A Novel Method of Transcriptional Response Analysis to Facilitate Drug Repositioning for Cancer Therapy

Guangxu Jin\textsuperscript{1,2}, Changhe Fu\textsuperscript{1}, Hong Zhao\textsuperscript{1,2}, Kemi Cui\textsuperscript{1,2}, Jenny Chang\textsuperscript{1,2,3}, and Stephen T.C. Wong\textsuperscript{1,2,3}

Abstract

Little research has been done to address the huge opportunities that may exist to reposition existing approved or generic drugs for alternate uses in cancer therapy. In addition, there has been little work on strategies to reposition experimental cancer agents for testing in alternate settings that could shorten their clinical development time. Progress in each area has lagged, in part, because of the lack of systematic methods to define drug off-target effects (OTE) that might affect important cancer cell signaling pathways. In this study, we addressed this critical gap by developing an OTE-based method to repurpose drugs for cancer therapeutics, based on transcriptional responses made in cells before and after drug treatment. Specifically, we defined a new network component called cancer-signaling bridges (CSB) and integrated it with a Bayesian factor regression model (BFRM) to form a new hybrid method termed CSB-BFRM. Proof-of-concept studies were conducted in breast and prostate cancer cells and in promyelocytic leukemia cells. In each system, CSB-BFRM analysis could accurately predict clinical responses to more than 90% of drugs approved by the U.S. Food and Drug Administration and more than 75% of experimental clinical drugs that were tested. Mechanistic investigation of OTEs for several high-ranking drug–dose pairs suggested repositioning opportunities for cancer therapy, based on the ability to enforce retinoblastoma-dependent repression of important E2F-dependent cell-cycle genes. Together, our findings establish new methods to identify opportunities for drug repositioning or to elucidate the mechanisms of action of repositioned drugs. *Cancer Res;* 72(1); 33–44. ©2011 AACR.

Major Findings

By developing a systematic approach to characterize off-target effects of drugs, this study may help the field learn how to reposition existing approved and generic drugs for alternate uses in cancer treatment.

Introduction

The study of drug repositioning has so far been limited to the “on-target repositioning” that applies a drug’s known pharmacologic mechanism to a different therapeutic indication, for example, comparing the structural similarities of small molecules (1, 2) or known side effects (3). In contrast, “off-target repositioning” attempts to describe the pharmacologic mechanisms still unclear for known molecules. A number of approaches have recently been developed for off-target repositioning with the use of gene signatures (4–7),—subsets of genes, or drug similarity networks (8), identified in the cancer transcriptional profiles following drug treatment. One common limitation of these methods is that they do not include the disease-specific prior knowledge or known mechanisms in the off-target repositioning process, so that they can be used to find similarities between the drugs but not the preference between them. Thus, we need to develop a method that incorporates prior knowledge of specific diseases to provide a more precise off-target drug repositioning.

A primary challenge of off-target repositioning is to address the off-target effects (OTE) of a drug on the proteins downstream in the signaling pathways and the genes that are regulated by those proteins. As an example, in breast cancer, raloxifene, tamoxifen, and fulvestrant are the pioneering drugs targeting the estrogen receptor (ER, ref. 9). The targeted proteins, however, often generate downstream effects on the linked signaling proteins and ultimately exert unexpected OTEs on cancer transcription (10–12). Creighton and colleagues (12) showed that tamoxifen together with estrogen deprivation (ED) can shut down classic estrogen signaling and activate alternative pathways such as HER2, which can also regulate gene expression. The unexpected downstream signaling proteins and altered cancer transcription can be considered as the off-targets of the treated drugs.
Quick Guide to Main Model Equations

The strategy for off-target drug repositioning is illustrated in Fig. 1. Facilitated by cancer-signaling bridges (CSB), we established a new method to facilitate drug repositioning for cancer therapy.

Major Assumptions of the Model

CSB definition

$S$ is denoted as a protein set of a signaling pathway [i.e., NCI-PID (Pathway Interaction Database) or BioCarta pathway (ref. 13)]. $C$ is denoted as a cancer protein set defined by the Online Mendelian Inheritance in Man database (14), in which each protein’s coding gene (or genes) has a close relationship with a cancer genetic disorder. II is the instance set of network motifs (15), and $\Pi^S C$ is a subset of II, where an instance comprises a set of proteins and a number of protein–protein interactions between them (Supplementary Data). Each CSB is a specific instance of 1 type of network motif; its protein set is denoted as $CSB_j (j \in \{1, 2, \ldots, |\Pi^S C|\})$. A CSB satisfies that

$$ |CSB \cap S| > 0, |CSB \cap C| > 0, \text{ and } |CSB| > |CSB \cap (S \cap C)|$$

(A)

