A Novel Group of Genes that Causes Endocrine Resistance in Breast Cancer Identified by Dynamic Gene Expression Analysis

Arvand Asghari

University of Houston

Katherine Wall

The University of Texas Health Science Center at Houston

Michael Gill

The University of Texas Health Science Center at Houston

Natascha Del Vecchio

University of Chicago

Farnaz Allahbakhsh

University of Houston

Jacky Wu

University of Houston

Nan Deng

The University of Texas MD Anderson Cancer Center

W. Jim Zheng

The University of Texas Health Science Center at Houston

Hulin Wu

The University of Texas Health Science Center at Houston

Michihisa Umetani (✉ mumetani@uh.edu)

University of Houston

Vahed Maroufy

The University of Texas Health Science Center at Houston

Research Article

Keywords: Breast cancer, Endocrine therapy, estrogen receptor, gene, TNBC

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1047808/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Oncotarget on April 6th, 2022. See the published version at <https://doi.org/10.18632/oncotarget.28225>.

Abstract

Breast cancer (BC) is the most common type of cancer and the second leading cancer-related cause of death in women worldwide. Endocrine therapy is an effective therapeutic approach for estrogen receptor (ER)-positive breast cancer; however, in many cases, tumor regrowth occurs after the therapy and the tumor becomes unresponsive anymore. While some gene mutations contribute to the resistance in some patients, the underlying causes of the resistance to endocrine therapies are mostly undetermined. In this study, we utilized our recently developed statistical approach to investigate the dynamic behavior of gene expression during the development of endocrine resistance and identified a novel group of genes that can be crucial to the development of resistance in BC. The expression of these genes is not only altered in cell models during the endocrine resistance development but also significantly changed in endocrine-resistant patients. Surprisingly, this group of genes was also identified as a group of key candidate genes in triple-negative breast cancer (TNBC), suggesting that endocrine resistance and TNBC share the same mechanisms during their development. Our findings explain some of the genetic underlying reasons for endocrine resistance and provide the potential to develop novel common therapeutic approaches against endocrine-resistant BC and TNBC.

Introduction

Breast cancer (BC) is the most prevalent type of cancer in women and the leading cause of cancer-related deaths among US women aged 20 to 59 ¹. More than 280 000 new US cases of breast cancer and 43 000 deaths are projected for 2021 ¹. While current five-year survival rates of BC have reached 90%, it is still the second leading cause of cancer-related deaths in women overall ¹. Treatment options for breast cancers typically involve surgical resection of the tumor followed by drug treatment based on the cancer type ². Breast cancers are categorized into subtypes based on hormone receptors (HR), namely estrogen receptor (ER) and progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression status. Nearly 70% of breast cancers express the estrogen receptor (ER⁺) without overexpression of HER2 (ER⁺/HER2-negative BC, luminal A breast cancer) ². While 13% of breast cancers are ER⁺/HER2⁺ (luminal B), more than 13% are ER⁻/HER2⁻, also referred to as triple-negative breast cancer (TNBC). While five-year survival rates of luminal A breast cancer are around 90%, the five-year survival rate of TNBC is much lower ².

ER exists in two isoforms, ER α (*ESR1*) and ER β (*ESR2*), with ER α as the dominant form in BC. ER acts as a transcription factor mediating gene expression and also as a signaling molecule, inducing kinase pathways and regulating cell growth in cultured breast cancer cells ³. Given that 80% of breast cancers express ER, ER-targeted endocrine therapies are a core component of systemic therapy. Endocrine therapies include selective ER modulators (SERMs) such as tamoxifen, selective ER down-regulators (SERDS) such as fulvestrant, and aromatase inhibitors (AIs) targeting estrogen biosynthesis. Endocrine therapies have been successful at improving cancer outcomes; however, the development of endocrine resistance, or resistance to inhibition of ER actions, remains a roadblock in breast cancer treatment. Many

patients have intrinsic resistance to endocrine therapies. Only 30% of patients with metastatic disease saw initial regression with endocrine therapies, however, in almost all of them the resistance developed eventually, and the tumor recurred not long after that. Moreover, more than 20% of patients who present with early breast cancer will develop endocrine resistance throughout treatment ^{4,5}.

Resistance to endocrine therapies has been traced to mutations in *ESR1*, alterations in receptor tyrosine kinases such as HER2, and alterations in signaling pathways such as the MAPK pathway. Point mutations in the ligand-binding domain (LBD) of *ESR1*, have been shown to cause endocrine resistance ⁶⁻⁸. The most common mutations, Y537 and D538, lead to constitutive ligand-independent activation of ER α ⁹. These mutants are less sensitive to fulvestrant or tamoxifen ⁷. LBD mutations in *ESR1* have been found in around 20% of metastatic ER⁺ cancers after endocrine therapy ^{7,10}. *ESR1* mutations in circulating tumor DNA (ctDNA) have been found in 36% of patients with metastatic tumors following AI treatment ¹¹. Gene fusions of the *ESR1* DNA binding domain to the C-terminus of other proteins, though rare, can promote ligand-independent ER α activity ¹².

