



# DeCoST: A New Approach in Drug Repurposing From Control System Theory

Thanh M. Nguyen<sup>1\*</sup>, Syed A. Muhammad<sup>2\*</sup>, Sara Ibrahim<sup>3</sup>, Lin Ma<sup>4</sup>, Jinlei Guo<sup>4</sup>, Baogang Bai<sup>4</sup> and Bixin Zeng<sup>5\*</sup>

<sup>1</sup> Department of Computer and Information Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States, <sup>2</sup> Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan,

<sup>3</sup> Department of Biology, School of Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States,

<sup>4</sup> The 1st School of Medicine and School of Information and Engineering, Wenzhou Medical University, Zhejiang, China,

<sup>5</sup> Institute of Lasers and Biomedical Photonics, Wenzhou Medical University, Wenzhou, China

## OPEN ACCESS

### Edited by:

Lixia Yao,  
Mayo Clinic, United States

### Reviewed by:

Yanshan Wang,  
Mayo Clinic, United States  
Fuhai Li,  
The Ohio State University,  
United States  
Zhichao Liu,  
National Center for Toxicological  
Research (FDA), United States

### \*Correspondence:

Thanh M. Nguyen  
thamnguy@iupui.edu  
Syed A. Muhammad  
aunmuhammad78@yahoo.com  
Bixin Zeng  
z\_bixin@163.com

### Specialty section:

This article was submitted to  
Translational Pharmacology,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 15 March 2018

**Accepted:** 15 May 2018

**Published:** 05 June 2018

### Citation:

Nguyen TM, Muhammad SA,  
Ibrahim S, Ma L, Guo J, Bai B and  
Zeng B (2018) DeCoST: A New  
Approach in Drug Repurposing From  
Control System Theory.  
Front. Pharmacol. 9:583.  
doi: 10.3389/fphar.2018.00583

In this paper, we propose DeCoST (Drug Repurposing from Control System Theory) framework to apply control system paradigm for drug repurposing purpose. Drug repurposing has become one of the most active areas in pharmacology since the last decade. Compared to traditional drug development, drug repurposing may provide more systematic and significantly less expensive approaches in discovering new treatments for complex diseases. Although drug repurposing techniques rapidly evolve from “one: disease-gene-drug” to “multi: gene, dru” and from “lazy guilt-by-association” to “systematic model-based pattern matching,” mathematical system and control paradigm has not been widely applied to model the system biology connectivity among drugs, genes, and diseases. In this paradigm, our DeCoST framework, which is among the earliest approaches in drug repurposing with control theory paradigm, applies biological and pharmaceutical knowledge to quantify rich connective data sources among drugs, genes, and diseases to construct disease-specific mathematical model. We use linear–quadratic regulator control technique to assess the therapeutic effect of a drug in disease-specific treatment. DeCoST framework could classify between FDA-approved drugs and rejected/withdrawn drug, which is the foundation to apply DeCoST in recommending potentially new treatment. Applying DeCoST in Breast Cancer and Bladder Cancer, we reprofiled 8 promising candidate drugs for Breast Cancer ER+ (Eribitux, Flutamide, etc.), 2 drugs for Breast Cancer ER- (Daunorubicin and Donepezil) and 10 drugs for Bladder Cancer repurposing (Zafirlukast, Tenofovir, etc.).

**Keywords:** drug repurposing, system control, breast cancer, bladder cancer, pathway, expression profile

## INTRODUCTION

Drug repurposing (also called drug repositioning) has become one of the most active areas in pharmacology since last decade (Oprea et al., 2011) because this approach could significantly reduce the cost and time to invent a new treatment. Before drug repurposing research became active, it was expected to take about 15 years and \$0.8–\$1 billion to bring a new drug into the market (Dimasi, 2001) due to many tests and clinical trials in order to be commercially approved by Food and Drug Administration (FDA) (USFDA, 2016). It is expected that the failure probability during

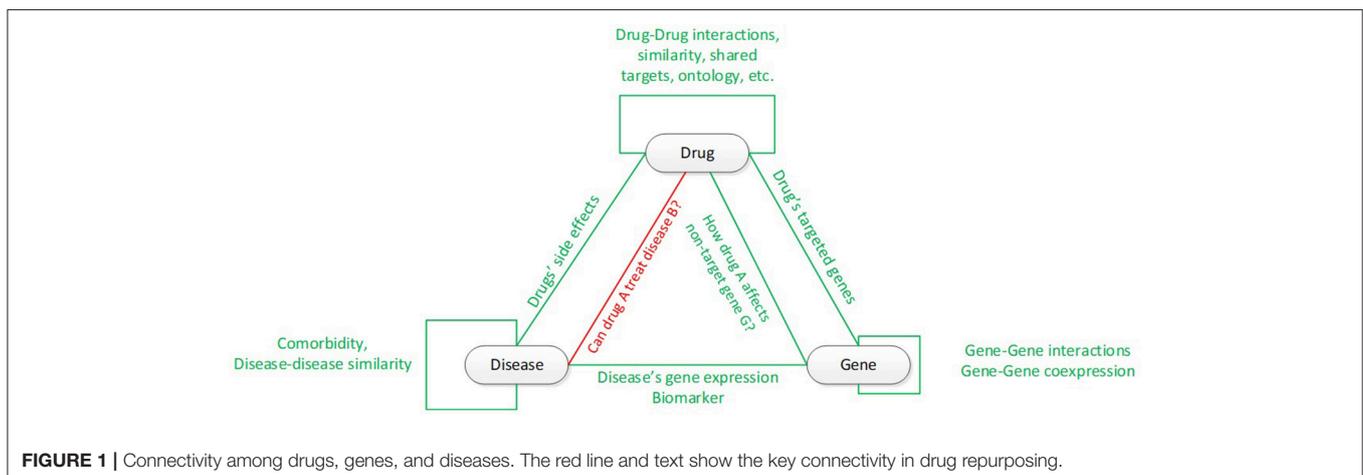
clinical trials is about 91.4% (Thomas et al., 2016). One of the key reasons for low productivity in traditional drug development is the lack of systematic evaluation of additional indications (Dudley et al., 2011), which may lead to unexpected side effects and low efficacy. Briefly, drug repurposing finds new indications for known drugs and compounds (Gupta et al., 2013) to reduce the risk of failure and shorten time of discovery. Drug repurposing applies modern computational techniques to digitalize genomic (Power et al., 2014), bioinformatics, chemical informatics (Bisson, 2012) and patients' individual health records (Xu et al., 2014) to offer more systematic evaluation of the chemical compound before entering the laboratory testing and clinical trial steps. In addition, drug repurposing could explore the large set of chemical compounds, which is estimated to be more than 90 million by PubChem statistics (Wang et al., 2014), to reduce the cost of synthesizing new compounds. Prominent successful examples for drug repurposing include Viagra, Avastin, and Rituxan (Dudley et al., 2011).

System biology (Pujol et al., 2010) plays an important role to in the evolvement of drug repurposing evolved from “one: disease-gene-drug” (Durrant et al., 2010) to “multi: gene, drug” (Chou, 2010; Medina-Franco et al., 2013) and from “lazy guilt-by-association” (Campillos et al., 2008; Keiser et al., 2009; Iorio et al., 2010; Gottlieb et al., 2011) to “systematic model-based pattern matching,” such as the Broad Institute’s Connectivity Maps (CMAP), C2MAP, etc. (Lamb et al., 2006; Hu and Agarwal, 2009; Huang et al., 2012; Jensen et al., 2012; Li and Lu, 2013; Subramanian et al., 2017). System biology reveals connectivity among drug, gene, and diseases (Figure 1). In this Figure, the green connectivity shows the types of connectivity for which drug repurposing could utilize to answer the key question: could drug A be re-indicated to treat disease B. The literature and public data sources for these types of connectivity have been thoroughly developed in the recent two decades, such as DrugBank (Law et al., 2013) and SFINX (Andersson et al., 2015) for drug-drug interaction; DrugBank (Law et al., 2013) and STITCH (Kuhn et al., 2012) for drug-gene/protein interaction; BioGRID (Chatr-Aryamontri et al., 2013), STRING (Szklarczyk et al., 2015), HAPPI (Chen et al., 2017), KEGG (Kanehisa et al.,

2017) and Reactome (Croft et al., 2011) for protein-protein interaction and human pathway; OMIM (Baxevanis, 2012) and GEO (Barrett et al., 2013) for disease-specific gene curation and analysis; the human disease network (Goh et al., 2007) for disease-disease connectivity; and SIDER for diseases’ drug-side-effect (Kuhn et al., 2016). The integration of rich data sources enable mathematical system modeling and analysis in system biology to deepen our understanding and predictive capability for biological processes, disease ontology (Hannon and Ruth, 2014; Goel and Richter-Dyn, 2016; Woodhead et al., 2016) and personalized medicine (Weston and Hood, 2004).

