BIOINFORMATICS

Vol. 27 ISMB 2011, pages i310–i316
doi:10.1093/bioinformatics/btr202

An enhanced Petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data

Guangxu Jin†, Hong Zhao†, Xiaobo Zhou and Stephen T. C. Wong*

Systems Medicine and Bioengineering Department, The Methodist Hospital Research Institute, Weill Medical College, Cornell University, Houston, TX 77030, USA

ABSTRACT

Motivation: Prediction of synergistic effects of drug combinations has traditionally been relied on phenotypic response data. However, such methods cannot be used to identify molecular signaling mechanisms of synergistic drug combinations. In this article, we propose an enhanced Petri-Net (EPN) model to recognize the synergistic effects of drug combinations from the molecular response profiles, i.e. drug-treated microarray data.

Methods: We addressed the downstream signaling network of the targets for the two individual drugs used in the pairwise combinations and applied EPN to the identified targeted signaling network. In EPN, drugs and signaling molecules are assigned to different types of places, while drug doses and molecular expressions are denoted by color tokens. The changes of molecular expressions caused by treatments of drugs are simulated by two actions of EPN: firing and blasting. Firing is to transit the drug and molecule tokens from one node or place to another, and blasting is to reduce the number of molecule tokens by drug tokens in a molecule node. The goal of EPN is to mediate the state characterized by control condition without any treatment to that of treatment and to depict the drug effects on molecules by the drug tokens.

Results: We applied EPN to our generated pairwise drug combination microarray data. The synergistic predictions using EPN are consistent with those predicted using phenotypic response data. The molecules responsible for the synergistic effects with their associated feedback loops display the mechanisms of synergism.

Availability: The software implemented in Python 2.7 programming language is available from request.

Contact: stwong@tmhs.org

1 INTRODUCTION

A combination of drugs, or a drug cocktail, is a common therapeutic strategy used in oncology. The strategy reduces drug toxicology caused by high doses of single drugs and generates enhanced effects with lower doses of combined drugs. A combination of two drugs may generate same effect (*simple additive*), blunted effect (*sub-additive*) and exaggerated effect (*super-additive* or *synergistic*). Drug combination focuses on the relationship between dosages and effects as well as the methodology distinguishing between additive and non-additive combinations. Methods of analysis for distinguishing between simple additivity and other non-additive outcomes include Lowe dose additivity (Chou and Talalay, 1983), Bliss independence (Bliss, 1939), Gaddum's non-interaction (highest single agent) (Berenbaum, 1989) and Potentiation (Lehar *et al.*, 2007, 2009). Among them, Lowe dose additivity is the most common model used in prediction of drug combination. Lowe dose additivity distinguishes additive and non-additive drug combinations by the *combination index* (*CI*), $\mathrm{CI} = C_{a,r}/\mathrm{IC}_{X,a} + C_{b,r}/\mathrm{IC}_{X,b}$ (Chou and Talalay, 1983), where in case of an inhibitory drug, X refers to a specific percent inhibition level (e.g. 50%), $C_{a,r}$ and $C_{b,r}$ are the concentration of drugs A and B given in a combination of the two drugs, and $\mathrm{IC}_{X,a}$ and $\mathrm{IC}_{X,b}$ are the concentration of drugs A and B yielding the same effect level, when treated alone, as the combination. If $\mathrm{CI} > 1$, the drug combination has a subadditive effect, and if $\mathrm{CI} < 1$, the drug combination has a synergistic effect. However, the existing methods for prediction of synergistic effects of drug combinations restrict to the phenotypic response data, and, therefore, cannot be used to describe the underlying mechanisms of signal transductions or signaling pathways.

The studies on signaling pathways and related interaction molecules would make it possible to recognize downstream effect of a drug or a drug combination on the targets. Exogenous signals are initially imposed on the receptors, which then carry the signals to second messengers, and eventually the signals are led to transcription-related molecules of DNA in the nucleus, such as transcription factors, RNA polymerases or histone modification complexes. Many drugs target on the upstream of signaling pathways, for example, membrane receptors or second messengers. They assert a long-term effect, called off-target effect (Keiser *et al.*, 2009; Lamb, 2007; Lamb *et al.*, 2006; MacDonald *et al.*, 2006), on gene transcription-related molecules through the downstream signals along the pathway. Since the studies on drug targets alone are not sufficient to indicate the off-target effects, we need to develop computational models in order to simulate the effects of drugs on the signaling pathways.

