A data driven approach to identify ZM-447439 as a potential repurposed drug to overcome tamoxifen-resistance in breast cancer

Fadhl M Alakwaa

University of Michigan

Paula Benny

University of Hawai'i Cancer Center

Kumardeep Chaudhary

Icahn School of Medicine at Mount Sinai

Lana X Garmire (✉ LGarmire@med.umich.edu)

University of Hawai'i Cancer Center <https://orcid.org/0000-0003-1672-6917>

Research article

Keywords: Breast cancer, drug repurposing, tamoxifen resistance, bioinformatics, estrogen receptor, LINC, MCF7-RR

Posted Date: August 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56103/v1>

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

[Read Full License](#)

A data driven approach to identify ZM-447439 as a potential repurposed drug to overcome tamoxifen-resistance in breast cancer

Fadhil M Alakwaa^{1,2*}, Paula Benny², Kumardeep Chaudhary^{2,3}, Lana X Garmire^{2,4*}

¹Current address: Department of Neurology, University of Michigan, Ann Arbor, MI 48109, USA

²University of Hawaii Cancer Center, Department of Epidemiology, 701 Ilalo Street, Honolulu, HI 96813, USA

³Current address: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

⁴Current address: Department of Computational Medicine and Bioinformatics, Building 520, 1600 Huron Parkway, Ann Arbor, MI 48109, USA

*Corresponding authors email: alakwaaf@med.umich.edu and LGarmire@med.umich.edu

Abstract:

Background: Tamoxifen is the most commonly used endocrine therapy (ET) for Breast Cancer (BC) patients expressing estrogen receptors (ER), representing almost 70% of all cases. However, one third of early stage BC patients demonstrate endocrine resistance to tamoxifen over the initial five-year treatment period, prompting significant research effort on identifying other drugs to alleviate tamoxifen resistance in ER+ BC patients.

Methods: we combined a total of 229 tamoxifen resistant and 363 tamoxifen sensitive tumors and tumor cell lines, to select genes that showed consistency as either up- or down- regulated differential expression among these datasets. We use these genes as the input to identify the drugs and compounds in Library of Integrated Network-Based Cellular Signatures (LINCS) database that can reverse the expression of these genes. With an innovative and comprehensive scoring system, we performed quality assessment on the results and prioritized drugs. Finally, we validated top five drugs using *in vitro* cell culture experiments.

Results: we identified that ZM-447439, an aurora kinase inhibitor, can reverse the gene signatures associated with tamoxifen resistance. This was accomplished by a novel bioinformatics approach and scoring system, screening from over 20,000 small molecules in the Library of Integrated Network-Based Cellular Signatures (LINCS) database. The *in vitro* cell culture experiments showed that ZM-447439 had high potency to reverse gene expression in the tamoxifen-resistance BC cell line (MCF7-RR).

Conclusion: we demonstrate the utility of a bioinformatics repurposing approach to identify candidate drug ZM-447439 with the potential to treat patients with acquired tamoxifen resistance.

KEYWORDS: Breast cancer, drug repurposing, tamoxifen resistance, bioinformatics, estrogen receptor, LINCS, MCF7-RR

Introduction

Breast cancer (BC) is one of the leading causes of cancer-related deaths in the United States (1). According to estimates reported by the American Cancer Society, 40,610 women died from BC in 2017. Approximately 70% of breast tumors express the estrogen receptor (ER), which binds and mediates many of the effects of the estrogen hormone (2). Inhibiting ER signaling using ER antagonists or anti-estrogens such as tamoxifen, the selective ER down-regulator flutemide, or the aromatase inhibitors such as letrozole, anastrozole, or exemestane have been reliable therapeutics since 1970 (3). In the adjuvant setting, tamoxifen reduces BC recurrence by 50% and mortality by 31% (4). However, around 40% of patients who respond to tamoxifen develop *de novo* resistance, which is not fully understood (5).

It is important to develop novel drugs to overcome tamoxifen resistance. Compounds that are able to restore endocrine sensitivity is a possible solution, based on the deep understanding of the molecular mechanism that drive endocrine resistance (6). For example, crosstalk between the ER and mammalian target of rapamycin (mTOR) signaling pathways is thought to play a role in the development of resistance to endocrine therapy (7), and combined inhibition of these pathways using everolimus and temsirolimus has been used to restore endocrine sensitivity (8). However, such combinatorial approaches have many challenges such as consumed time and cost (6). In all, overcoming tamoxifen resistance still represents a clinical unmet need.

Recently, repurposing drugs approved for other applications have been suggested as a solution to overcome tamoxifen resistance (9). This approach saves developmental costs and provides a much shorter time for drug approval and launch (10). In this paper, we have developed a novel bioinformatics pipeline to identify potential drug candidates to overcome tamoxifen resistance, by screening over 20,000 small molecules and compounds in the Library of Integrated Network-Based Cellular Signatures (LINCS) database (11). We

identified one kinase inhibitor, ZM-447439, that resensitized a tamoxifen resistance-BC cell line (MCF7-RR), which may be possible for treating patients with tamoxifen resistance.

Materials and Methods

Datasets

We used differential expressed genes (DEGs) obtained from five datasets from different platforms (*12-16*) that study the change of gene expression in the tamoxifen resistance tumors/cell lines, primarily available through public domain Gene Expression Omnibus (GEO) database. The detailed description of these datasets is described below:

Dataset 1: This dataset contains gene expression values of early stages BC ER+ tumors diagnosed in John Radcliffe, Oxford, UK and measured by Affymetrix HG-U133A microarray platform (*13*). This microarray dataset contains 85 tamoxifen resistance tumors and 91 sensitive tumors. It is publicly available from GEO accession number GSE6532. The study identified the 275 DEGs with 173 upregulated and 102 downregulated genes.

Dataset 2: This dataset contains gene expression values of 26 tamoxifen resistance and 138 tamoxifen sensitive tumors diagnosed between 1980 and 1995 in the Guys Hospital, London, UK (*14*). It is accessible from GEO accession number GSE9195. The study identified the 130 DEGs with 86 upregulated and 44 downregulated genes.