From the mathematical system-model-control-based point of view, there exist a mechanism regulating the gene expression profile. In the healthy condition, the gene expression stays in the stable equilibrium region such that  $\mathbf{x}(t) = f(\mathbf{x}(t-1)) \approx \mathbf{x}(t-1)$ , where  $f$  indicates the expression-regulating mechanism computed from data integration,  $\mathbf{x}$  stands for expression and  $t$  stands for time. In the disease state, the critical gene expression strays outside the stable region. In this case, without a control (treatment), the expression will be unbounded. The system control algorithms aim to find the sequence of control-treatment that optimally stabilize the expression back to the original equilibrium point, such as linear control (Willems, 1971; Chen et al., 2016), nonlinear control (Bardi and Capuzzo-Dolcetta, 2008; Falcone and Ferretti, 2013), adaptive neural network (Rovithakis and Christodoulou, 1994; Tong et al., 2014). By comparing the real drug treatments with the optimal control-treatment (also called hypo-treatment), we can evaluate the potential efficacy of the drug before being repurposed.

However, applying mathematical system modeling and control in drug repurposing is still in very early steps. There are three key challenges in applying system control approach. First, it is difficult to quantify the gene expression and real drug treatment, as there is very little literature discussing the “normal range” of each gene’s expression. Second, constructing a comprehensive and accurate mathematical model to simulate the gene expression change is complicated due to the diversity of gene-gene interaction mechanisms, mutation, and under-discovered data. Third, the biological systems are known for



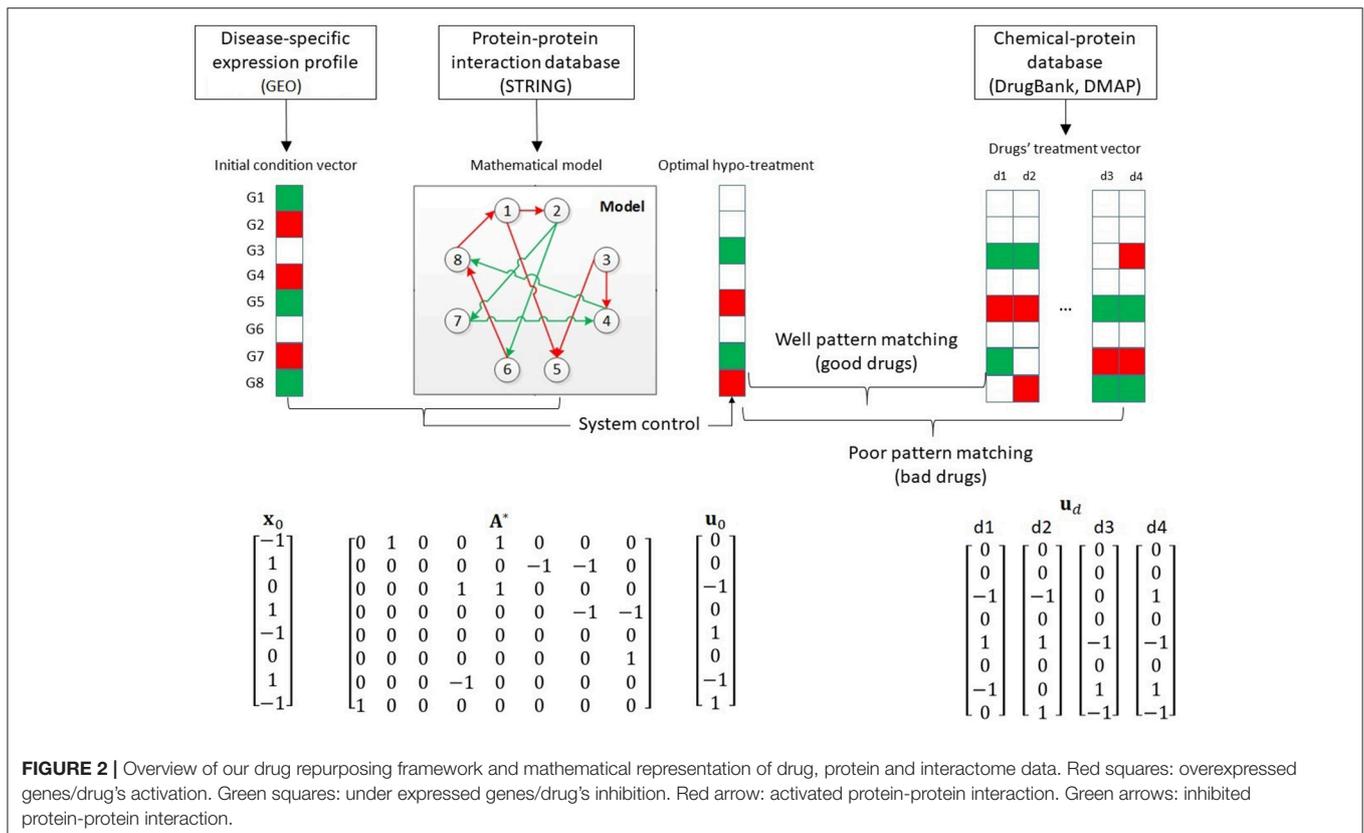
large scale for system control: there may be from hundreds to thousands of genes of interest in a specific disease or biological process.

In this paper, we propose DeCoST (Drug Repurposing from Control System Theory) to apply control system paradigm for drug repurposing purpose, with source code available at <https://github.com/thamnguy/DeCoST>. The DeCoST framework tackles these challenges above as follow. First, although we could not completely solve the “normal range” challenge, we discretized the gene expression and the connectivity data so that the control-system algorithm could be executed logically without the “normal range” impact. Second, to overcome the comprehensiveness challenge, we utilized the biological and pharmaceutical knowledge and public data sources to quantify the drug-protein interaction and disease-specific gene expression profile. We used the comprehensive public protein-protein databases to setup the mathematical model for the repurposing problem. Third, to reduce the complexity and high-dimensionality of the repurposing problem, we applied the linear-quadratic-regulator method, which is practical in large-scale system control, to compute the hypo-treatment and evaluate the drug therapy. We apply DeCoST in Breast Cancer and Bladder Cancer case studies. Among cancer diseases, Breast Cancer causes the most number of mortality women (Centers for Disease Control Prevention, 2013). Breast Cancer is also the most comprehensively studied disease among cancers, with nearly 20 approved drugs by Food and Drug Administration (FDA). In

addition, Breast Cancer has many subtypes, which is ideal for personalized drug repurposing. In contrast, FDA only approves 4 drugs for Bladder Cancer treatment although Bladder Cancer is the fourth most commonly diagnosed cancer in the United States (American Cancer Society, 2017). Therefore, drug development in Bladder Cancer is still an opened and attractive research area. From good performance when classifying between approved drugs and withdrawn drugs, we find 7 compounds that may be promising in Breast Cancer ER-positive subtype, 3 compounds in Breast Cancer ER-negative subtype and 10 compounds in Bladder Cancer for further drug repurposing *in-vivo* study.

## METHODS

We developed our drug repurposing framework from the system modeling and control points (Figure 2). The framework integrates three types of data. First, from the Disease-specific expression profile, we quantified the expression as the system initial condition vector, where each vector elements specified whether the corresponding gene was overexpressed (red), underexpressed (green) or normally expressed (white). Second, from the protein-protein interaction database, we built the mathematical system model in order to apply the system-control algorithm. The red arrows implies activative; and the green arrow implies inhibitive interactions. Third, from the chemical-protein interaction data, we quantified the treatment vector for each



drug for later ranking. Using the initial condition vector and the mathematical model, we computed the optimal hypo-treatment. By mapping the pattern of the optimal hypo-treatment and the drugs' treatment vectors, we could rank the drugs and suggest repurposed drugs.

## Retrieve the Expression Profile as the Initial Condition Vector

We used GEO2R service (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) to analyze GEO dataset for the initial condition vector. The GEO2R service runs on R 3.2.3 platform and utilizes the well-known bioinformatics packages Biobase 2.30.0 (Huber et al., 2015), GEOquery 2.40.0 (Davis and Meltzer, 2007), and limma 3.26.8 (Ritchie et al., 2015). In GEO2R's result, we filtered out genes whose adjusted  $p$ -values exceed 0.05. The filtered-out genes were marked with 0 in the initial condition vector. For genes, whose adjusted  $p$ -values are less than 0.05, we used the sign of base-10 logarithm fold-change (logFC) in the initial condition vector. In the other words, genes with logFC > 0, which implied that the genes were overexpressed in the disease condition, were marked by 1. Genes with logFC < 0, which implied that the gene were under expressed in the disease condition, were marked by -1.

We chose GSE10886 dataset for expression profile in Breast Cancer case study. GSE10886 is among the largest and most comprehensive Breast Cancer microarray in GEO at the tissue level. After the latest update in January 2013, GSE10886 has 226 samples and including 97 ER-positive-subtype samples, 69 ER-negative-subtype samples, and 32 control samples. We chose GSE31189 dataset for Bladder Cancer expression profile. This dataset contains 52 cancer samples and 40 control samples.