Many models have been established for network modeling on biological systems, such as ordinary differential equations (ODE), Fuzzy logic system (FLS) and Petri net (PN). ODE is to encode a network as a system of differential equations, which generates a detailed and biochemically realistic representation. However, the ODE model suffers the cost of many free parameters, which must be estimated. If the number of the nodes of the network reaches tens or hundreds, parameter estimation would become challenging. FLS modeling has the same parameter issue. Comparing with these quantitative models, PN model is a better choice to describe biological systems qualitatively. PN has been recently applied in metabolic (Reddy *et al.*, 1996; Voss *et al.*, 2003), genetic (Steggles *et al.*, 2007) and signaling networks (Ruths *et al.*, 2008) where few accurate kinetic data are accessible.

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The basic PN (Grunwald *et al*., 2008; Hardy and Robillard, 2008; Koch, 2011; Peterson, 1981; Ruths *et al.*, 2008) is composed of four parts: a set of *places* **P**, a set of *transitions* **T**, an *input* function **I** and an *output* function **O**. The input function **I** is a mapping from a transition t_j to a place $\mathbf{I}(t_j)$, known as the input place of the transition. The output function **O** maps a transition t_j to a place $\mathbf{O}(t_j)$, known as the output place of the transition. A PN structure, **C**, is defined as a four-tuple, $\mathbf{C}=(\mathbf{P},\mathbf{T},\mathbf{I},\mathbf{O})$.

2 APPROACH

In this study, we propose a new type of PN model, Enhanced PN (EPN), to predict the synergistic effects of drug combinations using the drug-treated microarray data. We extended the basic PN as multiple-color tokens, places and transitions comparing with the basic one-color PN (Fig. 1). The multiple-color tokens, places and transitions enable EPN to simulate the process in which the drugs impose their effects on molecules.

The goal of EPN is to find an optimal state in which the expression profile of genes exactly matches with the treated case. Once the optimal state is identified, we use the marking of one type of tokens, called effect tokens, to evaluate the effects of the treated drug on the molecules. In this way, we can predict the drug effects on the molecules. To verify whether a drug combination shows synergistic effect, we compare the identified drug effects from the microarray treated by pairwise combination of drugs A and B with those from the corresponding two microarrays treated by drugs A and B separately. Our hypothesis for synergism is that there exists at least one molecule that shows the enhanced effect in the pairwise combination comparing with the summation of the effects generated by the two drugs individually. The molecules with enhanced effects are used to predict whether the combination has synergistic effect.

To test our model, we generate a gene expression profile for breast cancer cell line MCF7 with 16 microarrays for gefitinib with the doses, 0, 5, 10 and 40 μMol and docetaxel with 0, 0.15, 0.6 and 1.2 μMol. The synergistic effect of the two drugs, gefitinib and docetaxel, has been confirmed by Takabatake *et al*. (2007). Then, we apply EPN model to the profile. The analysis shows that EPN predicts the synergistic pairwise drug combinations well. Most of the prediction results are consistent with the prediction results based on the phenotypic response data and the published literature.

Comparing with the existing methods, the advantage of using EPN in prediction of synergism is that it explains how two drugs could generate the synergistic effects. For the synergistic combinations of gefitinib and docetaxel, we found that docetaxel assists gefitinib in the synergy, while the synergistic molecules, such as KRT8, play a leading role in this process. The positive feed-forward loops between EGFR, target of gefitinib and KRT8 enable gefitinib to amplify its effect and dominate the synergism, whereas certain negative feedback loops between BCL2, target of docetaxel and KRT8 reduce its contribution to the synergism.

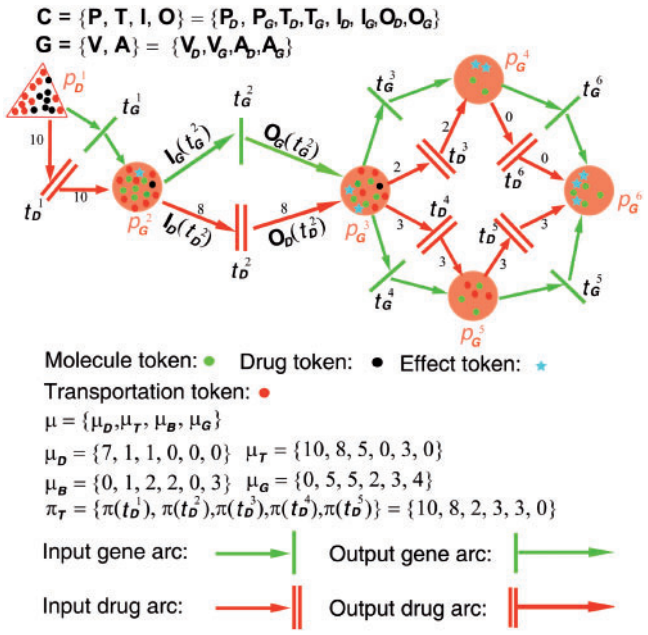

Fig. 1. The EPN model.