Off-Target Repositioning Method: CSB Bayesian Factor Regression Model

Bayesian factor regression model (BFRM; 16, 17) is applied to the off-target drug repositioning. BFRM deconvolutes the cancer transcriptional response data into signatures with a model of the form

$$X_i = \overline{X} \lambda_i + \phi_i (i = 1, 2, \ldots, m)$$

(B)

where $X_i$ is an $n$ dimension vector of fold-change (treatment vs. control) of drug $i$ in the cancer transcriptional response data; $X_{ij}, j = 1, 2, \ldots, n$, is the median value of fold-changes of gene $j$ in consideration of corresponding instances treated by drug $i$; $m$ is the number of drugs; and $n$ is the number of the coding genes for the CSB proteins expanded by the cancer proteins of a specific cancer type. $\overline{X} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is a sparse $n \times k$ matrix whose columns define the signatures $S_l, l = 1, 2, \ldots, k$, and each numerical value $\overline{X}_{ij}$ defines the weight of gene $j$ in the column of gene signature $S_l$. To address which parts of the cancer signals are responsible for the unknown pharmacologic mechanisms and to what extent they are targeted, the CSB-BFRM method needs to identify signatures (the targeted parts in the cancer signals) and effects (OTEs on the targeted parts; Fig. 1B). Thus, we define a weight matrix, $A$, as a combination of one output of BFRM, $\overline{X}$, and another matrix, $P = (p_1, p_2, \ldots, p_k)$, that contains the (sparse) probabilities that each gene is associated with each signature (refer to Materials and Methods). We call the matrix, $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, an effect matrix. Each numerical value, $\lambda_{ij}$, defines the effect of drug $i$ imposed on the gene signature, $S_p$. $\Psi = (\phi_1, \phi_2, \ldots, \phi_m)$ reflects measurement error and residual biologic noise.

Repositioning profile

The OTEs of a drug on a specific cancer are defined as a repositioning profile using $A$ and $\Lambda$ (Fig. 1C). A repositioning profile, $\Omega = (\omega_1, \omega_2, \ldots, \omega_m)^T$, is an $m \times k$ matrix to characterize the overall effects of $m$ drugs on $k$ signatures. The known drug targets are essential for identification of a repositioning profile. The targetable signatures are defined by the non-zero weights at the rows of the targets across signatures of $A$. We denote the targetable signatures for drug $i$ as a set $T_i$. For each targetable signature $t \in T_i$, we define the product between $R_t$ and the effect score $\Lambda_{ij}$ as the overall effect of drug $i$ imposed on signature $t$, $\Omega_{it} = R_t \times \Lambda_{ij}$, where $R_t = \sum^{m}_{j=1} A_{ij}$ denotes the response (or total weight) of the signature $t$ to the drug $i$. The repositioning profile for drug $i$, $\omega_i, i = 1, 2, \ldots, m$, is defined as

$$\omega_i = \begin{cases} R_t \times \Lambda_{ij}, & t \in T_i \\ 0, & t \notin T_i \end{cases}$$

(C)

The target information for certain drugs may be unavailable. To define repositioning profile for these drugs, we used a randomized process to simulate the targets of these drugs (refer to Materials and Methods). To reduce the computation bias, we repeated the randomized process 1,000 times and generated a sequence for repositioning profiles, $X = (\Omega^1, \Omega^2, \ldots, \Omega^{1000})$. 

Downloaded from cancerres.aacrjournals.org on April 15, 2017. © 2012 American Association for Cancer Research.
Off-Target Drug Repositioning Method for Cancer Therapy

Repositioning score

The identified repositioning profile is applied to define a numerical value, called repositioning score, to distinguish the OTEs of the drugs. We used a supervised regression model, support vector regression (SVR), to define the repositioning score. If a drug is approved by the U.S. Food and Drug Administration or undergoing clinical trials, the element of the label vector for prior knowledge, $L_h$, is equal to 1. SVR outputs a regression prediction vector, $P_h$, for each regression between repositioning profile $E_i$, $h = 1, 2, \ldots, 1,000$, and the label vector, $L$. $P_h$ is sorted in descending order. The drugs’ ranks in the sorted $P_h$ are recorded in a repositioning score vector, $\Theta^0$. Thus, we have a sequence for repositioning score,

$$\Theta = (\Theta^1, \Theta^2, \ldots, \Theta^{1,000})$$

(D)

The repositioning score for each drug is defined as mean ± standard variation across the 1,000 repositioning score vectors.

Off-targets and OTEs

The proposed repositioning score recognizes a drug’s activity from the OTEs on the targetable signatures that comprise a number of off-targets. The off-targets are identified as the CSB proteins whose OTEs are non-zero. For a drug $i$, its OTE on a CSB protein $j$, $j = 1, 2, \ldots, n$, in a targetable signature $t$, $t \in \{1, 2, \ldots, k\}$, is defined as the product of $A_{t,j}$ and $A_{t,j}$. Thus, the OTE of drug $i$ on the targetable signature $t$ is a vector,

$$E_{i,t} = (A_{1,t} \cdot A_{t,j}, A_{2,t} \cdot A_{t,j}, \ldots, A_{m,t} \cdot A_{t,j})^T$$

(Ea)

The OTE of drug $i$ on CSB protein $j$ is defined as the summation of all of $E_{i,t,j}$ across all of the targetable signatures $T_{e}$,

$$OTE_{ij} = \sum_{t=1}^{T_{e}} E_{i,t,j}$$

(Eb)