## Build Disease-Specific Mathematical System Model From Interactome Data

Due to the availability of public data sources for disease-specific pathway models, we built the disease-specific system model for Breast and Bladder Cancer differently. To avoid potential false-positive, which is a well-known issue in predictive data source, we preferred using the pathway data to construct the mathematical model. For Breast Cancer, we conducted literature search on public curated pathway databases Reactome (Croft et al., 2011) and Wikipathway (Pico et al., 2008) for human disease pathways. In these databases, we only select pathways where the disease name appears in the pathways' titles or description. As the result, we found the Integrated Breast Cancer Pathway (Ibrahim et al., 2015) on Wikipathway. This pathway is among the most comprehensive Breast Cancer human pathway in the literature, which covers 239 genes and 467 interactions. The pathway also integrates 24 Breast Cancer-related pathways, including several signaling network. The entire detail about this pathway could be found in Supplemental Table 2. However, we could not find any pathways having more than 50 genes for Bladder Cancer, which implied low coverage. Therefore, for the Bladder Cancer model, we queried Bladder-Cancer-associated genes from PubMed Gene (<https://www.ncbi.nlm.nih.gov/gene/>), one of the most comprehensive literature collection in biomedical and life

sciences. To filter the possible noise during the retrieval process, we used specific query in format: <Disease Name> AND "Homo sapiens"[porgn: \_\_txid9606]. After retrieving the Bladder-Cancer-associated genes, we converted the gene identification to UniProt Knowledge Base Reviewed identification (UniProt, 2013) to filter possible alias. We queried the STRING database v10 (Szklarczyk et al., 2015), one of the most comprehensive interactome databases to retrieve the interactions information among the candidate disease-specific proteins, especially the directionality and mechanism of interactions. To filter out possible noisy information, we limited the search results only on interaction with minimum of 500 confidence score. STRING database covers 7 types of mechanism: activation, expression, inhibition, catalysis, ptmod, binding, reaction.

After retrieving the disease-associated genes and interactions from these models above, we quantified the interactome to finalize the mathematical systems for these diseases. Among the interactions, activation and inhibitions are the mechanisms with the clearest and the most unambiguous impact/directionality. Thus, we quantified the activation mechanisms by +1 and the inhibition mechanisms by -1. For the other mechanisms, we quantified them by the default value of 0. The results of this step could be represented by adjacency matrices, as showed in Supplemental Figure 1.

## Retrieve Chemical-Protein Interaction for Treatment Vector

For each disease, we curated literature for two set of drugs. The positive set, denoted by D1, includes all drugs which are approved for treatment by Food and Drug Administration (FDA). The negative set, denoted by D2, includes drugs which are withdrawn from disease treatment, or withdrawn/terminated from disease-specific clinical trials due to toxic or inefficient issues. We query <https://clinicaltrials.gov/> for clinical trials information. To avoid the complexity of multi-drug and multi-disease treatment, we ignored literature mentioning more than one drug/disease during curation. We also ignored the biotech drugs since this type of drug does not target the molecular level, therefore it is difficult to setup the treatment vector for biotech drugs. **Table 1** summarizes the list of D1 and D2 drugs we curated for Breast Cancer and Bladder Cancer. For Breast Cancer, we found 16 D1 drugs and 7 D2 drugs. In addition, to examine the possible newly therapeutic drugs for Breast Cancer, we referred to 24 drug proposed by Huang et al. (2011) as D3, in which these drugs have been approved for some other diseases by never in trial for Breast Cancer. For Bladder Cancer, we found 3 D1 drugs and 2 D2 drugs. Since we could not find any repurposed drug list for Bladder Cancer in the literature, we selected all of the 421 FDA-approved drugs for non-Bladder-Cancer diseases, which have at least one drug-gene interaction with genes in Bladder Cancer model, as D3 for Bladder Cancer. The entire D3 drug lists for both Breast Cancer and Bladder Cancer could be found in Supplemental Table 1.

We queried the DrugBank (Law et al., 2013) and DMAP (Huang et al., 2015) database for the list of drug-protein

**TABLE 1** | Drug lists (D1 and D2) curated for Breast and Bladder cancer.

Disease	Drugs	Drug sets	Disease	Drugs	Drug sets
Breast cancer	Anastrozole	D1	Breast cancer	Trastuzumab	D1
Breast cancer	Cycloheximide	D1	Breast cancer	Vinblastine	D1
Breast cancer	Exemestane	D1	Breast cancer	Diethylstilbestrol	D2
Breast cancer	Fluorouracil	D1	Breast cancer	Dromostanolone	D2
Breast cancer	Fluoxymesterone	D1	Breast cancer	Formestane	D2
Breast cancer	Fulvestrant	D1	Breast cancer	Ixabepilone	D2
Breast cancer	Lapatinib	D1	Breast cancer	Avastin	D2
Breast cancer	Letrozole	D1	Breast cancer	Ethyl Carbamate	D2
Breast cancer	Miltefosine	D1	Breast cancer	Imetelstat	D2
Breast cancer	Paclitaxel	D1	Breast cancer	Tivozanib	D2
Breast cancer	Pamidronate	D1	Bladder cancer	Cisplatin	D1
Breast cancer	Raloxifene	D1	Bladder cancer	Doxorubicin HCl	D1
Breast cancer	Tamoxifen	D1	Bladder cancer	Thiotepa	D1
Breast cancer	Thiotepa	D1	Bladder cancer	Mitomycin C	D2
Bladder cancer	Gemcitabine	D2			

D1, FDA-approved drugs (positive/good drug set); D2, FDA-rejected/withdrawn drugs (negative/bad drug set).

interaction mechanism. DMAP and DrugBank covers 38 mechanisms of drug action. In DMAP, we filtered out interactions with confidence score less than 800 (over 1,000) to avoid noisy information. From biological knowledge, we quantified these mechanisms as showed in **Table 2**. Similar to quantification of protein-protein mechanism of action, an inhibited or similar action is map to  $-1$ ; and an activated or similar action is map to  $+1$ .

### Construct Disease-Specific Drugs' Therapeutic Scoring for Drug Repurposing Purpose

The key principle in applying system control to evaluate drugs' therapy relies in the following assumption: in disease condition, the gene expressions are derived away from the balanced level of 0. Therefore, a good treatment should reverse the gene expressions in disease condition and stabilize the expressions to the balance level. In **Figure 2**, we illustrate this principle and explain several mathematical notation in a toy example. Based on system biology literature (Alberghina, 2007), we assume that there exists a model governing the gene expressions, which allows us to model the expression using time-series perspective

$$\mathbf{x}(t) = f(\mathbf{x}(t - 1), \mathbf{u}(t - 1)) \tag{1}$$

where  $\mathbf{x} \in \mathbb{R}^N$  stands for the quantified gene expression of  $N$  genes,  $\mathbf{u} \in \mathbb{R}^N$  stands for the quantified treatment and  $t$  is the iteration and  $f$  is the arbitrary function controlling the expression change. The initial  $\mathbf{x}(0)$  is the quantified gene expression in disease condition. In this paper, we choose a linear model for  $f$ .

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t - 1) + \mathbf{u}(t - 1) \tag{2}$$

We chose the linear model because not only it is simple but also it has equilibrium point at the origin: if  $\mathbf{x}(t-1) = \mathbf{u}(t-1) = 0$

**TABLE 2** | Quantification of drug-protein mechanism of action in drug-protein interaction databases.

Mechanism of action	Quantification	Mechanism of action	Quantification
Activator	1	Ligand	0
Adduct	0.5	Metabolizer	0
Agonist	1	Modulator	0
Allosteric modulator	0	Multitarget	0
Antagonist	-1	Negative modulator	-1
Antibody	0	Neutralizer	0
Binder	0	Other	0
Chaperone	1	Other/unknown	0
Chelator	0	Partial agonist	1
Cleavage	-1	Partial antagonist	-1
Cofactor	1	Positive allosteric modulator	1
Component of	0	Potentiator	1
Cross-linking/alkylation	0	Product of	0
Incorporation into and destabilization	-1	Reducer	-1
Inducer	1	Stimulator	1
Inhibitor	-1	Suppressor	-1
Inhibitor, competitive	-1	Unknown	0
Inhibitory allosteric modulator	-1	Other terms	0
Intercalation	0	-	-

The Mechanism of Action terminologies are retrieved from drug-target annotation in DrugBank database. Quantification stands for the numerical representation of the Mechanism of Action in the modeling and computing steps.

then  $\mathbf{x}(t) = 0$ . This fact implies that when the gene expressions are already at the balance level, treatment is no longer needed. In addition, it is easier to setup a linear system with stability

(Chui and Chen, 2012)

$$\text{If } \|\mathbf{x}(0)\| < \varepsilon \text{ and } \mathbf{u} = 0 \text{ then } \|\mathbf{x}(t)\| < \varepsilon \forall t \quad (3)$$

where  $\|\mathbf{x}\|$  stands for the second norm of  $\mathbf{x}$  and  $\varepsilon$  is an arbitrary small number. This fact implies the self-adjustment of the gene expression at the control level. We setup matrix  $\mathbf{A}$  from quantification of protein-protein mechanism of interactions (section Methods). With temporal matrix  $\mathbf{A}^*$  as the result of section Methods

$$\mathbf{A}^*(i, j) = \begin{cases} -1 & \text{if protein } i \text{ inhibits protein } j \\ 1 & \text{if protein } i \text{ activates protein } j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Let  $\lambda$  be the eigenvalue of  $\mathbf{A}^*$  with the largest magnitude. By setting up  $\mathbf{A}$  as

$$\mathbf{A} = (1/\lambda) \mathbf{A}^* \quad (5)$$

We can guarantee the stability of system (2) (Chui and Chen, 2012).