HER2 mutations occur in 2.4% of primary tumors and 6.7% of metastatic tumors ⁹. Activating mutations of *HER2* can confer resistance to estrogen deprivation and fulvestrant treatment; fulvestrant sensitivity can be restored with HER2 inhibitors ¹³. HER2 amplification is less common, occurring in 0.8% of primary and 2.1% of metastatic tumors ⁹. Amplification of HER2 has been linked to tamoxifen resistance, through the hyperactivation of MAPK ¹⁴. Currently, ER⁺/HER2⁺ breast cancers are treated with both antiestrogens and HER2 inhibitors ⁹. However, for patients with endocrine-resistant breast cancers, treatments are limited as the development of endocrine resistance is not well understood. Existing scientific literature focuses on the role of *ESR1*, receptor tyrosine kinases, and their signaling pathways; however, genetic mutation of these genes comprises a small percentage of cases. In at least 60% of cases, additional factors must play important roles in the resistance development process as those cases show intact *ESR1* and no up- or down-regulation in other signaling pathways ¹⁰.

In this study, we explored the dynamic behavior of the entire gene population to identify novel genes that play fundamental roles in the development and progression of endocrine-resistant breast cancer. We used the (i) time-course gene expression patterns in estrogen deprived MCF7 cells, which are widely accepted as undergoing the process of developing endocrine resistance over time, (ii) patient data from endocrine-resistant tumors compared to endocrine-sensitive ones, and (iii) patient tumor data in TNBC compared to luminal A breast cancer ^{15,16} to compare the similarities among different types of BC and also to identify key candidate genes in such breast tumors.

Results

LTED Cells are the Cell-Based Model of Endocrine resistance Closest to Patients Data among Breast Cancer Cell Models

First, we identified the best cell-based model of endocrine resistance based on their similarities to patients' tumor data. For this purpose, we compared the gene expression patterns from cell models of endocrine resistance before and after establishment, utilizing datasets GSE20361 and GSE111151, and gene expression patterns of resistant and sensitive tumors from patients using the GSE87411 dataset (16). The GSE20361 dataset is characterized by acquiring endocrine resistance. By culturing MCF7 cells in the estrogen-deprived medium for a long period, cells developed resistance to endocrine therapies¹⁵. On the other hand, GSE111151 contains multiple models of tamoxifen resistance developed using several breast cancer cell lines (MCF7, T-47D, ZR-75-1, and BT-474) and the gene expression data is from before and after developing the resistance¹⁷. We used the whole gene expression patterns of these datasets to generate a heatmap and utilized hierarchical clustering to find the distance or similarity between each of these datasets. Amongst the cell line models, we found that the gene expression patterns of the long-term estrogen-deprived (LTED) cells are the closest to the patient data, visualized by the shortest distance between these two datasets in the column's dendrogram of the heatmap (Figure 1). ZR-75-1 tamoxifen-resistant cell lines are next based on their closeness to patient data, followed by BT-474, MCF7, and T-47D tamoxifen-resistant cell lines. Therefore, we decided to use LTED cells as the most suitable cell line model for studying endocrine resistance in breast cancer.

Estrogen Deprived MCF7 Cells Show Four Dominant Differential Expression Patterns with Unique Cell Functions

To better understand the process of acquiring endocrine resistance and its underlying gene expression patterns, we used the GSE20361 dataset¹⁵. In this work, researchers collected the RNA samples from the MCF7 cells with estrogen depletion after 0, 3, 15, 30, 90, 120, 150, and 180d. We applied our recently developed statistical pipeline to the dataset to find dynamically regulated genes employed in the process of endocrine-resistance development and progression¹⁸. The pipeline provides three main functions: First, statistical hypothesis testing determines a set of dynamic response genes (DRGs) that exhibit significant changes over time. Next, these DRGs are clustered into gene response modules (GRMs), subsets of DRGs with similar time-course expression patterns. Finally, the GRMs associations and regulatory effect are analyzed through a gene regulatory network built using ordinary differential equations.

Among 54 675 probes in the GSE 20361 microarray data, 14 693 genes were characterized as DRGs. The 14 693 DRGs were further ranked based on F-ratio, and the top 3 000 DRGs were selected for correlation-based iterative hierarchical clustering^{18,19}. The top 3 000 DRGs were clustered into 20 distinct GRMs with behavioral profiles corresponding to high degrees of correlation. Response module sizes ranged from 1 to 1 095 genes. Time-resolved expression of the top 3 000 DRGs and the module behavior for the 10 largest modules in the top 3 000 DRGs show a significant difference from one module to the other which emphasizes the resolution of our clustering method (Figure 2a, b)

Exploring the modules showed that the top four GRMs account for 2 748 (92% of the top 3 000) DRGs (Figure 2c). The 3 000 DRGs comprised 2 305 unique genes. The first GRM was the largest model, containing 1 095 genes characterized by a gradual downregulation in expression from day 0 to 30, at which point the pattern shifts, revealing sharp upregulation from day 30 to 150. The second module, comprised of 812 DRGs, showed nearly opposite behavior, demonstrating gradual upregulation until day 30, after which sharp downregulation was observed. Genes from individual GRMs were submitted to the Metascape gene annotation tool for gene enrichment analysis. The first module was significantly enriched for cell division, DNA replication, DNA-dependent DNA replication, and cell cycle pathways and ontologies. The interaction between these different gene ontologies is also visualized in a tree cluster in Figure 2e. The second GRM gene association included membrane trafficking, cellular protein catabolic processes, and macro-autophagy, while the third GRM enriched mostly for positive regulation of protein autophosphorylation and regulation of cellular ketone metabolic processes. The fourth GRM is mostly related to cell junction organization (Figure 2d, e). These results indicate that we identified the genes significantly changed during the development of endocrine resistance and found that each GRM encodes a unique cell function.