Dataset 3: This microarray dataset has 118 tamoxifen resistance and 134 tamoxifen sensitive tumors diagnosed between 1989 and 2001 in Cancer research center of Val d'Aurelle, Montpellier, France (*12*). It was measured using Qiagen-Operon Oligo set 2.1.3 platform and accessible from GEO accession number GSE9893. The study identified the 252 DEGs with 118 upregulated and 134 downregulated genes.

Tumors from each of the above three microarrays datasets were classified as tamoxifen resistance/tamoxifen sensitive if patients relapsed/still respond to tamoxifen within 5 years respectively.

Dataset 4 and 5: In these two independent studies, tamoxifen resistance and tamoxifen sensitive BC cell lines were sequenced using next generation sequencing (NGS) technology to identify the DEGs responsible

for tamoxifen resistance(15, 16). These studies identified the 272 DEGs with 141 upregulated and 131 downregulated genes.

Up- and down-regulated genes

We catalogued the DEGs from the above data resources which are available as supplementary tables S 1-3 and S7-8. The list of differentially expressed genes (DEGs) for each dataset was obtained from the supplementary files of the corresponding studies without further processing.

Pathway enrichment analysis

To find the functions and pathways the DEGs have in common, we performed pathway and gene ontology enrichment analysis using ConsensusPathDB resource (17) with the following selection criteria: adjusted p value <0.05 and minimum overlap with input list >10 genes.

Pathways/Gene Ontology (GOs) scoring:

We developed a pathway scoring system to select consistently enriched genes among the DEGs. We selected up and down regulated genes that are involved in highly scored pathways/GOs. The pathway scoring system is based on three metrics: the enrichment score (S_1), up/down regulated score (S_2), and the overlap score (S_3).

For each pathway/GO i , we calculated its total score ($S_{t_{pi}}$) as follows:

$$S_{t_{pi}} = S_{1_{pi}} + S_{2_{pi}} + S_{3_{pi}} \quad (1)$$

Where $S_{t_{pi}}$ is the total score of pathway/GO i

$S_{1_{pi}}$ is the enrichment score of pathway/GO i

$$S_{1_{pi}} = \frac{n_{pi}}{N_{pi}} \quad (1.1)$$

n_{pi} is the number of DEGs involved in pathway/GO i

N_{pi} is the total number of genes annotated in pathway/GO i

$S_{2_{pi}}$ is the up regulated score which equals the ratio between the number of up-regulated DEGs to the total number of DEGs involved in pathway/GO i

$$S_{2_{pi}} = \frac{n_{up_{pi}}}{n_{up_{pi}} + n_{down_{pi}}} \quad (1.2)$$

$n_{up_{pi}}$ is the number of up-regulated DEGs in pathway/GO i

$n_{down_{pi}}$ is the number of down-regulated DEGs in pathway/GO i

$S_{3_{pi}}$ is the overlap score which represents how many DEGs in pathway/GO i overlap with other pathways/GOs ($j=1, \dots, m$)

$$S_{3_{pi}} = \frac{1}{m-1} \sum_{j=1}^m \left(\frac{n_{pij}}{n_{pj}} \right) \quad (1.3)$$

n_{pij} is the number of overlapped DEGs between pathway/GO i and pathway/GO j

n_{pj} is the number of DEGs in pathway/GO j

m is the number of pathways and GOs term (85 pathways + 109 GOs = 194)

CLUE query system for LINCS data search

CLUE query system: To find the drugs that target selected DEGs, we used the cloud-based platform called CLUE (<https://clue.io/>) for the analysis of perturbational datasets generated using gene expression L1000 (11). The input of CLUE query app is a list of up- and down-regulated genes which represents the phenotype (tamoxifen resistance for this study). CLUE outputs a list of drugs that reverse these up and down regulated genes (29) (see CMAP technology). CLUE is hosted by the Library of Integrated Network-based Cellular Signatures (LINCS) project. The latter is funded by the National Institute of Health (NIH) that aims to map the change in gene expression, disease and the drugs. In other words, it measures the changes in gene expression that occurs when cells are exposed to many perturbing agents (e.g. drugs). We used CLUE version 1.1.1.1 that contains a total number of 250,479 signatures. For MCF7 cell line, CLUE has 8,668 signatures (2,892 compound, 2,160 gene over-expression and 3,616 gene knock-down). We used CLUE

API RESTful web services (<https://clue.io/api>) which offers programmatic access to annotations and perturbational signatures in the L1000 dataset.

LINC data: We used the version of LINCS data that contained 1,328,098 gene expression profiles as a result from applying 42,553 perturbagens (19,811 small molecule compounds, 18,493 short complimentary RNA molecules (shRNAs), 3,627 Complementary DNA (cDNAs), and 622 biologics) for a total of 476,251 signatures (consolidating replicates) (11). The raw data are available from GEO, accession number GSE92742.

Weighted connectivity score: Our input into the CLUE query system was a list of up (q_{up}) and down (q_{down}) regulated genes that were compared to the reference database (Touchstone, r) in LINCS using the weighted Kolmogorov-Smirnov enrichment statistic (29) as shown below:

$$w_{q,r} = \begin{cases} (ES_{up} - ES_{down})/2, & \text{if } \text{sgn}(ES_{up}) \neq \text{sgn}(ES_{down}) \\ 0, & \text{otherwise} \end{cases}$$

Where $w_{q,r}$ is the weighted connectivity score which represent the similarity measure between a query q (q_{up} , q_{down}) and a reference signature r . ES_{up} is the enrichment of q_{up} in r and ES_{down} is the enrichment of q_{down} in r .

