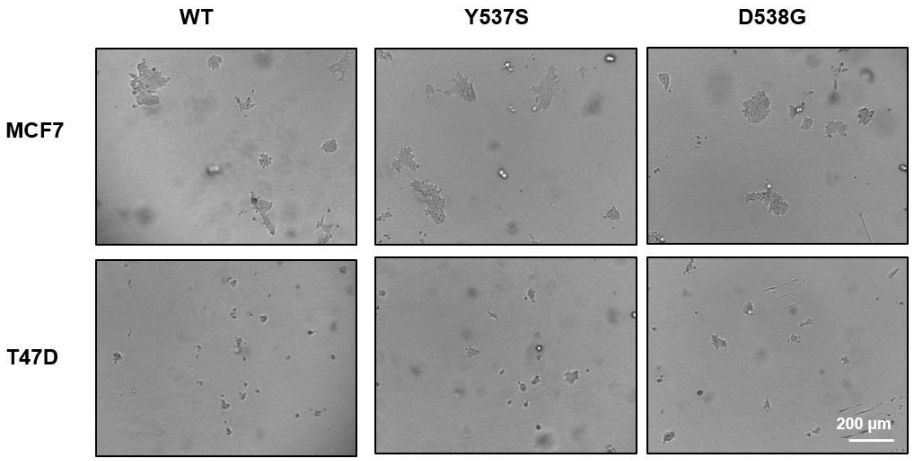



**Figure 104. *ESR1* mutant cells do not show different growth rates embending in collagen I.**

Representative images of the growth of *ESR1* WT and mutant cells in collagen I after 6 (MCF7) and 8 (T47D) days. This experiment was performed with biological triplicates. This experiment was done once.

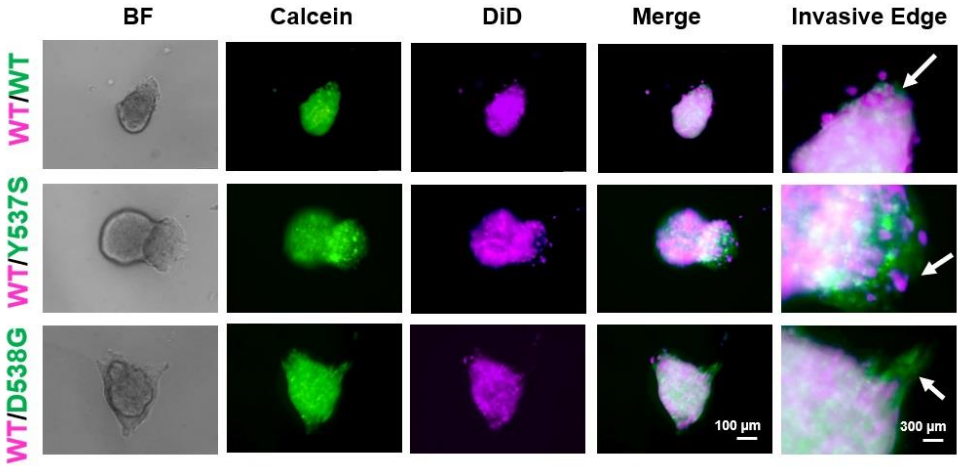



**Figure 105. T47D *ESR1* mutant cells show enhanced invasive properties in a co-culture system.**

Representative images of spheroid co-culture invasion in Type I collagen. T47D WT cells with DiD (pink) labelling were equally mixed with calcein (green) labelled T47D WT/Y537S/D538G cells. The mixed spheroids were formed in ULA plates and Collagen I was loaded into each well. Images were taken after 8 days. Invasive edges with GFP and Cy5 channels were pointed with white arrows. This experiment was done once.

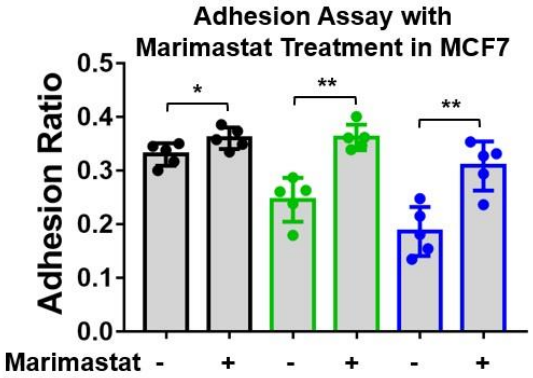



**Figure 106. MMP inhibition rescues the diminished adhesive properties in *ESR1* mutant cells.**

Quantification of adhesion ratios on collagen I of MCF7 *ESR1* WT and mutant cells with or without 30 µM Marimastat pretreatment for 24 hours. Bar graphs represent the mean ± SD with five biological replicates in each group. Student's t test was utilized to examine the effects of Marimastat treatment. This experiment was done once. (* $p<0.05$, ** $p<0.01$)

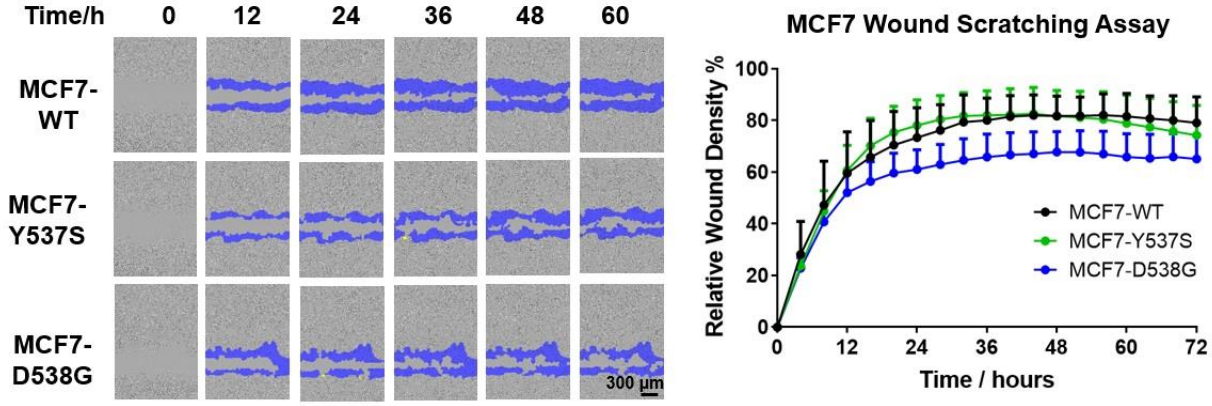



**Figure 107. MCF7 *ESR1* mutant cells do not show altered migratory property.**

IncuCyte wound scratching assay on MCF7 WT and mutant cells using IncuCyte living imaging system for 72 hours. Cell migration rates were quantified based on wound closure density. Pairwise two-way ANOVA between WT and each mutant was performed. This experiment was done once.

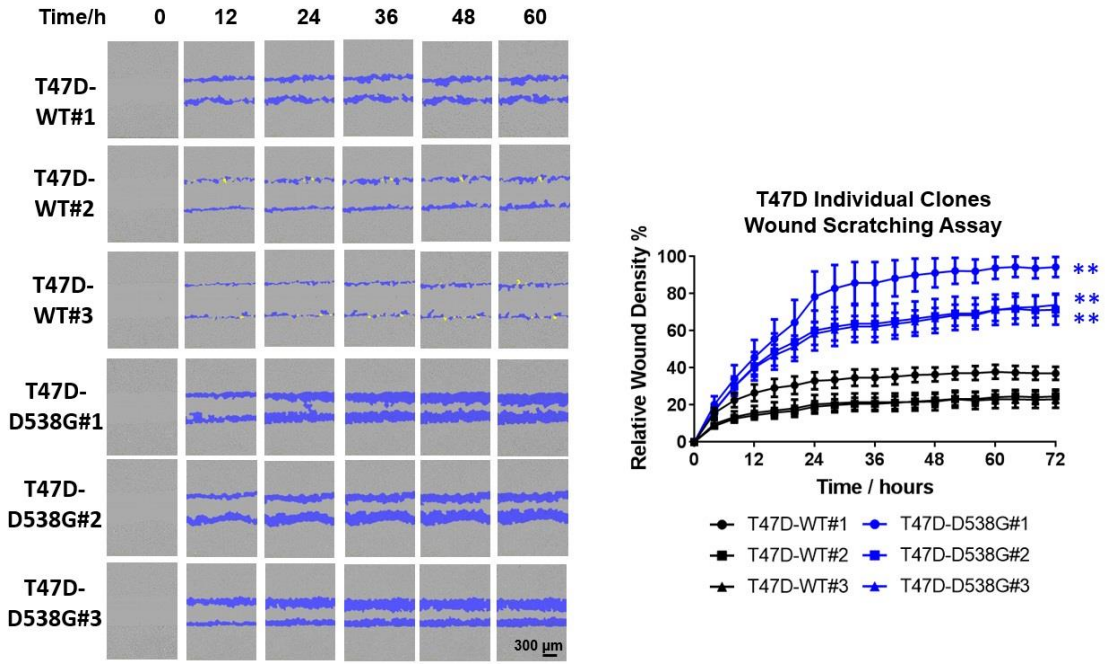



**Figure 108. Enhanced migration is reproducible in T47D D538G individual clones.**

Left panel: Confirmation of migratory alterations with T47D WT and D538G individual clones for 72 hours. Pairwise Two-way ANOVA was used to test the significance between mean of WT cells and each D538G clone. Right panel: Representative images of T47D WT and D538G migration in individual clone manner. The migratory parts are labelled in blue. This experiment was done once. (** $p<0.01$)

*This experimeny was operated by Jian Chen.

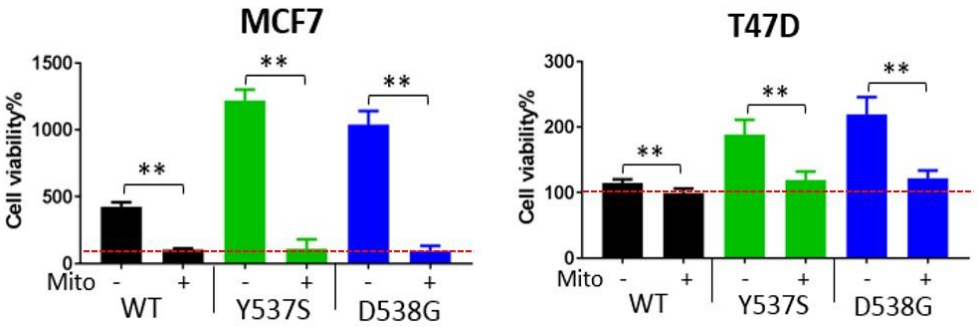



**Figure 109. Mitomycin C blocks proliferation of *ESR1* mutant cells with in 3 days.**

MCF7 and T47D were seeded into 96-well plates and treated with or without 5ug/ml Mitomycin C for 3 days. Cell numbers were quantified using FluoReporter dsDNA quantification kit. Students' t-test was used to compare the Mitomycin C effects for each cell line. This experiment was done once. (** $p<0.01$)

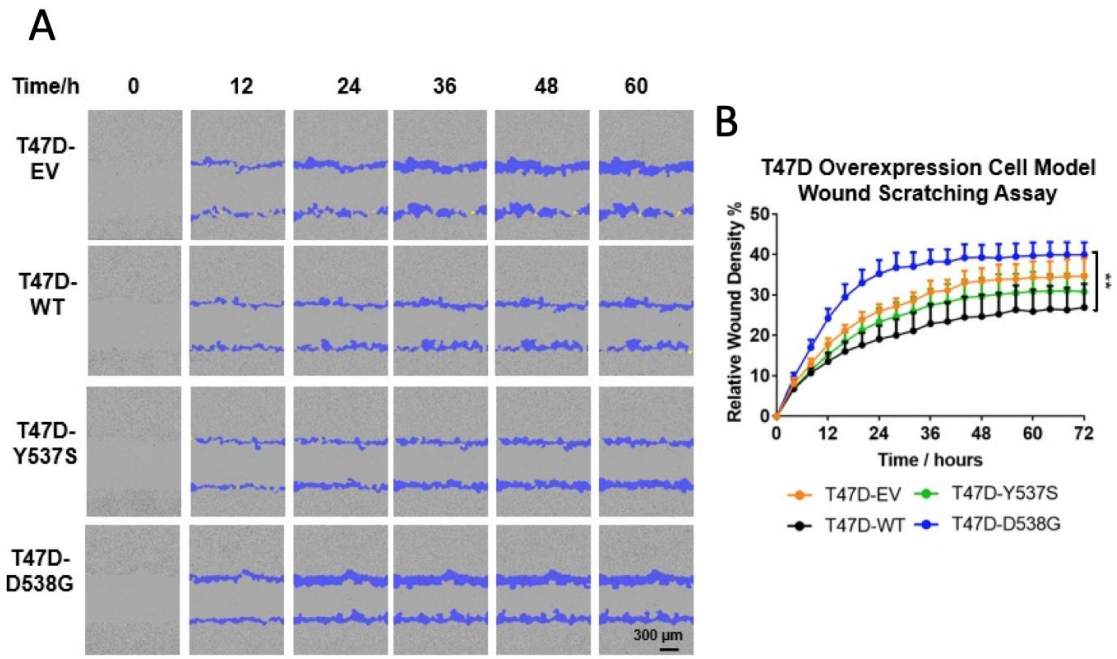



**Figure 110. T47D-D538G overexpression models exhbit enhanced migration.**

A. IncuCyte wound scratching assay on MCF7 WT and mutant cells using IncuCyte living imaging system for 72 hours. B. Cell migration rates were quantified based on wound closure density. Pairwise two-way ANOVA between WT and each mutant was performed. This experiment was done once. (** $p<0.01$)

*This experiment was reproduced by Jian Chen.

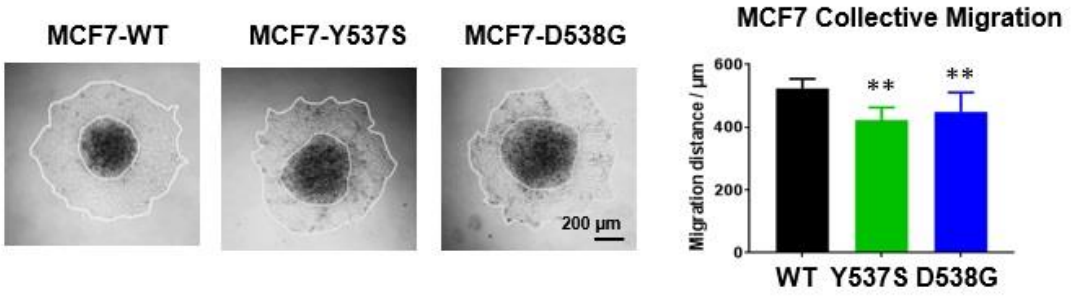



**Figure 111. MCF7 *ESR1* mutant cells do not show enhanced collective migration.**

Representative images (Left panel) and quantification (Right panel) of spheroid collective migration of MCF7 mutant cells. 3000 MCF7 cells were initially seeded into 96-well round bottom ULA plates to form spheroids, which were then transferred onto Collagen I coated plates. Collection migration were measured within 4 days. The objectives were 4X. Migratory distance were calculate based on the mean radius of each spheroid normalized to the original areas. Dunnett's test was used. This experiment was repeated for three times and data from one representative experiment is shown here. (** $p<0.01$)

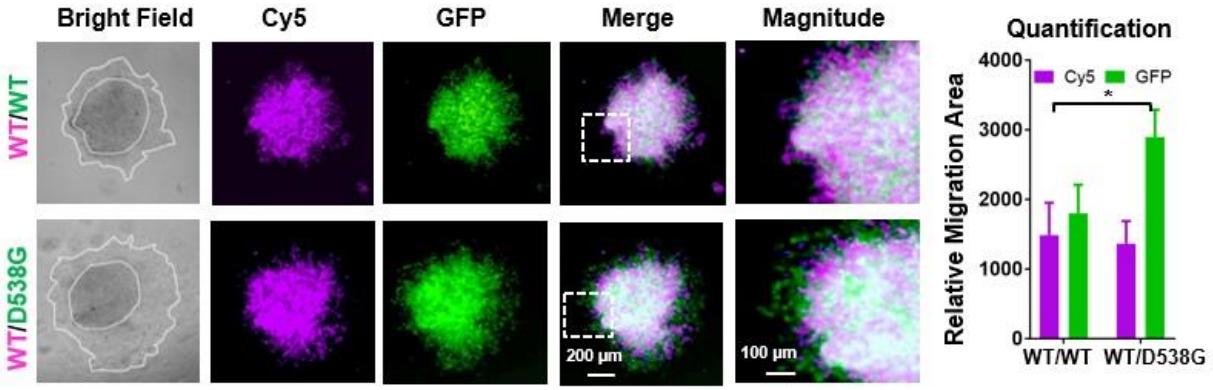



**Figure 112. T47D-D538G cells display enahnced collective migration in a co-culture system.**

Representative images of spheroid co-culture collective migration in Type I collagen. T47D WT cells with DiD (pink) labelling were equally mixed with calcein (green) labelled T47D WT /D538G cells in a 96-well round bottom ULA-plate to form spheroids for 2 days. Mixed spheroids were then transferred to Collagen I coated plate for collective migration. The objectives were 4X for the entire spheroids and 10X for the migratory edge. Migration distances of Cy5 and GFP signals were calculated separately. Students' t test was used. This experiment was done once. (* $p<0.05$)

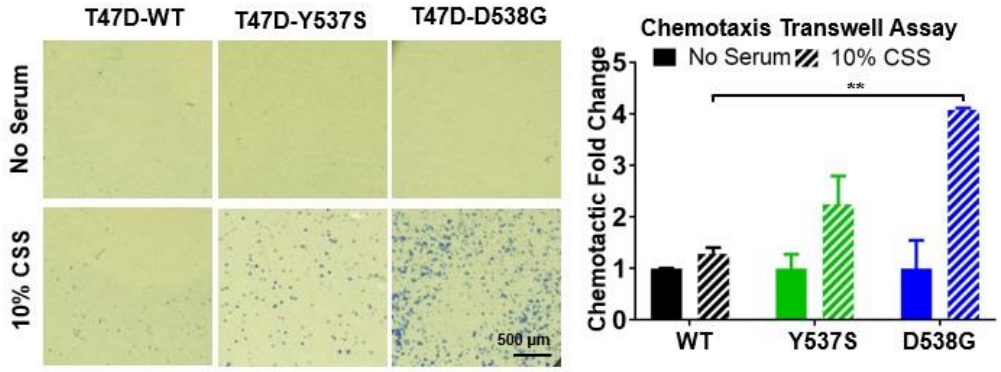



**Figure 113. T47D-D538G cells show increased chemotaxis.**

Representative images (left panel) and quantification (right panel) of chemotaxis assay on T47D-*ESR1* mutant cells. The chemotaxis assay was performed after hormone deprivation, and 10% CSS was used the chemotaxis for 72 hours. Serum free medium group was set as negative control. Cells were stained using crystal violet and quantification were conducted after dissolving the staining and measure the absorbance under OD 450. Dunnett's test was used to compare the difference between WT and each mutant cell. This experiment was repeated twice, data from one representative experiment is shown here. (** $p<0.01$)

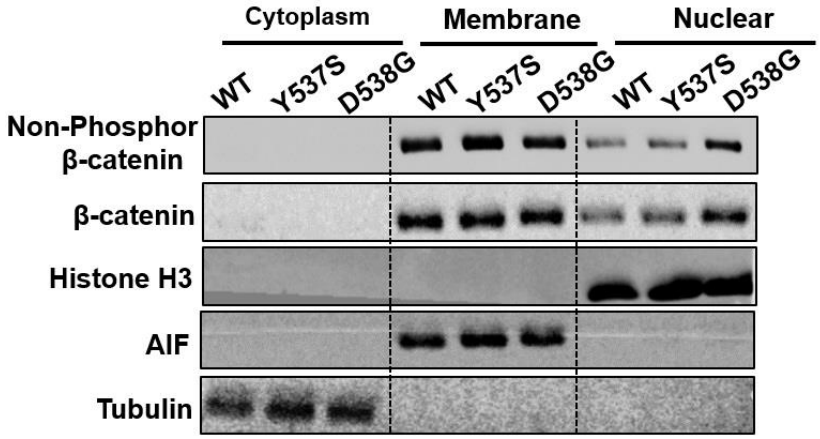



**Figure 114. T47D-D538G cells show more active β-catenin proetin levels in the nuclei portion.**

Immunoblot showing the active β-catenin and total β-catenin in different fractionations of T47D *ESR1* mutant cells. Histone H3, AIF and tubulin were used for loading controls towards nuclei, membrane and cytoplasm portions. This experiment was repeated twice, blots from one representative experiment is shown here.