Dynamic Gene Expression Analysis of LTED Data Overlaps with Endocrine-Resistant Patient Data and also with Genes Regulated in TNBC

To further narrow down our search for identifying the candidate genes with active roles associated with endocrine resistance, we also analyzed GSE87411, an endocrine-resistant/sensitive patient dataset; the only patient dataset publicly available to the best of our knowledge¹⁶. A group of 109 breast cancer patients was treated with aromatase inhibitors for 2-4 weeks and their response to treatment was evaluated¹⁶. Following the evaluation of patient response to treatment, they were categorized into endocrine-resistant and sensitive patients. We used the gene expression profile of these patients to find the genes significantly regulated in endocrine-resistant breast cancer tumors as compared to endocrine-sensitive breast cancer tumors. We compiled patient data, analyzed the microarrays as two groups, and compared the results to find the significantly differentially expressed genes²⁰. Among these genes, 984 genes were significantly upregulated in endocrine-resistant patients while 621 genes were significantly downregulated. We then compared these significantly regulated genes in the patient dataset to the DRGs found from our dynamic gene expression analysis and selected those with highly correlated expression patterns between the two datasets for further analysis, which returned 319 genes (Figure 3a).

TNBC accounts for around 15% of breast cancer cases, characterized as one of the most aggressive types of breast cancer with no established therapy options yet². Similar to endocrine-resistant breast cancer patients, TNBC patients are also resistant to conventional endocrine therapies due to its nature of the lack of hormone receptors. To examine whether these two groups share certain gene expression patterns or not, we explored the gene expression similarities between the development of endocrine resistance and TNBC. First, we identified the 1 497 genes significantly and differentially expressed in TNBC compared to luminal A breast cancer. This list was compared to the 2 305 DRGs including 319 commonly shared genes between the DRG pipeline analysis and endocrine-resistant patient data.

Surprisingly, 97.8 % of genes significantly expressed in the TNBC vs. Luminal A gene set were included in the DRGs, and more than 80% of the genes commonly shared among the DRGs and endocrine-resistant patient data were also significant in the TNBC vs. Luminal A gene set. This common set between the three sets of analyses is comprised of 254 genes. We posit these genes are important to the development of endocrine resistance, estrogen deprivation, and TNBC (Figure 3a).

Gene ontology analysis on the shared 254 genes showed significant enrichment for cell cycle, cell division, and DNA repair pathways, all signatures of genes affected in cancer cells (Figure 3b). Genes related to DNA repair pathways were included in both TNBC and endocrine-resistant breast cancer, indicating the importance of the DNA repair system to account for mutations in the DNA in their cancers²¹. Other important genes include ribonuclease H2 subunit A (*RNASEH2A*), a gene that is known to be upregulated in cancers. *RNASEH2A* is a mediator of the removal of lagging-strand Okazaki fragment RNA primers, thus it can be integral in the proliferation of both triple-negative and endocrine-resistant breast cancers²². Moreover, most of the common genes are regulated by transcription factors such as E2F1, MYCN, and TFDP1, which are all important transcription factors in TNBC (figure 3c). This finding confirms the notion that TNBC tumors share significant similarities with endocrine-resistant breast cancer.

MCM family, RAD51, CAV1, and CCNE1 are Among the Top Candidate Genes Responsible for Endocrine-Resistant Development

To further identify the genes important for the endocrine resistance development and progression, we utilized several bioinformatic approaches designated to ranking and prioritizing genes. We used several tools to generate gene-gene and protein-protein interaction networks based on the candidate genes in each of the modules²³⁻²⁶. These tools rank or prioritize genes according to the functional similarities of the genes, direct and indirect interactions amongst them, and their roles in the same pathway to cluster the genes or proteins together. For tools that require a training gene set, we gathered known genes for endocrine-resistance development from literature and submitted them as the training set. Using generated networks, we selected the genes with the highest number of neighbors in the network as the master regulators of genes, which will be our main candidates (Figure 4). These core genes may have the most essential roles in the development of endocrine resistance in LTED-MCF7 cells. Thereby, we were able to narrow the list of candidates to 34 genes from major four modules, mostly from modules 1 and 2 (n = 31) (Table S1).

Among the genes in module 1, we found *PARP1* and *E2F1*, recently discovered important genes for endocrine resistance in breast cancer, indicating the validity of our approach. We also found minichromosome maintenance (MCM) family genes as important genes, namely *MCM2* and *MCM7*, *RAD51*, and *TCF3* (Figure 4), which have not been studied thoroughly²⁷⁻²⁹. MCM family genes (*MCM2-7*) form an MCM complex protein that functions as a DNA replication licensing factor and plays a central role in eukaryotic DNA replication³⁰. Recent evidence suggests that blocking the expression of these genes can lead to the inhibition of the growth of tamoxifen-resistant cancer cells. This evidence perfectly

matches the expression pattern of these genes in our dataset, as these genes are significantly upregulated during the endocrine-resistance process in breast cancer cells³¹. RAD51 overexpression, a key protein of homologous recombination, is also linked to overall poor survival and endocrine resistance in breast cancer, although the exact underlying signaling pathways are not well understood yet³². As for TCF3, while there is no study on its role in endocrine resistance, it is important for breast cancer differentiation, development, and prognosis³³. All the genes in module two were slightly upregulated by day 30 and significantly downregulated afterward (Figure 5). In line with our data, there are reports showing downregulation of CAV1 as an important step in breast cancer development and resistance to endocrine therapies³⁴, and further studies are warranted. Some of the other important candidate genes were selected from modules 3 and 4, including *ATG3*, *CCNE1*, and *MFAP4*.