Drug ranking score: To rank drugs, we empirically developed a novel drug prioritization system based on four metrics: CLUE connectivity score in MCF7 cell line subject to drug treatment ($S_{1_{di}}$), drug selectivity score ($S_{2_{di}}$), genetic perturbant score ($S_{3_{di}}$), and drug class score ($S_{4_{di}}$).

$$S_{t_{di}} = S_{1_{di}} + S_{2_{di}} + S_{3_{di}} + S_{4_{di}} \quad (2)$$

Where $S_{t_{di}}$ is the total score of drug i

$S_{1_{di}}$ is the potency score of drug i to reverse the gene expression of the DEGs in MCF7 cell line.

$$S_{1_{di}} = \frac{CS_{CLUE\ i}}{-100} \quad (2.1)$$

CS_{CLUE} is the connectivity score of the drug i in MCF7 cell line generated from the CLUE server (11).

S_{2di} is the selectivity score of drug i

$$S_{2di} = 1 - TAS_i \quad (2.2)$$

TAS_i is the average transcriptional impact score of drug i which describes its activity, relative to all other drugs, as derived from its replicate reproducibility and magnitude of differential gene expression (11).

S_{3di} is the genetic perturbation score that represents the averaged connectivity score of knocking down the target genes ($1, \dots, j$) of drug i

$$S_{3di} = \frac{AVG(CS_{(TG1)i}, \dots, CS_{(TGj)i})}{-100} \quad (2.3)$$

$CS_{(TGj)i}$ is the connectivity score of knocking down target gene j by drug i

S_{4di} is the class score of drug i . It represents the number of drugs that have negative connectivity scores and belong to the same class as drug i .

$$S_{4di} = \frac{negR_{class\ of\ drug\ i}}{100} \quad (2.4)$$

$negR_{class\ of\ drug\ i}$ is the ratio between number of drugs that have negative score and the total number of drugs belonging to the class of drug i .

$$negR_{class\ of\ drug\ i} = \frac{n_{NEG}}{n_{NEG} + n_{POS}}$$

n_{NEG} is number of drugs that have negative connectivity score and belong to class of drug i . And n_{POS} is number of drugs that have positive connectivity score and belong to class of drug i .

Cell culture validation

Tamoxifen resistance cell line (MCF7-RR) was purchased from EMD Millipore (Burlington, MA) and the MCF7 cell line was a gift from the HTS lab (Cancer Centre-University of Hawaii, HI). The BC cell lines were maintained at 37°C in a humidified atmosphere (5% CO₂) in phenol-red free IMEM media containing

5% fetal bovine serum and 1% penicillin-streptomycin. All media and supplements were obtained from Gibco-Invitrogen, USA.

Cells were expanded in T-75 flasks and when confluent, were trypsinized and seeded in a 96-well plate (BD Biosciences) at a concentration of 5,000 cells in 200 μ L of growth media per well. Cells were left to adhere overnight and the following day treated with the drugs in fivefold serial dilutions ranging from 100 to 0.0064 μ M. After 72 hours of treatment, old media was replaced with fresh media and proliferation was assessed with CellTiter-Glo Luminescent Cell Viability Assay Kit (Promega). After adding CellTiter reagent, plates were placed on an orbital shaker for 2 minutes and further incubated for 10 minutes. Luminescence was read using Victor V3 plate-reader (Perkin Elmer).

We purchased compounds ZM-447439, palbociclib, raltitrexed, oxindole-I and tamoxifen from Selleckchem (Houston, TX), and BMS-754807 from Sigma (North Liberty, IA). We reconstituted the drugs according to the manufacturer's instructions and stored the stock solutions at -20°C.

MCF7-RR were exposed to varying concentrations of tamoxifen (10 nM, 100 nM, 1000 nM) to determine the synergistic effect of the drugs. 300,000 cells were seeded and allowed to adhere overnight at 37°C. The following day, the five drugs were added and cells were incubated for 72 hours with a media change every 2 days. At the end of incubation, cells were counted using trypan blue stain and a hemocytometer. The IC50 was determined by the concentration at which half the cells were killed.

qPCR validation

Extracted RNA from MCF7-RR was reverse transcribed into cDNA and quantified using qPCR using SYBR Green RT-PCR kit on Step One Plus. PCR reactions were subject to 40 cycles of denaturation (95°C, 30s). Standard curve for each primer was constructed using serial dilutions of cDNA and normalized to GAPDH housekeeping gene. Primers; (5'-3') AurkA: ATT GCA GAT TTT GGG TGG T, AAA CTT

CAG TAG CAT GTT CCT GTC, AurkB: GTC CTC CGG AAA GAG CCT GT, TTG AGC GCC ACG
ATG AAA TG.

Drug synergetic effect calculation

To investigate the *in vitro* pharmaco-dynamic interactions of combining ZM-447439 with tamoxifen, we used CompuSyn v1.0 (22) software to calculate the combination index (CI) based on the median-effect principle (30), where $CI < 1$, $= 1$, and > 1 indicates synergism, additive effect and antagonism, respectively.

Results

Drug reposition bioinformatics pipeline

The overall design of our proposed bioinformatics pipeline is shown in Figure 1. We used five datasets that interrogated the gene expression changes between tamoxifen resistant tamoxifen resistance and sensitive tumors (12-14) and cell lines (15, 16). Briefly, we first used the list of significant differentiated expressed genes (DEGs, adjusted p-value < 0.05) between tamoxifen resistant and sensitive tumors/cell lines as reported in the previous five studies (12-16), with a total of 229 tamoxifen resistant tumors, 363 tamoxifen sensitive tumors, as well as tamoxifen sensitive or resistant cell lines. Then we selected the 864 genes that showed consistency as either up- or down- regulated DEGs among these datasets. Next we identified the drugs and compounds that can reverse the expression of the consistent DEGs using CLUE platform (11), with an innovative and comprehensive scoring system (see Materials and Methods section). We also performed quality assessment on the results and prioritized drugs based on our proposed drugs ranking system. Finally, we validated top five drugs using *in vitro* cell culture experiments.