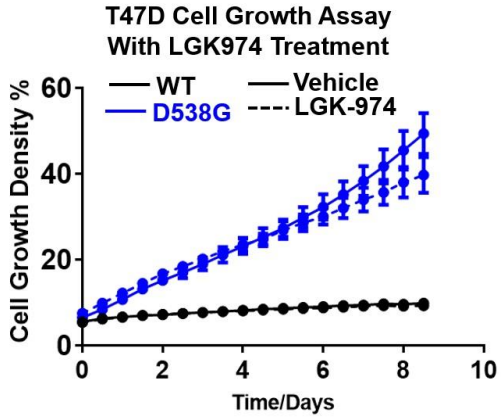



**Figure 115. LGK974 treatment does not block proliferation of T47D-WT and D538G cells.**

Growth curve of T47D WT and D538G cells under CSS condition with or without the treatment of 5μM LGK974 over 9 days. Cell growth were quantified using cell density with IncuCyte living image system. Two-way ANOVA was performed to compare the effects of LGK974 treatment. This experiment was done once.

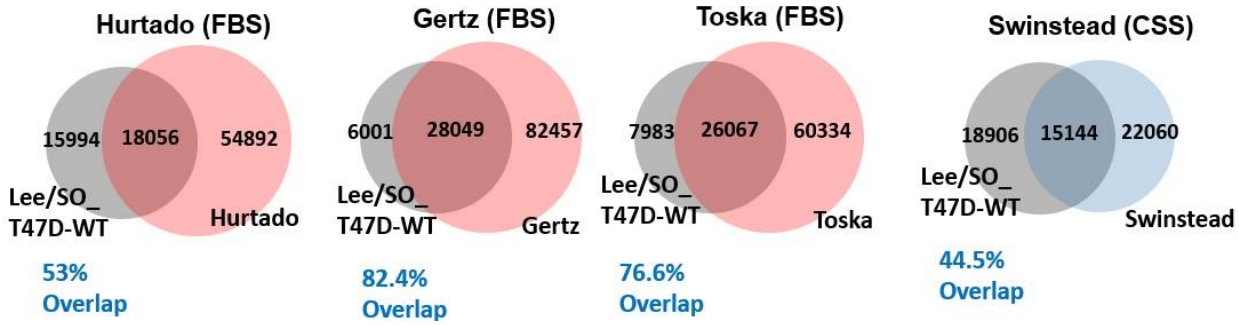



**Figure 116. Intersection of FOXA binding peaks to other public availbale FOXA1 ChIP-seq data sets in T47D cell line.**

Venn diagrams showed the intersection of FOXA1 binding peaks in T47D-WT-veh group with other public available FOXA1 ChIP-seq data sets in T47D parental cells in FBS (Hurtado, Gertz and Toska) and CSS (Swinstead) conditions. All data sets were downloaded and reprocessed, MACS2 was used for peak calling with threshold of $q<0.05$. Overlap percentage of FOXA1 peaks from this study was labelled below each diagram.

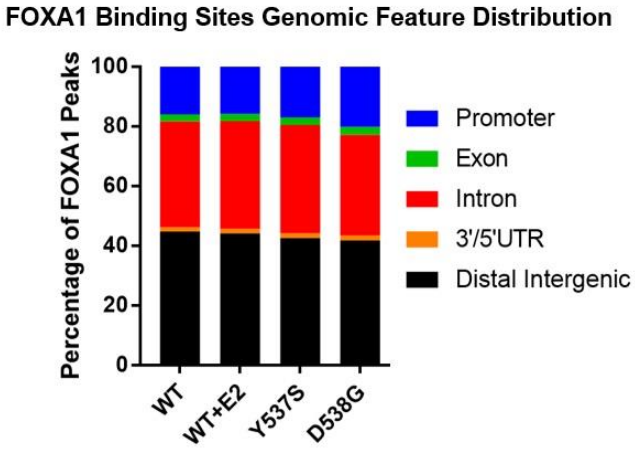



**Figure 117. Genomic feature distribution of FOXA1 peaks in T47D *ESR1* mutant cell models.**

Stacked plot showing the genomic feature distribution patterns of all FOXA1 binding sites in the the four groups. The promoter regions are defined as +/- 3 kb around the transcriptional start sites.

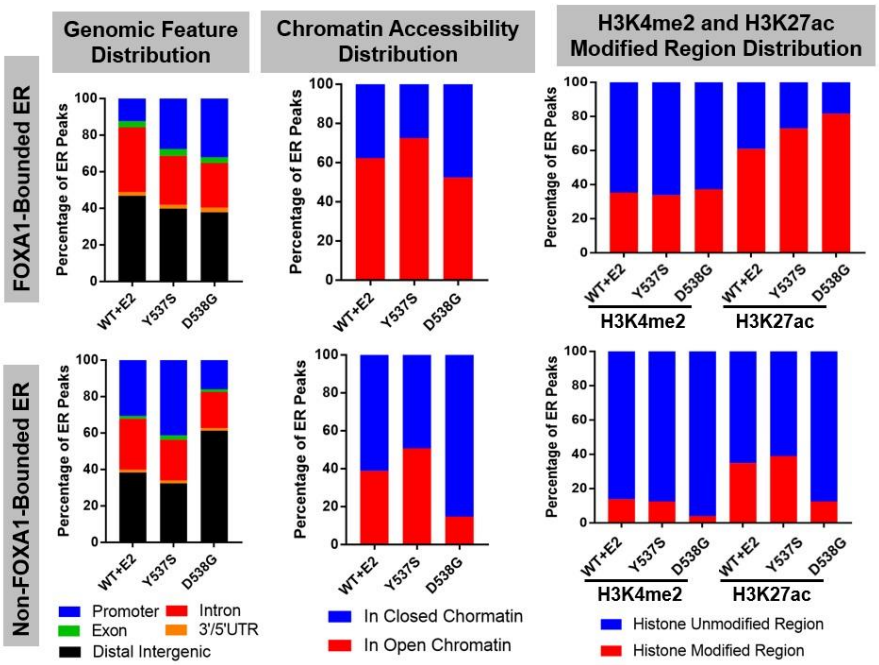



**Figure 118. D538G FOXA1-unbound but not bound ER binds at intergenic regions with poor chromatin accessibility.**

Stacked plots showing the genomic feature distribution (left panel), intersection with ATAC-seq-called peaks (middle panel) and intersection with H3K4me2 and H3K27ac peaks (right panel) of FOXA1-bound or non-FOXA1 bound ER peaks in T47D *ESR1* mutant cell models.

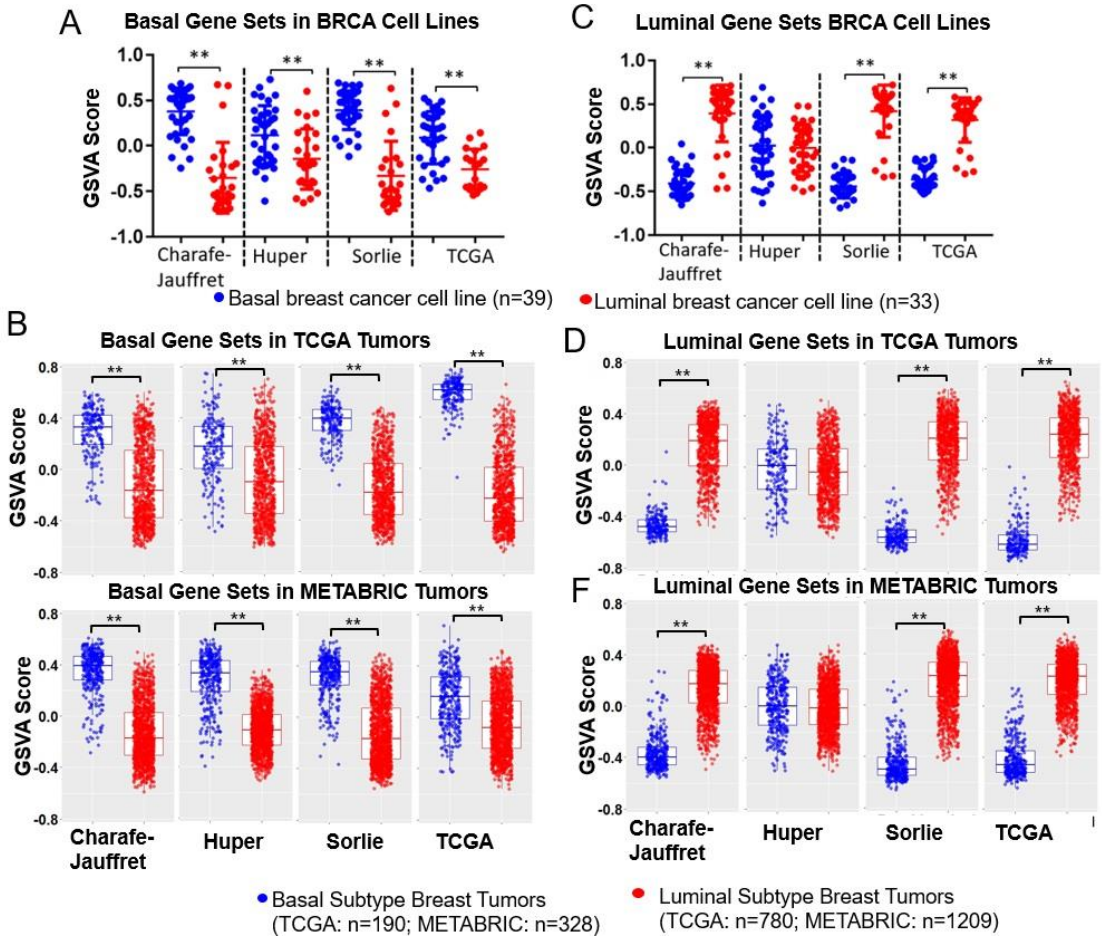



**Figure 119. Quality control of the four pairs of luminal/basal gene sets.**

A and C. Dot plots showing GSVA score of the four pairs of basal (A) /luminal (C) marker gene sets enrichment in luminal and basal breast cancer cells as quality controls. Each plot represents mean ± SD from GSVA score. Mann Whitney U test was used within each group. B and D, Box plots showing GSVA score of the four pairs of four pairs of basal (B) /luminal (D) marker gene sets enrichment in luminal and basal breast cancer tumors from TCGA and METABRIC as quality controls. Each plot represents median ± SD from GSVA score. Mann Whitney U test was used within each group. (* $p<0.05$, ** $p<0.01$).

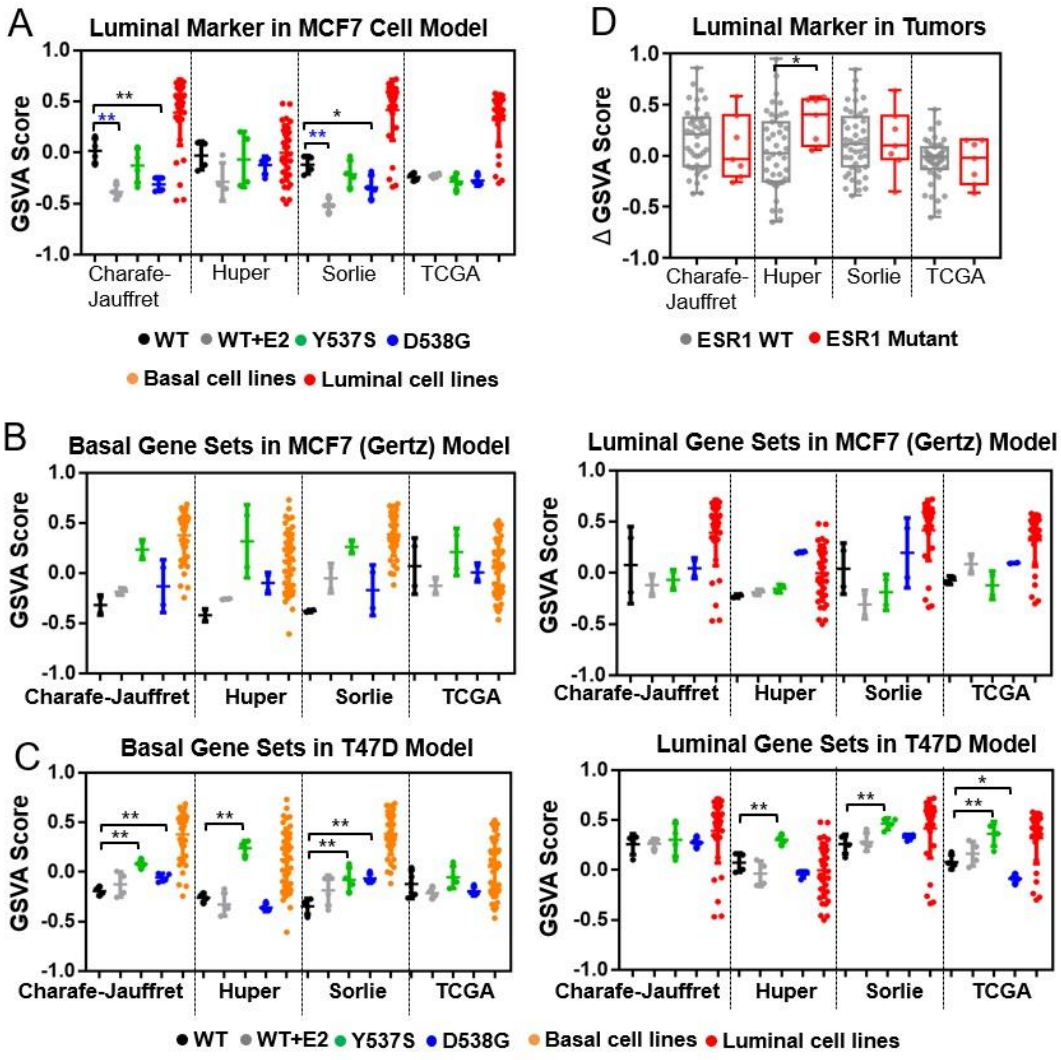



**Figure 120. *ESR1* mutant cell show increased basal marker enrichment.**

A. Dot plots showing GSVA score of the four pairs of luminal marker gene sets enrichment in MCF7 genome-edited cell models. Scores from luminal and basal breast cancer cell lines were used as positive controls. Dunnett's test was used within each group. B and C, Dot plots showing GSVA score of the four pairs of basal/luminal marker gene sets enrichment in another MCF7 (Gertz) (B) and T47D (C) genome-edited cell models. Scores from luminal and basal breast cancer cell lines were used as positive controls. Dunnett's test was used within each group. D, Box plots representing luminal marker enrichments in primary-matched paired metastatic samples. Delta GSVA score of each sample was calculated by subtracting the scores of primary tumors from the matched metastatic tumors. Mann-Whitney U test was performed to compare the Delta GSVA scores between WT (N=44) or *ESR1* mutation-harboring (N=7) paired tumors. (* $p<0.05$, ** $p<0.01$).