Long-term Estrogen Deprived MCF7 Cells Gained Resistance at Day 90

To validate the bioinformatic findings, we established the LTED MCF7 cell model by culturing human breast cancer MCF7 cells in estrogen-depleted growth media for an extended period. This cell model recapitulated acquired resistance to aromatase inhibitors in postmenopausal women and the same experimental settings as the study on which we used for our gene expression analysis with an extended sample collection period¹⁵. After the initial quiescence state of around 30 days when the cell's growth rate was minimal, they started to adapt with estrogen depletion and regain their growth ability while losing their responsiveness to 17 β -estradiol (E_2). At around day 90, LTED cells showed complete loss of response to E_2 , characterized by their loss of ability to respond to estrogen stimulation to increase their growth rate, while their basal growth rate reached similar levels as those of parental MCF7 cells cultured in estrogen-rich media (Figure 5a). We then continued to maintain the LTED cells for another 6 months. After a period of super-sensitivity to estrogen at low doses (10^{-11} to 10^{-13} M), which happened around day 150, they became completely unresponsive to estrogen treatments at all concentrations by day 250 (Figure 5b). These results are consistent with the previous reports and follow a similar timeline^{15,35-38}.

To validate the gene expression profiles of the candidate genes, we collected samples of LTED cells at different time points (from day 0 to day 135) and measured the expression of some of the candidate genes using qRT-PCR (Figure 5c). The expression profile of these genes closely mimicked their gene expression patterns from the microarray studies, which further confirms microarray results and our downstream bioinformatics analysis. Furthermore, as our qRT-PCR analysis contains much more data points compared to the previous study, the expression patterns of the genes are more detailed than the microarray results.

Discussion

In the present study, we utilized a newly developed statistical and computational pipeline to examine the process of endocrine resistance in breast cancer and found novel underlying gene associations. First, we compared the data from patients resistant and sensitive to the endocrine therapies with publicly available

gene expression data from cell-based models. We found LTED MCF7 cell model is the closest to the patients' data among cell models analyzed based on their whole gene expression patterns. Next, we reanalyzed the gene expression in LTED cells during the process of acquiring endocrine resistance. Using our statistical and computational pipeline, which is designed for identifying dynamically significant genes and capable of comparing genes and gene clusters based on their time-resolved expression^{18,19}, we grouped the genes into different modules, or DRGs, based on their expression patterns in LTED cells. Then, we compared the DRGs to the gene expressions of patients' tumor samples from endocrine-resistant or sensitive tumors and found a group of 319 genes as potential drivers of endocrine-resistance development. Finally, using multiple bioinformatics approaches, we narrowed down the candidate genes list to 34 genes from major four modules, primarily from modules 1 and 2. These genes were selected based on gene-gene network analysis and literature analysis. Analyzing the gene-gene networks enabled us to select the genes with a higher probability of being more integral to the networks. These candidate genes are potential targets for developing potent therapies for endocrine-resistant breast cancer. The expression patterns of some of these genes were further validated in biological settings by developing an LTED cell line.

Twenty out of the 34 genes were from the first module, which showed a slight decrease in expression by day 30, followed by significant upregulation afterward, which correlates to the growth rates of MCF7 cells under the development of endocrine resistance. As anticipated, the genes in module 1 were primarily associated with DNA replication and repair mechanisms, emphasizing the importance of these mechanisms in the replication and growth of cancer cells. Minichromosome maintenance complex genes *MCM2*, *3*, *4*, *6*, and *7* were identified in Module 1. Overexpression of *MCM2*, *MCM3*, *MCM4*, and *MCM6* is associated with luminal B, HER2⁺, and triple-negative breast cancers^{39,40}. *RAD51*, one of the genes in module 1 is known to be related to overall poor survival and endocrine resistance in breast cancer, while it is also linked to regulation of metastasis in TNBC^{32,41}. Among other genes in module 1 is replication factor C subunit 3 (*RFC3*), which is essential for the homologous DNA pairing and strand exchange. Downregulation of *RFC3* has been shown to attenuate cell proliferation, migration, and invasion in TNBC^{42,43}. The cyclin-dependent kinases (CDKs) are critical regulatory enzymes governing cell cycle transitions and play an essential role in the development of endocrine resistance in some breast cancer cases^{10,44}. While CDK 4/6 are the most studied members of the CDK family in relation to endocrine-resistant breast cancer, our analysis revealed another member of this family, *CDK1* to be important for both endocrine resistance and TNBC. Notably, CDK1 is required for the initiation of mitosis and cell proliferation, and its inhibition in TNBC led to a decrease in cell viability and an increase in cell apoptosis⁴⁵.

Fourteen genes are related to the rest of the modules, with 11 of them related to module 2. The majority of these genes are already known to be related to breast cancer in some way, yet their connection to endocrine resistance and TNBC has not been discovered yet and calls for further experiments. Among the genes in module 2, the network analysis suggested *ELAVL1*, *GABARAPL2*, and *CAV1* as the main regulators of all other genes. Protein Phosphatase 1A (PPM1A) is also among the genes in module 2, which means that it was downregulated in endocrine-resistant breast cancer, as well as TNBC. This

protein is a member of the protein phosphatase 2C family of Ser/Thr protein phosphatases and has been shown to regulate mitogen-activated protein kinase cellular signaling pathways as well as proliferation, cell invasion, and migration ^{46,47}. Recently, Mazumdar et al. showed that this protein is significantly downregulated in ER-negative breast cancers and that its upregulation suppresses *in vitro* and *in vivo* growth of TNBC cells ⁴⁸, which supports our findings.