Bioinformatics analysis results on differentially expressed genes

We first identified up-regulated and down-regulated genes using the five datasets (Table 1). All together we identified a total of 864 unique DEGs (467 up and 397 down-regulated) between tamoxifen resistance and sensitive tumors/cell lines, after FDR correction (adjusted p-values <0.05). Table S1 in the supplementary file listed the details of up-regulated and down regulated genes in each dataset. To understand the biological functions involved in tamoxifen resistance, we performed a pathway and gene ontology enrichment analysis on the DEGs above, using the ConsensusPathDB resource (17). We mapped 607 genes out of 864 DEGs (70%) to the database of the ConsensusPathDB (17). The pathway enrichment analysis identified 85 significant pathways (adjusted p value <0.05, Figure 2 and Table S2) in six different popular databases (Wikipathways (12), Reactome (40), Netpath (2), INOH (1), PID (10) and KEGG (20)). Additionally, the GO enrichment analysis identified 109 GO entities: 97 biological processes, 6 molecular functions and 6 cellular components (Table S3). We ranked these significant pathways by the rate of enrichment, that is, the ratio between the number of DEGs and the total number of genes involved in a specific pathway (Figure 2). Foxm1 transcription factor network and Aurora B signaling pathways have the two highest enrichment rates of 25% each (Figure 2); 12 DEGs are detected out of 43 annotated genes in these two pathways, according to pathway interaction database (PID) (18). Thus genes in these specific pathways/GOs are good candidates to be selected to reverse the tamoxifen resistance phenotype.

Table S4 shows the number of up and down-regulated DEGs per pathway/GO. Interestingly, most of the pathway/GO that have a high enrichment rate also have a larger number of corresponding up-regulated genes despite that the enrichment rate is normalized by gene set size. Each dataset contributes differently with the number of genes to these pathways. For example, the Foxm1 transcription factor network pathway has 4, 3, 3, 0, 2 up-regulated genes from dataset 1, 2, 3, 4 and 5 respectively. This emphasizes the feasibility of aggregating independent datasets.

In order to select genes as the input for CLUE query system (www.clue.io) to identify the candidate drugs, we developed a comprehensive scoring system (see the methods section) to rank the gene list. This scoring system takes into account of the drug connectivity score of MCF7 cell line subject to drug treatment in

CLUE server (S_1), drug selectivity score (S_2), genetic perturbing score (S_3), and drug class score (S_4), and combine them into a single total score (S_t). We selected the 150 most up-regulated genes from the high ranked pathways/GOs. Table S4 shows the list of pathways/GOs ranked by its total score (S_t). Pathway APC/C mediated degradation of cell cycle proteins has the highest total score ($S_t=1.37$). It has a score of $S_1=0.24$ as enrichment score, $S_2=1$ with 11 up-regulated DEGs, and $S_3=0.12$ as many of these 11 genes involved in other pathways/GOs. The final list of 150 up-regulated DEGs is listed in Table S5. 72, 44, 35, 4, 28 genes were selected from Datasets 1-5, respectively. We used the same approach above to select the 150 down-regulated DEGs (Table S5), except that we changed S_2 to the down/up rate score instead of up/down rate score. Table S6 shows the pathways/GO ranked based on its total score (S_t). Pathway Beta1 integrin cell surface interactions had the highest total score ($S_t=1.08$), with 11 down-regulated genes. The names of the 150 down-regulated genes are shown in Table S5. 65, 5, 49, 1, 34 genes were selected from Datasets 1-5 respectively.

Filtering drug candidates with potential to re-sensitize tamoxifen resistance

We had 190 drugs and 376 genetic perturbants that have connectivity score < -90 (Table S7 and Table S8) in the MCF7 cell line. From Table S7, palbociclib, an CDK inhibitor, had the best connectivity score ($Cs=-99.96$) which is consistent with reports of the potential of CDK inhibitors to significantly reduce tamoxifen resistance (5, 19, 20).

Within the 190 hits, it is possible that false positive and inconsistent drugs exist, as such, it is critical to evaluate multiple replicates of the same compound. For example, some drugs were purchased from different vendors and produced inconsistent connectivity scores. As another example, the drug simvastatin (Broad ID #; BRD-KA81772229) produced negative score ($Cs=-99.62$) however, its identical structure from another vendor (BRD-K22134346) produced positive results ($Cs=30.20$). Such discrepancy may be due to the difference in drug purity from different vendors (personal communication with CLUE team). We removed 30 drugs that have inconsistent scores, such as simvastatin.

Another method to improve the biological relevance of predicted drug responses is to integrate the connectivity score of the drugs (Table S7) with its genetic perturbants (Table S8). We assumed that knocking down the target genes of the effective drugs ($Cs < -90$) should also produce negative connectivity score. For example, palbociclib ($Cs = -99.96$) has three target genes (CDK4, CDK6, and CCND3). Knocking down of palbociclib's target genes produced negative scores -88, -59 and -67 respectively, indicating the credibility of the connectivity score of palbociclib. Due to this concern, we removed additional 22 drugs that did not have any target genes. Finally, among the drugs each of which has a single target gene, we removed 38 that have scores inconsistent with those by knocking down its target gene. For example, the connectivity score of a JAK3-inhibitor drug (BRD-K04546108) is -99.85, however the score of knocking down JAK3 gene is 80.76, which indicates predicated response may not be reliable. After applications of these quality control steps, our final list contains 100 drugs and compounds (Figure 3).

Prioritizing drug selection with a scoring system

The ultimate goal of this study is to find candidate drugs that overcome tamoxifen resistance. To achieve this goal, we developed a drug prioritization system to rank the final list of 100 drugs (Figure 3) based on the summarization of four scores: CLUE connectivity score in MCF7 cell line ($S_{1_{di}}$), drug selectivity score ($S_{2_{di}}$), genetic perturbant score ($S_{3_{di}}$), and the drug class score ($S_{4_{di}}$). We hypothesized that the best drug had to meet four criteria: high efficacy; high negative connectivity score ($S_{1_{di}}$, produced from CLUE server) to reverse DEGs in MCF7 cell line; high selectivity ($S_{2_{di}}$), consistent with knocking down its target genes ($S_{3_{di}}$); and most of the drugs belonging to its drug class have negative connectivity score ($S_{4_{di}}$).

The 100 drugs belong to 123 unique classes. For some classes, such as mTOR inhibitors, all the drugs that belong to it (10 drugs), produced negative connectivity score (Figure 4, Table S9). This indicates that the drugs in this class have the potential to reverse tamoxifen resistance.