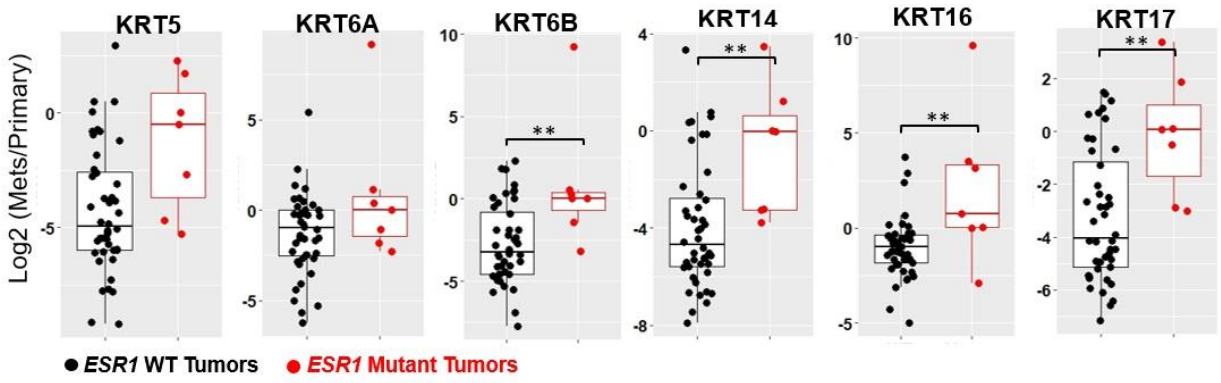



**Figure 121. Basal cytokeratins are increased in *ESR1* mutant metastatic lesions.**

Box plots representing the six basal cytokeratin expression in primary-matched paired metastatic samples. Log2 (CPM+1) values were used, and expression levels in each metastatic tumor were normalized to the matched primary tumor. Mann-Whitney U test was performed to compare the expression between WT (N=44) or *ESR1* mutation-harboring (N=7) paired tumors. (* $p<0.05$, ** $p<0.01$)

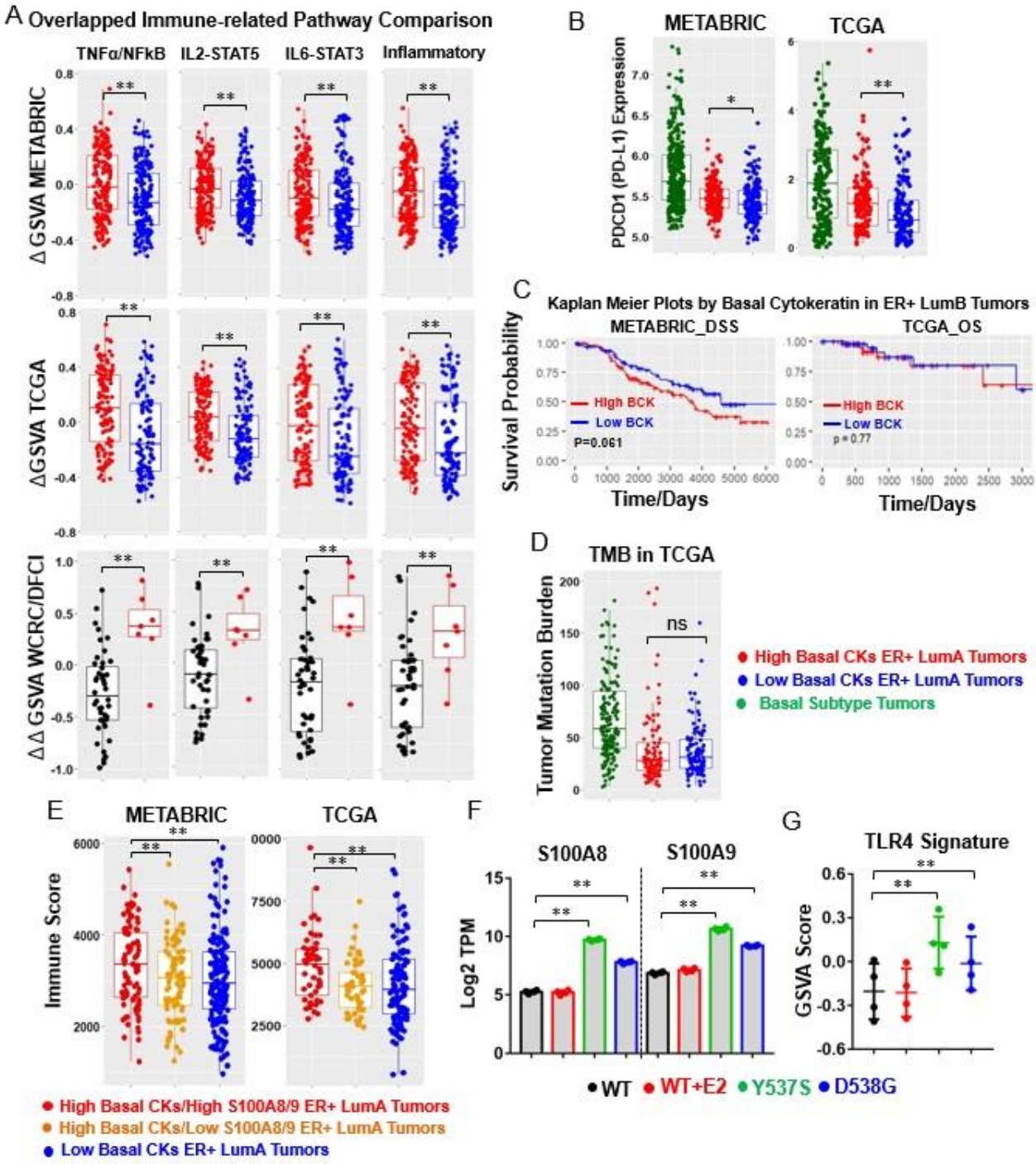



**Figure 122. Basal cytokeratins predict good prognosis and correlate to enhanced immune response.**

A. Box plots showing the enrichment score of the four overlapped immune-related pathways across basal tumors and BCK-high and low subsets in ER+ LumA tumors in METABRIC and TCGA (Top two panels), and between *ESR1* mutant and WT metastatic lesions (Bottom panel).

Mann Whitney U test was used for each comparison. B and D. Box plots comparing PDCD1 expression (B) and tumor mutation burdens (D) across basal tumors and BCK-high and low subsets in ER+ LumA tumors in METABRIC and TCGA. Mann Whitney U test was used between BCKs-high and low groups. C. Kaplan-Meier plots showing the disease-specific (METABRIC) and overall (TCGA) survival between BCKs high and low subsets of ER+ LumB patients. BCKs high and low were defined by the upper and bottom quartiles of each subset. Censored patients were labelled in cross symbols. Log rank test was used and hazard ratio with 95% CI were labelled. E. ER+ LumA patients subset of BCKs-high quantile were further divided by the mean expression of S100A8 and S100A9. ESTIMATE immune scores were compared across BCKs-high S100A8/9 low and high groups as well as BCKs-low groups. Pair-wised Mann Whitney U test was performed. F. Bar graph representing the Log2 (TPM+1) expression of S100A8 and S100A9 from MCF7 *ESR1* WT and mutant cells. Each bar represents mean ± SD with four biological replicates. G. Dot plots showing GSVA score of TLR4 signature gene set enrichment in MCF7 *ESR1* mutant cell models. Each plot represents mean ± SD with four biological replicates from RNA-seq. Dunnett's test was used. (* $p<0.05$, ** $p<0.01$)

## Appendix D Supplementary Tables

**Table 3. The sequence of sgRNA and oligos used to generate T47D *ESR1* mutant cell lines via CRISPR**

| Name | Length(bp) | Sequence |
|---|---|---|
| sgRNA targeting ESR1 gene | 20 | TCTCCAGCAGCAGGTCATAG |
| Oligo for Y537S | 70 | GCGGTGGGCGTCCAGCATCTCCAGCAGCAGG TCAGAGAGGGGCACCACGTTCTTGCACTTCATGCTGTAC |
| Oligo for D538G | 70 | GTAGGCGGTGGGCGTCCAGCATCTCCAGCAGCAG GCCATAGAGGGGCACCACGTTCTTGCACTTCATGCT |

**Table 4. DNA sequence of the oligos used used to generate MCF7 *ESR1* mutant cell lines via AAV**

| **Homology Arm Cloning Primers** | | |
|---|---|---|
| Targeting Vector | Homology Arm | Forward/Reverse Primer |
| ESR1 exon 10 | 5' | GCAGAGTTGTGGCTAGTGGA/ AAGCTGAGGGCTTTCAGAAG |
| | 3' | TCCCAGCTCCCATCCTAAAGTG/ AAAGGATGCATTGCCATAGG |
| **Pre-Cre Screening Primers** | | |
| Targeting Vector | Homology Arm | Forward/Reverse Primer |
| ESR1 exon 10 | 5' | GGCAAGTCTCCAACTTGAGC/ GCAGACAGCGAATTAATTCC |
| | 3' | TTAAGGTACCACTGTGCATATG/ CCGGGAAGATCCAAGTACAG |
| **Post Cre Screening Primers** | | |
| Targeting Vector | Forward/Reverse Primer | |
| ESR1 exon 10 | GGCAAGTCTCCAACTTGAGC/ CATATGCACAGTGGTACCTTAA | |
| **Bi-Allelic Sequencing Primers** | | |
| Targeting Vector | Forward/Reverse Primer | Nested Sequencing |
| ESR1 exon 10 | TCCCAGCTCCCATCCTAAAGTG/ AAAGGATGCATTGCCATAGG | CCCCTTCTAGGGATTTCAGC |
| **Targeted Allele Sequencing Primers** | | |
| Targeting Vector | Forward/Reverse Primer | Nested Sequencing |
| ESR1 exon 10 | TTAAGGTACCACTGTGCATATG/ AGAGGCAGAGCTTTCAGCAC | TCCCAGCTCCCATCCTAAAGTG |
| **Mutagenesis Primers** | | |
| Mutation | Forward/Reverse Primer | |
| ESR1 Y537S | CCTCTCTGACCTGCTGCTGGA/ TCCAGCAGCAGGTCAGAGAGG | |
| ESR1 D538G | CCTCTATGGCCTGCTGCTGGA/ TCCAGCAGCAGGCCATAGAGG | |

**Table 5. ddPCR amplification primers and probe sequences for Y537S and D538G *ESR1* mutations.**

| ddPCR Amplification Primers Sequences | | |
|---|---|---|
| Mut_ID | Primer_F | Primer_R |
| ESR1_Y537S | CAGCATGAAGTGCAAGAACGT | TGGGCGTCCAGCATCTC |
| ESR1_D538G | GCATGAAGTGCAAGAACGTG | AAGTGGCTTTGGTCCGTCT |
| ddPCR Probe Sequences | | |
| Mut_ID | Probe_WT | Probe_mut |
| ESR1_Y537S | CCCCTCTATGACCTGC (VIC) | CCCTCTCTGACCTGC (FAM) |
| ESR1_D538G | TCTATGACCTGCTGCTGGAGATGCT (FAM) | TCTATGGCCTGCTGCTGGAGATGCT (HEX) |

**Table 6. Detailed information of antibody used in Chapter 2**

| Target | Manufacturer, catalog # | Species | monoclonal or polyclonal | Dilution used |
|---|---|---|---|---|
| **Estrogen Receptor α** | Cell Signaling Technology (#8644) | Rabbit | Monoclonal | 1 to 1000 |
| **Phospho-Estrogen Receptor-a (Ser118)** | Signalway Antibody (#11072) | Rabbit | Polyclonal | 1 to 500 |
| **Phospho-IGF-I Receptor β (Tyr1135)** | Cell Signaling Technology (#3918) | Rabbit | Monoclonal | 1 to 200 |
| **IGF-I Receptor β (D23H3)** | Cell Signaling Technology (#9750) | Rabbit | Monoclonal | 1 to 1000 |
| **Phospho-IRS-1 (Ser302)** | Cell Signaling Technology (#2384) | Rabbit | Polyclonal | 1 to 500 |
| **Phospho-Akt (Thr308)** | Cell Signaling Technology (#13038) | Rabbit | Monoclonal | 1 to 1000 |
| **Phospho-Akt (Ser473)** | Cell Signaling Technology (#4060) | Rabbit | Monoclonal | 1 to 1000 |
| **Akt (pan) (40D4) Mouse mAb** | Cell Signaling Technology (#2920) | Mouse | Monoclonal | 1 to 1000 |
| **Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204)** | Cell Signaling Technology (#4377) | Rabbit | Monoclonal | 1 to 1000 |
| **p44/42 MAPK (Erk1/2)** | Cell Signaling Technology (#9102) | Rabbit | Polyclonal | 1 to 1000 |
| **IRS-1** | Santa Cruz sc7200 | Rabbit | Polyclonal | 1 to 1000 |
| **Insulin Receptor β** | Cell Signaling Technology (#3025) | Rabbit | Monoclonal | 1 to 1000 |
| **β-Actin** | Thermo Fisher (A3853) | Mouse | Monoclonal | 1 to 10000 |

**Table 7. Sequences of qRT-PCR primers used in Chapter 2**

| Target | Forward | Reverse |
|---|---|---|
| RPLP0 | TAAACCCTGCGTGGCAATC | TTGTCTGCTCCCACAATGAAA |
| ESR1 | GAGTATGATCCTACCAGACCCTTC | CCTGATCATGGAGGGTCAAATC |
| IGFBP4 | ACGAGGACCTCTACATCATCC | GTCCACACACCAGCACTTG |
| GREB1 | GGTTCTTGCCAGATGACAATGG | CTTGGGTTGAGTGGTCAGTTTC |
| PGR | TCGCCTTAGAAAGTGCTGTC | GCTTGGCTTTCATTTGGAACG |
| IRS1 | TCTGCTCAGCGTTGGTG | GTGCATGCTCTTGGGTTTG |

**Table 8. Sequences of ChIP-qPCR primers used in Chapter 2**

| Target | Forward | Reverse |
|---|---|---|
| IGFBP4_ChIP | GGGTTGGGCAAGGAAAAGTT | CTTCTCTGCACCGTGGTTTGT |
| GREB1_ChIP | GTGGCAACTGGGTCATTCTGA | CGACCCACAGAAATGAAAAGG |

**Table 9. Disease and Function pathways enriched in mutant cells in the absence of estrogen**

| Cell line | Diseases or Functions Annotation | p-Value | Activation z-score | #genes involved |
|---|---|---|---|---|
| T47D | cancer | 5.27E-14 | 0.316 | 394 |
| | cell movement | 5.50E-14 | 0.446 | 121 |
| | tumorigenesis of tissue | 2.96E-13 | 1.357 | 375 |
| | neoplasia of epithelial tissue | 6.30E-13 | 0.951 | 371 |
| | epithelial cancer | 9.39E-13 | 0.470 | 369 |
| MCF7 | migration of cells | 5.11E-12 | 2.852 | 86 |
| | cell movement | 2.01E-11 | 2.702 | 91 |
| | cancer | 1.09E-10 | 0.538 | 290 |
| | invasion of cells | 1.67E-10 | 1.406 | 49 |
| | malignant solid tumor | 2.37E-10 | -0.628 | 287 |

**Table 10. Sequences of qRT-PCR primers used in Chapter 3**

| Targets | Forward | Reverse |
|---|---|---|
| RPLP0 | TAAACCCTGCGTGGCAATC | TTGTCTGCTCCCACAATGAAA |
| DSC1 | CGCCCCATATTTTGAACACAG | ACGAGTATGGAGAGTGTCAGG |
| DSC2 | TCGATGCTAGAAAACTCCTTGG | GTCAACTCCAGGACCTCTTATG |
| DSG1 | TCAATCCGAAGGCAGAAACG | TGCGGTATGTAACTTGCTGG |
| DSG2 | GAATCGTATCTCTGGAGCCTG | TGCCATCTCTTGCTTCTACTG |
| PKP1 | GAAGTCCAAGTCTTCCCAGTC | GATATCCCCATGAGCCATTCC |
| GJA1 | GGATCGGGTTAAGGGAAAGAG | AGGAGACATAGGCGAGAGG |
| GJB2 | AGGAGATCAAAACCCAGAAGG | AAGACGTACATGAAGGCGG |
| GJB5 | GAAGAGGCACCGAGAAGC | TGGAACACATAGAGAAAGGCG |
| TIMP3 | TGATGGCAAGATGTACACGG | GAAGTCACAAAGCAAGGCAG |

**Table 11. Detailed information of antibodies used in Chapter 3**

| Target | Manufacturer, catalog # | Species | monoclonal or polyclonal | Dilution used |
|---|---|---|---|---|
| **desmocollin 1 (dsc1)** | Santa Cruz sc398590 | Mouse | Monoclonal | 1 to 500 |
| **desmoglein 2 (dsg2)** | Santa Cruz sc80663 | Mouse | Monoclonal | 1 to 500 |
| **plakophilin (pkp1)** | Santa Cruz sc33636 | Mouse | Monoclonal | 1 to 500 |
| **connexin 26** | Santa Cruz sc7261 | Goat | Polyclonal | 1 to 500 |
| **connexin 43** | Sigma-Aldrich C6219 | Rabbit | Polyclonal | 1 to 500 |
| **Phospho-GSK3β (Ser9)** | Cell Signaling Technology (#5558) | Rabbit | Monoclonal | 1 to 1000 |
| **Phospho-GSK3α (Ser21)** | Cell Signaling Technology (#9316) | Rabbit | Monoclonal | 1 to 1000 |
| **GSK3β** | Cell Signaling Technology (#12456) | Mouse | Monoclonal | 1 to 1000 |
| **GSK3α** | Cell Signaling Technology (#4337) | Rabbit | Monoclonal | 1 to 1000 |
| **β-catenin** | BD #610154 | Mouse | Polyclonal | 1 to 1000 |
| **Non-phospho β-Catenin (Ser45)** | Cell Signaling Technology (#19807) | Rabbit | Monoclonal | 1 to 500 |
| **Estrogen Receptor α** | Cell Signaling Technology (#8644) | Rabbit | Monoclonal | 1 to 1000 |
| **FOXA1** | Abcam ab23738 | Mouse | Monoclonal | 1 to 1000 |
| **c-FOS** | Santa Cruz sc52 | Rabbit | Polyclonal | 1 to 1000 |
| **TIMP3** | Abcam ab39184 | Mouse | Monoclonal | 1 to 500 |
| **Tubulin** | Sigma-Aldrich T6557 | Mouse | Monoclonal | 1 to 10000 |
| **AIF** | Cell Signaling Technology (#5318) | Rabbit | Monoclonal | 1 to 1000 |
| **Histone H3** | Cell Signaling Technology (#4499) | Rabbit | Monoclonal | 1 to 1000 |

**Table 12. Informaton of *ESR1* mutant metastatic samples in WCRC/Charite cohort**

| Sample ID | Metastatic Site | Mutation Type | Mutation Frequency % |
|---|---|---|---|
| BoM05 | Bone | D538G | 1.15 |
| BoM07 | Bone | D538G | 2.9 |
| BoM30 | Bone | D538G | 14.6 |
| BoM35 | Bone | Y537S | 1.98 |
| BoM43 | Bone | D538G | 5.3 |
| BrM10 | Brain | D538G | 0.18 |
| BrM16 | Brain | D538G | 35.4 |
| BrM20 | Brain | D538G | 45.8 |
| BrM72 | Brain | Y537C | 0.26 |
| BrM74 | Brain | D538G | 31.8 |
| | | Y537S | 0.3 |
| GM4B | GI | D538G | 10.8 |
| | | Y537S | 12.4 |
| GM9 | GI | D538G | 0.34 |