Interestingly, the genes we identified important for the resistance to endocrine therapies include most genes that are also significantly regulated in TNBC compared to luminal A breast cancer (Figure 3a), suggesting that these genes are both of great importance in the endocrine-resistance process as well as triple-negative breast cancer. We addressed the gap in the knowledge regarding the genetic factors underlying endocrine-resistant as well as triple-negative breast cancers. These two subtypes of breast cancer are the most fatal types of breast cancer with no known effective therapeutic approaches available to date. Thus, research on the underlying genetic factors of these cancers is of great importance. Here, we found a group of candidate genes that are significantly regulated in the process of endocrine resistance, and with further studies, they will be potential targets for the treatment of endocrine-resistant breast cancers. Moreover, the majority of these genes are also significantly regulated in TNBC compared to luminal A breast cancer, suggesting that endocrine-resistant breast cancer and TNBC share significant similarities in their mechanisms. Therefore, these genes we identified can also be potential therapeutic targets for treating TNBC. Still, more than 1 200 genes were significantly regulated in TNBC and endocrine resistance development in LTED cells but not in endocrine-resistant breast cancer patients. This can be because data from endocrine-resistant patients only show the comparisons between resistant and sensitive tumors, and thereby, does not represent the genes that are important for establishing endocrine resistance. Since we cannot determine whether the endocrine-resistant patients had already gained resistance or not at the time of the endocrine therapy, we could not track the development of endocrine resistance from patients' datasets. In contrast, our approach analyzed the process of developing endocrine resistance, not just the endpoints of endocrine resistance. Our analysis found several novel genes with potential significance in endocrine-resistant breast cancer as well as TNBC and opened new doors toward designing new therapeutic approaches for endocrine-resistant breast cancer and TNBC.

Our novel statistical and computational analysis approach, combined with the use of multiple types of data from cell models to patients, revealed a new potential to develop effective therapeutic approaches toward various diseases. Further biological experiments are warranted to confirm a measurable degree of importance of these genes. While we did our best to use both cell line data and patient data to find the genes with the most physiological and clinical significance, more datasets for both patient data and dynamic gene expression data will enhance the validation and increase the reliability of our findings.

Methods

Data Acquisition and Processing

The microarray data of GSE20361, GSE87411, and GSE111151 were downloaded from GEO, the public functional genomics repository ¹⁵⁻¹⁷. Data from GSE20361 were analyzed using a Matlab-based pipeline to identify DRGs described as below, while data from GSE111151 and GSE87411 were used in conjunction with Limma. Using default options, we identified the differentially expressed genes from both datasets for downstream analysis. To generate the heatmap, the log-fold-change values of the datasets and the “ComplexHeatmap” package in R were used. To identify the genes significantly regulated in TNBC compared to luminal A breast cancer, we used UALCAN ⁴⁹. Each gene was checked for its expression in the TNBC subtype compared to luminal A subtype, and those with significant differential expression, using a significance level of 0.05, were selected for further analysis. To generate the gene-gene and protein-protein interaction networks, gene lists were analyzed using NetworkAnalyst, ToppGene, ToppFun, and Funcoup, using the default parameters. Metascape tool was used for all the gene enrichment analysis throughout the paper using the default parameters.

Dynamic Gene Expression Analysis Pipeline

As part of the analysis, we used a pipeline tool we developed ^{18,19}. One key assumption of the methodology is that a small fraction of the genes responds to external stimuli. Further, we assume that the larger fraction of the genes maintains constant expression over time. Thus, spline smoothing was applied to the expression patterns using a smoothing regularity parameter obtained by minimizing generalized cross-validation of the top 200 most-responsive probes and applied to all expression profiles. A subset of the genes that exhibit time course patterns that have relatively smooth trajectories was ranked by their interquartile range, and the top genes for the estimation subset were selected. Then, the DRGs with similar expression patterns over time are clustered into temporal GRMs using Iterative Hierarchical Clustering (IHC).

LTED Cell Line Development

MCF7 cells were cultured in estrogen-depleted growth media (phenol red-free RPMI1690 medium (Thermo Fisher Scientific, Waltham, MA) supplemented with 5% charcoal-stripped serum (CSS) and 10 mg/ml of Insulin (Thermo Fisher Scientific) for extended periods. Cell samples were collected every 7 days and used for downstream analysis.

Proliferation Assay

Both parental and LTED MCF7 cells were cultured in a 96-well plate. After 24 hours, the cells were treated in triplicates with E₂ in RPMI1690 plus 5% CSS. The treatment was repeated every other day and at the end of the seventh day, then the plates were frozen for the assay. FluoReporter blue fluorometric dsDNA quantification kit (Invitrogen, Waltham, MA) was used per the manufacturer’s protocol. The excitation/emission wavelength of 360/460 nm were read and normalized to the control wells.

qRT-PCR Analysis

mRNA abundance was evaluated by qRT-PCR as described¹⁹. Relative mRNA levels were calculated by using the comparative CT method normalized to cyclophilin. The primers were designed using Primer Express Software (Applied Biosystems, Waltham, MA) as shown in Table S2.

Data Availability

All data are available in the main text or Supplementary Information.

Declarations

Acknowledgments

The authors acknowledge the kind helps of Linh Bui and Robert Stephen from the University of Houston. This work was supported by HEALTH Research Institute Pilot Funding for Basic Biomedical Science (MU), Cancer Prevention and Research Institute of Texas grant RP170668 (WJZ), and National Institutes of Health grants 1UL1TR003167, R01AG066749 (WJZ), and R01AI087135 (HW).