The final list of 100 drugs ranked by its total score (S_t) is shown in Table S10. Among them, 13 drugs have total scores (S_t) > 0.7, which belong to the following classes: Aurora kinase inhibitor (ZM-447439 (S_t = 0.83), alisertib (0.81), barasertib (0.7)); CDK inhibitor (palbociclib (0.82); Ribonucleotide reductase inhibitor (gemcitabine(0.79)); HDAC inhibitor (depudecin (0.76)); Thymidylate synthase inhibitor (raltitrexed (0.75)); MTOR inhibitor (temsirolimus (0.75)); IGF-1 inhibitor (BMS-754807(0.74), linsitinib (0.70)); MEK inhibitor (selumetinib (0.72)); IMPDH inhibitor (azathioprine (0.72)); VEGFR inhibitor (oxindole-I (0.71)).

Experimental validation of top 5 ranked drug and compounds

We validated five drugs out of the list of 100 drugs (Table S10). We validated two top scoring drugs (ZM-447439 and palbociclib) as well as the three other drug candidates that had not been tested in clinical trials (raltitrexed, BMS-754807 and oxindole-I). We evaluated these five drugs for cytotoxic effects in ER-positive tamoxifen-resistant MCF7-RR breast cancer cell line *in vitro* via cell viability assays, with or without the presence of tamoxifen (Figure 5A-E).

Cell viability assay results reveal that each of the five drugs was able to induce pronounced cell death on its own after reaching a certain concentration (red curves). However, cell viability dropped even faster when tamoxifen was added in combination with one of the five drug candidates (green, blue and purple curves). For example, adding increasing concentrations of tamoxifen to ZM-447439 further reduced cell viability (increased cell death) in a dose-dependent manner, indicating that ZM-447439 was able to re-sensitize tamoxifen-resistant breast cancer cells to tamoxifen antitumoral effects. The IC₅₀ of ZM-447439, palbociclib, raltitrexed, BMS-754807 and oxindole-I combined with tamoxifen (10 nM) were 2.43, 12.3, 8.3, 16.981, and 11.13 μ M respectively (Figure 5F), demonstrating that ZM-447439 is most potent among all five selected candidates. Subsequent quantitative real-time PCR experiments confirmed that AURKA gene expression significantly dropped (~10 fold change) after one hour of treatment using 10 nM tamoxifen

and 0.5 μ M ZM-447439 (Figure 5G). Further reduction in Aurora A/B mRNA levels was noticed when 0.5 μ M ZM-447439 was combined with 100nM tamoxifen. Relative to AURKA, AURKB has less significant fold changes. This trend is supported by previous study that showed AURKA was a determinant of tamoxifen sensitivity (21). The in vitro experimental results provide the preliminary evidence that ZM-447439, a known Aurora kinase inhibitor, is an alternative drug candidate that can re-sensitize BC patients toward tamoxifen treatment.

Lastly, we also explored the pharmaco-dynamic interactions between ZM-447439 and tamoxifen. Towards this, we used software CompuSyn v1.0 (22) to compute the combination index (CI) (Figure 5H). $CI < 1$, $= 1$, and > 1 indicate synergism, additive effect and antagonism, respectively. Since the CI values between ZM-447439 and tamoxifen ranging from 0.75 to 0.02 for $fa = 0.07 \sim 0.99$ in the Fa-CI plot, it strongly suggests that these two drugs are synergistic.

Discussion

In this study, we developed a systematic computational drug repositioning strategy to rapidly identify FDA-approved or investigational drugs that have the potential to overcome tamoxifen resistance in breast cancer patients. We identified ZM-447439, a known AURORA kinase inhibitor, as a potential drug candidate for treatment of patients with tamoxifen-resistant breast tumors.

BC remains the most prevalent cancer in women, with the majority of these tumors expressing ER α . Resistance to endocrine therapies remains a critical limitation towards curing a significant subset of ER+ patients. A large number of gene expression signatures pertaining to tamoxifen resistance and the behavior of drugs in many BC cell lines were generated although few examples have been translated into effective clinical treatments. A key challenge that has been addressed in this paper is the extraction of meaningful information from the big data that has been generated to accelerate the translation into effective medical treatments for BC patients with tamoxifen resistance. To this end, we have succeeded in solving two significant challenges. The first challenge consisted in the absence of reliable overlapped biomarker within

the list of DEGs reported by independent datasets used in this study (lack of reproducibility at individual gene level). Our solution was to find a group of functionally related genes rather than a single gene that contributes significantly to the resistance of tamoxifen. To achieve this, we developed a pathway ranking system to score pathway and GO terms (Equation 1). This highly selective method of DEGs involving stringent selection criteria improved the accurate identification of target genes with likely relevance to tamoxifen resistance. Second, a large number of drugs that have high negative connectivity score from CLUE platform may be false positive. We used a novel drug scoring system to score drugs not based only on CLUE score but based on target genes perturbation score and drug class score.

Our rigorous bioinformatics pipeline, followed by *in vitro* viability experiments, successfully identified five novel compounds of distinct chemical structure and primary mechanisms, as being able to reverse tamoxifen resistance gene expression. Among them, ZM-447439 is most potent in this regard. The strong synergism effect ($CI < 0.1$) when combined with tamoxifen, adds such confidence. Additionally, many other previous clinical or experimental results support the potential role of ZM-447439 proposed here. ZM-447439 was in phase 1 clinical trials for other gynecologic cancers (23). It also appeared to sensitize resistance cancer cells to cisplatin in treating cervical cancers (24). ZM-447439 showed both anti-proliferative and pro-apoptotic properties in gastroenteropancreatic neuroendocrine tumor cells (25). It likely reduces cell proliferation by triggering aberrant cell division and mitotic defects (23). It was also reported to reduce histone H3 phosphorylation at Ser10 (H3S10ph) in Hep2 carcinoma cells (26). H3S10ph is known to be mediated by Aurora kinase, the known target of ZM-447439, to regulate chromatin condensation during cell division (Sawicka and Seiser, 2012, PMID 22564826). Therefore, the molecular mechanisms of anti-proliferative and pro-apoptotic effects for ZM-447439 are likely complex, and Aurora kinase inhibition is part of it.