**Table 13. Gene sets used in GSVA for comprehensive cell-cell adhesome analysis.**

| Desmosome | Gap Junction | Tight Junction | Adherens Junction |
|---|---|---|---|
| DSC1 | GJA1 | CLDN1 | CDH1 |
| DSC2 | GJA3 | CLDN2 | CDH2 |
| DSC3 | GJA4 | CLDN3 | CDH3 |
| DSG1 | GJA5 | CLDN4 | CDH4 |
| DSG2 | GJA6P | CLDN5 | CDH5 |
| DSG3 | GJA8 | CLDN6 | CDH6 |
| DSG4 | GJA9 | CLDN7 | CDH7 |
| PKP1 | GJA10 | CLDN8 | CDH8 |
| PKP2 | GJB1 | CLDN9 | CDH9 |
| PKP3 | GJB2 | CLDN10 | CDH10 |
| PKP4 | GJB3 | CLDN11 | CDH11 |
| JUP | GJB4 | CLDN12 | CDH12 |
| DSP | GJB5 | CLDN14 | CDH13 |
| KRT1 | GJB6 | CLDN15 | CDH15 |
| KRT2 | GJB7 | CLDN16 | CDH16 |
| KRT3 | GJC1 | CLDN17 | CDH17 |
| KRT4 | GJC2 | CLDN18 | CDH18 |
| KRT5 | GJC3 | CLDN19 | CDH19 |
| KRT6A | GJD2 | CLDN20 | CDH20 |
| KRT6B | GJD3 | CLDN22 | CDH22 |
| KRT7 | GJD4 | CLDN23 | CDH23 |
| KRT8 | GJE1 | OCLN | CDH24 |
| KRT9 | | TJP1 | CTNNA1 |
| KRT10 | | TJP2 | CTNNB1 |
| KRT12 | | TJP3 | VCL |
| KRT13 | | F11R | |
| KRT14 | | JAM2 | |
| KRT15 | | JAM3 | |
| KRT16 | | CGN | |
| KRT17 | | | |
| KRT18 | | | |
| KRT19 | | | |
| KRT20 | | | |

**Table 14. Detailed information of 51 paired metastatic breast cancer tumor samples merged from WCRC and DFCI cohrots.**

| Pair Number | Primary Tumor ID | Metastatic Tumor ID | Cohort | Metastatic Site | *ESR1* Genotype in metastatic lesions | Mutation Subtype |
|---|---|---|---|---|---|---|
| **1** | 43P | 43M | WCRC | Bone | Mutation | D538G |
| **2** | BP72 | BM72 | WCRC | Brain | Mutation | Y537C & L536Q |
| **3** | 0031.50T | 0006.56M | WCRC | Ovary | Mutation | Y537N |
| **4** | GP4 | GM4B | WCRC | GI | Mutation | Y537S & D538G |
| **5** | 043_Archival1 | 043_Prospective1 | DFCI | Liver | Mutation | D538G |
| **6** | 295_Archival1 | 295_T2 | DFCI | Liver | Mutation | E380Q |
| **7** | 325_Archival1 | 325_T1 | DFCI | Liver | Mutation | D538G |
| **8** | 19P | 19M | WCRC | Bone | WT | |
| **9** | 22P | 22M | WCRC | Bone | WT | |
| **10** | 31P | 31M | WCRC | Bone | WT | |
| **11** | 34P | 34M | WCRC | Bone | WT | |
| **12** | 44P | 44M | WCRC | Bone | WT | |
| **13** | 48P | 48M | WCRC | Bone | WT | |
| **14** | 55P | 55M | WCRC | Bone | WT | |
| **15** | 60P | 60M | WCRC | Bone | WT | |
| **16** | A25P | A25M | WCRC | Bone | WT | |
| **17** | 2P_RCS | 2M_RCS | WCRC | Brain | WT | |
| **18** | 3P_RCS | 3M_RCS | WCRC | Brain | WT | |
| **19** | 4P_RCS | 4M_RCS | WCRC | Brain | WT | |
| **20** | 6P_RCS | 6M_RCS | WCRC | Brain | WT | |
| **21** | BP7 | BM7 | WCRC | Brain | WT | |
| **22** | BP17 | BM17 | WCRC | Brain | WT | |
| **23** | BP47 | BM47 | WCRC | Brain | WT | |
| **24** | BP51 | BM51 | WCRC | Brain | WT | |
| **25** | BP62 | BM62 | WCRC | Brain | WT | |
| **26** | 0029.50T | 0003.56M | WCRC | Ovary | WT | |
| **27** | 0030.50T | 0004.56M | WCRC | Ovary | WT | |
| **28** | 0032.50T | 0008.56M | WCRC | Ovary | WT | |
| **29** | 0033.50T | 0009.56M | WCRC | Ovary | WT | |
| **30** | 0034.50T | 0014.56M | WCRC | Ovary | WT | |
| **31** | 0035.50T | 0018.56M | WCRC | Ovary | WT | |
| **32** | OP1 | OM1 | WCRC | Ovary | WT | |
| **33** | OP5 | OM5A | WCRC | Ovary | WT | |
| **34** | OP5 | OM5B | WCRC | Ovary | WT | |
| **35** | OP8 | OM8 | WCRC | Ovary | WT | |

| Table 14 Continued | | | | | | |
|---|---|---|---|---|---|---|
| **36** | GP2 | GM2A | WCRC | GI | WT | |
| **37** | GP2 | GM2B | WCRC | GI | WT | |
| **38** | GP4 | GM4A | WCRC | GI | WT | |
| **39** | GP7 | GM7 | WCRC | GI | WT | |
| **40** | 069_Archival1 | 069_Prospective1 | DFCI | Bone | WT | |
| **41** | 074_Archival1 | 074_Prospective1 | DFCI | Liver | WT | |
| **42** | 076_Archival1 | 076_Prospective1 | DFCI | Liver | WT | |
| **43** | 188_Archival1 | 188_Prospective1 | DFCI | Bone | WT | |
| **44** | 195_Archival1 | 195_Prospective1 | DFCI | Liver | WT | |
| **45** | 291_Archival1/ 291_Archival2 | 291_T1 | DFCI | Liver | WT | |
| **46** | 295_Archival1 | 295_Archival2 | DFCI | Bone | WT | |
| **47** | 306_Archival2 | 306_T1 | DFCI | Breast | WT | |
| **48** | 307_A1 | 307_T1 | DFCI | Liver | WT | |
| **49** | 348_A1 | 348_T1 | DFCI | Liver | WT | |
| **50** | 381_A1 | 381_T1 | DFCI | Liver | WT | |
| **51** | 495_A1 | 495_T1 | DFCI | Liver | WT | |

**Table 15. *ESR1* mutant cell ER ChIP-seq reads alignment and peak calling information.**

| Cell model | Groups | #reads | %mapped (hg38) | # ER binding peaks (p<10-5) |
|---|---|---|---|---|
| MCF7 | WT-Input | 16856221 | 94.21 | |
| | WT-veh | 23447868 | 85 | 125 |
| | WT-E2 | 16375519 | 89.15 | 12472 |
| | Y537S-Input | 17529190 | 92.47 | |
| | Y537S-veh | 20516994 | 90.35 | 657 |
| | Y537S-E2 | 20680234 | 89.81 | 1847 |
| | D538G-Input | 20702360 | 82.44 | |
| | D538G-veh | 20222723 | 87.33 | 1016 |
| | D538G-E2 | 20606853 | 91.4 | 5403 |
| T47D | WT-Input | 18730431 | 95.97 | |
| | WT-veh | 21211902 | 90.02 | 615 |
| | WT-E2 | 22925286 | 91.73 | 1724 |
| | Y537S-Input | 17630142 | 93.54 | |
| | Y537S-veh | 19794083 | 87.65 | 1096 |
| | Y537S-E2 | 19581111 | 88.2 | 3300 |
| | D538G-input | 19130173 | 94.4 | |
| | D538G-veh | 21516548 | 89.4 | 1468 |
| | D538G-E2 | 20555516 | 86.55 | 9552 |

**Table 16. Gene sets used in Chapter 3 section 3.3.**

| Collagen I | Collagen II | Collagen IV | Fibronectin | Laminin | Tenascin | Vitronectin | Integrin Gene Set |
|---|---|---|---|---|---|---|---|
| COL1A1 | COL2A1 | COL4A1 | FN | LAMA1 | TNC | VTN | ITGA1 |
| COL1A2 | | COL4A2 | | LAMA2 | | | ITGA2 |
| | | COL4A3 | | LAMA3 | | | ITGA3 |
| | | COL4A4 | | LAMA4 | | | ITGA4 |
| | | COL4A5 | | LAMA5 | | | ITGA5 |
| | | COL4A6 | | LAMB1 | | | ITGA6 |
| | | | | LAMB2 | | | ITGA7 |
| | | | | LAMB3 | | | ITGA8 |
| | | | | LAMB4 | | | ITGA9 |
| | | | | LAMC1 | | | ITGA10 |
| | | | | LAMC2 | | | ITGA11 |
| | | | | LAMC3 | | | ITGAD |
| | | | | | | | ITGAE |
| | | | | | | | ITGAL |
| | | | | | | | ITGAM |
| | | | | | | | ITGAV |
| | | | | | | | ITGA2B |
| | | | | | | | ITGAX |
| | | | | | | | ITGB1 |
| | | | | | | | ITGB2 |
| | | | | | | | ITGB3 |
| | | | | | | | ITGB4 |
| | | | | | | | ITGB5 |
| | | | | | | | ITGB6 |
| | | | | | | | ITGB7 |
| | | | | | | | ITGB8 |

**Table 17. 84 ECM adhesion genes examined in qPCR array.**

| 84 ECM Adhesion Genes for qPCR Array | | | | | | | House-keeping Genes |
|---|---|---|---|---|---|---|---|
| ADAMTS1 | COL1A1 | ECM1 | ITGAL | LAMB1 | MMP3 | SPP1 | ACTB |
| ADAMTS13 | COL4A2 | FN1 | ITGAM | LAMB3 | MMP7 | TGFB1 | B2M |
| ADAMTS8 | COL5A1 | HAS1 | ITGAV | LAMC1 | MMP8 | THBS1 | GAPDH |
| CD44 | COL6A1 | ICAM1 | ITGB1 | MMP1 | MMP9 | THBS2 | HPRT1 |
| CDH1 | COL6A2 | ITGA1 | ITGB2 | MMP10 | NCAM1 | THBS3 | RPLP0 |
| CLEC3B | COL7A1 | ITGA2 | ITGB3 | MMP11 | PECAM1 | TIMP1 | |
| CNTN1 | COL8A1 | ITGA3 | ITGB4 | MMP12 | SELE | TIMP2 | |
| COL11A1 | CTGF | ITGA4 | ITGB5 | MMP13 | SELL | TIMP3 | |
| COL12A1 | CTNNA1 | ITGA5 | ANOS1 | MMP14 | SELP | TNC | |
| COL14A1 | CTNNB1 | ITGA6 | LAMA1 | MMP15 | SGCE | VACM1 | |
| COL15A1 | CTNND1 | ITGA7 | LAMA2 | MMP16 | SPARC | VCAN | |
| COL16A1 | CTNND2 | ITGA8 | LAMA3 | MMP2 | SPG7 | VTN | |

**Table 18. Significantly enriched Hallmark pathways in T47D-Y537S and D538G mutant cells versus WT.**

| Models | Gene Set | Enrichment Score | FDR value |
|---|---|---|---|
| Enrichemd pathways in T47D-Y537S vs WT | E2F_TARGETS | 1.426 | 0.118 |
| | ESTROGEN_RESPONSE_EARLY | 1.453 | 0.128 |
| | MYC_TARGETS_V1 | 1.371 | 0.149 |
| | ESTROGEN_RESPONSE_LATE | 1.387 | 0.158 |
| | INTERFERON_ALPHA_RESPONSE | 1.468 | 0.161 |
| | G2M_CHECKPOINT | 1.316 | 0.233 |
| Enrichemd pathways in T47D-D538G vs WT | E2F_TARGETS | 1.428 | 0.122 |
| | MYC_TARGETS_V1 | 1.354 | 0.193 |
| | ANGIOGENESIS | 1.239 | 0.233 |
| | WNT_BETA_CATENIN_SIGNALING | 1.224 | 0.240 |
| | ESTROGEN_RESPONSE_EARLY | 1.310 | 0.240 |
| | OXIDATIVE_PHOSPHORYLATION | 1.247 | 0.246 |

**Table 19. Peak calling information for ATAC-seq of T47D *ESR1* mutant cells.**

| T47D Cell Model | Peak # | |
|---|---|---|
| | Vehicle | E2 |
| WT#1 | 60589 | 70929 |
| WT#2 | 53323 | 66580 |
| Y537S#1 | 65910 | 81482 |
| Y537S#2 | 65496 | 73340 |
| D538G#1 | 30466 | 33760 |
| D538G#2 | 15783 | 14068 |
| D538G#3 | 33061 | 19314 |

**Table 20. Peak calling and genome alignment information for FOXA1 ChIP-seq in T47D *ESR1* mutant cells.**

| Sample | Total reads | Mapped reads | Map% | Peak Numbers |
|---|---|---|---|---|
| WT-Input | 76202236 | 72202531 | 94.75% | |
| WT-FOXA1 | 43525486 | 40330154 | 92.66% | 34462 |
| WT-E2-Input | 76202236 | 72202531 | 94.75% | |
| WT-E2-FOXA1 | 51126633 | 47928963 | 93.75% | 30904 |
| Y537S-Input | 82405537 | 75587701 | 91.73% | |
| Y537S-FOXA1 | 60862536 | 56738010 | 93.22% | 61934 |
| D538G-Input | 75891650 | 70857546 | 93.37% | |
| D538G-FOXA1 | 50559987 | 47646723 | 94.24% | 54766 |

**Table 21. Novel target genes that potentially regulated by FOXA1-associated open chromatin in T47D-Y537S and D538G cells.**

| T47D-Y537S | | T47D-D538G |
|---|---|---|
| FIBCD1 | SPARC | SLC27A2 |
| SLC4A8 | AHNAK2 | PRKG1 |
| AFF1 | SYTL2 | GFRA1 |
| C5orf58 | TMTC1 | SATB1 |
| GPR158 | RMST | TCF4 |
| PKNOX2 | ADPRH | FDXR |
| PPP2R2A | LIFR | CPNE8 |
| GALNT6 | FGD5 | FEM1C |
| LYST | C14orf132 | FARP1 |
| ANKS1B | SORBS1 | SULF2 |
| FEM1C | S100A2 | ARFGEF3 |
| DCAF4 | VASH1 | TPCN1 |
| ROBO2 | PTPRU | MACROD2 |
| IQCJ-SCHIP1 | ADAMTS15 | BRD3 |
| FAM102A | CLSTN2 | QSOX1 |
| CLIP2 | MBP | THBS1 |
| TMEM187 | ALDH5A1 | ZHX2 |
| STK10 | SYNE1 | CLSTN2 |
| IRF1 | FLVCR2 | ERN1 |
| EMD | CECR2 | CDHR3 |
| MAN1A1 | PCED1B | WWOX |
| RNF207 | KITLG | MEIS2 |
| NR4A1 | IDS | SEMA3C |
| MAN1C1 | SPTSSA | SIPA1L2 |
| FASN | CRABP2 | RGS6 |
| PADI3 | C19orf33 | LMO2 |
| CSAD | MCIDAS | |
| GPC6 | MIR205HG | |
| APOD | CYP7B1 | |
| SETBP1 | RGL1 | |
| EGFR | | |

**Table 22. Sequence of qRT-PCR primers used in Chapter 4.**

| Gene | Forward | Reverse |
|---|---|---|
| *RPLP0* | TAAACCCTGCGTGGCAATC | TTGTCTGCTCCCACAATGAAA |
| *ESR1* | CGCCCCATATTTTGAACACAG | ACGAGTATGGAGAGTGTCAGG |
| *PGR* | TCGATGCTAGAAAACTCCTTGG | GTCAACTCCAGGACCTCTTATG |
| *KRT5* | TCAATCCGAAGGCAGAAACG | TGCGGTATGTAACTTGCTGG |
| *KRT14* | GAATCGTATCTCTGGAGCCTG | TGCCATCTCTTGCTTCTACTG |
| *KRT16* | GAAGTCCAAGTCTTCCCAGTC | GATATCCCCATGAGCCATTCC |
| *KRT17* | GGATCGGGTTAAGGGAAAGAG | AGGAGACATAGGCGAGAGG |

**Table 23. Sequence of ChIP-qPCR primers used in Chapter 4.**

| Binding Sites | Forward | Reverse |
|---|---|---|
| CTCF_Peak1 | ACTGTGGTTTCTCTGACGC | TGAACAAGAGCCTATAAAACCCC |
| CTCF_Peak2 | AAAAGCTCTCGTGGGTTCC | AGCAGGGACGAAGTGAAAC |