Author Contributions

AA, VM, and MU conceptualized the idea. AA, KW, MG, NDV, FA, JW, ND, WJZ, and HW contributed the experiments. AA, KW, MG, MU, and VM wrote the manuscript, and all authors reviewed and edited the manuscript.

Competing Interests

Authors declare that they have no competing interests.

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians* **71**, 7-33, doi:10.3322/caac.21654 (2021).
2. DeSantis, C. E. *et al.* Breast Cancer Statistics, 2019. *CA: A Cancer Journal for Clinicians* **69**, 438-451, doi:10.3322/caac.21583 (2019).
3. Coleman, K., M. Intracellular Signaling Pathways: Nongenomic Actions of Estrogens and Ligand-independent Activation of Estrogen Receptors. *Frontiers in Bioscience* **6**, d1379, doi:10.2741/Coleman (2001).
4. Osborne, C. K. & Schiff, R. Mechanisms of Endocrine Resistance in Breast Cancer. *Annual Review of Medicine* **62**, 233-247, doi:10.1146/annurev-med-070909-182917 (2011).
5. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 Mutations—A Mechanism for Acquired Endocrine Resistance in Breast Cancer. *Nature Reviews Clinical Oncology* **12**, 573-583, doi:10.1038/nrclinonc.2015.117 (2015).

6. Robinson, D. R. *et al.* Activating ESR1 Mutations in Hormone-Resistant Metastatic Breast Cancer. *Nature Genetics* **45**, 1446-1451, doi:10.1038/ng.2823 (2013).
7. Jeselsohn, R. *et al.* Emergence of Constitutively Active Estrogen Receptor- Mutations in Pretreated Advanced Estrogen Receptor-Positive Breast Cancer. *Clinical Cancer Research* **20**, 1757-1767, doi:10.1158/1078-0432.CCR-13-2332 (2014).
8. Merenbakh-Lamin, K. *et al.* D538G Mutation in Estrogen Receptor- : A Novel Mechanism for Acquired Endocrine Resistance in Breast Cancer. *Cancer Research* **73**, 6856-6864, doi:10.1158/0008-5472.CAN-13-1197 (2013).
9. Hanker, A. B., Sudhan, D. R. & Arteaga, C. L. Overcoming Endocrine Resistance in Breast Cancer. *Cancer Cell* **37**, 496-513, doi:10.1016/j.ccell.2020.03.009 (2020).
10. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438.e426, doi:10.1016/j.ccell.2018.08.008 (2018).
11. Schiavon, G. *et al.* Analysis of ESR1 Mutation in Circulating Tumor DNA Demonstrates Evolution During Therapy for Metastatic Breast Cancer. *Science Translational Medicine* **7**, 313ra182-313ra182, doi:10.1126/scitranslmed.aac7551 (2015).
12. Lei, J. T., Gou, X. & Ellis, M. J. ESR1 Fusions Drive Endocrine Therapy Resistance and Metastasis in Breast Cancer. *Molecular & Cellular Oncology* **5**, e1526005, doi:10.1080/23723556.2018.1526005 (2018).
13. Croessmann, S. *et al.* Combined Blockade of Activating ERBB2 Mutations and ER Results in Synthetic Lethality of ER+/HER2 Mutant Breast Cancer. *Clinical Cancer Research* **25**, 277-289, doi:10.1158/1078-0432.CCR-18-1544 (2019).
14. Kurokawa, H. *et al.* Inhibition of HER2/neu (erbB-2) and Mitogen-activated Protein Kinases Enhances Tamoxifen Action Against HER2-Overexpressing, Tamoxifen-resistant Breast Cancer Cells. *Cancer Research* **60**, 5887-5894 (2000).
15. Aguilar, H. *et al.* Biological Reprogramming in Acquired Resistance to Endocrine Therapy of Breast Cancer. *Oncogene* **29**, 6071-6083, doi:10.1038/onc.2010.333 (2010).
16. Ellis, M. J. *et al.* Ki67 Proliferation Index as a Tool for Chemotherapy Decisions During and After Neoadjuvant Aromatase Inhibitor Treatment of Breast Cancer: Results From the American College of Surgeons Oncology Group Z1031 Trial (Alliance). *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **35**, 1061-1069, doi:10.1200/JCO.2016.69.4406 (2017).
17. Hultsch, S. *et al.* Association of Tamoxifen Resistance and Lipid Reprogramming in Breast Cancer. *BMC cancer* **18**, 850, doi:10.1186/s12885-018-4757-z (2018).
18. Carey, M., Ramírez, J. C., Wu, S. & Wu, H. A Big Data Pipeline: Identifying Dynamic Gene Regulatory Networks from Time-course *Gene Expression Omnibus* Data with Applications to Influenza Infection. *Statistical Methods in Medical Research* **27**, 1930-1955, doi:10.1177/0962280217746719 (2018).
19. Maroufy, V. *et al.* Gene expression dynamic analysis reveals co-activation of Sonic Hedgehog and epidermal growth factor followed by dynamic silencing. *Oncotarget* **11**, 1358-1372, doi:10.18632/oncotarget.27547 (2020).