Repositioning score

We applied CSB-BFRM to 3 cancer transcriptional response data from Connectivity Map 02 (CMAP 02; ref. 4). There are 6,100 types of cancer cell lines: MCF7 breast cancer cell line, PC3 prostate cancer cell line, and HL60 promyelocytic leukemia cell line. Each instance has a treatment case for 1 drug with 1 dosage and variable numbers of controls (1, 5, or 6). There are 3,095, 1,742, and 1,229 instances designed for MCF7, PC3, and HL60 cell lines, respectively. The transcriptional response data were derived from the Developmental Therapeutics Program (DTP) of the National Cancer Institute/National Institutes of Health (NCI/NIH).

Materials and Methods

The drug-treated transcriptional response data were derived from Connectivity Map 02 (CMAP 02; ref. 4). There are 6,100 treatment instances, in which 6,066 instances were treated on 3 types of cancer cell lines: MCF7 breast cancer cell line, PC3 prostate cancer cell line, and HL60 promyelocytic leukemia cell line. Each instance has a treatment case for 1 drug with 1 dosage and variable numbers of controls (1, 5, or 6). There are 3,095, 1,742, and 1,229 instances designed for MCF7, PC3, and HL60 cell lines, respectively. The transcriptional response data...
of MCF7 include 3,628 gene microarrays for 1,198 single-dose drugs, 96 multiple dose drugs, and 1,390 drug–dose pairs. The transcriptional response data of PC3 have 2,017 gene microarrays for 1,150 single-dose drugs, 31 multiple dose drugs, and 1,215 drug–dose pairs. The transcriptional response data of HL60 comprise 1,406 gene microarrays for 1,061 single-dose drugs, 17 multiple dose drugs, and 1,099 drug–dose pairs. Additional data used in this article can be found in the Supplementary Data.

Figure 1 illustrates the strategies used in our off-target repositioning method, CSB-BFRM, and the Quick Guide provides an overview of the key definitions and modeling components.

Figure 1A shows the advantage of combining CSB and BFRM (16–19) to reposition drugs that cater not only to the treatment–response but also to the expanded cancer signaling mechanisms, making it feasible for off-target repositioning for cancers. In Fig. 1B, the input to CSB-BFRM is a treatment–response matrix $X$ ($n \times m$) whose $m$ columns correspond to the treated drugs and $n$ rows correspond to the coding genes for the identified CSB proteins for the cancer type of interest. The statistical factor analysis, BFRM, decomposes the treatment–response matrix $X$ into another 2 matrices, weight matrix $A$ ($n \times k$) and effect matrix $L$ ($k \times m$). A weight matrix, $A$ ($n \times k$), is a sparse matrix (most of the elements are zero, as indicated by white color) whose columns define $k$ signatures and their non-zero elements indicate which proteins are included in the signatures. BFRM imposes a sparse prior on the association of the genes to the signatures. Another matrix, $P = (\rho_1, \rho_2, \ldots, \rho_k)$, contains the (sparse) probabilities that each gene is associated with each factor. The cutoff point for each element, $P_{i,j}$ of $P$ matrix was chosen as the mean of all the non-zero values in the $P$ matrix. If $P_{i,j}$ is higher than the cutoff point, the corresponding value, $A_{i,j}$ of weight matrix $A$ will be kept, and else, $A_{i,j}$ is set as zero. An effect matrix, $L$ ($k \times m$), shows the effects of the $m$ drugs imposed on the $k$ signatures. BFRM model applies hierarchical priors for values of the non-zero elements in $A$ and gets posterior via
Markov chain Monte Carlo (MCMC). MCMC analysis for the posterior simulation is implemented in a Gibbs sampling manner. The BFRM model is implemented with a software package, BFRM 2.0 (16, 17). The number of signatures, \( k \), is determined by an evolution algorithm in the BFRM 2.0 software.

In Fig. 1C, the repositioning profile definition takes advantage of the identification of the targetable signatures. If a drug’s target information is available, the targetable signatures are defined by the non-zero weights at the rows of the targets across signatures of \( A \). The proteins in each targetable signature are determined by the non-zero elements in each corresponding column of \( A \). For each targetable signature, the total of the non-zero weights is used to evaluate the response of the signature to the drug. In \( A \), the score corresponding to the row of the signature and the column of the drug shows the effect of the drug on the signature. The OTE that the drug imposes on the signature is defined as a weighted score obtained by multiplying the response of the signature to the drug by the effect of the drug on the signature. The repositioning profile is used to illustrate the OTEs of the drug on all of the signatures, in which the OTEs for the targetable signatures are defined as the weighted scores, whereas those for the untargetable signatures are zeros.