**Table 24. Top 50 consistently upregulated basal marker genes and their fold changes in MCF7 Y537S and D538G *ESR1* mutant cell models and *ESR1* mutant metastatic tumors.**

| Rank | Gene | Log2FC (*ESR1* Mutant Tumor/WT Tumor) | Log2FC (MCF7 Y537S /MCF7 WT) | Log2FC (MCF7 D538G /MCF7 WT) |
|---|---|---|---|---|
| 1 | KRT6B | 3.495429414 | 6.141875551 | 4.124631713 |
| 2 | KRT16 | 3.024548328 | 5.632080601 | 3.836623101 |
| 3 | KRT17 | 3.049716004 | 5.838076563 | 3.594592205 |
| 4 | KRT6A | 2.075352248 | 5.781916196 | 3.351235535 |
| 5 | IFI27 | 1.805867673 | 3.401325813 | 4.651018751 |
| 6 | PI3 | 0.024422421 | 6.208238492 | 2.885062442 |
| 7 | KRT5 | 2.939032272 | 3.687206913 | 2.340135001 |
| 8 | SLPI | 1.410330444 | 4.000419294 | 2.495206183 |
| 9 | S100A2 | 0.286351096 | 4.307547678 | 1.866178248 |
| 10 | AKR1C2 | 2.054279145 | 2.793986351 | 1.224445332 |
| 11 | IFIT3 | 0.429042944 | 2.428806761 | 3.011858425 |
| 12 | SERPINB5 | 0.539979062 | 3.289655861 | 1.389504703 |
| 13 | TUBA4A | 0.935368253 | 2.086581101 | 1.133115486 |
| 14 | KRT14 | 3.136969294 | 0.552338967 | 0.409255672 |
| 15 | CTSC | 0.84204988 | 1.828784876 | 1.221442376 |
| 16 | SLC16A1 | 0.053386619 | 1.861596644 | 1.908296644 |
| 17 | C1R | 2.021204311 | 0.885578759 | 0.717318078 |
| 18 | TRIM29 | 1.200345669 | 1.61322296 | 0.712885135 |
| 19 | CASP4 | 1.184695996 | 0.882917141 | 1.187304271 |
| 20 | C3 | 2.173705812 | 0.824046367 | 0.253216719 |
| 21 | MT2A | 0.968764069 | 1.194632893 | 1.015503923 |
| 22 | IGFBP6 | 1.013261307 | 0.997450498 | 1.021831237 |
| 23 | MT1X | 1.570141459 | 0.734572997 | 0.710361888 |
| 24 | ANXA1 | 0.988407082 | 0.956821402 | 1.067216361 |
| 25 | SGK1 | 0.630464393 | 1.274709744 | 1.072899867 |
| 26 | LAMB3 | 1.01041695 | 0.965640718 | 0.891674952 |
| 27 | PLSCR1 | 0.083219847 | 1.253570553 | 1.526927652 |
| 28 | IL18 | 0.065975109 | 1.691162472 | 1.095860331 |
| 29 | PDZK1IP1 | 0.476064519 | 1.686988639 | 0.668525288 |
| 30 | BTN3A2 | 0.435186771 | 1.1024978 | 1.21076719 |
| 31 | PTRF | 1.37549488 | 0.412432181 | 0.944591998 |
| 32 | SP100 | 0.268669945 | 0.977791813 | 1.335170811 |
| 33 | TUBB6 | 1.093434073 | 0.585726826 | 0.89129392 |
| 34 | B2M | 0.889771444 | 0.800044887 | 0.830191855 |

| Table 24 Continued | | | | |
|---|---|---|---|---|
| 35 | FSCN1 | 0.373880313 | 1.111697975 | 1.00508699 |
| 36 | GPX8 | 0.362528006 | 0.934793665 | 1.187248666 |
| 37 | TAP2 | 0.989719891 | 0.707706057 | 0.650096653 |
| 38 | CCNA1 | 0.070921455 | 0.7990588 | 1.351502225 |
| 39 | HTRA1 | 0.311878466 | 0.926179248 | 0.96069683 |
| 40 | PSAT1 | 0.299571759 | 0.881318035 | 1.010756342 |
| 41 | C1S | 1.455400905 | 0.548887814 | 0.178911911 |
| 42 | GJB3 | 1.09024293 | 0.444264205 | 0.63727444 |
| 43 | TMEM154 | 0.263687802 | 1.255521249 | 0.647841246 |
| 44 | NXN | 1.041980339 | 0.38673064 | 0.723455788 |
| 45 | AKR1C3 | 1.676831714 | 0.338352419 | 0.083890877 |
| 46 | CLMP | 0.874760249 | 0.561368283 | 0.653387695 |
| 47 | KLK5 | 0.540557154 | 1.370733869 | 0.10261943 |
| 48 | GAS1 | 1.678084381 | 0.138290048 | 0.170820813 |
| 49 | ADA | 0.197822161 | 0.88037713 | 0.889130662 |
| 50 | CDC42EP3 | 0.232383503 | 0.837192011 | 0.885951764 |
| 51 | RGS2 | 1.056426879 | 0.468992326 | 0.428213261 |

**Table 25. Top 50 consistently upregulated basal marker genes and their fold changes in T47D Y537S and D538G *ESR1* mutant cell models.**

| Rank | Gene | Log2FC (T47D Y537S /T47D WT) | Log2FC (T47D D538G /T47D WT) |
|---|---|---|---|
| 1 | WLS | 1.35360068 | 1.067051838 |
| 2 | HTRA1 | 1.076922283 | 1.053304444 |
| 3 | PSAT1 | 2.029619804 | 0.092129681 |
| 4 | IGF2BP2 | 1.265785768 | 0.663456294 |
| 5 | AKR1C2 | 0.464019802 | 1.1045904 |
| 6 | TUBB6 | 0.759223607 | 0.665802393 |
| 7 | KIRREL | 0.734386185 | 0.689607431 |
| 8 | DMD | 1.157707303 | 0.255161959 |
| 9 | ATP1B3 | 0.928814702 | 0.330269719 |
| 10 | FZD6 | 0.263432741 | 0.98303621 |
| 11 | SFN | 0.34321561 | 0.900142684 |
| 12 | AKR1C1 | 0.586722712 | 0.609476914 |
| 13 | ZDHHC2 | 0.684354492 | 0.445340601 |
| 14 | AMD1 | 0.456297285 | 0.668593844 |
| 15 | CORO1C | 0.669219087 | 0.441420822 |
| 16 | CLMP | 0.320590804 | 0.780122414 |
| 17 | STK17A | 0.410839544 | 0.634954052 |
| 18 | LARP6 | 0.376631222 | 0.584892508 |
| 19 | PLS3 | 0.382697753 | 0.51707012 |
| 20 | FAM83D | 0.281161677 | 0.607955108 |
| 21 | FSCN1 | 0.589949484 | 0.274022199 |
| 22 | FERMT1 | 0.309810849 | 0.535982386 |
| 23 | LAMB3 | 0.799619322 | 0.034719097 |
| 24 | CXCL1 | 0.827563621 | 0.001480047 |
| 25 | TKT | 0.205997424 | 0.618186653 |
| 26 | OSMR | 0.600503506 | 0.222194545 |
| 27 | CASP1 | 0.712034947 | 0.099411853 |
| 28 | BMP1 | 0.215309961 | 0.588147063 |
| 29 | ACTN1 | 0.739860851 | 0.058424218 |
| 30 | ANXA2 | 0.146349239 | 0.616638058 |
| 31 | OSBPL3 | 0.152722358 | 0.601511708 |
| 32 | PGM2 | 0.118557625 | 0.623031034 |
| 33 | GPSM2 | 0.383629843 | 0.33476749 |
| 34 | LOX | 0.360382257 | 0.338013248 |

| Table 25 Continued | | | |
|---|---|---|---|
| 35 | MT2A | 0.573954748 | 0.092486672 |
| 36 | ANTXR1 | 0.28181458 | 0.367424056 |
| 37 | CFL2 | 0.510797771 | 0.132589905 |
| 38 | SLC1A3 | 0.136463671 | 0.47632872 |
| 39 | LAMC2 | 0.440756292 | 0.17137219 |
| 40 | ARNTL2 | 0.580564993 | 0.020257406 |
| 41 | CXCL3 | 0.539864871 | 0.055029805 |
| 42 | SH3D19 | 0.341060493 | 0.243783108 |
| 43 | TWSG1 | 0.262413101 | 0.316726754 |
| 44 | DGKA | 0.276394576 | 0.29869685 |
| 45 | S100A10 | 0.197251624 | 0.325269434 |
| 46 | FBLIM1 | 0.311721134 | 0.207709929 |
| 47 | AKR1C3 | 0.165906393 | 0.349439013 |
| 48 | YBX1 | 0.095275332 | 0.416258017 |
| 49 | TBC1D1 | 0.317829189 | 0.192376681 |
| 50 | ADORA2B | 0.415433079 | 0.09213304 |

**Table 26. CTCF gene signature derived from MCF7.**

| | | | | |
|---|---|---|---|---|
| SNORA76C | BATF2 | IFI27 | RNVU1-19 | TAP1 |
| SNORA52 | HLA-A | OAS3 | DDX58 | LGALS9 |
| SNORA38B | RARRES3 | IFNB1 | ISG20 | HLA-C |
| SNORA44 | RSAD2 | SAMD9 | BST2 | C19orf66 |
| SNORA63 | SELL | APOL3 | ISG15 | CCL5 |
| SNORA71E | PARP14 | TRIM22 | IFNL2 | HLA-H |
| CHD3 | HLA-F | B2M | APOL6 | TRIM21 |
| OAS2 | IFITM1 | IFITM3 | TMEM140 | IFI6 |
| PSIP1 | TRANK1 | USP18 | ERAP1 | DEF8 |
| XAF1 | OASL | GBP3 | IFNL3 | IFIH1 |
| PSMB8 | OAS1 | MIR6753 | DENND4B | BTN3A3 |
| IFI44 | DDX60L | DDX60 | PARP9 | CFB |
| HLA-B | IFIT1 | PARP10 | LGALS3BP | |
| IFIT3 | APOL2 | PLEKHA4 | B4GALNT1 | |
| UBE2L6 | IFNL1 | MX2 | PCDHB8 | |
| PSMB9 | GBP1 | MX1 | C4orf33 | |
| CMPK2 | TAPSAR1 | OPTN | HERC5 | |
| IFIT2 | IFI35 | SCARNA1 | PLSCR1 | |
| HCP5 | IFI44L | PARP12 | PPA2 | |
| APOL1 | KRT83 | SAMHD1 | SNORA81 | |
| UBA7 | RTP4 | NLRC5 | IFITM2 | |
| TAP2 | CD74 | KRT80 | CTCF | |

**Table 27. Significantly enriched Hallmark pathways in TCGA/METABRIC basal CK high vs. low ER+ LumA tumors or *ESR1* mutant vs. WT metastatic tumors.**

| Cohorts | Pathway | FDR | Delta (BCKs-H-Low) / Delta (*ESR1* Mut-WT) |
|---|---|---|---|
| METABRIC | APICAL_JUNCTION | 5.83E-17 | 0.165737086 |
| | KRAS_SIGNALING_DN | 1.47E-08 | 0.063705276 |
| | COAGULATION | 1.97E-08 | 0.121500258 |
| | KRAS_SIGNALING_UP | 2.64E-07 | 0.131011177 |
| | MYOGENESIS | 2.80E-07 | 0.09880153 |
| | EPITHELIAL_MESENCHYMAL_TRANSITION | 4.95E-06 | 0.179430643 |
| | P53_PATHWAY | 6.14E-06 | 0.074673107 |
| | TNFA_SIGNALING_VIA_NFKB | 0.000246843 | 0.106823929 |
| | APICAL_SURFACE | 0.000246843 | 0.074190922 |
| | UV_RESPONSE_DN | 0.000273545 | 0.080000313 |
| | WNT_BETA_CATENIN_SIGNALING | 0.000388505 | 0.085470945 |
| | ESTROGEN_RESPONSE_LATE | 0.000476716 | 0.064901185 |
| | NOTCH_SIGNALING | 0.000604983 | 0.072911441 |
| | HYPOXIA | 0.000750845 | 0.061767452 |
| | IL2_STAT5_SIGNALING | 0.000871736 | 0.066543077 |
| | INFLAMMATORY_RESPONSE | 0.002189939 | 0.075654576 |
| | APOPTOSIS | 0.002833201 | 0.055982856 |
| | IL6_JAK_STAT3_SIGNALING | 0.005612071 | 0.063165342 |
| | TGF_BETA_SIGNALING | 0.00921643 | 0.064138132 |
| TCGA | KRAS_SIGNALING_DN | 6.36E-27 | 0.218665024 |
| | P53_PATHWAY | 1.41E-22 | 0.196074636 |
| | APICAL_JUNCTION | 1.20E-16 | 0.246090673 |
| | MYOGENESIS | 1.82E-13 | 0.235564311 |
| | COAGULATION | 8.05E-11 | 0.214338817 |
| | APICAL_SURFACE | 4.24E-10 | 0.159208523 |
| | HYPOXIA | 3.75E-09 | 0.148011399 |
| | WNT_BETA_CATENIN_SIGNALING | 3.93E-09 | 0.182519418 |
| | XENOBIOTIC_METABOLISM | 3.93E-09 | 0.119355883 |
| | NOTCH_SIGNALING | 7.25E-08 | 0.149981221 |
| | TNFA_SIGNALING_VIA_NFKB | 4.26E-07 | 0.204986489 |
| | KRAS_SIGNALING_UP | 1.29E-06 | 0.174943391 |
| | EPITHELIAL_MESENCHYMAL_TRANSITION | 3.82E-06 | 0.242480057 |
| | ESTROGEN_RESPONSE_LATE | 7.52E-06 | 0.10960509 |
| | IL2_STAT5_SIGNALING | 2.83E-05 | 0.119562833 |
| | APOPTOSIS | 0.000243387 | 0.09458993 |
| | ANGIOGENESIS | 0.000504489 | 0.14272267 |
| | PANCREAS_BETA_CELLS | 0.000642597 | 0.081838617 |
| | IL6_JAK_STAT3_SIGNALING | 0.001452421 | 0.121061723 |
| | HEDGEHOG_SIGNALING | 0.003349106 | 0.101208044 |

| | Table 27 Continued | | |
|---|---|---|---|
| | INFLAMMATORY_RESPONSE | 0.003567269 | 0.111540487 |
| | ALLOGRAFT_REJECTION | 0.009617507 | 0.108742751 |
| **WCRC/DFCI** | INTERFERON_GAMMA_RESPONSE | 0.037165375 | 0.615810891 |
| | TNFA_SIGNALING_VIA_NFKB | 0.037165375 | 0.614327554 |
| | IL6_JAK_STAT3_SIGNALING | 0.037165375 | 0.612826937 |
| | APOPTOSIS | 0.037165375 | 0.42066377 |
| | ALLOGRAFT_REJECTION | 0.039918033 | 0.610716922 |
| | REACTIVE_OXIGEN_SPECIES_PATHWAY | 0.039918033 | 0.469116263 |
| | IL2_STAT5_SIGNALING | 0.039918033 | 0.410461232 |
| | INFLAMMATORY_RESPONSE | 0.049453246 | 0.505107181 |
| | COMPLEMENT | 0.049453246 | 0.473673716 |
| | HYPOXIA | 0.049453246 | 0.387105337 |
| | P53_PATHWAY | 0.049453246 | 0.346993328 |

**Table 28. Fold changes of individual genes of ESTIMATE immune gene sets bwteen BCK hign vs low in TCGA and METABRIC ER+ LumA tumors and *ESR1* Mutant vs WT tumors from WCRC/DFCI cohort.**

| Rank | Genes | Log2FC in TCGA BCKs-High/Low | Log2FC in METABRIC BCKs-High/Low | Log2FC_WCRC-DFCI *ESR1* mut-WT |
|---|---|---|---|---|
| 1 | S100A9 | 1.601316269 | 1.077278992 | 1.53758206 |
| 2 | CD247 | 0.413538915 | 0.313209501 | 3.260166253 |
| 3 | S100A8 | 1.108094443 | 0.816634628 | 1.041584967 |
| 4 | IL32 | 0.447953288 | 0.173507576 | 2.269010156 |
| 5 | ZAP70 | 0.30933052 | 0.064415322 | 2.166670167 |
| 6 | IL7R | 0.494608888 | 0.492498407 | 1.550098637 |
| 7 | CD3D | 0.661204552 | 0.357350022 | 1.375990823 |
| 8 | LTB | 0.445871997 | 0.365330838 | 1.493328748 |
| 9 | GZMB | 0.390867579 | 0.286498881 | 1.541813814 |
| 10 | CD48 | 0.368438939 | 0.119951501 | 1.700054467 |
| 11 | CD52 | 0.786586045 | 0.354687773 | 1.04158955 |
| 12 | GBP2 | 0.688515323 | 0.477552994 | 1.01367327 |
| 13 | GNLY | 0.409300091 | 0.234401794 | 1.526046023 |
| 14 | ITK | 0.246050512 | 0.123968063 | 1.57280554 |
| 15 | TNFRSF1B | 0.396925588 | 0.065722408 | 1.471862944 |
| 16 | CCL5 | 0.518146036 | 0.398993117 | 0.949595289 |
| 17 | PRF1 | 0.334532898 | 0.089608258 | 1.382305512 |
| 18 | CORO1A | 0.285491461 | 0.009621764 | 1.505964093 |
| 19 | TRAF3IP3 | 0.065459507 | 0.085650471 | 1.623056979 |
| 20 | IL18RAP | 0.143439114 | 0.156641602 | 1.472153035 |
| 21 | LCK | 0.429308439 | 0.022671296 | 1.311784854 |
| 22 | NKG7 | 0.514854459 | 0.164698597 | 1.075522109 |
| 23 | FGR | 0.219103963 | 0.102644336 | 1.428534766 |
| 24 | PSTPIP1 | 0.321019731 | -0.002051992 | 1.423255236 |
| 25 | IL2RG | 0.49631738 | -0.00866051 | 1.204081073 |
| 26 | SRGN | 0.367878196 | 0.15625081 | 1.166245092 |
| 27 | IRF8 | 0.092805316 | 0.105755298 | 1.473277818 |
| 28 | CCDC69 | 0.414863541 | -0.015222741 | 1.269999748 |
| 29 | IL4R | 0.569515762 | 0.244280913 | 0.838528235 |
| 30 | PTPRCAP | 0.759690674 | 0.080668949 | 0.811921336 |
| 31 | GZMK | 0.620498773 | 0.219371657 | 0.809961166 |