Aurora kinase family has three members: Aurora kinase A, B, and C. It has been reported that inhibition of AURKB with barasertib (drug #14 in Table S10) could be a candidate to treat BC patients with tamoxifen resistance (27). In addition, AURKA was overexpressed in NSCLC and contributed to cisplatin-based

chemotherapy resistance (28). To our knowledge, the combined inhibition AURKA and AURKB with ZM-447439 has not been described to restrict growth of antiestrogen resistant BC cells elsewhere. Our in vitro data indicate that ZM-based treatment alone or in combination with tamoxifen may induce a regression of tamoxifen-resistant tumors in breast cancer patients in vivo. Co-treatments of tamoxifen with aurora kinase inhibitor such as ZM-447439 could be especially favorable *in vivo*, as we have shown the combination to be synergistic, that is, the antitumor property of their mixture is likely to be greater than either drug separately.

Conclusions

we propose a generalizable, systematical bioinformatics pipeline to connect tamoxifen resistance based on consistent gene expression in different datasets with drug perturbation expression profiles. We successfully and preliminarily demonstrated that drugs that have high scores from such a pipeline show potency to reverse tamoxifen resistance, with ZM-447439 being the most promising candidate.

Ethical Approval and Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of supporting data

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Disclosure of Potential Conflicts of Interest

The authors declare no conflicts of interest

Authors' Contributions

Supervision of the work: L.X. Garmire

Conception and design: F.M Alakwaa

Development of methodology: F.M Alakwaa

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P Benny, F.M Alakwaa

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L.X Garmire, F.M Alakwaa, K Chaudhary, P Benny

Writing, review, and/or revision of the manuscript: L.X Garmire, P Benny, K Chaudhary, F.M Alakwaa

Competing interests

The author(s) declare no competing interests.

Funding

This research was supported by grants K01ES025434 awarded by NIEHS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (<http://datascience.nih.gov/bd2k>), P20 COBRE GM103457 awarded by NIH/NIGMS, and R01 LM012373 and R01 LM012907 awarded by NLM, R01 HD084633 awarded by NICHD to LX Garmire.

Acknowledgements

We gratefully acknowledge Dr Herbert Yu from the University of Hawaii-Cancer Center for introducing the research question addressed in the paper.

Tables

Table 1: A list of microarray and RNA-seq datasets used in this study

Microarray	Studies			Number of tumor samples		Number of DEGs	
	No	Ref	Year	Tamoxifen resistance	Tamoxifen sensitive	Up	Down
	1	(13)	2007	85	91	173	102
	2	(14)	2008	26	138	86	44
	3	(12)	2008	118	134	118	134
RNA-seq	Studies			Number of replicates		Number of DEGs	
	No	Ref	Year	Tamoxifen resistance	Tamoxifen sensitive	Up	Down
	4	(16)	2012	3	3	10	10
	5	(15)	2011	3	3	131	121

Figure Legends

Figure 1: Bioinformatics analysis pipeline to identify drugs that overcome tamoxifen resistance. In step 1, we identified the significant differentiated genes (DEGs) (864 unique genes, 467 up and 397 down genes) between tamoxifen resistance and tamoxifen sensitive tumors/cell lines among five datasets used in this study (Table 1). The first 3 studies (D1-D3) are primary BC patient samples, and the study 4 and 5 (D4-5) are BC cell line samples. In step 2, we selected 300 genes that are consistent among the five datasets using

our proposed pathway scoring system (equation 1, Materials and Methods). In step 3, we used CLUE query system (www.clue.io) to identify the candidate drugs (190 drugs) that can reverse the gene expression of the 300 consistent genes. In step 4, we removed the false positive drugs and ranked drugs based on our novel drugs scoring system (equation 2, Materials and Methods). In step 5, we experimentally validated five top drug candidates with *in vitro* cell culture.

Figure 2: The ratio of enrichment (x-axis) of the significant pathways (adjusted p value <0.05, y-axes). We performed pathway enrichment analysis on the significant differentiated genes “DEGs” (864 unique genes) between tamoxifen resistance (tamoxifen resistance) and sensitive (tamoxifen sensitive) tumors/cell lines in five datasets used in this study. Ratio of enrichment represents the ratio between the number of DEGs and the total number of annotated genes in a specific pathway (Equation 1.1). Color represents the source of the pathway databases: Wikipathways (31), Reactome (32), Netpath (33), INOH (34), PID (18) and KEGG (35) .

Figure 3: A flow chart of drugs filtration (A) and ranking (B). (A) We performed filtering on 2,773 drugs based on the following criteria. First, we selected 190 drugs out of 2,873 that had been tested in the Library of Integrated Network Cellular Signatures (LINCS) database (11) and had negative connectivity score ($C_s < -90$). Second, we removed 30 drugs if the same drug was purchased from different vendors but one of them produced positive score. Third, we removed 22 drugs that have no annotated target genes. Finally, we removed 38 drugs that have inconsistent scores with its target genes. (B) We ranked the final 100 drugs based on the summarized average of four criteria: S_1 is the drug connectivity score in MCF7 cell line calculated by CLUE system [10]; S_2 is the drug selectivity score; S_3 is the drug consistency score between the connectivity score of the drug and its target genes; S_4 is the score of drug class which represents the proportion of drugs with negative connectivity score in that drug class.

Figure 4: The score of drug class. This score represents the ratio between the drugs that have negative connectivity scores and the total number of drugs belonging to this class. For example, the score of mTOR inhibitor class is 100 as all drugs in this class have negative scores.