The objective of the linear control is to find a sequence of  $\mathbf{u}(t)$  such that

$$\mathbf{x}(t) \rightarrow 0 \text{ as } t \rightarrow \infty \quad (6)$$

Optimal control considers not only how to stabilize  $\mathbf{x}$  quickly but also consider the cost-effective of the treatment  $\mathbf{u}$ . Regarding this point, the optimal linear control aims to minimize

$$J(\mathbf{x}(0)) = \sum_{t=0}^{\infty} (\mathbf{x}(t)^T \mathbf{x}(t) + \mathbf{u}(t)^T \mathbf{u}(t)) \quad (7)$$

To solve the optimization problem (2-7) we solved the corresponding Riccati equation (Arnold and Laub, 1984)

$$\mathbf{A}^T \mathbf{P} \mathbf{A} - \mathbf{P} - \mathbf{A}^T \mathbf{P} (\mathbf{P} + \mathbf{I})^{-1} \mathbf{P} \mathbf{A} + \mathbf{I} = 0 \quad (8)$$

using DARE algorithm (Arnold and Laub, 1984) in Matlab (<https://www.mathworks.com/help/control/ref/dare.html>). In (8),  $\mathbf{P}$  is just an intermediate result containing no biological representation. We compute the treatment vector  $\mathbf{u}(t)$  as follow

$$\mathbf{u}(t) = -(\mathbf{I} + \mathbf{P})^{-1} \mathbf{P} \mathbf{A} \mathbf{x}(t) \quad (9)$$

In system control practice, since  $\mathbf{u}(t)$  often converges to 0 quickly (Bemporad et al., 2002), the first treatment vector  $\mathbf{u}(0) = -(\mathbf{I} + \mathbf{P})^{-1} \mathbf{P} \mathbf{A} \mathbf{x}(0)$  often plays the most important role in optimally stabilizing the system (2). Therefore, we can consider  $\mathbf{u}(0)$  as the optimal hypo-treatment. We compare the similarity between the real drug treatment ( $\mathbf{u}_d$ ) and the hypo-treatment as the therapeutic score  $T(d)$  for each drug  $d$  as follow

$$T_d = |\mathbf{u}_d^T \text{sign}(\mathbf{u}(0))| / |\text{abs}(\mathbf{u}_d)^T \text{abs}(\text{sign}(\mathbf{u}(0)))| \quad (10)$$

where  $\text{abs}$  stand for the absolute value function. Here,  $T_d$  ranges between  $-1$  and  $1$ . The numerator  $|\mathbf{u}_d^T \text{sign}(\mathbf{u}(0))|$  is the matching function between drug  $d$  and the optimal hypo-treatment, which is incremented when  $\mathbf{u}_d(i)$  and  $\mathbf{u}(0)(i)$  are non-zero analog, and decremented when  $\mathbf{u}_d(i)$  and  $\mathbf{u}(0)(i)$  are opposite. We measured the impact of  $T_d$  score by the receiver operating characteristic when we use  $T_d$  to classify D1 drugs vs. D2 drugs.

## RESULTS

### Therapeutic Scores for Breast Cancer Drugs

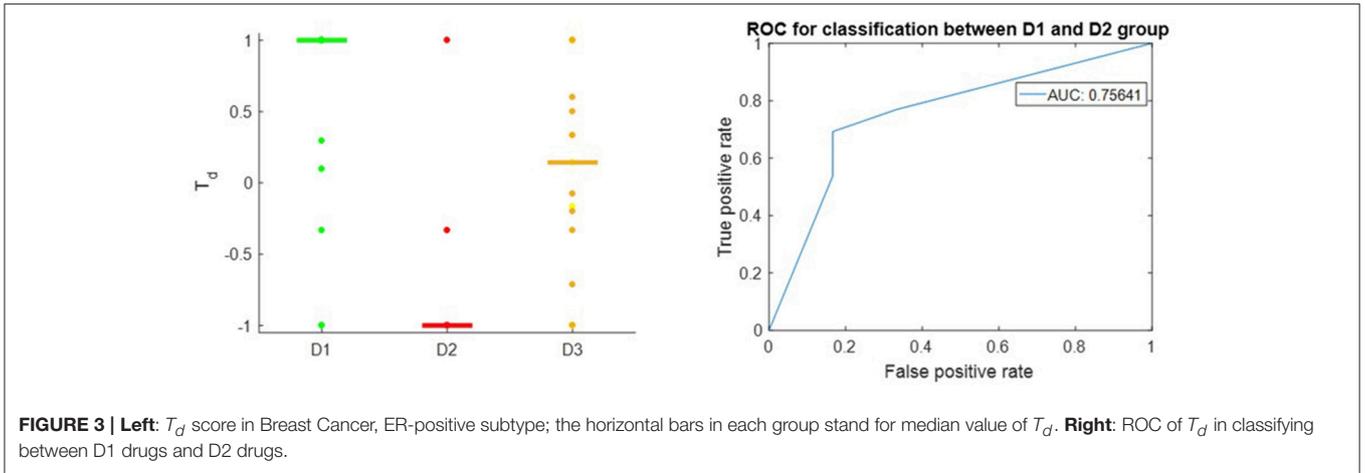
From the Integrated Breast Cancer Pathway (Ibrahim et al., 2015) on Wikipathway (section Methods) and the Breast Cancer drug list in Supplemental Table 3, we queried 222 drug-protein interactions for the drugs' treatment vectors (Supplemental Table 4). Supplemental Table 5 contains the initial condition vector from GEO2R expression analysis.

Figure 3 shows that the  $T_d$  score is able to give appropriate ranking for most of the well-known therapeutic drugs and suggest candidate drugs for repurposing in Breast Cancer ER-positive case.  $T_d$  score reflexes the difference between the D1 and D2 drugs with receiver operator characteristic (Hanley and McNeil, 1982) area under the curve (AUC) of 0.76. This result is comparable to the overall result queried from Broad Institute CMAP (Subramanian et al., 2017) on MCF-7, the Breast Cancer ER+ cell line, using the Touchstone tool (<https://clue.io/touchstone>). Especially on the drugs covered in CMAP, DeCoST achieves AUC of 0.91, which is much higher than the AUC achieved by CMAP (0.79), as showed in the Supplemental Text 1. We did not setup training set and test set for classification because the model construction and  $T_d$  calculation does not need the drug categories. The  $T_d$  scores for D1 drugs in Breast Cancer ER-negative case are relatively lower than the scores for ER-positive case (Figure 4). Comparison detail has been shown in Supplemental Table 5. Using  $T_d$  for classifying D1 and D2 drugs yields AUC of 0.68. In fact, clinical trials and literature have showed several drugs which are effective in ER-positive treatment but show little or no impact in ER-negative treatment. For example, Tamoxifen ( $T_d$  ER-positive: 0.294,  $T_d$  ER-negative: 0.176), which is a selective estrogen receptor modulator, does not prevent ER-negative Breast Cancer, when the estrogen receptor genes do not express (Fabian, 2007; Uray and Brown, 2011).

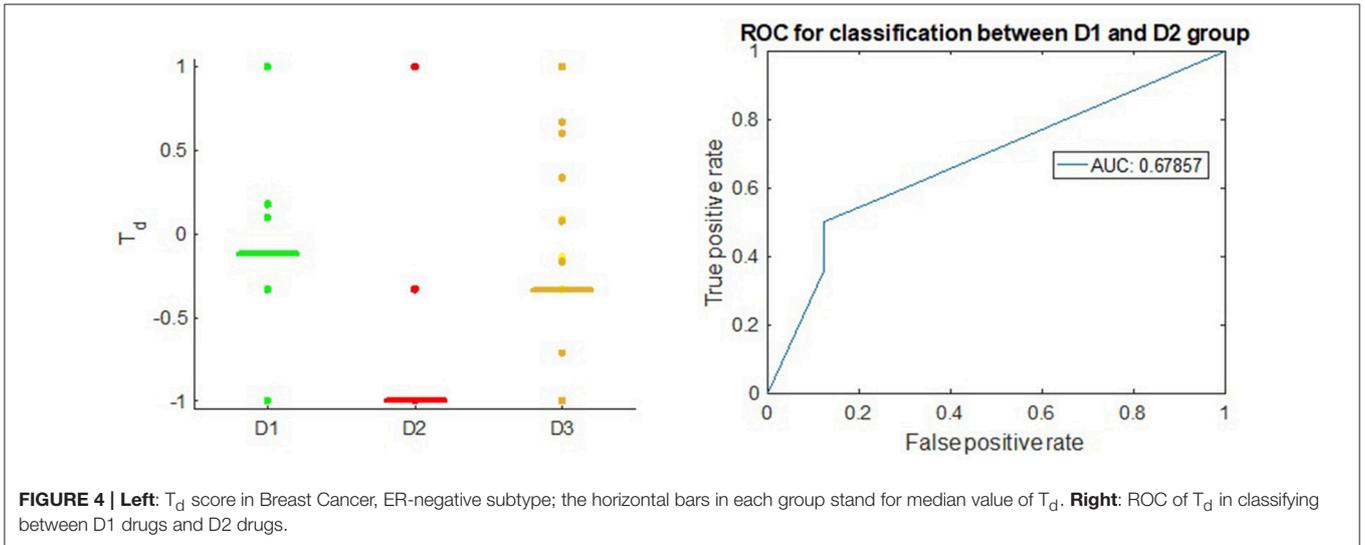
### Therapeutic Scores for Bladder Cancer Drugs

Since we could not find any human pathway with sufficient coverage for Bladder Cancer, our Bladder Cancer system model retrieved the Bladder-Cancer-specific genes from PubMed Gene server. The model contains 738 proteins and 1,241 protein-protein interactions. From 6 drugs in the Bladder Cancer case-study, we retrieved 48 drug-protein interactions for drugs' treatment vector. From GSE31189 gene expression dataset, we found 221 genes whose expression differs from the balance level. Details about the Bladder Cancer system could be found in Supplemental Tables 6-8.