| Table 28 Continued | | | | |
|---|---|---|---|---|
| 32 | CD69 | 0.426890693 | 0.026873741 | 1.193873714 |
| 33 | IL2RB | 0.072850229 | 0.174489495 | 1.357467094 |
| 34 | GMFG | 0.406831292 | 0.061980758 | 1.071485896 |
| 35 | CD2 | 0.436604239 | 0.270858867 | 0.814391975 |
| 36 | SELL | 0.27056852 | 0.126182829 | 1.079933706 |
| 37 | CCR7 | 0.523943424 | 0.198223411 | 0.753540883 |
| 38 | ITGB2 | -0.036451122 | -0.157794345 | 1.580813656 |
| 39 | RAC2 | 0.157589317 | 0.003528168 | 1.215357595 |
| 40 | LSP1 | 0.524954383 | 0.035907728 | 0.814152879 |
| 41 | MAFB | 0.422932974 | 0.146418418 | 0.79816295 |
| 42 | CD300A | 0.179850428 | -0.074569617 | 1.252541962 |
| 43 | KLRB1 | 0.763638645 | 0.400174101 | 0.188117369 |
| 44 | IL10RA | 0.001399892 | -0.005414777 | 1.350997914 |
| 45 | TAP1 | -0.039853169 | 0.042975626 | 1.318112202 |
| 46 | LST1 | 0.183991611 | -0.017446104 | 1.139036473 |
| 47 | TNFAIP3 | 0.240427094 | 0.11503523 | 0.918483104 |
| 48 | BCL2A1 | 0.384121807 | 0.001384382 | 0.878105278 |
| 49 | PTGER4 | 0.435353694 | 0.300976139 | 0.52416248 |
| 50 | IKZF1 | 0.093746249 | 0.081175821 | 1.084681183 |
| 51 | NCF4 | 0.196904796 | -0.002033078 | 1.058477077 |
| 52 | SELPLG | 0.022315529 | 0.006553574 | 1.169622954 |
| 53 | GZMH | 0.237348895 | 0.04134614 | 0.918964537 |
| 54 | HCK | -0.027015241 | -0.066014915 | 1.282120014 |
| 55 | SLA | 0.1688816 | -0.055418535 | 1.058068604 |
| 56 | LYZ | 0.164614559 | 0.121589451 | 0.849608571 |
| 57 | AOAH | -0.009678118 | 0.016284818 | 1.10723055 |
| 58 | HCLS1 | 0.161194138 | 0.021240663 | 0.925041434 |
| 59 | MSN | 0.147625576 | 0.067493467 | 0.875931609 |
| 60 | VAV1 | 0.038196608 | -0.05830908 | 1.094222521 |
| 61 | CD27 | 0.517945982 | 0.18992917 | 0.357055412 |
| 62 | CST7 | 0.552554503 | 0.095074567 | 0.352610006 |
| 63 | LGALS9 | 0.184304935 | -0.041708154 | 0.852539399 |
| 64 | VNN2 | -0.009733622 | 0.049256839 | 0.9509395 |
| 65 | EMP3 | 0.32725366 | 0.103869164 | 0.553635445 |
| 66 | LILRB2 | -0.110929568 | -0.057834864 | 1.136743429 |
| 67 | PLEK | -0.112302291 | -0.013070276 | 1.092491425 |
| 68 | ITGAL | -0.103595932 | -0.019702411 | 1.082769911 |
| 69 | CTSS | -0.15830633 | -0.05436723 | 1.164196917 |

| Table 28 Continued | | | | |
|---|---|---|---|---|
| 70 | GBP1 | 0.306642428 | 0.158010094 | 0.403435211 |
| 71 | NFKBIA | 0.244094096 | 0.043184478 | 0.578960614 |
| 72 | NCF2 | -0.053214081 | -0.017011721 | 0.934133142 |
| 73 | ADCY7 | -0.225493543 | -0.043171947 | 1.10158569 |
| 74 | WIPF1 | 0.166616961 | 0.058131724 | 0.601931364 |
| 75 | PTPRC | -0.086664214 | -0.000961853 | 0.848432265 |
| 76 | LY96 | 0.348222564 | 0.114996243 | 0.291827752 |
| 77 | CASP1 | 0.435685028 | 0.189975294 | 0.115356266 |
| 78 | BIN2 | -0.040076799 | -0.029664101 | 0.777286549 |
| 79 | PTPRE | 0.052307732 | 0.095929 | 0.55050269 |
| 80 | CD74 | 0.133992398 | -0.198348195 | 0.750915634 |
| 81 | DOCK2 | -0.148049725 | 0.028265491 | 0.751353988 |
| 82 | RGS1 | 0.410653222 | -0.110762176 | 0.321482591 |
| 83 | CD37 | 0.234075434 | -3.91E-06 | 0.378071758 |
| 84 | FCER1G | -0.165823364 | -0.137424113 | 0.905932481 |
| 85 | NCKAP1L | -0.167669879 | -0.111917447 | 0.882094989 |
| 86 | ADAM8 | -0.002578961 | -0.194937423 | 0.797556497 |
| 87 | IFI30 | 0.006959484 | -0.077487486 | 0.657006339 |
| 88 | LAIR1 | -0.109098299 | -0.053715198 | 0.703123245 |
| 89 | GMIP | -0.100471328 | -0.092916337 | 0.733619771 |
| 90 | CSTA | 0.831003213 | 0.025626805 | -0.352344022 |
| 91 | MICAL1 | -0.008580726 | -0.021722342 | 0.529356413 |
| 92 | TPP1 | -0.280443451 | -0.054567391 | 0.824654519 |
| 93 | CLEC2B | 0.385162981 | 0.031271805 | 0.073019901 |
| 94 | MYO1F | 0.024727532 | -0.015895136 | 0.477090705 |
| 95 | LILRB1 | -0.307436362 | 0.022646993 | 0.76805104 |
| 96 | GIMAP6 | 0.110554952 | 0.079743687 | 0.290189763 |
| 97 | CD53 | -0.019857942 | -0.037148079 | 0.511512591 |
| 98 | RASSF2 | 0.049744332 | -0.000285605 | 0.372763125 |
| 99 | RHOG | 0.159421653 | -0.030066877 | 0.265582503 |
| 100 | ALOX5AP | 0.01328876 | 0.084680576 | 0.257365815 |
| 101 | TCIRG1 | -0.09636752 | -0.294741868 | 0.732745882 |
| 102 | ARHGDIB | -0.005735016 | -0.107223773 | 0.4210453 |
| 103 | LPXN | -0.078881899 | -0.061030167 | 0.444070371 |
| 104 | LAPTM5 | -0.071383578 | -0.172590804 | 0.492103876 |
| 105 | ARHGAP15 | 0.166546341 | 0.036998446 | 0.033545917 |
| 106 | TPST2 | -0.060419552 | -0.277387023 | 0.57017827 |
| 107 | RNASE6 | 0.201820994 | 0.067995816 | -0.050496219 |

| Table 28 Continued | | | | |
|---|---|---|---|---|
| 108 | RAB27A | 0.037966255 | 0.117942743 | 0.048989411 |
| 109 | GIMAP4 | 0.146370323 | 0.09609722 | -0.06958235 |
| 110 | TYROBP | 0.096245207 | -0.150705218 | 0.208646306 |
| 111 | P2RY14 | 0.147893573 | 0.067277576 | -0.072024598 |
| 112 | GLRX | -0.235186071 | -0.310606809 | 0.668636716 |
| 113 | FYB | -0.23401642 | -0.090215527 | 0.441744351 |
| 114 | ITGA4 | -0.310385139 | -0.016782332 | 0.425359126 |
| 115 | SH2B3 | -0.253741255 | -0.123459462 | 0.444794294 |
| 116 | SAMHD1 | -0.263447731 | 0.013525164 | 0.294389029 |
| 117 | GPR65 | -0.236780289 | -0.133856668 | 0.412992526 |
| 118 | LCP2 | -0.127656186 | -0.017474997 | 0.182661868 |
| 119 | FLI1 | 0.171513857 | -0.030283454 | -0.106536506 |
| 120 | FGL2 | 0.089111467 | 0.152217724 | -0.254571763 |
| 121 | PTGER2 | 0.10249198 | 0.05757992 | -0.216474682 |
| 122 | MFSD1 | -0.281408984 | -0.218998951 | 0.362680941 |
| 123 | CYBB | -0.367756483 | -0.143036303 | 0.326398739 |
| 124 | CLEC4A | -0.023339541 | -0.037636507 | -0.27639 |
| 125 | RHOH | -0.052271631 | 0.021300047 | -0.316333661 |
| 126 | MNDA | -0.223893536 | -0.007124488 | -0.195601995 |
| 127 | RABGAP1L | -0.206707047 | -0.028284854 | -0.502761181 |
| 128 | EVI2B | -0.049447081 | -0.007165865 | -0.763019152 |
| 129 | ARHGEF6 | -0.193224295 | 0.00361767 | -0.689086054 |

## Bibliography


1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2020.* CA: a cancer journal for clinicians, 2020. **70**(1): p. 7-30.
2. Brody, J.G., et al., *Environmental pollutants and breast cancer: epidemiologic studies.* Cancer: Interdisciplinary International Journal of the American Cancer Society, 2007. **109**: p. 2667-2711.
3. Cancer, C.G.o.H.F.i.B., *Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies.* The lancet oncology, 2012. **13**(11): p. 1141-1151.
4. Picon‐Ruiz, M., et al., *Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention.* CA: a cancer journal for clinicians, 2017. **67**(5): p. 378-397.
5. Brewer, H.R., et al., *Family history and risk of breast cancer: an analysis accounting for family structure.* Breast cancer research and treatment, 2017. **165**(1): p. 193-200.
6. Michailidou, K., et al., *Association analysis identifies 65 new breast cancer risk loci.* Nature, 2017. **551**(7678): p. 92.
7. Chen, W.Y., et al., *Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk*. Jama, 2011. **306**(17): p. 1884-1890.
8. Sánchez-Zamorano, L.M., et al., *Healthy lifestyle on the risk of breast cancer.* Cancer Epidemiology and Prevention Biomarkers, 2011. **20**(5): p. 912-922.
9. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-752.
10. Sørlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proceedings of the National Academy of Sciences, 2001. **98**(19): p. 10869-10874.
11. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC genomics, 2006. **7**(1): p. 96.

12. Dai, X., et al., *Breast cancer intrinsic subtype classification, clinical use and future trends.* American journal of cancer research, 2015. **5**(10): p. 2929.
13. Basudan, A., et al., *Frequent ESR1 and CDK pathway copy-number alterations in metastatic breast cancer.* Molecular Cancer Research, 2019. **17**(2): p. 457-468.
14. Zhang, M.H., et al., *Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials.* Biomedical reports, 2014. **2**(1): p. 41-52.
15. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
16. Seol, H., et al., *Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance.* Modern pathology, 2012. **25**(7): p. 938-948.
17. Bertucci, F., P. Finetti, and D. Birnbaum, *Basal breast cancer: a complex and deadly molecular subtype.* Current molecular medicine, 2012. **12**(1): p. 96-110.
18. Network, C.G.A., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61.
19. Karn, T., et al., *Association between genomic metrics and immune infiltration in triple-negative breast cancer.* JAMA oncology, 2017. **3**(12): p. 1707-1711.
20. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* Journal of Clinical Oncology, 2009. **27**(8): p. 1160.
21. Makki, J., *Diversity of breast carcinoma: histological subtypes and clinical relevance.* Clinical Medicine Insights: Pathology, 2015. **8**: p. CPath. S31563.
22. Corben, A.D., *Pathology of invasive breast disease.* Surgical Clinics, 2013. **93**(2): p. 363-392.
23. Malhotra, G.K., et al., *Histological, molecular and functional subtypes of breast cancers.* Cancer biology & therapy, 2010. **10**(10): p. 955-960.
24. Weigelt, B., F.C. Geyer, and J.S. Reis-Filho, *Histological types of breast cancer: how special are they?* Molecular oncology, 2010. **4**(3): p. 192-208.
25. Ciriello, G., et al., *Comprehensive molecular portraits of invasive lobular breast cancer.* Cell, 2015. **163**(2): p. 506-519.

26. Berx, G. and F. Van Roy, *The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression.* Breast Cancer Research, 2001. **3**(5): p. 289.
27. Sobinsky, J.D., et al., *Unusual metastatic patterns of invasive lobular carcinoma of the breast*. Case reports in oncological medicine, 2013. **2013**.
28. Borst, M.J. and J.A. Ingold, *Metastatic patterns of invasive lobular versus invasive ductal carcinoma of the breast*. Surgery, 1993. **114**(4): p. 637-642.
29. Weigelt, B., J.L. Peterse, and L.J. Van't Veer, *Breast cancer metastasis: markers and models*. Nature Reviews Cancer, 2005. **5**(8): p. 591-602.
30. Redig, A.J. and S.S. McAllister, *Breast cancer as a systemic disease: a view of metastasis*. Journal of internal medicine, 2013. **274**(2): p. 113-126.
31. Weigelt, B., J.L. Peterse, and L.J. Van't Veer, *Breast cancer metastasis: markers and models*. Nature Reviews Cancer, 2005. **5**(8): p. 591.
32. Tomaskovic-Crook, E., E.W. Thompson, and J.P. Thiery, *Epithelial to mesenchymal transition and breast cancer*. Breast Cancer Research, 2009. **11**(6): p. 213.
33. Brabletz, T., et al., *EMT in cancer*. Nature Reviews Cancer, 2018. **18**(2): p. 128.
34. Kotiyal, S. and S. Bhattacharya, *Breast cancer stem cells, EMT and therapeutic targets*. Biochemical and biophysical research communications, 2014. **453**(1): p. 112-116.
35. Ye, X., et al., *Upholding a role for EMT in breast cancer metastasis*. Nature, 2017. **547**(7661): p. E1-E3.
36. Bill, R. and G. Christofori, *The relevance of EMT in breast cancer metastasis: Correlation or causality?* FEBS letters, 2015. **589**(14): p. 1577-1587.
37. Chaffer, C.L., et al., *EMT, cell plasticity and metastasis*. Cancer and Metastasis Reviews, 2016. **35**(4): p. 645-654.
38. Massagué, J. and A.C. Obenauf, *Metastatic colonization by circulating tumour cells*. Nature, 2016. **529**(7586): p. 298.
39. Cheung, K.J. and A.J. Ewald, *A collective route to metastasis: Seeding by tumor cell clusters*. Science, 2016. **352**(6282): p. 167-169.
40. Pantel, K. and M. Speicher, *The biology of circulating tumor cells*. Oncogene, 2016. **35**(10): p. 1216.

41. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis.* Cell, 2014. **158**(5): p. 1110-1122.
42. Cheung, K.J., et al., *Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters.* Proceedings of the National Academy of Sciences, 2016. **113**(7): p. E854-E863.
43. Liu, X., et al., *Homophilic CD44 interactions mediate tumor cell aggregation and polyclonal metastasis in patient-derived breast cancer models.* Cancer discovery, 2019. **9**(1): p. 96-113.
44. Gkountela, S., et al., *Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding.* Cell, 2019. **176**(1-2): p. 98-112. e14.
45. Mohme, M., S. Riethdorf, and K. Pantel, *Circulating and disseminated tumour cells—mechanisms of immune surveillance and escape.* Nature reviews Clinical oncology, 2017. **14**(3): p. 155.
46. Aggouraki, D., et al., *Correlation of circulating tumor cells (CTCs) expressing stemness and EMT phenotypes with immunosuppressive cells in metastatic breast cancer patients*. 2014, AACR.
47. Taranova, A.G., et al., *Allergic pulmonary inflammation promotes the recruitment of circulating tumor cells to the lung.* Cancer research, 2008. **68**(20): p. 8582-8589.
48. Strell, C., et al., *Neutrophil granulocytes promote the migratory activity of MDA-MB-468 human breast carcinoma cells via ICAM-1*. Experimental cell research, 2010. **316**(1): p. 138-148.
49. Spicer, J.D., et al., *Neutrophils promote liver metastasis via Mac-1 mediated interactions with circulating tumor cells.* Cancer research, 2012: p. canres. 2393.2011.
50. Huh, S.J., et al., *Transiently entrapped circulating tumor cells interact with neutrophils to facilitate lung metastasis development*. Cancer research, 2010: p. 0008-5472. CAN-09-4442.
51. Soysal, S.D., A. Tzankov, and S.E. Muenst, *Role of the tumor microenvironment in breast cancer.* Pathobiology, 2015. **82**(3-4): p. 142-152.

52. Yoneda, T. and T. Hiraga, *Crosstalk between cancer cells and bone microenvironment in bone metastasis.* Biochemical and biophysical research communications, 2005. **328**(3): p. 679-687.

53. Chen, Q., et al., *Carcinoma–astrocyte gap junctions promote brain metastasis by cGAMP transfer.* Nature, 2016. **533**(7604): p. 493-498.

54. Wang, H., et al., *The osteogenic niche promotes early-stage bone colonization of disseminated breast cancer cells.* Cancer cell, 2015. **27**(2): p. 193-210.

55. Hall, J.M., J.F. Couse, and K.S. Korach, *The multifaceted mechanisms of estradiol and estrogen receptor signaling.* Journal of Biological Chemistry, 2001. **276**(40): p. 36869-36872.

56. Jia, M., K. Dahlman-Wright, and J.-Å. Gustafsson, *Estrogen receptor alpha and beta in health and disease.* Best practice & research Clinical endocrinology & metabolism, 2015. **29**(4): p. 557-568.

57. Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835.

58. Kumar, R., et al., *The dynamic structure of the estrogen receptor.* Journal of amino acids, 2011. **2011**.

59. Margeat, E., et al., *The human estrogen receptor α dimer binds a single SRC-1 coactivator molecule with an affinity dictated by agonist structure.* Journal of molecular biology, 2001. **306**(3): p. 433-442.

60. Keller, H., et al., *Signaling cross-talk between peroxisome proliferator-activated receptor/retinoid X receptor and estrogen receptor through estrogen response elements.* Molecular Endocrinology, 1995. **9**(7): p. 794-804.

61. Seol, W., et al., *Inhibition of estrogen receptor action by the orphan receptor SHP (short heterodimer partner).* Molecular Endocrinology, 1998. **12**(10): p. 1551-1557.

62. Cowley, S.M., et al., *Estrogen receptors α and β form heterodimers on DNA.* Journal of Biological Chemistry, 1997. **272**(32): p. 19858-19862.

63. Schwabe, J.W., et al., *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements.* Cell, 1993. **75**(3): p. 567-578.

64. Yaşar, P., et al., *Molecular mechanism of estrogen–estrogen receptor signaling.* Reproductive medicine and biology, 2017. **16**(1): p. 4-20.
65. Shiau, A.K., et al., *The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen.* Cell, 1998. **95**(7): p. 927-937.
66. Montano, M.M., et al., *The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists.* Molecular Endocrinology, 1995. **9**(7): p. 814-825.
67. Carroll, J.S., et al., *Genome-wide analysis of estrogen receptor binding sites.* Nature genetics, 2006. **38**(11): p. 1289-1297.
68. Hurtado, A., et al., *FOXA1 is a key determinant of estrogen receptor function and endocrine response.* Nature genetics, 2011. **43**(1): p. 27.
69. Wright, T.M., et al., *Delineation of a FOXA1/ERα/AGR2 Regulatory Loop That Is Dysregulated in Endocrine Therapy–Resistant Breast Cancer.* Molecular Cancer Research, 2014. **12**(12): p. 1829-1839.
70. Fu, X., et al., *FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer.* Proceedings of the National Academy of Sciences, 2016. **113**(43): p. E6600-E6609.
71. Li, G., et al., *Chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing technology and application.* BMC genomics, 2014. **15**(12): p. S11.
72. Fullwood, M.J., et al., *An oestrogen-receptor-α-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.
73. Smith, I.E. and M. Dowsett, *Aromatase inhibitors in breast cancer.* New England Journal of Medicine, 2003. **348**(24): p. 2431-2442.
74. Muss, H.B., *Endocrine therapy for advanced breast cancer: a review.* Breast cancer research and treatment, 1992. **21**(1): p. 15-26.
75. Johnston, S.R. and M. Dowsett, *Aromatase inhibitors for breast cancer: lessons from the laboratory.* Nature Reviews Cancer, 2003. **3**(11): p. 821-831.
76. Lønning, P., *Lack of complete cross-resistance between different aromatase inhibitors; a real finding in search for an explanation?* European journal of cancer, 2009. **45**(4): p. 527-535.