Sometimes, the target information of a drug may be unavailable. We designed a randomized process to find these targetable signatures. In the randomized process, a number of proteins randomly chosen from the CSB proteins were considered as the candidates for drug targets. The hypothesis is that the drug generates OTEs on the CSB proteins even if it does not target the CSB proteins directly. The number of proteins chosen is determined by a random numerical value drawn from a uniform distribution between 1 and \( \mu \), where \( \mu \) is the mean value of the targets for the drugs whose targets are known. The randomized process is repeated 1,000 times for those drugs with unknown targets to reduce the computational bias in the identification of their candidate targets or off-targets. Still, some drugs have known drug targets that are not included in the CSB protein set. These targets are led to the CSB proteins, using the shortest paths in the protein–protein interaction network. The CSB proteins identified are considered as the targets or off-targets of these drugs.

To rank the activities of drugs, we propose a single numerical value, repositioning score, for each drug. In this study, because a number of drugs are known to be FDA-approved or undergoing clinical trials for breast cancer, prostate cancer, and promyelocytic leukemia, we used the supervised regression model to define the repositioning score (Fig. 1D). For other cancer types, the FDA approval and clinical trial information may be unavailable. To apply the CSB-BFRM method to these cancer types, the supervised method should be replaced by an unsupervised data mining method, such as clustering. The SVR algorithm is implemented in R, using the package “e1071.” All of the parameters are used as default except that the parameter “\( C \)” for cross-validation is set to be 5. We specified cross-validation as 5-fold.

All of the materials and methods reported in this article are included in a Web-based tool, R2D2-CSB, which is available at http://r2d2drug.org/Software/csbc/csba.aspx.

Results

CSBs expand the signaling proteins to cancer proteins

To investigate the off-target drug repositioning for cancers, we introduced the new network elements, CSBs, that can be used to extend the known canonical signaling pathways (13, 20) to the proteins whose coding genes have a close relationship with cancer genetic disorders (14, 21), or for short, cancer proteins (Fig. 2A). The data sources for definition of CSBs are listed in Supplementary Table S1. CSBs are the instances of network motifs (15, 22, 23), or building blocks, of the protein interaction networks (refs. 24–28; Supplementary Table S2).

Besides being able to link many previously unrelated cancer proteins to a known signaling pathway of interest (Supplementary Fig. S1), CSBs have the following 4 characteristics that determine their important role in off-target drug repositioning: (i) CSBs are significantly enriched in the connections between oncogenic signaling pathways and cancer proteins (Supplementary Table S3 and Data); (ii) most CSBs, nearly 70%, are not shared by multiple types of cancers but are specific to 1 cancer type (Fig. 2B); (iii) signaling proteins and cancer proteins linked by CSBs are significantly more likely to be targeted by known anticancer drugs (Fig. 2D); and (iv) although most known anticancer drugs select the proteins in signaling pathways as their targets (Fig. 2C), they still generate relatively high effects, transmitted by CSBs, onto cancer proteins (Fig. 2E; Supplementary Data).

Application of CSB-BFRM to cancer transcriptional response data

We applied the proposed off-target repositioning method, CSB-BFRM, to 3 cancer transcriptional response data sets of MCF7 breast cancer cell line, PC3 prostate cancer cell line, and HL60 promyelocytic leukemia cell line (Supplementary Data). The inputs and outputs of CSB-BFRM are shown in Supplementary Tables S4 to S12. We tested the performance of CSB-BFRM to predict the activities of FDA-approved drugs and drugs undergoing clinical trials for breast cancer, prostate cancer, and promyelocytic leukemia and used the identified off-targets and OTEs to explain the mechanisms of action of repositioned drugs.

Performance of repositioning prediction

To evaluate the performance of CSB-BFRM in prediction of drug activities based on the identified repositioning profiles of drugs, we used the receiver operating characteristic (ROC) method. The area under the ROC curve (AUC) illustrates how useful the repositioning profiles are for prediction of the known data of FDA approval and clinical trials information. In Fig. 3A and B, we show the ROC curves for the predictions of the activities of FDA-approved and clinical trial breast cancer drugs. The AUCs for the ROCs in Fig. 3A are 0.94 ± 0.02 (\( P < 10^{-4} \); Fisher exact 2-tailed test) and those for the ROCs in Fig. 3B are 0.79 ± 0.04 \( P < 10^{-4} \);
Fisher exact 2-tailed test). Because the FDA approval information for prostate cancer and promyelocytic leukemia is limited (Supplementary Table S13), it was merged with clinical trial information to do the repositioning predictions. The performance of the prediction on the FDA-approved and clinical trial prostate cancer drugs is indicated by the ROC curve shown in Fig. 3C. The AUC of the ROC curve in Fig. 3C is 0.78 ± 0.03 (P < 10^{-5}; Fisher exact 2-tailed test). The ROC curve for the prediction on the FDA-approved and clinical trial promyelocytic leukemia drugs is shown in Fig. 3D and its AUC is 0.91 ± 0.06 (P < 10^{-5}; Fisher exact 2-tailed test). The results indicate that the activities of the FDA-approved and clinical trial drugs for breast cancer, prostate cancer, and promyelocytic leukemia are accurately predicted by the CSB-BFRM.