3 METHODS

3.1 RNA isolation and gene expression profiling

MCF-7 cells were seeded at 2×10^5/ml in 6-well plates. After confluence was achieved, the cells were treated with different concentrations of the two drugs or Dimethyl sulfoxide (DMSO) in duplicate for 48 h. RNA was extracted from the cells using RNAeasy mini kit (Qiagen). The quality of the RNA samples was monitored using 2100 bioanalyzer (Agilent) before gene expression profiling with Agilent human 4×44 k microarrays.

3.2 EPN model

An EPN structure, **C**, is defined as a four-tuple, $\mathbf{C}=(\mathbf{P},\mathbf{T},\mathbf{I},\mathbf{O})$, where $\mathbf{P}=\{\mathbf{P}_D,\mathbf{P}_G\}$, $\mathbf{P}_D$ is a set for the places of drugs, $\mathbf{P}_G$ is a set of places for molecules, such as genes or proteins, sometimes, called as gene places; $\mathbf{T}=\{\mathbf{T}_D,\mathbf{T}_G\}$, $\mathbf{T}_D$ is a set of transitions for drug tokens, $\mathbf{T}_G$ is a set of transitions for gene tokens; $\mathbf{I}=\{\mathbf{I}_D,\mathbf{I}_G\}$, $\mathbf{I}_D$ is an input mapping between a drug transition t_D^j and the place $\mathbf{I}(t_D^j)$, $\mathbf{I}_G$ is an input mapping between a gene transition t_G^j and the place $\mathbf{I}(t_G^j)$; $\mathbf{O}=\{\mathbf{O}_D,\mathbf{O}_G\}$, $\mathbf{O}_D$ is an output mapping between a drug transition t_D^j and the place $\mathbf{O}(t_D^j)$, $\mathbf{O}_G$ is an output mapping between a gene transition t_G^j and the place $\mathbf{O}(t_G^j)$.

An EPN graph has four types of nodes. A circle ○ represents a gene place, a triangle **Δ** represents a drug place, a bar | represents a gene transition and a bar || represents a drug transition (Fig. 1). The arcs $\mathbf{A}_D$ and $\mathbf{A}_G$ have two colors, $\mathbf{A}_D$ is for drugs and $\mathbf{A}_G$ is for genes. Green arcs $\mathbf{A}_G$ transit gene tokens from input place to transitions or transit gene tokens from transitions to output place. Red arcs $\mathbf{A}_D$ transit drug tokens from input place $\mathbf{I}(t_D^j)$ to transition t_D^j or transit drug tokens from transitions t_D^j to output place $\mathbf{O}(t_D^j)$. EPN has four types of tokens, that is, gene tokens, drug tokens, effect tokens (blasting tokens) and drug transportation tokens. A marking μ of an EPN, $\mathbf{C}=(\mathbf{P},\mathbf{T},\mathbf{I},\mathbf{O})$, is a function from the set of places **P** to the non-negative integer space, $\mathbf{N}^4, \mu:\mathbf{P}\to\mathbf{N}^4$. The marking $\mu=(\mu_D,\mu_T,\mu_B,\mu_G)$ can also be defined as a $4n$-vector, $\mu=(\mu_D^1,\mu_D^2,\ldots,\mu_D^n,\mu_T^1,\mu_T^2,\ldots,\mu_T^n,\mu_B^1,\mu_B^2,\ldots,\mu_B^n,\mu_G^1,\mu_G^2,\ldots,\mu_G^n)$, where $n=|\mathbf{P}|$ and each $\mu_D^i\in\mathbf{N}$, $\mu_T^i\in\mathbf{N}\mu_B^i\in\mathbf{N}$, $\mu_G^i\in\mathbf{N}$, $i=1,2,\ldots,n$. Another making π_T of the $\mathbf{T}_D$ is a function from the set of drug transitions $\mathbf{T}_D$ to the non-negative integers **N**, $\pi:\mathbf{T}\to\mathbf{N}$. The marking π_T can be defined as an m-vector, $\pi=(\pi_T^1,\pi_T^2,\ldots,\pi_T^m)$, where $m=|\mathbf{T}_D|$ and each $\pi_D^i\in N$, $i=1,2,\ldots,m$.