77. McDonnell, D.P., S.E. Wardell, and J.D. Norris, *Oral selective estrogen receptor downregulators (SERDs), a breakthrough endocrine therapy for breast cancer*. 2015, ACS Publications.

78. Osborne, C., A. Wakeling, and R. Nicholson, *Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action.* British journal of cancer, 2004. **90**(1): p. S2-S6.

79. Weir, H.M., et al., *AZD9496: an oral estrogen receptor inhibitor that blocks the growth of ER-positive and ESR1-mutant breast tumors in preclinical models.* Cancer research, 2016. **76**(11): p. 3307-3318.

80. Garner, F., et al., *RAD1901: a novel, orally bioavailable selective estrogen receptor degrader that demonstrates antitumor activity in breast cancer xenograft models.* Anti-Cancer Drugs, 2015. **26**(9): p. 948.

81. Lai, A., et al., *Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts.* Journal of medicinal chemistry, 2015. **58**(12): p. 4888-4904.

82. Dutertre, M. and C.L. Smith, *Molecular mechanisms of selective estrogen receptor modulator (SERM) action.* Journal of Pharmacology and Experimental Therapeutics, 2000. **295**(2): p. 431-437.

83. Puhalla, S., S. Bhattacharya, and N.E. Davidson, *Hormonal therapy in breast cancer: a model disease for the personalization of cancer care.* Molecular oncology, 2012. **6**(2): p. 222-236.

84. Jordan, V.C., *Tamoxifen: a most unlikely pioneering medicine.* Nature reviews Drug discovery, 2003. **2**(3): p. 205-213.

85. Osborne, C.K. and R. Schiff, *Mechanisms of endocrine resistance in breast cancer.* Annual review of medicine, 2011. **62**: p. 233-247.

86. Musgrove, E.A. and R.L. Sutherland, *Biological determinants of endocrine resistance in breast cancer.* Nature Reviews Cancer, 2009. **9**(9): p. 631-643.

87. Normanno, N., et al., *Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer.* Endocrine-related cancer, 2005. **12**(4): p. 721-747.

88. Sighoko, D., et al., *Discordance in hormone receptor status among primary, metastatic, and second primary breast cancers: biological difference or misclassification?* The oncologist, 2014. **19**(6): p. 592.
89. Ellis, M.J., et al., *Outcome prediction for estrogen receptor–positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics.* Journal of the National Cancer Institute, 2008. **100**(19): p. 1380-1388.
90. Lei, J.T., et al., *Functional annotation of ESR1 gene fusions in estrogen receptor-positive breast cancer.* Cell reports, 2018. **24**(6): p. 1434-1444. e7.
91. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer.* Nature genetics, 2013. **45**(12): p. 1439.
92. Formisano, L., et al., *Aberrant FGFR signaling mediates resistance to CDK4/6 inhibitors in ER+ breast cancer.* Nature communications, 2019. **10**(1): p. 1-14.
93. Drago, J.Z., et al., *FGFR1 Amplification Mediates Endocrine Resistance but Retains TORC Sensitivity in Metastatic Hormone Receptor–Positive (HR+) Breast Cancer.* Clinical Cancer Research, 2019. **25**(21): p. 6443-6451.
94. Nayar, U., et al., *Acquired HER2 mutations in ER+ metastatic breast cancer confer resistance to estrogen receptor–directed therapies.* Nature genetics, 2019. **51**(2): p. 207-216.
95. Foley, J., et al. *EGFR signaling in breast cancer: bad to the bone*. in *Seminars in cell & developmental biology*. 2010. Elsevier.
96. Zhang, Y., et al., *Elevated insulin-like growth factor 1 receptor signaling induces antiestrogen resistance through the MAPK/ERK and PI3K/Akt signaling routes.* Breast Cancer Research, 2011. **13**(3): p. R52.
97. Murphy, C.G. and M.N. Dickler, *Endocrine resistance in hormone-responsive breast cancer: mechanisms and therapeutic strategies.* Endocrine-related cancer, 2016. **23**(8): p. R337-R352.
98. Fu, X., C.K. Osborne, and R. Schiff, *Biology and therapeutic potential of PI3K signaling in ER+/HER2-negative breast cancer.* The Breast, 2013. **22**: p. S12-S18.

99. Beaver, J.A. and B.H. Park, *The BOLERO-2 trial: the addition of everolimus to exemestane in the treatment of postmenopausal hormone receptor-positive advanced breast cancer.* Future oncology, 2012. **8**(6): p. 651-657.

100. Lange, C.A. and D. Yee, *Killing the second messenger: targeting loss of cell cycle control in endocrine-resistant breast cancer.* Endocrine Related Cancer, 2011. **18**(4): p. C19.

101. Sutherland, R.L. and E.A. Musgrove, *Cyclins and breast cancer.* Journal of mammary gland biology and neoplasia, 2004. **9**(1): p. 95-104.

102. Butt, A.J., et al., *Downstream targets of growth factor and oestrogen signalling and endocrine resistance: the potential roles of c-Myc, cyclin D1 and cyclin E.* Endocrine-related cancer, 2005. **12**(Supplement_1): p. S47-S59.

103. Abukhdeir, A.M. and B.H. Park, *P21 and p27: roles in carcinogenesis and drug resistance.* Expert reviews in molecular medicine, 2008. **10**.

104. Thangavel, C., et al., *Therapeutically activating RB: reestablishing cell cycle control in endocrine therapy-resistant breast cancer.* Endocrine-related cancer, 2011. **18**(3): p. 333-345.

105. Wardell, S.E., et al., *Efficacy of SERD/SERM Hybrid-CDK4/6 inhibitor combinations in models of endocrine therapy–resistant breast cancer.* Clinical Cancer Research, 2015. **21**(22): p. 5121-5130.

106. Maycotte, P. and A. Thorburn, *Targeting autophagy in breast cancer.* World journal of clinical oncology, 2014. **5**(3): p. 224.

107. Achinger-Kawecka, J., et al., *Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer.* Nature communications, 2020. **11**(1): p. 1-17.

108. Toy, W., et al., *Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists.* Cancer discovery, 2017. **7**(3): p. 277-287.

109. Harrod, A., et al., *Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer.* Oncogene, 2017. **36**(16): p. 2286.

110. Bahreini, A., et al., *Mutation site and context dependent effects of ESR1 mutation in genome-edited breast cancer cell models.* Breast Cancer Research, 2017. **19**(1): p. 60.

111. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer*. Nature genetics, 2013. **45**(12): p. 1439-1445.
112. Spoerke, J.M., et al., *Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant*. Nature communications, 2016. **7**.
113. Chu, D., et al., *ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients*. Clinical Cancer Research, 2016. **22**(4): p. 993-999.
114. Guttery, D.S., et al., *Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor–positive metastatic breast cancer*. Clinical chemistry, 2015. **61**(7): p. 974-982.
115. Fribbens, C., et al., *Plasma ESR1 mutations and the treatment of estrogen receptor–positive advanced breast cancer*. Journal of clinical oncology, 2016. **34**(25): p. 2961-2968.
116. Razavi, P., et al., *The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers*. Cancer cell, 2018. **34**(3): p. 427-438. e6.
117. Jeselsohn, R., et al., *Allele-specific chromatin recruitment and therapeutic vulnerabilities of ESR1 activating mutations*. Cancer cell, 2018. **33**(2): p. 173-186. e5.
118. Robinson, D.R., et al., *Integrative clinical genomics of metastatic cancer*. Nature, 2017. **548**(7667): p. 297.
119. Katzenellenbogen, J.A., et al., *Structural underpinnings of oestrogen receptor mutations in endocrine therapy resistance*. Nature Reviews Cancer, 2018. **18**(6): p. 377-388.
120. Oesterreich, S. and N.E. Davidson, *The search for ESR1 mutations in breast cancer*. Nature genetics, 2013. **45**(12): p. 1415-1416.
121. Jeselsohn, R., et al., *ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer*. Nature reviews Clinical oncology, 2015. **12**(10): p. 573.
122. Fuqua, S.A., G. Gu, and Y. Rechoum, *Estrogen receptor (ER) α mutations in breast cancer: hidden in plain sight*. Breast cancer research and treatment, 2014. **144**(1): p. 11-19.
123. Li, S., et al., *Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts*. Cell reports, 2013. **4**(6): p. 1116-1130.

124. Jeselsohn, R., et al., *Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor–positive breast cancer.* Clinical Cancer Research, 2014. **20**(7): p. 1757-1767.

125. Robinson, D.R., et al., *Activating ESR1 mutations in hormone-resistant metastatic breast cancer.* Nature genetics, 2013. **45**(12): p. 1446.

126. Merenbakh-Lamin, K., et al., *D538G mutation in estrogen receptor-α: A novel mechanism for acquired endocrine resistance in breast cancer.* Cancer research, 2013. **73**(23): p. 6856-6864.

127. Mao, C., et al., *Antiestrogen Resistant Cell Lines Expressing Estrogen Receptor α Mutations Upregulate the Unfolded Protein Response and are Killed by BHPI.* Scientific Reports, 2016. **6**.

128. Harrod, A., et al., *Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer.* Oncogene, 2016.

129. Yu, L., et al., *Estrogen-independent Myc overexpression confers endocrine therapy resistance on breast cancer cells expressing ERαY537S and ERαD538G mutations.* Cancer letters, 2019. **442**: p. 373-382.

130. Gelsomino, L., et al., *ESR1 mutations affect anti-proliferative responses to tamoxifen through enhanced cross-talk with IGF signaling.* Breast cancer research and treatment, 2016. **157**(2): p. 253-265.

131. Wang, P., et al., *Sensitive detection of mono-and polyclonal ESR1 mutations in primary tumors, metastatic lesions, and cell-free DNA of breast cancer patients.* Clinical cancer research, 2016. **22**(5): p. 1130-1137.

132. Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-918.

133. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-1183.

134. Kumar, V. and M. Jain, *The CRISPR–Cas system for plant genome editing: advances and opportunities.* Journal of experimental botany, 2015. **66**(1): p. 47-57.

135. Wang, P., et al., *Sensitive detection of mono-and polyclonal ESR1 mutations in primary tumors, metastatic lesions and cell free DNA of breast cancer patients.* Clinical Cancer Research, 2015: p. clincanres. 1534.2015.

136. Gustin, J.P., et al., *Knockin of mutant PIK3CA activates multiple oncogenic pathways.* Proceedings of the National Academy of Sciences, 2009. **106**(8): p. 2835-2840.

137. Rasband, W.S., *ImageJ*. 1997, Bethesda, MD.

138. Creighton, C.J., et al., *Insulin-like growth factor-I activates gene transcription programs strongly associated with poor breast cancer prognosis.* Journal of Clinical Oncology, 2008. **26**(25): p. 4078-4085.

139. Sikora, M.J., et al., *Endocrine response phenotypes are altered by charcoal-stripped serum variability.* Endocrinology, 2016. **157**(10): p. 3760-3766.

140. Fanning, S.W., et al., *Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation.* Elife, 2016. **5**: p. e12792.

141. Spoerke, J.M., et al., *Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant.* Nature communications, 2016. **7**(1): p. 1-10.

142. Fribbens, C., et al., *Plasma ESR1 mutations and the treatment of estrogen receptor-positive advanced breast cancer.* 2016.

143. Mauro, L., et al., *Estradiol increases IRS-1 gene expression and insulin signaling in breast cancer cells.* Biochemical and biophysical research communications, 2001. **288**(3): p. 685-689.

144. Molloy, C.A., F.E. May, and B.R. Westley, *Insulin receptor substrate-1 expression is regulated by estrogen in the MCF-7 human breast cancer cell line.* Journal of Biological Chemistry, 2000. **275**(17): p. 12565-12571.

145. Cicenas, J., *The potential role of Akt phosphorylation in human cancers.* Int J Biol Markers, 2008. **23**(1): p. 1-9.

146. Robey, R.B. and N. Hay. *Is Akt the "Warburg kinase"?—Akt-energy metabolism interactions and oncogenesis*. in *Seminars in cancer biology*. 2009. Elsevier.

147. Parisot, J., et al., *Altered expression of the IGF-1 receptor in a tamoxifen-resistant human breast cancer cell line.* British journal of cancer, 1999. **79**(5-6): p. 693.

148. Fagan, D.H., et al., *Acquired resistance to tamoxifen is associated with loss of the type I insulin-like growth factor receptor: implications for breast cancer treatment.* Cancer research, 2012. **72**(13): p. 3372-3380.

149. Chen, D., et al., *Phosphorylation of human estrogen receptor α at serine 118 by two distinct signal transduction pathways revealed by phosphorylation-specific antisera.* Oncogene, 2002. **21**(32): p. 4921-4931.

150. Chandarlapaty, S., et al., *Prevalence of ESR1 mutations in cell-free DNA and outcomes in metastatic breast cancer: a secondary analysis of the BOLERO-2 clinical trial.* JAMA oncology, 2016. **2**(10): p. 1310-1315.

151. O'Leary, B., et al., *The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial.* Cancer discovery, 2018. **8**(11): p. 1390-1403.

152. Ross, R.L., J.M. Askham, and M.A. Knowles, *PIK3CA mutation spectrum in urothelial carcinoma reflects cell context-dependent signaling and phenotypic outputs.* Oncogene, 2013. **32**(6): p. 768-776.

153. Jeselsohn, R., et al., *Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor–positive breast cancer.* Clinical Cancer Research, 2014. **20**(7): p. 1757-1767.

154. Jeselsohn, R., et al., *ESR1 mutations [mdash] a mechanism for acquired endocrine resistance in breast cancer.* Nature reviews Clinical oncology, 2015. **12**(10): p. 573-583.

155. Kotin, R.M., et al., *Site-specific integration by adeno-associated virus.* Proceedings of the National Academy of Sciences, 1990. **87**(6): p. 2211-2215.

156. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nature biotechnology, 2014. **32**(4): p. 347-355.

157. Manning, B.D. and L.C. Cantley, *AKT/PKB signaling: navigating downstream.* Cell, 2007. **129**(7): p. 1261-1274.

158. !!! INVALID CITATION !!!

159. Peterson, T.R., et al., *DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival.* Cell, 2009. **137**(5): p. 873-886.
160. Hunzicker-Dunn, M.E., et al., *PKA and GAB2 play central roles in the FSH signaling pathway to PI3K and AKT in ovarian granulosa cells.* Proceedings of the National Academy of Sciences, 2012. **109**(44): p. E2979-E2988.
161. Baserga, R., F. Peruzzi, and K. Reiss, *The IGF-1 receptor in cancer biology.* International journal of cancer, 2003. **107**(6): p. 873-877.
162. Alessi, D.R., et al., *Mechanism of activation of protein kinase B by insulin and IGF-1.* The EMBO journal, 1996. **15**(23): p. 6541.
163. Mulvihill, M.J., et al., *Discovery of OSI-906: a selective and orally efficacious dual inhibitor of the IGF-1 receptor and insulin receptor.* Future, 2009. **1**(6): p. 1153-1171.
164. DeSantis, C.E., et al., *Breast cancer statistics, 2017, racial disparity in mortality by state.* CA: a cancer journal for clinicians, 2017. **67**(6): p. 439-448.
165. Turner, N.C., et al., *Advances in the treatment of advanced oestrogen-receptor-positive breast cancer.* The Lancet, 2017. **389**(10087): p. 2403-2414.
166. Spoerke, J.M., et al., *Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant.* Nature communications, 2016. **7**: p. 11579.
167. Robinson, D.R., et al., *Activating ESR1 mutations in hormone-resistant metastatic breast cancer.* Nature genetics, 2013. **45**(12): p. 1446-1451.
168. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Science translational medicine, 2015. **7**(313): p. 313ra182-313ra182.
169. Fanning, S.W., et al., *Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation.* Elife, 2016. **5**.
170. Yu, L., et al., *Estrogen-Independent Myc Overexpression Confers Endocrine Therapy Resistance on Breast Cancer Cells Expressing ERαY537S and ERαD538G Mutations.* Cancer letters, 2018.

171. Zhang, K., et al., *Clinical value of circulating ESR1 mutations for patients with metastatic breast cancer: a meta-analysis.* Cancer management and research, 2018. **10**: p. 2573.
172. Varešlija, D., et al., *Transcriptome characterization of matched primary breast and brain metastatic tumors to detect novel actionable targets.* JNCI: Journal of the National Cancer Institute, 2018.
173. Priedigkeit, N., et al., *Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases.* JCI insight, 2017. **2**(17).
174. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nature methods, 2017. **14**(4): p. 417.
175. Soneson, C., M.I. Love, and M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.* F1000Research, 2015. **4**.
176. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-140.
177. Yu, M., et al., *Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility.* Science, 2014. **345**(6193): p. 216-220.
178. Miles, F.L., J.E. Lynch, and R.A. Sikes, *Cell-based assays using calcein acetoxymethyl ester show variation in fluorescence with treatment conditions.* Journal of biological methods, 2015. **2**(3).
179. Jimenez, M.A., et al., *Quantitative microfluidic fluorescence microscopy to study vaso-occlusion in sickle cell disease.* haematologica, 2015. **100**(10): p. e390-e393.
180. Jiang, S., et al., *Scaffold attachment factor SAFB1 suppresses estrogen receptor α-mediated transcription in part via interaction with nuclear receptor corepressor.* Molecular Endocrinology, 2006. **20**(2): p. 311-320.
181. Benjamin, J.M., et al., *αE-catenin regulates actin dynamics independently of cadherin-mediated cell–cell adhesion.* The Journal of cell biology, 2010. **189**(2): p. 339-352.
182. Sikora, M.J., et al., *Invasive lobular carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and altered tamoxifen response.* Cancer research, 2014: p. canres. 2779.2013.