For the repositioning on the MCF7 breast cancer cell line, we listed the first 22 drugs with the highest repositioning scores in Table 1 and showed the ranks for all the 1,390 drugs in Supplementary Table S14. These first 22 drugs predict all 14 FDA-approved drugs (with drug dosages) from the 1,390 drugs (P < 10^{-10}, hypergeometric test). Furthermore, we have listed the repositioned drugs with their repositioning scores for PC3 prostate cancer cell line and HL60 promyelocytic leukemia in Supplementary Tables S15 and S16. The relatively small numbers of drugs with the highest repositioning scores predict the FDA-approved drugs and clinical trial drugs for prostate cancer and promyelocytic leukemia (PC3: P < 10^{-4}, hypergeometric test; HL60: P < 10^{-2}, hypergeometric test).

### OTEs and off-targets

We investigated 8 pairs of drug doses with relatively high repositioning scores repositioned for MCF7 breast cancer cell line, as shown in Table 2. The 8 drug–dose pairs are raloxifene at 0.1 and 7.8 μmol/L, tamoxifen at 1 and 7 μmol/L, paclitaxel at 4.6 and 1 μmol/L, and fulvestrant at 1 and 0.01 μmol/L. We identified the off-targets for the 8 drug-dose pairs and listed them with their OTEs in Supplementary Table S17. To remove the redundant off-targets with relatively lower OTEs, we used the mean of the absolute values, [OTE], of all off-targets as a threshold, δ. We chose those off-targets whose OTEs are higher than the threshold δ or lower than −δ for the following analysis. We conducted the gene set enrichment analysis (GSEA; ref. 29) on the off-targets of each drug. The off-targets of each drug are significantly enriched in 2 important cellular functions, cell cycle (P < 10^{-5}, hypergeometric test) and apoptosis of cells (P < 10^{-20}, hypergeometric test). The enrichment P values for all of the 8 drug-dose pairs are shown in Supplementary Table S18. We also conducted pathway analysis on the identified off-targets, using Ingenuity Pathway Analysis (IPA; Ingenuity Systems, Inc.) software. Subsequently, 2 important signaling pathways related to cell cycle and apoptosis were identified, namely, cell-cycle G1–S checkpoint and p53 signaling pathways.
Mechanisms of repositioned drugs

The Rb-dependent repression of E2F-mediated transcription (30) is the key to understanding the mechanisms of the 8 repositioned drug-dose pairs. We summarized the signal cascade as follows:

\[
\text{INK4} \rightarrow \text{CDK4/6} \downarrow \\
\text{SCF} \rightarrow \text{Cyclin D/E} \downarrow \\
\rightarrow \text{Rb}^\uparrow \rightarrow \text{Rb repression complex} \rightarrow \begin{cases} \text{E2F} \\ \text{DP-1} \end{cases} \Rightarrow \text{Cell cycle}
\]

The OTEs and off-targets of the 8 repositioning drug-dose pairs for this signaling cascade are shown in Table 2. To better illustrate the drugs' effects on the signaling cascade, their OTEs and off-targets are also displayed in Fig. 4. All 8 drug-dose pairs inhibit the core part of the signaling cascade, the heterodimer of E2F and DP-1. The inhibition of either E2F or DP-1 ensures that the gene expression is repressed even if Rb is phosphorylated. Still, some drug-dose pairs targeting other parts of the signaling cascade enforce the transcriptional repression. Paclitaxel at 4.6 \( \mu \text{mol/L} \) has a relatively high positive OTE (higher than 0) on the RBL1 protein (a member of the Rb protein family), which increases the expression of RBL1 protein and strengthens the recruitment of histone deacetylases and other nuclear factors to repress gene expression. Paclitaxel at 4.6 \( \mu \text{mol/L} \) also has positive OTEs on the INK4 (p15) and SCF (BTRC) proteins, which enhances the inhibition of
CDK4/6 and cyclin D/E as well as phosphorylation of Rb, so that the association of Rb family members with both histone deacetylases and E2Fs is enhanced and gene expression is repressed. Fulvestrant at 1 and 0.01 μmol/L have negative OTEs (lower than 0) on the kinase CDK2 and decrease its expression, which in turn, reduce the phosphorylated Rb and enhance the Rb-dependent repression of E2F-mediated transcription. Thus, by various means, these drugs enforce the transcriptional repression of key cell-cycle genes.

Consistency with dose–response curves
We used the dose–response data derived from the DTP of NCI/NIH (31) to validate our hypothesis. Checking the dose–response curves for raloxifene, tamoxifen, paclitaxel, and fulvestrant (Fig. 5), we found that all of the 4 drugs with considered dosages (lower than 10 μmol/L) have a significant inhibition on cell growth. This result is consistent with the predicted OTEs enhancing the Rb-dependent repression of E2F-mediated transcription of the key genes for cell-cycle progression.