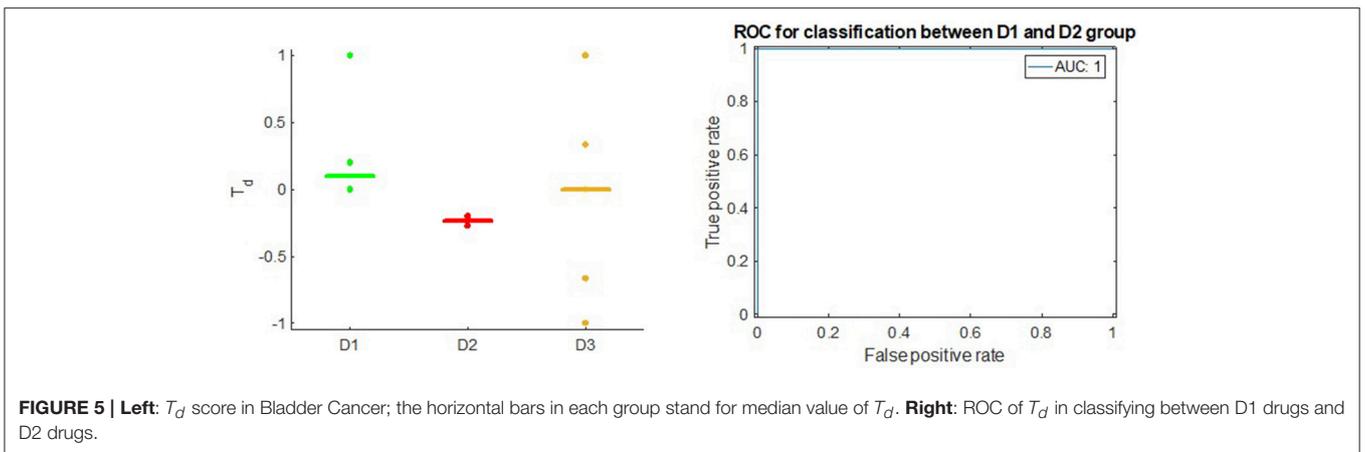
We observed AUC of 1.0 (Figure 5) when we used  $T_d$  score to classify between D1 and D2 drugs in Bladder Cancer. Here, all of the D1 drugs receive non-negative  $T_d$  scores: Cisplatin receives the score of 0.2, Doxorubicin Hydrochloride receives the score of 0.0 and Thiotepa receives the score of 1.0. All of the D2 drugs receive negative  $T_d$  scores: Mitomycin C receives the score of  $-0.2$  and Gemcitabine receives the score of  $-0.09$ .



**FIGURE 3 |** Left:  $T_d$  score in Breast Cancer, ER-positive subtype; the horizontal bars in each group stand for median value of  $T_d$ . Right: ROC of  $T_d$  in classifying between D1 drugs and D2 drugs.



**FIGURE 4 |** Left:  $T_d$  score in Breast Cancer, ER-negative subtype; the horizontal bars in each group stand for median value of  $T_d$ . Right: ROC of  $T_d$  in classifying between D1 drugs and D2 drugs.



**FIGURE 5 |** Left:  $T_d$  score in Bladder Cancer; the horizontal bars in each group stand for median value of  $T_d$ . Right: ROC of  $T_d$  in classifying between D1 drugs and D2 drugs.

## Potential Drugs for Breast Cancer Studies and Biological Insights

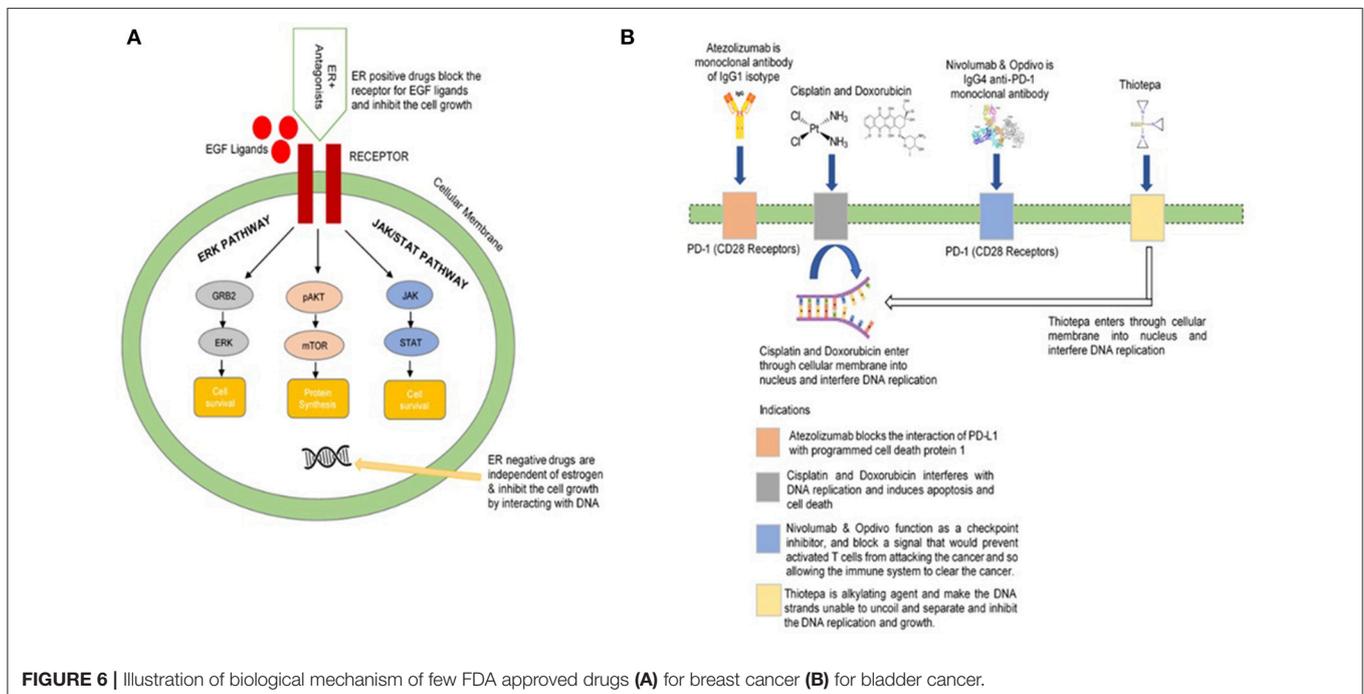
From the  $T_d$  scores for D3 drugs, our framework suggests 8 drugs (Eribitux, Flutamide, Medryson, Methylprednisolone, Norethindrone, Prednisolone, Prednisonea, and Vandetanib) with high potential efficacy in Breast Cancer ER+ drug repurposing. Significantly, these drugs do not directly target Estrogen receptor, which is the most well-known approach in Breast Cancer ER+ drug design. Tamoxifen is a typical example of Breast Cancer drugs which slows cancer process by blocking estrogen hormone receptors, preventing hormones from binding to them. About 80% of all breast cancers are ER+: the cancer cells grow in response to the hormone estrogen (Bulut and Altundag, 2015). About 65% of the ER+ cases grow in response to another hormone, progesterone (Hefti et al., 2013). Tumors in ER/PR-positive cases are much more likely to respond to hormone therapy than tumors that are ER/PR-negative. ER+ breast cancer entirely depends on the estrogen for growth and propagation involving genomic and non-genomic pathways. Epidermal growth factor receptor (EGFR) is a receptor found on both normal and tumor cells that is important for cell growth (Herbst, 2004; Khoo et al., 2015). ER-positive (ER+) drugs recommended for repurposing in this framework block the activities and growth of EGFR (Figure 6A). These drugs show different mechanism of action with the common objective of the inhibition of the growth of cancerous cells. By adjusting and modifying the known biases of the interactomic networks, our procedure would help to reveal the therapeutic effect of drugs along with effective treatments.

For Breast Cancer ER- case, our framework suggests Daunorubicin and Donepezil as the repurposing candidates. These drugs are independent of estrogen and usually inhibit the

cell growth by either interacting with DNA or inhibiting Cholinesterases. Daunorubicin interacts with DNA by intercalation and inhibition of macromolecular biosynthesis (Momparler et al., 1976). This inhibits the progression of the enzyme topoisomerase II, and thereby stopping the process of replication. Donepezil is in a class of cholinesterase inhibitor that improves mental function and fatigue in cancer. The current research focused on recent large-scale efforts to systematically find repositioning candidates and elucidate individual disease mechanisms in cancer (Bruera et al., 2007). Personalized medicine and repositioning both aim to improve the productivity of current drug discovery pipelines. Standard drug discovery strategies can also lead to repositioning opportunities. D1, D2, and D3 drugs (Table 1) found to potentially modulate the desired activity are repositioning candidates.