Figure 5: In vitro experimental validation of top drug candidates predicted. (A-E) Drug efficacy in MCF7-RR cell line measured by cell proliferation assay after 72 h treatment. MCF7-RR cells were treated with (vehicle, 0.0064, 0.032, 0.16, 0.8, 4, 20, 100) μ M of five different drugs and/or various doses (vehicle, 10 nM, 100 nM, and 1000 nM) of tamoxifen for 72 h. The drugs are: ZM-447439 (A), BMS-754807 (B), Palbociclib (C), oxindole-1 (D) and raltitrexed (E). (F) IC-50 (μ M) of five drugs combined with different concentrations (10 nM, 100 nM, and 1000 nM) of tamoxifen. (G) qPCR fold change of AURKA (blue) gene and AURKB (red) between treated vs untreated MCF7-RR cell line. cell line was treated with combinations of ZM (0.5 μ M) and three different concentrations of tamoxifen (10 nm, 100 nm and 1000 nm). (H) Combination index (CI) plot of tamoxifen and ZM-447439 as a function of effect level (fa) on the x-axis. Combination index (y-axis) is a quantitative measure of the degree of drug interaction in terms of synergism and antagonism for a given endpoint of the effect measurement. CI=1 indicates additive effect, CI<1 indicates synergism and CI>1 indicates antagonism. This figure was generated using CompuSyn v1.0 (22).

Supplementary Materials

Supplementary Table 1: A list of up-regulated and down-regulated significant differentiated genes (DEGs) among the five different datasets used in this study (Table 1).

Supplementary Table 2: A list of 85 significant pathways as outcome of pathway analysis of the 864 unique DEGs genes between tamoxifen resistance and sensitive tumor/cell.

Supplementary Table 3: A list of 109 significant gene ontology (GOs) terms as outcome of GO analysis of the 864 unique DEGs genes between tamoxifen resistance and sensitive tumor/cell.

Supplementary Table 4: Pathways and Gene Ontology (GOs) terms ranked by its enrichment rate, up/down score and overlapped score

Supplementary Table 5: A list of selected 150 up-regulated and 150 down-regulated genes based on pathways scoring system

Supplementary Table 6: Pathways and GOs ranked by its enrichment rate, down/up score and overlapped score

Supplementary Table 7: A list of drugs ranked by the connectivity score of MCF7 cell line

Supplementary Table 8: A list of genetic perturbants (knockdown and over-expressed genes) ranked by the connectivity score in MCF7 cell line

Supplementary Table 9: A list of drug classes ranked by the score of drug classes that represents the ratio between number of drugs that have negative connectivity score and total number of drugs belong to the same class. For example, the score of MTOR class is 100 because all drugs (10 drugs) belong to this class produce negative connectivity score.

Supplementary Table 10: A list of drugs ranked by its total score (S_T)

1. U. S. D. o. H. H. Services. (2019).
2. S. Saha Roy, R. K. Vadlamudi, Role of estrogen receptor signaling in breast cancer metastasis. *International journal of breast cancer* **2012**, 654698-654698 (2012).
3. W. Fan, J. Chang, P. Fu, Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies. *Future medicinal chemistry* **7**, 1511-1519 (2015).
4. S. Ali *et al.*, Molecular mechanisms and mode of tamoxifen resistance in breast cancer. *Bioinformation* **12**, 135-139 (2016).

5. R. B. Riggins, R. S. Schrecengost, M. S. Guerrero, A. H. Bouton, Pathways to tamoxifen resistance. *Cancer letters* **256**, 1-24 (2007).
6. A. Milani, E. Geuna, G. Mittica, G. Valabrega, Overcoming endocrine resistance in metastatic breast cancer: Current evidence and future directions. *World journal of clinical oncology* **5**, 990-1001 (2014).
7. M. Zhao, B. Ramaswamy, Mechanisms and therapeutic advances in the management of endocrine-resistant breast cancer. *World journal of clinical oncology* **5**, 248-262 (2014).
8. I. Mayer, Role of mTOR inhibition in preventing resistance and restoring sensitivity to hormone-targeted and HER2-targeted therapies in breast cancer. *Clinical advances in hematology & oncology : H&O* **11**, 217-224 (2013).
9. L. Huang, S. Zhao, J. M. Frasier, Y. Dai, An Integrated Bioinformatics Approach Identifies Elevated Cyclin E2 Expression and E2F Activity as Distinct Features of Tamoxifen Resistant Breast Tumors. *PLOS ONE* **6**, e22274 (2011).
10. S. Pushpakom *et al.*, Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **18**, 41 (2018).
11. A. Subramanian *et al.*, A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e1417 (2017).
12. M. Chanrion *et al.*, A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 1744-1752 (2008).
13. S. Loi *et al.*, Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
14. S. Loi, M. Piccart, C. Sotiriou, The use of gene-expression profiling to better understand the clinical heterogeneity of estrogen receptor positive breast cancers and tamoxifen response. *Critical Reviews in Oncology/Hematology* **61**, 187-194 (2007).
15. K. J. Huber-Keener *et al.*, Differential gene expression in tamoxifen-resistant breast cancer cells revealed by a new analytical model of RNA-Seq data. *PloS one* **7**, e41333-e41333 (2012).
16. A. M. Mendes-Pereira *et al.*, Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2730-2735 (2012).
17. A. Kamburov *et al.*, ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Research* **39**, D712-D717 (2010).
18. C. F. Schaefer *et al.*, PID: the Pathway Interaction Database. *Nucleic acids research* **37**, D674-D679 (2009).
19. S. Pernas, S. M. Tolaney, E. P. Winer, S. Goel, CDK4/6 inhibition in breast cancer: current practice and future directions. *Ther Adv Med Oncol* **10**, 1758835918786451-1758835918786451 (2018).
20. A. AlFakeeh, C. Brezden-Masley, Overcoming endocrine resistance in hormone receptor-positive breast cancer. *Curr Oncol* **25**, S18-S27 (2018).
21. X. Q. Zheng *et al.*, Aurora-A is a determinant of tamoxifen sensitivity through phosphorylation of ER α in breast cancer. *Oncogene* **33**, 4985-4996 (2014).
22. T.-C. a. M. Chou. (ComboSyn, Inc. Paramus, NJ, 2007).
23. K. Umene *et al.*, Aurora kinase inhibitors: Potential molecular-targeted drugs for gynecologic malignant tumors. *Biomed Rep* **1**, 335-340 (2013).