The 4 repositioned drugs not only generate OTEs on the cell-cycle G1–S checkpoint signaling pathway but also impose OTEs on the p53 signaling pathway (Table 2 and Fig. 4). The OTEs on the p53 signaling pathway are helpful to understand why raloxifene, tamoxifen, and paclitaxel induce apoptosis at higher dosages, whereas fulvestrant does not induce any cell death on MCF7 (Fig. 5). In a comparison of the OTEs of raloxifene at lower and higher dosages, these 2 OTEs appear to be opposite to each other. At the lower dosage, the negative OTE decreases TP53 and blocks apoptosis, whereas at the higher dosage it increases apoptotic cell death. This is also seen with tamoxifen. On the other hand, paclitaxel is predicted to increase the expression of TP53 and induce apoptosis of cells at both lower and higher dosages. Several experiments on paclitaxel with dosages between 10^{-7} and 10^{-5} mol/L induce cell death. In contrast, fulvestrant decreases the expression of TP53 at both lower and higher dosages and cannot induce the apoptosis of cells at any of the considered dosages.

Discussion
In summary, we presented a new computational method for off-target drug repositioning using cancer transcriptional response data before and after treatment. Facilitated by the new network elements, CSBs, we have shown the potential of the proposed method, CSB-BFRM, in the repositioning of drugs for specific cancer types. CSB-BFRM performs well in predicting the activities of FDA-approved drugs and clinical trial drugs for breast cancer, prostate cancer, and

### Table 1. The activities of drugs predicted by repositioning scores for MCF7 breast cancer cell line

<table>
<thead>
<tr>
<th>Predicted rank</th>
<th>Repositioned drugs (conc.(^{a}))</th>
<th>Repositioning score (mean ± SD)</th>
<th>Status(^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Raloxifene (0.0000001)</td>
<td>0.85 ± 0.77</td>
<td>FDA CT</td>
</tr>
<tr>
<td>2</td>
<td>Paclitaxel (0.0000046)</td>
<td>2.75 ± 1.01</td>
<td>FDA CT</td>
</tr>
<tr>
<td>3</td>
<td>Tamoxifen (0.000001)</td>
<td>3.43 ± 1.97</td>
<td>FDA CT</td>
</tr>
<tr>
<td>4</td>
<td>Paclitaxel (0.0000001)</td>
<td>5.46 ± 33.73</td>
<td>FDA CT</td>
</tr>
<tr>
<td>5</td>
<td>Fulvestrant (0.000001)</td>
<td>6.31 ± 0.98</td>
<td>FDA CT</td>
</tr>
<tr>
<td>6</td>
<td>Exemestane (0.00000001)</td>
<td>21.82 ± 86.28</td>
<td>FDA CT</td>
</tr>
<tr>
<td>7</td>
<td>Letrozole (0.000014)</td>
<td>26.43 ± 96.52</td>
<td>FDA CT</td>
</tr>
<tr>
<td>8</td>
<td>Sulindac (0.000003)</td>
<td>29.02 ± 18.10</td>
<td>CT</td>
</tr>
<tr>
<td>9</td>
<td>Fulvestrant (0.00000001)</td>
<td>31.01 ± 42.76</td>
<td>FDA CT</td>
</tr>
<tr>
<td>10</td>
<td>Daunorubicin (0.00000007)</td>
<td>37.95 ± 83.48</td>
<td>FDA CT</td>
</tr>
<tr>
<td>11</td>
<td>Clomifene (0.00000066)</td>
<td>38.97 ± 20.41</td>
<td>FDA CT</td>
</tr>
<tr>
<td>12</td>
<td>Sulindac (0.0001)</td>
<td>46.48 ± 19.31</td>
<td>CT</td>
</tr>
<tr>
<td>13</td>
<td>Estradiol (0.00000001)</td>
<td>53.47 ± 79.65</td>
<td>FDA CT</td>
</tr>
<tr>
<td>14</td>
<td>Imatinib (0.000001)</td>
<td>55.64 ± 68.28</td>
<td>CT</td>
</tr>
<tr>
<td>15</td>
<td>Estradiol (0.00000001)</td>
<td>58.60 ± 30.09</td>
<td>FDA CT</td>
</tr>
<tr>
<td>16</td>
<td>Methotrexate (0.0000088)</td>
<td>59.80 ± 127.82</td>
<td>FDA CT</td>
</tr>
<tr>
<td>17</td>
<td>Bezafibrate (0.000011)</td>
<td>68.08 ± 139.31</td>
<td>FDA CT</td>
</tr>
<tr>
<td>18</td>
<td>Doxorubicin (0.00000068)</td>
<td>70.22 ± 218.72</td>
<td>FDA CT</td>
</tr>
<tr>
<td>19</td>
<td>Valproic acid (0.000005)</td>
<td>76.00 ± 69.52</td>
<td>CT</td>
</tr>
<tr>
<td>20</td>
<td>Raloxifene (0.00000078)</td>
<td>80.74 ± 84.14</td>
<td>FDA CT</td>
</tr>
<tr>
<td>21</td>
<td>Amlodipine (0.0000132)</td>
<td>113.04 ± 193.55</td>
<td>FDA CT</td>
</tr>
<tr>
<td>22</td>
<td>Tamoxifen (0.0000007)</td>
<td>124.33 ± 137.36</td>
<td>FDA CT</td>
</tr>
</tbody>
</table>

\(^{a}\)Values in parentheses indicate concentrations (conc.) in mol/L.