## Potential Drugs for Bladder Cancer Studies and Biological Insights

From the list of 143 FDA-approved drug with high  $T_d$  score, we found 10 candidates drugs (with  $T_d = 1$ ) whose mechanisms are promising for Bladder Cancer repurposing. The  $T_d$  scores for all Bladder Cancer drugs could be found in Supplemental Table 9. The prevalence of drug-repositioning studies has resulted in a variety of innovative computational methods for the identification of new opportunities for the use of old drugs. We sorted the potential list of drugs against bladder cancer. The reprofiling of these drugs followed the same biological mechanisms. For example, Zafirlukast antagonizes ATP-binding cassette and may improve the efficacy of anticancer effects (Sun et al., 2012). Similarly, Tenofovir may reduce the risk of bladder or others cancers while dopamine receptor antagonist Thioridazine inhibits tumor growth (Yin et al., 2015). Losartan



is an angiotensin II receptor (AT-II-R) blocker that is widely used by human for blood pressure regulation but it also shows antitumor property (Barreras and Gurk-Turner, 2003). Ciclopirox was first marketed in 1982 as an antifungal agent found in several topical drug products. However, further research demonstrated that it was able to kill bladder cancer cells (Weir et al., 2011). The Atezolizumab, Cisplatin, Doxorubicin, Nivolumab, Opdivo, Thiotepe, and others (Figure 6B) are FDA approved drugs which are recommended for bladder cancer.

## DISCUSSION

The applications of drug-repositioning studies have brought a variety of new *in silico* approaches in drug designing and development. In most of the studies, the anticancer effect of newly designed drugs usually has been presented *in vitro* as clinical trials are very expensive and time consuming, but remain the only way to validate drug efficiency *in vivo*. Therefore, to establish accurate and effective drug-repositioning framework needs development of new computational techniques. In this work, we discuss and demonstrate the application of control system theory as a computational method to evaluate drug efficacy and repurposing from integrated system biology data. The capability in classification between approved and withdrawn drugs is the fundamental foundation for our framework in drug repurposing. It is important to note that although our AUC of 0.76 and 0.68 in Breast Cancer is inferior compared to the state-of-the-art methods (Cheng et al., 2012; Zheng et al., 2015), our validation is conducted from the pharmaceutical knowledge of drug's efficacy on treatment at the system-pathway level; meanwhile, the other methods often validate at the targeted molecular level. In addition, we set strict criteria in choosing the negative set by only choosing drugs that are rejected or withdrawn from disease-specific clinical trials and treatments. The state-of-the-art methods tend to be more relaxed on the negative set by choosing drug not being used in disease-specific drugs, which may have limitation on repurposing options. In addition, the appropriate assessment of tamoxifen efficacy between Breast Cancer ER+ and Breast Cancer ER- highlights the potential advantages of our framework in personalized drug repurposing. Compare to the approved drugs, the candidate drugs suggested in this work show different promising drug mechanisms which may be useful in future drug design.

In our work, although the number of target may be among the key difference between the D1 drugs and the D2 drugs, our analysis shows that the number of drugs' targeted genes and the targeted genes are not the only factors affecting the clinical outcome and predictive results in drug repurposing. As showed in Supplemental Table 3, D1 drugs, on the average, has more targets than D2 drugs. However, D1 drugs for Breast Cancer (average number of targets: 4.8) include both single-target (such as Anastrozole, Exemestane, and Fluorouracil) and multi-target (such as Tamoxifen, Paclitaxel, and Cycloheximide) ones. D2 (average number of targets: 3.3) drugs also contains the single-target (such as Ixabepilone and Avastin) and the multi-target (such as Imetelstat and Diethylstilbestrol). In the result section, DeCoST's evaluation for these drugs showed above is appropriate for their clinical outcome. In addition, drugs

targeting the same marker genes do not necessary have the same outcome. For example, both Tamoxifen and Diethylstilbestrol target the estrogen receptors ESR1 and ESR2, which are the marker in Breast Cancer ER+ (Yip and Rhodes, 2014). However, their clinical outcomes and DeCoST's evaluation are opposite, primarily because they have opposite mechanisms on the same targets of estrogen receptors: Tamoxifen is the estrogen inhibitor while Diethylstilbestrol is the estrogen activator. Since Breast Cancer ER+ is strongly associated with the overexpression of estrogen receptors (Yip and Rhodes, 2014), Tamoxifen could have therapeutic outcome because it reverses the disease signature. Meanwhile, Diethylstilbestrol should have poor outcome because it shows the analog to the disease signature.

In this work, we have showed the results between DeCoST and the Broad Institute CMAP, which is among the most well-known and comprehensive platforms for drug repurposing. In addition, our strategy of repurposing is similar to CMAP. Although Supplemental Text 1 shows that our DeCoST has higher AUC than CMAP does, it is inappropriate to conclude that DeCoST is better than the CMAP. There are fundamental differences in conducting experiment making comparison not totally solid. First, the expression profiles acquired by CMAP are at the cell line level; meanwhile, in this work DeCoST acquires the expression profile at the tissue level, which is closer to *in-vivo* studies. Second, due to several factors in experimental design, CMAP does not contains cell line for Breast Cancer ER- and Bladder Cancer. CMAP also covered less number of drugs, compared to the drug list evaluated in this work. Therefore, the key point in comparative evaluation should be on the repurposing hypotheses suggested by these platforms in future *in-vivo* studies and the biological insights of these hypotheses. In our results, we have offered several biological explanations why drugs recommended by DeCoST could be repurposed. Unfortunately, we could not compare between CMAP and DeCoST at this point. DeCoST focuses primarily on recommending drugs that have never been in disease-specific clinical trials; meanwhile, CMAP (<https://clue.io/repurposing-app>) primarily reports on drugs that has been under early phases of clinical trials. Therefore, we believe that DeCoST could provide complimentary advantages, in addition to CMAP.

The advantages of our framework are established not only by advanced computational method but also by two layers of personalized system (Li and Jones, 2012). In the first layer, the disease-specific gene expression could differ among different patients and subtypes, which results in different initial state condition. In the second layer, different types of disturbance among molecular-molecular interactions could be discovered and represented differently in the system modeling step. In our results, we show that Tamoxifen, which is approved to treat Breast Cancer, may not be effective in treating Breast Cancer ER-. The strong support from literature to this evaluation is a good example of the personalized medicine characteristics. In addition, our framework could easily integrate the results from many other state-of-the-art repurposing approaches such as molecular docking and gene-set enrichment analysis to refine the efficacy prediction. The main idea in this framework, which is based on control system theory, could be applied in many other bioinformatics problem, such as target prioritization and

discovering new combination of treatments. In addition, our framework could easily be extended to evaluate combination of treatment, with careful preprocessing the drug-drug interaction data (Ayvaz et al., 2015; Wang et al., 2017).

In addition, our framework shows repurposing capacity at both target level and pathway level. At the target level, we show typical examples for EGFR-targeted and ACHE-targeted drugs. Patients being considered for anti-epidermal EGFR therapy are often screened for mutations in the oncogene KRAS (Hoorens et al., 2010) because a constitutively active KRAS gene downstream of EGFR would not be affected by EGFR inhibition. Many diseases have approved combination regimens, such as metastatic colorectal and bladder cancer and its four-drug FOLFIRI (folinic acid, 5-fluorouracil, irinotecan) with cetuximab regimen (Raoul et al., 2009). Losartan is an angiotensin II receptor (AT-II-R) blocker and this angiotensin-converting enzyme inhibitors (ACE) may have a protective role in bladder and other cancers (Yazdannejat et al., 2016). In the other hand, a typical example at the pathway level is Thioridazine. Thioridazine-induced effects are associated with inhibition of the canonical NF $\kappa$ B pathway.

The limitations in this work are the method to quantify the categorical data from public genomic/proteomic databases and the simplicity of linear system control. First, all of the data are discretized into only three values:  $-1$ ,  $0$ , and  $1$ , which could lower the resolution of the final drug therapeutic score. Second, the linear system control approach needs to assume that the gene expression transition could be approximate closely by a linear equation, which is still unverified due to the scarcity of time-series gene expression data. Therefore, when applying into another repurposing problem, biologists and pharmacologists should apply deeper domain knowledge to increase the resolution of discrete quantification. Furthermore, mathematical nonlinear system identification and reinforcement learning, which are popular approach in unknown system control, could be used to increase the accuracy of system modeling and make the system more personalized. Integration of other resources, such as drugs, genes, and systems associated with side-effects (Kuhn et al., 2016; Maier et al., 2018) and high-throughput screening (Deftereos et al., 2011; Macarron et al., 2011) would also be valuable expansions of this work in the future. Also, the computational complexity of DeCoST is generally high (expected  $O(n^8)$ , where  $n$  is the number of genes in the model). This complexity is manageable with most of the existing biological pathway model (expect about 400 genes). However, this could be a bottleneck if the number of genes raises to several thousands.