183. Langmead, B., *Aligning short sequencing reads with Bowtie*. Current protocols in bioinformatics, 2010. **32**(1): p. 11.7. 1-11.7. 14.
184. Feng, J., et al., *Identifying ChIP-seq enrichment using MACS*. Nature protocols, 2012. **7**(9): p. 1728.
185. Stark, R. and G. Brown, *DiffBind: differential binding analysis of ChIP-Seq peak data*. R package version, 2011. **100**: p. 4.3.
186. Lerdrup, M., et al., *An interactive environment for agile analysis and visualization of ChIP-sequencing data*. Nature structural & molecular biology, 2016. **23**(4): p. 349.
187. Yu, G., L.-G. Wang, and Q.-Y. He, *ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization*. Bioinformatics, 2015. **31**(14): p. 2382-2383.
188. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. Bioinformatics, 2010. **26**(6): p. 841-842.
189. Bailey, T.L., et al., *The MEME suite*. Nucleic acids research, 2015. **43**(W1): p. W39-W49.
190. Anderson, A.P. and A.G. Jones, *erefinder: Genome‐wide detection of oestrogen response elements*. Molecular ecology resources, 2019. **19**(5): p. 1366-1373.
191. Blanchard, Z., et al., *Estrogen-independent molecular actions of mutant estrogen receptor alpha in endometrial cancer*. bioRxiv, 2019: p. 369165.
192. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. Genome biology, 2014. **15**(12): p. 550.
193. Krämer, A., et al., *Causal analysis approaches in ingenuity pathway analysis*. Bioinformatics, 2013. **30**(4): p. 523-530.
194. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
195. Hänzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC bioinformatics, 2013. **14**(1): p. 7.
196. Kassambara, A., *ggpubr:"ggplot2" based publication ready plots*. R package version 0.1, 2017. **6**.
197. Chen, H. and P.C. Boutros, *VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R*. BMC bioinformatics, 2011. **12**(1): p. 35.

198. Ianevski, A., et al., *SynergyFinder: a web application for analyzing drug combination dose–response matrix data.* Bioinformatics, 2017. **33**(15): p. 2413-2415.
199. Soneson, C., M.I. Love, and M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.* F1000Res, 2015. **4**: p. 1521.
200. Zundelevich, A., et al., *ESR1 mutations are frequent in newly diagnosed metastatic and loco-regional recurrence of endocrine-treated breast cancer and carry worse prognosis.* Breast Cancer Research, 2020. **22**(1): p. 1-11.
201. Priedigkeit, N., et al., *Intrinsic subtype switching and acquired ERBB2/HER2 amplifications and mutations in breast cancer brain metastases.* JAMA oncology, 2017. **3**(5): p. 666-671.
202. Saias, L., et al., *Cell–cell adhesion and cytoskeleton tension oppose each other in regulating tumor cell aggregation.* Cancer research, 2015. **75**(12): p. 2426-2433.
203. Luo, H., et al., *Cancer-associated fibroblasts: a multifaceted driver of breast cancer progression.* Cancer letters, 2015. **361**(2): p. 155-163.
204. Ungefroren, H., et al., *Interaction of tumor cells with the microenvironment*. Cell Communication and Signaling, 2011. **9**(1): p. 18.
205. Friedl, P., et al., *Classifying collective cancer cell invasion.* Nature cell biology, 2012. **14**(8): p. 777.
206. Kang, S.-A., S. Bajana, and T. Tanaka, *In vitro Flow Adhesion Assay for Analyzing Shear-resistant Adhesion of Metastatic Cancer Cells to Endothelial Cells.* Bio-protocol, 2016. **6**(4).
207. Sáez, J.C. and L. Leybaert, *Hunting for connexin hemichannels.* FEBS letters, 2014. **588**(8): p. 1205-1211.
208. Nekrasova, O. and K.J. Green, *Desmosome assembly and dynamics.* Trends in cell biology, 2013. **23**(11): p. 537-546.
209. Levine, K.M., et al., *FGFR4 overexpression and hotspot mutations in metastatic ER+ breast cancer are enriched in the lobular subtype.* npj Breast Cancer, 2019. **5**(1): p. 19.
210. Cohen, O., et al., *Abstract S1-01: Whole exome and transcriptome sequencing of resistant ER+ metastatic breast cancer*. 2017, AACR.

211. Joseph, R., et al., *Integrative model of genomic factors for determining binding site selection by estrogen receptor-α*. Molecular systems biology, 2010. **6**(1): p. 456.

212. Oyamada, M., K. Takebe, and Y. Oyamada, *Regulation of connexin expression by transcription factors and epigenetic mechanisms.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2013. **1828**(1): p. 118-133.

213. Medvedeva, Y.A., et al., *EpiFactors: a comprehensive database of human epigenetic factors and complexes.* Database, 2015. **2015**.

214. Nigam, A., et al., *Loss of cell-cell and cell-matrix adhesion molecules in colorectal cancer.* British journal of cancer, 1993. **68**(3): p. 507.

215. Rege, T.A. and J.S. Hagood, *Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis.* The FASEB journal, 2006. **20**(8): p. 1045-1054.

216. Sparano, J.A., et al., *Randomized phase III trial of marimastat versus placebo in patients with metastatic breast cancer who have responding or stable disease after first-line chemotherapy: Eastern Cooperative Oncology Group trial E2196.* Journal of Clinical Oncology, 2004. **22**(23): p. 4683-4690.

217. Miller, K., et al., *A randomized phase II pilot trial of adjuvant marimastat in patients with early-stage breast cancer.* Annals of oncology, 2002. **13**(8): p. 1220-1224.

218. Nyberg, P., et al., *MMP-9 activation by tumor trypsin-2 enhances in vivo invasion of human tongue carcinoma cells.* Journal of dental research, 2002. **81**(12): p. 831-835.

219. Mitchell, C.B. and G.M. O'Neill, *Cooperative cell invasion: matrix metalloproteinase–mediated incorporation between cells.* Molecular biology of the cell, 2016. **27**(21): p. 3284-3292.

220. Zardavas, D., J. Baselga, and M. Piccart, *Emerging targeted agents in metastatic breast cancer.* Nature reviews Clinical oncology, 2013. **10**(4): p. 191.

221. Liu, J., et al., *Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974.* Proceedings of the National Academy of Sciences, 2013. **110**(50): p. 20224-20229.

222. Arnesen, S., et al., *Estrogen receptor alpha mutations in breast cancer cells cause gene expression changes through constant activity and through secondary effects.* bioRxiv, 2020.

223. Swinstead, E.E., et al., *Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions.* Cell, 2016. **165**(3): p. 593-605.
224. Gertz, J., et al., *Distinct properties of cell-type-specific and shared transcription factor binding sites.* Molecular cell, 2013. **52**(1): p. 25-36.
225. Toska, E., et al., *PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D.* Science, 2017. **355**(6331): p. 1324-1330.
226. Li, Z., et al., *Upregulation of IRS1 enhances IGF1 response in Y537S and D538G ESR1 mutant breast cancer cells.* Endocrinology, 2017. **159**(1): p. 285-296.
227. Zhang, Q., et al., *Abstract P4-01-04: ESR1 mutation in cell free DNA (cfDNA) is associated with significantly increased circulating tumor cell (CTC)-clusters and progress in stage III/IV breast cancer after systemic treatments*. 2019, AACR.
228. Aasen, T., et al., *Gap junctions and cancer: communicating for 50 years.* Nature Reviews Cancer, 2016. **16**(12): p. 775.
229. Solan, J.L. and P.D. Lampe, *Key connexin 43 phosphorylation events regulate the gap junction life cycle.* Journal of Membrane Biology, 2007. **217**(1-3): p. 35-41.
230. Solan, J.L. and P.D. Lampe, *Connexin43 phosphorylation: structural changes and biological effects.* Biochemical Journal, 2009. **419**(2): p. 261-272.
231. Solan, J.L. and P.D. Lampe, *Specific Cx43 phosphorylation events regulate gap junction turnover in vivo.* FEBS letters, 2014. **588**(8): p. 1423-1429.
232. Xu, X., et al., *Antimalarial drug mefloquine inhibits nuclear factor kappa B signaling and induces apoptosis in colorectal cancer cells.* Cancer science, 2018. **109**(4): p. 1220-1229.
233. Yulyana, Y., et al., *Carbenoxolone enhances TRAIL-induced apoptosis through the upregulation of death receptor 5 and inhibition of gap junction intercellular communication in human glioma.* Stem cells and development, 2013. **22**(13): p. 1870-1882.
234. Murphy, P.A., S. Begum, and R.O. Hynes, *Tumor angiogenesis in the absence of fibronectin or its cognate integrin receptors.* PloS one, 2015. **10**(3): p. e0120872.
235. Chen, C., et al., *The biology and role of CD44 in cancer progression: therapeutic implications.* Journal of hematology & oncology, 2018. **11**(1): p. 64.

236. Jia, S., et al., *Clinically Observed Estrogen Receptor Alpha Mutations within the Ligand-Binding Domain Confer Distinguishable Phenotypes.* Oncology, 2018. **94**(3): p. 176-189.
237. Simpson, K.J., et al., *Identification of genes that regulate epithelial cell migration using an siRNA screening approach.* Nature cell biology, 2008. **10**(9): p. 1027.
238. Tie, F., et al., *CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing.* Development, 2009. **136**(18): p. 3131-3141.
239. Dorighi, K.M., et al., *Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation.* Molecular cell, 2017. **66**(4): p. 568-576. e4.
240. Reis-Filho, J.S. and L. Pusztai, *Gene expression profiling in breast cancer: classification, prognostication, and prediction.* The Lancet, 2011. **378**(9805): p. 1812-1823.
241. Arpino, G., et al., *Gene expression profiling in breast cancer: a clinical perspective.* The Breast, 2013. **22**(2): p. 109-120.
242. Thorsson, V., et al., *The immune landscape of cancer.* Immunity, 2018. **48**(4): p. 812-830. e14.
243. Kennecke, H., et al., *Metastatic behavior of breast cancer subtypes.* Journal of Clinical Oncology, 2010. **28**(20): p. 3271-3277.
244. Lee, U., et al., *A prognostic gene signature for metastasis-free survival of triple negative breast cancer patients.* PLoS one, 2013. **8**(12).
245. Velasco-Velázquez, M. and R.G. Pestell, *The CCL5/CCR5 axis promotes metastasis in basal breast cancer.* Oncoimmunology, 2013. **2**(4): p. e23660.
246. Changavi, A.A., A. Shashikala, and A.S. Ramji, *Epidermal growth factor receptor expression in triple negative and nontriple negative breast carcinomas.* Journal of laboratory physicians, 2015. **7**(2): p. 79.
247. de Graauw, M., et al., *Annexin A1 regulates TGF-β signaling and promotes metastasis formation of basal-like breast cancer cells.* Proceedings of the National Academy of Sciences, 2010. **107**(14): p. 6340-6345.
248. Soliman, H., F. Khalil, and S. Antonia, *PD-L1 expression is increased in a subset of basal type breast cancer cells.* PLoS one, 2014. **9**(2).

249. Emens, L.A., et al., *IMpassion130: a phase III randomized trial of atezolizumab with nab-paclitaxel for first-line treatment of patients with metastatic triple-negative breast cancer (mTNBC)*. 2016, American Society of Clinical Oncology.
250. Hollestelle, A., et al., *Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines.* Breast cancer research and treatment, 2010. **121**(1): p. 53-64.
251. Adélaïde, J., et al., *Integrated profiling of basal and luminal breast cancers.* Cancer research, 2007. **67**(24): p. 11565-11575.
252. Kim, C., et al., *Changes in intrinsic subtype of breast cancer during tumor progression in the same patient*. International journal of clinical and experimental pathology, 2015. **8**(11): p. 15184.
253. Cicatiello, L., et al., *Estrogen receptor α controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs.* The American journal of pathology, 2010. **176**(5): p. 2113-2130.
254. Jambal, P., et al., *Estrogen switches pure mucinous breast cancer to invasive lobular carcinoma with mucinous features.* Breast cancer research and treatment, 2013. **137**(2): p. 431-448.
255. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603.
256. Marcotte, R., et al., *Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance.* Cell, 2016. **164**(1-2): p. 293-309.
257. Daemen, A., et al., *Modeling precision treatment of breast cancer.* Genome biology, 2013. **14**(10): p. R110.
258. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 2005, Springer. p. 397-420.
259. Kassambara, A., et al., *Package 'survminer'*. 2017.
260. Soetaert, K., *plot3D: Tools for plotting 3-D and 2-D data.* R package version, 2014: p. 10-2.
261. Wang, X. and M. Li, *Correlate tumor mutation burden with immune signatures in human cancers.* BMC immunology, 2019. **20**(1): p. 4.

262. Mayakonda, A., et al., *Maftools: efficient and comprehensive analysis of somatic variants in cancer.* Genome research, 2018. **28**(11): p. 1747-1756.
263. Saltz, J., et al., *Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images.* Cell reports, 2018. **23**(1): p. 181-193. e7.
264. Charafe-Jauffret, E., et al., *Gene expression profiling of breast cell lines identifies potential new basal markers.* Oncogene, 2006. **25**(15): p. 2273.
265. Huper, G. and J.R. Marks, *Isogenic normal basal and luminal mammary epithelial isolated by a novel method show a differential response to ionizing radiation.* Cancer research, 2007. **67**(7): p. 2990-3001.
266. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-1740.
267. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(14): p. 8418-8423.
268. Need, E.F., et al., *The unique transcriptional response produced by concurrent estrogen and progesterone treatment in breast cancer cells results in upregulation of growth factor pathways and switching from a Luminal A to a Basal-like subtype.* BMC cancer, 2015. **15**(1): p. 791.
269. Creighton, C.J., et al., *Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors.* Genome biology, 2006. **7**(4): p. R28.
270. Sikora, M.J., et al., *Invasive lobular carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and altered tamoxifen response.* Cancer research, 2014. **74**(5): p. 1463-1474.
271. Fiorito, E., et al., *CTCF modulates Estrogen Receptor function through specific chromatin and nuclear matrix interactions.* Nucleic acids research, 2016. **44**(22): p. 10588-10602.
272. Tang, Z., et al., *CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription.* Cell, 2015. **163**(7): p. 1611-1627.

273. Kabos, P., et al., *Cytokeratin 5 positive cells represent a steroid receptor negative and therapy resistant subpopulation in luminal breast cancers.* Breast cancer research and treatment, 2011. **128**(1): p. 45-55.

274. Axlund, S.D., et al., *Progesterone-inducible cytokeratin 5-positive cells in luminal breast cancer exhibit progenitor properties.* Hormones and Cancer, 2013. **4**(1): p. 36-49.

275. Yin, P., et al., *Genome-wide progesterone receptor binding: cell type-specific and shared mechanisms in T47D breast cancer cells and primary leiomyoma cells.* PLoS one, 2012. **7**(1): p. e29021.

276. Wardell, S.E., et al., *Partial agonist activity of the progesterone receptor antagonist RU486 mediated by an amino-terminal domain coactivator and phosphorylation of serine400.* Molecular Endocrinology, 2010. **24**(2): p. 335-345.

277. Zhang, S., J. Jonklaas, and M. Danielsen, *The glucocorticoid agonist activities of mifepristone (RU486) and progesterone are dependent on glucocorticoid receptor levels but not on EC50 values.* Steroids, 2007. **72**(6-7): p. 600-608.

278. Alshareeda, A.T., et al., *Characteristics of basal cytokeratin expression in breast cancer.* Breast cancer research and treatment, 2013. **139**(1): p. 23-37.

279. Abd El‐Rehim, D.M., et al., *Expression of luminal and basal cytokeratins in human breast carcinoma.* The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 2004. **203**(2): p. 661-671.

280. Rakha, E., et al., *Breast carcinoma with basal differentiation: a proposal for pathology definition based on basal cytokeratin expression.* Histopathology, 2007. **50**(4): p. 434-438.

281. Davoli, T., et al., *Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy.* Science, 2017. **355**(6322): p. eaaf8399.

282. Tamborero, D., et al., *A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations.* Clinical Cancer Research, 2018. **24**(15): p. 3717-3728.

283. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nature communications, 2013. **4**: p. 2612.

284. Gebhardt, C., et al., *S100A8 and S100A9 in inflammation and cancer.* Biochemical pharmacology, 2006. **72**(11): p. 1622-1631.

285. Ryckman, C., et al., *Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion.* The Journal of Immunology, 2003. **170**(6): p. 3233-3242.

286. Cheriyath, V., D.W. Leaman, and E.C. Borden, *Emerging roles of FAM14 family members (G1P3/ISG 6–16 and ISG12/IFI27) in innate immunity and cancer.* Journal of Interferon & Cytokine Research, 2011. **31**(1): p. 173-181.

287. Lee, H.R., et al., *Brahma-related gene 1-associated expression of 9-27 and IFI-27 is involved in acquired cisplatin resistance of gastric cancer cells.* Molecular medicine reports, 2013. **8**(3): p. 747-750.

288. Yu, K.S., et al., *The protease inhibitor, elafin, induces p53 - dependent apoptosis in human melanoma cells.* International journal of cancer, 2010. **127**(6): p. 1308-1320.

289. Ariazi, E.A., et al., *A new role for ERα: Silencing via DNA methylation of basal, stem cell, and EMT genes.* Molecular Cancer Research, 2017. **15**(2): p. 152-164.

290. Heizmann, C.W., G. Fritz, and B. Schafer, *S100 proteins: structure, functions and pathology.* Front Biosci, 2002. **7**(1): p. 1356-68.

291. Sinha, P., et al., *Proinflammatory S100 proteins regulate the accumulation of myeloid-derived suppressor cells.* The Journal of Immunology, 2008. **181**(7): p. 4666-4675.

292. Wang, L., et al., *Increased myeloid-derived suppressor cells in gastric cancer correlate with cancer stage and plasma S100A8/A9 proinflammatory proteins.* The Journal of Immunology, 2013. **190**(2): p. 794-804.

293. Björk, P., et al., *Identification of human S100A9 as a novel target for treatment of autoimmune disease via binding to quinoline-3-carboxamides.* PLoS biology, 2009. **7**(4).