\(^{b}\)CT, clinical trial breast cancer drug; FDA, FDA-approved breast cancer drug.
Table 2. The predicted off-targets and OTEs for the repositioned drugs on MCF7 breast cancer cell line

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Off-targets</th>
<th>OTEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raloxifene</td>
<td>Cell-cycle G₁-S</td>
<td></td>
</tr>
<tr>
<td>(0.1 μmol/L)</td>
<td>TFDP1 Transcription factor Dp-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E2F1 E2F transcription factor 1</td>
<td>-0.0433</td>
</tr>
<tr>
<td></td>
<td>E2F2 E2F transcription factor 2</td>
<td>-0.0268</td>
</tr>
<tr>
<td></td>
<td>E2F4 E2F transcription factor 4,</td>
<td>-0.0206</td>
</tr>
<tr>
<td></td>
<td>p107/p130-binding</td>
<td>-0.0392</td>
</tr>
<tr>
<td></td>
<td>p53 signaling pathway</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TP53 Tumor protein p53</td>
<td>-0.0216</td>
</tr>
<tr>
<td></td>
<td>MDM2 Mdm2 p53-binding protein</td>
<td></td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>Cell-cycle G₁-S</td>
<td></td>
</tr>
<tr>
<td>(1 μmol/L)</td>
<td>TFDP1 Transcription factor Dp-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TP53 Tumor protein p53</td>
<td>-0.0138</td>
</tr>
<tr>
<td></td>
<td>MDM2 Mdm2 p53-binding protein</td>
<td>-0.0157</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>Cell-cycle G₁-S</td>
<td></td>
</tr>
<tr>
<td>(4.6 μmol/L)</td>
<td>TFDP1 Transcription factor Dp-1</td>
<td>-0.0216</td>
</tr>
<tr>
<td></td>
<td>RBL1 Retinoblastoma-like 1 (p107)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDK2 Cyclin-dependent kinase inhibitor 2B p15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDK2 Cyclin-dependent kinase inhibitor 2B p15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDM2 Mdm2 p53-binding protein</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p53 signaling pathway</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TP53 Tumor protein p53</td>
<td>-0.0138</td>
</tr>
<tr>
<td></td>
<td>MDM2 Mdm2 p53-binding protein</td>
<td>-0.0157</td>
</tr>
<tr>
<td>Fulvestrant</td>
<td>Cell-cycle G₁-S</td>
<td></td>
</tr>
<tr>
<td>(1 μmol/L)</td>
<td>TFDP1 Transcription factor Dp-1</td>
<td>-0.0192</td>
</tr>
<tr>
<td></td>
<td>E2F1 E2F transcription factor 1</td>
<td>-0.0134</td>
</tr>
<tr>
<td></td>
<td>E2F2 E2F transcription factor 2</td>
<td>-0.0184</td>
</tr>
<tr>
<td></td>
<td>E2F4 E2F transcription factor 4,</td>
<td>-0.0142</td>
</tr>
<tr>
<td></td>
<td>p107/p130-binding</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p53 signaling pathway</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TP53 Tumor protein p53</td>
<td>-0.0192</td>
</tr>
<tr>
<td></td>
<td>MDM2 Mdm2 p53-binding protein</td>
<td>-0.0115</td>
</tr>
</tbody>
</table>

(Continued on following page)
promyelocytic leukemia, using the corresponding transcription response datasets. The predicted OTEs and off-targets help to better understand the mechanisms of action of repositioned drugs.

In Table 1, the repositioning list for MCF7 breast cancer cell line includes all of the FDA-approved breast cancer drugs targeting the ER, which appear in the 1,390-drug set. The drugs are raloxifene at 0.1 and 7.8 μmol/L, tamoxifen at 1 and 7 μmol/L, fulvestrant at 1 and 0.01 μmol/L, and estradiol at 0.01 and 0.1 μmol/L. The repositioning result is consistent with the fact that MCF7 is an ER+ breast cancer cell line. In addition, raloxifene and tamoxifen are the selective ER modulators (SERM; ref. 32). These SERMs function as pure antagonists when acting through ER-β on genes containing estrogen response elements but can function as partial agonists when acting on them through ER-α. The repositioning results for raloxifene and tamoxifen are consistent with the "partial agonist" property of raloxifene and tamoxifen. These 2 drugs generate higher effects at lower dosages. Raloxifene at 0.1 μmol/L has a higher repositioning rank than raloxifene at 7.8 μmol/L, and tamoxifen at 1 μmol/L has a higher repositioning rank than tamoxifen at 7 μmol/L.