In addition, the advantages of our framework in personalized medicine may associate with the reproducibility issues (Draghici et al., 2006; Frye et al., 2015). As mentioned, the disease-specific gene expression could differ among different patients

and subtypes. Therefore, we could not completely guarantee that applying our framework on different gene expression data and on different interactome data sources (Chatr-Aryamontri et al., 2013; Szklarczyk et al., 2015) would return the same result. Therefore, by reproducibility, we can only guarantee that given a specific gene expression profile and an interactome data source, we can always produce the same result. In this work, we have tried to tackle the reproducibility issue by using tight criteria to select the positive/negative drug set, by maintaining the relevance and coverage of the disease-specific model, and by choosing the expression data set with high number of samples.

## CONCLUSION

In this work, we have developed DeCoST, one of the first techniques from system control paradigm, to tackle the drug repurposing challenges. We showed that DeCoST could appropriately retrieve the clinical outcomes of drugs treating personalized Breast Cancer and Bladder Cancer. From the good retrieval result, DeCoST suggests repurposing 8-candidate drugs for Breast and 10 drugs for Bladder Cancer with biological insights. This framework would be promising to discover new therapeutic strategies to treat other cancer diseases.

## AUTHOR CONTRIBUTIONS

TN designed the study (including the mathematical details), curated the Bladder Cancer drug dataset and analyzed the computational results. SM validated and provided biological insights for the results. SI constructed the Breast Cancer pathway model and collected the drug clinical outcomes for Breast Cancer. LM processed the expression and protein-protein interaction data for Bladder Cancer. LM, JG, BB, and BZ implemented the system control algorithm used in the paper. All authors contributed to the manuscript writing and edition.

## ACKNOWLEDGMENTS

The authors thank Dr. Jake Chen from Informatics Institute—School of Medicine—the Alabama University of Birmingham for helpful comments in the study design. The authors also thank Wenzhou Medical University, Zhejiang, China for covering the publication cost for this paper.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.00583/full#supplementary-material>

## REFERENCES

- Alberghina, L. (2007). *Systems Biology: Definitions and Perspectives*. Berlin: Springer Science and Business Media.
- American Cancer Society (2017). *Key Statistics for Bladder Cancer*. American Cancer Society.

- Andersson, M. L., Bottiger, Y., Bastholm-Rahmner, P., Ovesjo, M. L., Veg, A., and Eiermann, B. (2015). Evaluation of usage patterns and user perception of the drug-drug interaction database SFINX. *Int. J. Med. Inform.* 84, 327–333. doi: 10.1016/j.ijmedinf.2015.01.013
- Arnold, W. F. III, and Laub, A. J. (1984). Generalized eigenproblem algorithms and software for algebraic Riccati

- equations. *Proc. IEEE* 72, 1746–1754. doi: 10.1109/PROC.1984.13083
- Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, N. P., et al. (2015). Toward a complete dataset of drug-drug interaction information from publicly available sources. *J. Biomed. Inform.* 55, 206–217. doi: 10.1016/j.jbi.2015.04.006
- Bardi, M., and Capuzzo-Dolcetta, I. (2008). *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Boston, MA: Springer Science and Business Media.
- Barreras, A., and Gurk-Turner, C. (2003). Angiotensin II receptor blockers. *Proc. Bayl. Univ. Med. Cent.* 16, 123–126. doi: 10.1080/08998280.2003.11927893
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Baxevanis, A. D. (2012). Searching online mendelian inheritance in man (OMIM) for information on genetic loci involved in human disease. *Curr. Protoc. Hum. Genet.* 13, 11–10. doi: 10.1002/0471250953.bi0102s27
- Bemporad, A., Morari, M., Dua, V., and Pistikopoulos, E. N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica* 38, 3–20. doi: 10.1016/S0005-1098(01)00174-1
- Bisson, W. H. (2012). Drug repurposing in chemical genomics: can we learn from the past to improve the future? *Curr. Top Med. Chem.* 12, 1883–1888. doi: 10.2174/156802612804547344
- Bruera, E., El Osta, B., Valero, V., Driver, L. C., Pei, B. L., Shen, L., et al. (2007). Donepezil for cancer fatigue: a double-blind, randomized, placebo-controlled trial. *J. Clin. Oncol.* 25, 3475–3481. doi: 10.1200/JCO.2007.10.9231
- Bulut, N., and Altundag, K. (2015). Does estrogen receptor determination affect prognosis in early stage breast cancers? *Int. J. Clin. Exp. Med.* 8, 21454–21459.
- Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science* 321, 263–266. doi: 10.1126/science.1158140
- Centers for Disease Control and Prevention (2013). *Breast Cancer Statistics, October 23, 2013*. Available online at: <http://www.cdc.gov/cancer/breast/statistics/>
- Chatr-Aryamontri, A., Breitkreutz, B. J., Heinicke, S., Boucher, L., Winter, A., and Tyers, M. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41, D816–D823. doi: 10.1093/nar/gks1158
- Chen, J., Pandey, R., and Nguyen, T. M. (2017). HAPPI-2: a comprehensive and high-quality map of human annotated and predicted protein interactions. *BMC Genomics* 18:182 doi: 10.1186/s12864-017-3512-1
- Chen, M. Z., Zhang, L., Su, H., and Chen, G. (2016). Stabilizing solution and parameter dependence of modified algebraic Riccati equation with application to discrete-time network synchronization. *IEEE Trans. Automat. Contr.* 61, 228–233. doi: 10.1109/TAC.2015.2434011
- Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., et al. (2012). Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* 8:e1002503. doi: 10.1371/journal.pcbi.1002503
- Chou, T.-C. (2010). Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res.* 70, 440–446. doi: 10.1158/0008-5472.CAN-09-1947
- Chui, C. K., and Chen, G. (2012). *Linear Systems and Optimal Control*. Berlin: Springer Science and Business Media.
- Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39, D691–D697. doi: 10.1093/nar/gkq1018
- Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi: 10.1093/bioinformatics/btm254
- Deftereos, S. N., Andronis, C., Friedla, E. J., Persidis, A., and Persidis, A. (2011). Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 323–334. doi: 10.1002/wsbm.147
- Dimasi, J. A. (2001). New drug development in the United States from 1963 to 1999. *Clin. Pharmacol. Ther.* 69, 286–296. doi: 10.1067/mcp.2001.115132
- Draghici, S., Khatri, P., Eklund, A. C., and Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* 22, 101–109. doi: 10.1016/j.tig.2005.12.005
- Dudley, J. T., Deshpande, T., and Butte, A. J. (2011). Exploiting drug-disease relationships for computational drug repositioning. *Brief. Bioinformatics* 12, 303–311. doi: 10.1093/bib/bbr013
- Durrant, J. D., Amaro, R. E., Xie, L., Urbaniak, M. D., Ferguson, M. A., Haapalainen, A., et al. (2010). A multidimensional strategy to detect polypharmacological targets in the absence of structural and sequence homology. *PLoS Comput. Biol.* 6:e1000648. doi: 10.1371/journal.pcbi.1000648
- Fabian, C. J. (2007). The what, why and how of aromatase inhibitors: hormonal agents for treatment and prevention of breast cancer. *Int. J. Clin. Pract.* 61, 2051–2063. doi: 10.1111/j.1742-1241.2007.01587.x
- Falcone, M., and Ferretti, R. (2013). *Semi-Lagrangian Approximation Schemes for Linear and Hamilton—Jacobi Equations*. Philadelphia, PA: SIAM.
- Frye, S. V., Arkin, M. R., Arrowsmith, C. H., Conn, P. J., Glicksman, M. A., Hull-Ryde, E. A., et al. (2015). Tackling reproducibility in academic preclinical drug discovery. *Nat. Rev. Drug Discov.* 14, 733–734. doi: 10.1038/nrd4737
- Goel, N. S., and Richter-Dyn, N. (2016). *Stochastic Models in Biology*. London, UK: Elsevier.
- Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabasi, A. L. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8685–8690. doi: 10.1073/pnas.0701361104
- Gottlieb, A., Stein, G. Y., Ruppin, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* 7:496. doi: 10.1038/msb.2011.26
- Gupta, S. C., Sung, B., Prasad, S., Webb, L. J., and Aggarwal, B. B. (2013). Cancer drug discovery by repurposing: teaching new tricks to old dogs. *Trends Pharmacol. Sci.* 34, 508–517. doi: 10.1016/j.tips.2013.06.005
- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747
- Hannon, B., and Ruth, M. (2014). *Modeling Dynamic Biological Systems*. London, UK: Springer.
- Hefti, M. M., Hu, R., Knoblauch, N. W., Collins, L. C., Haibe-Kains, B., Tamimi, R. M., et al. (2013). Estrogen receptor negative/progesterone receptor positive breast cancer is not a reproducible subtype. *Breast Cancer Res.* 15:R68. doi: 10.1186/bcr3462
- Herbst, R. S. (2004). Review of epidermal growth factor receptor biology. *Int. J. Radiat. Oncol. Biol. Phys.* 59(Suppl.), 21–26. doi: 10.1016/j.ijrobp.2003.11.041
- Hoorens, A., Jouret-Mourin, A., Sempoux, C., Demetter, P., De Hertogh, G., and Teugels E. (2010). Accurate KRAS mutation testing for EGFR-targeted therapy in colorectal cancer: emphasis on the key role and responsibility of pathologists. *Acta Gastroenterol. Belg.* 73, 497–503.
- Hu, G., and Agarwal, P. (2009). Human disease-drug network based on genomic expression profiles. *PLoS ONE* 4:e6536. doi: 10.1371/journal.pone.0006536
- Huang, H., Nguyen, T., Ibrahim, S., Shantharam, S., Yue, Z., and Chen, J. Y. (2015). DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinformatics* 16(Suppl. 13):S4. doi: 10.1186/1471-2105-16-S13-S4
- Huang, H., Wu, X., Pandey, R., Li, J., Zhao, G., Ibrahim, S., et al. (2012). C(2)Maps: a network pharmacology database with comprehensive disease-gene-drug connectivity relationships. *BMC Genomics* 13(Suppl. 6):S17. doi: 10.1186/1471-2164-13-S6-S17
- Huang, H., Xiaogang, W., Ibrahim, S., Kenzie, M. M., and Chen, J. Y. (2011). “Predicting drug efficacy based on the integrated breast cancer pathway model,” in *2011 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, (San Antonio, TX), 42–45.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121. doi: 10.1038/nmeth.3252
- Ibrahim, S., Hanspers, K., Willighagen, E., Wagle, P., Chen, J., Digles, D., et al. (2015). *Integrated Breast Cancer Pathway (Homo sapiens)*. Wikipathway.org, Wikipathway.org.
- Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14621–14626. doi: 10.1073/pnas.1000138107
- Jensen, P. B., Jensen, L. J., and Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* 13, 395–405. doi: 10.1038/nrg3208
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

- Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., and Roth, B. L. (2009). Predicting new molecular targets for known drugs. *Nature* 462, 175–181. doi: 10.1038/nature08506
- Khoo, C., Rogers, T. M., Fellowes, A., Bell, A., and Fox, S. (2015). Molecular methods for somatic mutation testing in lung adenocarcinoma: EGFR and beyond. *Transl Lung Cancer Res.* 4, 126–141. doi: 10.3978/j.issn.2218-6751.2015.01.10
- Kuhn, M., Letunic, I., Jensen, L. J., and Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Res.* 44, D1075–D1079. doi: 10.1093/nar/gkv1075
- Kuhn, M., Szklarczyk, D., Franceschini, A., von Mering, C., Jensen, L. J., and Bork, P. (2012). STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Res.* 40, D876–D880. doi: 10.1093/nar/gkr1011
- Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., and Golub, T. R. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi: 10.1126/science.1132939
- Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., and Wishart, D. S. (2013). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42, D1091–D1097. doi: 10.1093/nar/gkt1068
- Li, J., and Lu, Z. (2013). Pathway-based drug repositioning using causal inference. *BMC Bioinformatics* 14(Suppl. 16):S3. doi: 10.1186/1471-2105-14-S16-S3
- Li, Y. Y., and Jones, S. J. (2012). Drug repositioning for personalized medicine. *Genome Med.* 4:27. doi: 10.1186/gm326
- Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., and Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195. doi: 10.1038/nrd3368
- Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E., et al. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* 555, 623–628. doi: 10.1038/nature25979
- Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S., and Houghten, R. A. (2013). Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov. Today* 18, 495–501. doi: 10.1016/j.drudis.2013.01.008
- Momparler, R. L., Karon, M., Siegel, S. E., and Avila, F. (1976). Effect of adriamycin on DNA, RNA, and protein synthesis in cell-free systems and intact cells. *Cancer Res.* 36, 2891–2895.
- Oprea, T. I., Bauman, J. E., Bologa, C. G., Buranda, T., Chigaev, A., and Sklar, L. A. (2011). Drug repurposing from an academic perspective. *Drug Discov. Today Ther. Strateg.* 8, 61–69. doi: 10.1016/j.ddstr.2011.10.002
- Pico, A. R., Kelder, T., van Iersel, M. P., Hanspers, K., Conklin, B. R., and Evelo, C. (2008). WikiPathways: pathway editing for the people. *PLoS Biol.* 6:e184. doi: 10.1371/journal.pbio.0060184
- Power, A., Berger, A. C., and Ginsburg, G. S. (2014). Genomics-enabled drug repositioning and repurposing: insights from an IOM Roundtable activity. *JAMA* 311, 2063–2064. doi: 10.1001/jama.2014.3002
- Pujol, A., Mosca, R., Farres, J., and Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci.* 31, 115–123. doi: 10.1016/j.tips.2009.11.006
- Raoul, J. L., Van Laethem, J. L., Peeters, M., Brezault, C., Husseini, F., Cals, L., et al. (2009). Cetuximab in combination with irinotecan/5-fluorouracil/folinic acid (FOLFIRI) in the initial treatment of metastatic colorectal cancer: a multicentre two-part phase I/II study. *BMC Cancer* 9:112.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rovithakis, G. A., and Christodoulou, M. A. (1994). Adaptive control of unknown plants using dynamical neural networks. *IEEE Trans. Syst. Man Cybernet.* 24, 400–412. doi: 10.1109/21.278990
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Golub, T. R. et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437.e17–1452.e17. doi: 10.1016/j.cell.2017.10.049
- Sun, Y. L., Kathawala, R. J., Singh, S., Zheng, K., Talele, T. T., Chen, Z. S., et al. (2012). Zafirlukast antagonizes ATP-binding cassette subfamily G member 2-mediated multidrug resistance. *Anticancer Drugs* 23, 865–873. doi: 10.1097/CAD.0b013e328354a196
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., von Mering, C., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Thomas, D., Burns, J., Audette, J., Carrol, A., Dow-Hygelund, C., and Hay, M. (2016). *Clinical development success rates 2006–2015*. San Diego, CA: Biomedtracker; Washington, DC: BIO/Bend: Amplion.
- Tong, S., Wang, T., Li, Y., and Zhang, H. (2014). Adaptive neural network output feedback control for stochastic nonlinear systems with unknown dead-zone and unmodeled dynamics. *IEEE Trans. Cybern.* 44, 910–921. doi: 10.1109/TCYB.2013.2276043
- UniProt, C. (2013). Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* 41, D43–D47. doi: 10.1093/nar/gks1068
- Uray, I. P., and Brown, P. H. (2011). Chemoprevention of hormone receptor-negative breast cancer: new approaches needed. *Recent Results Cancer Res.* 188, 147–162. doi: 10.1007/978-3-642-10858-7\_13
- USFDA. (2016). *Step 3: Clinical Research*. Available online at: <http://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm>
- Wang, Y., Liu, S., Rastegar-Mojarad, M., Wang, L., Shen, F., Liu, F., et al. (2017). “Dependency and AMR embeddings for drug-drug interaction extraction from biomedical literature,” in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (Boston, MA: ACM).
- Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., et al. (2014). PubChem BioAssay: 2014 update. *Nucleic Acids Res.* 42, D1075–D1082. doi: 10.1093/nar/gkt978
- Weir, S. J., Patton, L., Castle, K., Rajewski, L., Kasper, J., and Schimmer, A. D. (2011). The repositioning of the anti-fungal agent ciclopirox olamine as a novel therapeutic agent for the treatment of hematologic malignancy. *J. Clin. Pharm. Ther.* 36, 128–134. doi: 10.1111/j.1365-2710.2010.01172.x
- Weston, A. D., and Hood, L. (2004). Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* 3, 179–196. doi: 10.1021/pr0499693
- Willems, J. (1971). Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Automat. Contr.* 16, 621–634. doi: 10.1109/TAC.1971.1099831
- Woodhead, J. L., Watkins, P. B., Howell, B. A., Siler, S. Q., and Shoda, L. K. (2016). The role of quantitative systems pharmacology modeling in the prediction and explanation of idiosyncratic drug-induced liver injury. *Drug Metab. Pharmacokinet.* 32, 40–45. doi: 10.1016/j.dmpk.2016.11.008
- Xu, H., Aldrich, M. C., Chen, Q., Liu, H., Peterson, N. B., Ruan, X., et al. (2014). Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J. Am. Med. Assoc.* 311:002649. doi: 10.1136/amiajnl-2014-002649
- Yazdannejat, H., Hosseinimehr, S. J., Ghasemi, A., Pourfallah, T. A., and Rafiei, A. (2016). Losartan sensitizes selectively prostate cancer cell to ionizing radiation. *Cell Mol. Biol.* 62, 30–33.
- Yin, T., He, S., Shen, G., Ye, T., Guo, F., and Wang, Y. (2015). Dopamine receptor antagonist thioridazine inhibits tumor growth in a murine breast cancer model. *Mol. Med. Rep.* 12, 4103–4108. doi: 10.3892/mmr.2015.3967
- Yip, C. H., and Rhodes, A. (2014). Estrogen and progesterone receptors in breast cancer. *Future Oncol.* 10, 2293–2301. doi: 10.2217/fon.14.110
- Zheng, C., Guo, Z., Huang, C., Wu, Z., Li, Y., Chen, X., et al. (2015). Large-scale direct targeting for drug repositioning and discovery. *Sci. Rep.* 5:11970. doi: 10.1038/srep11970

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer YW and handling Editor declared their shared affiliation.

Copyright © 2018 Nguyen, Muhammad, Ibrahim, Ma, Guo, Bai and Zeng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.