24. L. Zhang, S. Zhang, ZM447439, the Aurora kinase B inhibitor, suppresses the growth of cervical cancer SiHa cells and enhances the chemosensitivity to cisplatin. *J Obstet Gynaecol Res* **37**, 591-600 (2011).
25. I. Georgieva *et al.*, ZM447439, a novel promising aurora kinase inhibitor, provokes antiproliferative and proapoptotic effects alone and in combination with bio- and chemotherapeutic agents in gastroenteropancreatic neuroendocrine tumor cell lines. *Neuroendocrinology* **91**, 121-130 (2010).
26. Z.-J. Long *et al.*, ZM 447439 inhibition of aurora kinase induces Hep2 cancer cell apoptosis in three-dimensional culture. *Cell Cycle* **7**, 1473-1479 (2008).
27. S. L. Larsen *et al.*, Aurora kinase B is important for antiestrogen resistant cell growth and a potential biomarker for tamoxifen resistant breast cancer. *BMC Cancer* **15**, 239-239 (2015).
28. A. Tang *et al.*, Aurora kinases: novel therapy targets in cancers. *Oncotarget* **8**, 23937-23954 (2017).
29. J. Lamb *et al.*, The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* **313**, 1929 (2006).
30. T.-C. Chou, Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacological Reviews* **58**, 621 (2006).
31. D. N. Slenter *et al.*, WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research* **46**, D661-D667 (2017).
32. D. Croft *et al.*, Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research* **39**, D691-D697 (2011).
33. K. Kandasamy *et al.*, NetPath: a public resource of curated signal transduction pathways. *Genome Biology* **11**, R3 (2010).
34. S. Yamamoto *et al.*, INOH: ontology-based highly structured database of signal transduction pathways. *Database (Oxford)* **2011**, bar052-bar052 (2011).
35. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).

Figure 1

Figure 2

Figure 3

(A)

(B)

Figure 4

Figure 5

Figures

Figure 1

Bioinformatics analysis pipeline to identify drugs that overcome tamoxifen resistance. In step 1, we identified the significant differentiated genes (DEGs) (864 unique genes, 467 up and 397 down genes) between tamoxifen resistance and tamoxifen sensitive tumors/cell lines among five datasets used in this study (Table 1). The first 3 studies (D1-D3) are primary BC patient samples, and the study 4 and 5 (D4-5) are BC cell line samples. In step 2, we selected 300 genes that are consistent among the five datasets using our proposed pathway scoring system (equation 1, Materials and Methods). In step 3, we used CLUE query system (www.clue.io) to identify the candidate drugs (190 drugs) that can reverse the gene expression of the 300 consistent genes. In step 4, we removed the false positive drugs and ranked drugs based on our novel drugs scoring system (equation 2, Materials and Methods). In step 5, we experimentally validated five top drug candidates with in vitro cell culture.

Figure 2

The ratio of enrichment (x-axis) of the significant pathways (adjusted p value <0.05, y-axes). We performed pathway enrichment analysis on the significant differentiated genes “DEGs” (864 unique genes) between tamoxifen resistance (tamoxifen resistance) and sensitive (tamoxifen sensitive) tumors/cell lines in five datasets used in this study. Ratio of enrichment represents the ratio between the number of DEGs and the total number of annotated genes in a specific pathway (Equation 1.1). Color represents the source of the pathway databases: Wikipathways (31), Reactome (32), Netpath (33), INOH (34), PID (18) and KEGG (35) .

Figure 3

A flow chart of drugs filtration (A) and ranking (B). (A) We performed filtering on 2,773 drugs based on the following criteria. First, we selected 190 drugs out of 2,873 that had been tested in the Library of Integrated Network Cellular Signatures (LINCS) database (11) and had negative connectivity score ($C_s < -90$). Second, we removed 30 drugs if the same drug was purchased from different vendors but one of them produced positive score. Third, we removed 22 drugs that have no annotated target genes.

Finally, we removed 38 drugs that have inconsistent scores with its target genes. (B) We ranked the final 100 drugs based on the summarized average of four criteria: S1 is the drug connectivity score in MCF7 cell line calculated by CLUE system [10]; S2 is the drug selectivity score; S3 is the drug consistency score between the connectivity score of the drug and its target genes; S4 is the score of drug class which represents the proportion of drugs with negative connectivity score in that drug class.

Figure 4

The score of drug class. This score represents the ratio between the drugs that have negative connectivity scores and the total number of drugs belonging to this class. For example, the score of mTOR inhibitor class is 100 as all drugs in this class have negative scores.

Figure 5

In vitro experimental validation of top drug candidates predicted. (A-E) Drug efficacy in MCF7-RR cell line measured by cell proliferation assay after 72 h treatment. MCF7-RR cells were treated with (vehicle, 0.0064, 0.032, 0.16, 0.8, 4, 20, 100) μ M of five different drugs and/or various doses (vehicle, 10 nM, 100

nM, and 1000 nM) of tamoxifen for 72 h. The drugs are: ZM-447439 (A), BMS-754807 (B), Palbociclib (C), oxindole-I (D) and raltitrexed (E). (F) IC-50 (μ M) of five drugs combined with different concentrations (10 nM, 100 nM, and 1000 nM) of tamoxifen. (G) qPCR fold change of AURKA (blue) gene and AURKB (red) between treated vs untreated MCF7-RR cell line. cell line was treated with combinations of ZM (0.5 μ M) and three different concentrations of tamoxifen (10 nm, 100 nm and 1000 nm). (H) Combination index (CI) plot of tamoxifen and ZM-447439 as a function of effect level (fa) on the x-axis. Combination index (y-axis) is a quantitative measure of the degree of drug interaction in terms of synergism and antagonism for a given endpoint of the effect measurement. CI=1 indicates additive effect, CI<1 indicates synergism and CI>1 indicates antagonism. This figure was generated using CompuSyn v1.0 (22).

Supplementary Files

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

- [TMSTablesnew.xlsx](#)