The identified off-targets and OTEs display the complexity of the drugs' activities. On one hand, some drugs at higher dosages have their own specific off-targets or OTEs. In the repositioning for the MCF7 breast cancer cell line, paclitaxel at 4.6 μmol/L has extra positive OTEs on the RBL1 protein, Rb's kinase (CDK2), the cyclin proteins' inhibitor (SCF), and the inhibitor of kinase CDK4/6, INK4 (p15), which are absent at the lower dosage. These OTEs ensure that paclitaxel can strengthen the transcription repression on the key genes regulating the cell cycle. On the other hand, at different dosages, the same drug would generate different effects on its specific off-targets and signaling pathways. For example, at the higher dosages, raloxifene and tamoxifen have positive OTEs on p53 protein while exhibiting negative OTEs on p53 protein at the lower dosages. Because the complexity of the drugs' activities is not easily explained by "on-target" studies, the OTEs on the downstream signaling proteins have to be identified and linked to transcription, rather than simple analysis of the effects on known drug targets.

BFRM plays a central role in recognizing the OTEs of repositioned drugs. It factorizes the response (fold change of expression) of a molecule into different component values according to the latent factors (signatures). The CSB-BFRM recognizes the essential latent factors (targetable signatures) and factorized component values (OTEs) for these signatures. For the repositioning on the MCF7 breast cancer cell line, we compared the original response (fold change) on off-targets in cell-cycle G1–S checkpoint and p53 signaling pathways with the recognized OTEs on these targets (Supplementary Fig. S2). The data scale is changed. Fold changes of the molecules are between 0.4 and 1.6, whereas OTEs are between −0.10 and 0.10. The factorized OTEs allow easy recognition of positive and negative effects. For instance, all of the original fold changes of tamoxifen at 1 μmol/L are higher than 1, whereas the OTEs are between −0.05 and 0.05.
If we use the original fold changes, we cannot tell the difference between OTEs on the heterodimer of E2F and DP-1 (negative) and those for p53 (positive). The recognized OTEs are better in reflecting the mechanism of action of repositioned drugs.

The proposed off-target drug repositioning method, CSB-BFRM, takes advantage of the availability of disease-specific prior knowledge. For example, the definition of CSBs as shown in Fig. 1 requires prior knowledge of the cancer genes that have genetic disorders associated with the cancer type of interest. However, CSB-BFRM would face difficulties in repositioning drugs for rare cancer types, as prior knowledge is often unavailable. Fortunately, with the rapid development of next-generation sequencing, more people will be willing to study these rare cancer types and to generate corresponding genetic mutation data. The identification of key genes with genetic disorders would allow further identification of CSBs for these cancer types. Thus, we believe that using CSB-BFRM for repositioning drugs for rare cancer types will also be feasible.

Figure 4. OTEs and off-targets of raloxifene, tamoxifen, paclitaxel, and fulvestrant on the cell-cycle G1–S checkpoint and p53 signaling pathways. Right, the signal cascade in the cell-cycle G1–S checkpoint signaling pathway. Left, the p53 signaling pathway. The drug–dose pairs are listed in the middle. The drug-targeted pathway was generated using IPA software.

Figure 5. Dose–response curves for raloxifene, tamoxifen, fulvestrant, and paclitaxel. The value of dots between 0% and 100% means the drug inhibits the cell growth. The growth percentage of −100 means all cells are killed. The dose–response curves for raloxifene and tamoxifen imply that they induce the cell death at higher dosages, whereas the curve for paclitaxel shows that some experiments also cause the cell death. In contrast, the curve for fulvestrant indicates that at higher dosages, fulvestrant does not induce the cell death. The data source is DTP of NCI/NIH (31).
Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

G. Jin and S.T.C. Wong initiated the idea for applying systems biology in cancer drug repurposing as well as the CSB-BFRM model. G. Jin and C. Fu processed the data used in the analysis and developed the source code for CSB-BFRM model. G. Jin did the repositioning on MCF7, PC3, and HLEo. C. Fu designed the Web-based interface for CSB-BFRM model. G. Jin, H. Zhao, J. Chang, and C. K. Cui discussed the biologic interpretation of the repositioning results. S.T.C. Wong supervised this work. G. Jin, H. Zhao, J. Chang, and S.T.C. Wong wrote the manuscript.

Acknowledgments

The authors appreciate the discussion and advice of Suzanne Faqua, Michael Lewis, and Rachel Schiff from Lester and Sue Smith Breast Center, Baylor College of Medicine; Xiang-Sun Zhang from Institute of Applied Mathematics, Chinese Academy of Sciences; Luonan Chen from Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; as well as colleagues in the Cancer Systems Biology Laboratory, The Methodist Hospital Research Institute.

Grant Support

This work was supported by NIH grant U54CA149196 and John S. Dunn Research Foundation to S.T.C. Wong.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 13, 2011; revised October 17, 2011; accepted November 9, 2011; published OnlineFirst November 22, 2011.

References

A Novel Method of Transcriptional Response Analysis to Facilitate Drug Repositioning for Cancer Therapy

Guangxu Jin, Changhe Fu, Hong Zhao, et al.

Cancer Res 2012;72:33-44. Published OnlineFirst November 22, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-2333

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2011/11/22/0008-5472.CAN-11-2333.DC1

Cited articles
This article cites 32 articles, 19 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/1/33.full.html#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
/http://cancerres.aacrjournals.org/content/72/1/33.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.