20. Ritchie, M. E. *et al.* Limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies. *Nucleic Acids Research* **43**, e47-e47, doi:10.1093/nar/gkv007 (2015).
21. Fang, C.-B., Wu, H.-T., Zhang, M.-L., Liu, J. & Zhang, G.-J. Fanconi Anemia Pathway: Mechanisms of Breast Cancer Predisposition Development and Potential Therapeutic Targets. *Frontiers in Cell and Developmental Biology* **8**, 160, doi:10.3389/fcell.2020.00160 (2020).
22. Shen, J., Lin, S., Liu, L. & Wang, C. Ribonuclease H2 Subunit A impacts invasiveness and chemoresistance resulting in poor survivability of breast cancer in ER dependent manner. *American Journal of Translational Research* **12**, 2281-2294 (2020).
23. Xia, J., Benner, M. J. & Hancock, R. E. W. NetworkAnalyst - Integrative Approaches for Protein–protein Interaction Network Analysis and Visual Exploration. *Nucleic Acids Research* **42**, W167-W174, doi:10.1093/nar/gku443 (2014).
24. Zhou, G. *et al.* NetworkAnalyst 3.0: A Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-analysis. *Nucleic Acids Research* **47**, W234-W241, doi:10.1093/nar/gkz240 (2019).
25. Ogris, C., Guala, D., Kaduk, M. & Sonnhhammer, E. L. L. FunCoup 4: New Species, Data, and Visualization. *Nucleic Acids Research* **46**, D601-D607, doi:10.1093/nar/gkx1138 (2018).
26. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for Gene List Enrichment Analysis and Candidate Gene Prioritization. *Nucleic Acids Research* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).
27. Pulliam, N. *et al.* Poly-ADP-Ribosylation of Estrogen Receptor-Alpha by PARP1 Mediates Antiestrogen Resistance in Human Breast Cancer Cells. *Cancers* **11**, 43, doi:10.3390/cancers11010043 (2019).
28. Lee, K. Y. *et al.* PI3-Kinase/p38 Kinase-dependent E2F1 Activation Is Critical for Pin1 Induction in Tamoxifen-resistant Breast Cancer Cells. *Molecules and Cells* **32**, 107-111, doi:10.1007/s10059-011-0074-y (2011).
29. Louie, M. C., Zou, J. X., Rabinovich, A. & Chen, H.-W. ACTR/AIB1 Functions as an E2F1 Coactivator To Promote Breast Cancer Cell Proliferation and Antiestrogen Resistance. *Molecular and Cellular Biology* **24**, 5157-5171, doi:10.1128/MCB.24.12.5157-5171.2004 (2004).
30. Forsburg, S. L. Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiology and Molecular Biology Reviews* **68**, 109-131, doi:10.1128/MMBR.68.1.109-131.2004 (2004).
31. Liang, Z. *et al.* Simvastatin Suppresses the DNA Replication Licensing Factor MCM7 and Inhibits the Growth of Tamoxifen-resistant Breast Cancer Cells. *Scientific Reports* **7**, 41776, doi:10.1038/srep41776 (2017).
32. Jia, Y. *et al.* Aberrant Regulation of RAD51 Promotes Resistance of Neoadjuvant Endocrine Therapy in ER-positive Breast Cancer. *Scientific Reports* **9**, 12939, doi:10.1038/s41598-019-49373-w (2019).
33. O'Brien, C. S., Farnie, G., Howell, S. J. & Clarke, R. B. Breast Cancer Stem Cells and Their Role in Resistance to Endocrine Therapy. *Hormones & Cancer* **2**, 91-103, doi:10.1007/s12672-011-0066-6 (2011).

34. Thomas, N. B. P. *et al.* Growth of Hormone-dependent MCF-7 Breast Cancer Cells Is Promoted by Constitutive Caveolin-1 Whose Expression Is Lost in an EGF-R-Mediated Manner During Development of Tamoxifen Resistance. *Breast Cancer Research and Treatment* **119**, 575-591, doi:10.1007/s10549-009-0355-8 (2010).
35. Shaw, L. E., Sadler, A. J., Pugazhendhi, D. & Darbre, P. D. Changes in Oestrogen Receptor- α and - β During Progression to Acquired Resistance to Tamoxifen and Fulvestrant (Faslodex, ICI 182,780) in MCF7 Human Breast Cancer Cells. *The Journal of Steroid Biochemistry and Molecular Biology* **99**, 19-32, doi:10.1016/j.jsbmb.2005.11.005 (2006).
36. Chan, C. M. W., Martin, L.-A., Johnston, S. R. D., Ali, S. & Dowsett, M. Molecular Changes Associated with the Acquisition of Oestrogen Hypersensitivity in MCF-7 Breast Cancer Cells on Long-term Oestrogen Deprivation. *The Journal of Steroid Biochemistry and Molecular Biology* **81**, 333-341, doi:10.1016/s0960-0760(02)00074-2 (2002).
37. Jeng, M. H. *et al.* Estrogen Receptor Expression and Function in Long-term Estrogen-deprived Human Breast Cancer Cells. *Endocrinology* **139**, 4164-4174, doi:10.1210/endo.139.10.6229 (1998).
38. Yue, W., Wang, J.-P., Conaway, M. R., Li, Y. & Santen, R. J. Adaptive Hypersensitivity Following Long-term Estrogen Deprivation: Involvement of Multiple Signal Pathways. *The Journal of Steroid Biochemistry and Molecular Biology* **86**, 265-274, doi:10.1016/S0960-0760(03)00366-2 (2003).
39. Issac, M. S. M., Yousef, E., Tahir, M. R. & Gaboury, L. A. MCM2, MCM4, and MCM6 in Breast Cancer: Clinical Utility in Diagnosis and Prognosis. *Neoplasia* **21**, 1015-1035, doi:10.1016/j.neo.2019.07.011 (2019).
40. Yousef, E. M. *et al.* MCM2: An Alternative to Ki-67 for Measuring Breast Cancer Cell Proliferation. *Modern Pathology* **30**, 682-697, doi:10.1038/modpathol.2016.231 (2017).
41. Wiegman, A. P. *et al.* Rad51 Supports Triple Negative Breast Cancer Metastasis. *Oncotarget* **5**, 3261-3272, doi:10.18632/oncotarget.1923 (2014).
42. He, Z.-Y. *et al.* Up-Regulation of RFC3 Promotes Triple Negative Breast Cancer Metastasis and is Associated With Poor Prognosis Via EMT. *Translational Oncology* **10**, 1-9, doi:10.1016/j.tranon.2016.10.004 (2017).
43. Li, Y. *et al.* Multifaceted Regulation and Functions of Replication Factor C Family in Human Cancers. *American Journal of Cancer Research* **8**, 1343-1355 (2018).
44. Spring, L., Bardia, A. & Modi, S. Targeting the Cyclin D-cyclin-dependent Kinase (CDK) 4/6-Retinoblastoma Pathway with Selective CDK 4/6 Inhibitors in Hormone Receptor-positive Breast Cancer: Rationale, Current Status, and Future Directions. *Discovery Medicine* **21**, 65-74 (2016).
45. Liu, Y. *et al.* Triple Negative Breast Cancer Therapy with CDK1 siRNA Delivered by Cationic Lipid Assisted PEG-PLA Nanoparticles. *Journal of Controlled Release* **192**, 114-121, doi:10.1016/j.jconrel.2014.07.001 (2014).
46. Lammers, T. & Lavi, S. Role of Type 2C Protein Phosphatases in Growth Regulation and in Cellular Stress Signaling. *Critical Reviews in Biochemistry and Molecular Biology* **42**, 437-461, doi:10.1080/10409230701693342 (2007).

47. Zhang, B. *et al.* Protein Phosphatase 1A (PPM1A) Is Involved in Human Cytotrophoblast Cell Invasion and Migration. *Histochemistry and Cell Biology* **132**, 169-179, doi:10.1007/s00418-009-0601-5 (2009).
48. Mazumdar, A. *et al.* The Phosphatase PPM1A Inhibits Triple Negative Breast Cancer Growth by Blocking Cell Cycle Progression. *npj Breast Cancer* **5**, 22, doi:10.1038/s41523-019-0118-6 (2019).
49. Chandrashekar, D. S. *et al.* UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia (New York, N.Y.)* **19**, 649-658, doi:10.1016/j.neo.2017.05.002 (2017).

Figures

Figure 1

Heatmap of the global gene expression patterns between cell-based models of endocrine resistance and patient data. Data for the LTED cell model was obtained from GSE20361. Data for patient's responses to hormonal therapies was obtained from GSE87411. The rest of the cell model data was obtained from GSE111151. For cell models ZR75-1, BT474, and T47D the data consisted of two populations of tamoxifen-resistant cell lines both of which are included here in the heatmap. The dendrogram represents the closeness of the datasets, thus the LTED cell line is the most similar to the patient's data followed by ZR-75-1 tamoxifen-resistant cells.

Figure 2

Dynamic gene expression analysis results, gene expression trends of top four modules, and network analysis of the top four modules DRGs using Metascape. a) Presents the individual time-resolved expression trends for the top 3 000 dynamic response gene probes. These 3 000 probes are comprised of 2 305 identifiable unique genes. Y-axis represents the normalized expression of genes as described in the methods section. b) Illustrates the multiple-sampling method (MSM) results of mean expression trends

for the 10 largest modules. c) The time-resolved expression trends for the top four largest modules are shown. Each blue dotted curve represents one gene, and the middle orange line illustrates the mean expression at each time point. d) Presents the significantly enriched terms in modules 1-4 based on the corresponding Fisher's Exact Test p-values. e) Presents the statistically enriched terms in module 1 hierarchically clustered into a tree based on Kappa-statistical similarities among gene memberships.

Figure 3

Similar gene expression patterns between endocrine-resistant breast cancer and triple-negative breast cancer. a) Venn diagram depicts the shared significantly regulated genes between endocrine-resistant and triple-negative breast cancers and patient’s data. b) Gene ontology analysis showing the significantly enriched ontologies and pathways for the common genes. c) Transcription factor (TF) analysis of the common genes shows the majority of them are regulated by E2F1, MYCN, and TP53 TFs.

Figure 4

Tissue-specific protein-protein interaction network for modules 1 and 2 candidate genes. Candidate genes from modules 1 and 2 were fed to NetworkAnalyst and the generated interaction networks are shown below. Red and orange circles represent the candidate proteins while the yellow ones are other proteins that interact with our candidate proteins. The bigger the size of the circle, the more interactions it has. The scale of the module one and two figures is different as module 1 contains significantly more genes that made the visualization harder.

Figure 5

Proliferation assay and qRT-PCR results confirm the bioinformatics findings. a) Proliferation assay of the LTED cells at day 90 and normal MCF7 cells under different treatment doses of E2 (from 10^{-14} to 10^{-7} M) showed that LTED cells were not responsive to estrogen treatment anymore. b) Proliferation assay of the LTED cells at different days under several doses of E2 (from 10^{-15} to 10^{-7} M) showed that cells completely lost their responsiveness to E2 treatment by day 250. c) qRT-PCR results for MCM2, CDK1, and

CAV1 confirmed that the gene expression patterns in vitro match their expected gene expression patterns from microarray bioinformatics analysis. All values represent mean $\pm$ SEM (n = 3 - 6).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AsghariSupplementaryInformation.pdf](#)