EPN executes by *firing* of transitions and *blasting* of drug tokens. A drug transition fires by removing drug tokens from its input places and creating new drug tokens which are distributed to its output places. A drug transition may fire if it is *enabled*. A drug transition $t_j\in\mathbf{T}_D$ in a marked EPN $\mathbf{C}=(\mathbf{P},\mathbf{T},\mathbf{I},\mathbf{O})$ with marking μ is enabled if for all $p_i\in\mathbf{P}$, $\mu_D(p_i)\geq\#(p_i,\mathbf{I}_D(t_j))$,

where $\#(p_i, \mathbf{I}_D(t_j))$ is the token number transited from p_i to $\mathbf{I}_D(t_j)$. Firing an enabled transition t_j results in a new marking μ' defined by $\mu'_D(p_i) = \mu_D(p_i) - \#(p_i, \mathbf{I}_D(t_j))$ $\mu'_D(\mathbf{O}_D(t_j)) = \mu_D(\mathbf{O}_D(t_j)) + \#(p_i, \mathbf{O}_D(t_j))$, $\mu'_T(p_i) = \mu_T(p_i) + \#(p_i, \mathbf{I}_D(t_j))$, and $\pi'_T(t_j) = \pi_T(t_j) + \#(p_i, \mathbf{I}_D(t_j))$, where the markings $\mu'_T(p_i)$ and $\pi'_T(t_i)$ can record the number of drug tokens that passed by the place p_i and transition t_j. In the same way, a gene transition fires by removing gene tokens from its input places and creating new gene tokens which are distributed to its output places. A gene transition may fire if it is *enabled*. A gene transition $t_j \in \mathbf{T}_G$ in a marked EPN $\mathbf{C} = (\mathbf{P}, \mathbf{T}, \mathbf{I}, \mathbf{O})$ with marking μ is enabled if for all $p_i \in \mathbf{P}$, $\mu_G(p_i) \geq \#(p_i, \mathbf{I}_G(t_j))$. Firing an enabled transition t_j results in a new marking μ' defined by $\mu'_G(p_i) = \mu_G(p_i) - \#(p_i, \mathbf{I}_G(t_j))$ and $\mu'_G(\mathbf{O}_G(t_j)) = \mu_G(\mathbf{O}_G(t_j)) + \#(p_i, \mathbf{O}_G(t_j))$.

Blasting only executes on the drug tokens in the output places. The effect tokens are also called as bombs. After blasting in the output place, a bomb can destroy b gene tokens and the bomb is consumed. A blasting may occur if it is *enabled*. A blasting of B bombs in the output place p_j of a marked PN $\mathbf{C} = (\mathbf{P}, \mathbf{T}, \mathbf{I}, \mathbf{O})$ with marking μ is enabled if $\mu_D(p_j) \geq B$ and $\mu_G(p_j) \geq B \times b$. Blasting B bombs in a place p_j results in a new marking μ' defined by $\mu'_D(p_j) = \mu_D(p_j) - B$, $\mu'_G(p_j) = \mu_G(p_j) - B \times b$, and $\mu'_B(p_j) = \mu_B(p_j) + B$.

The state of EPN is defined by its gene marking μ_G. The firing of a transition and blasting of bombs represent a change in the state of the EPN by a change in the gene marking of the EPN. The state space of an EPN with n places is the set of all gene markings, that is, $\mathbf{N}^n$. The change in state caused by firing a transition and blasting a place is defined by two change functions δ_T and δ_B called the next-state functions. Since t_j can be fired only if it is enabled, $\delta_T(\mu, t_j)$ is undefined if t_j is not enabled in marking μ, If t_j is enabled, then $\delta_T(\mu, t_j) = \mu'$, where μ' is the marking which results from removing tokens from the input of t_j and adding tokens to the outputs of t_j. If blasting in place p_j is enabled, $\delta_B(\mu, p_j)$ can be defined, which satisfies that $\delta_B(\mu, p_j) = \mu''$, where μ'' is the marking which results from the consuming of bombs and reducing tokens from the output place p_j.

3.3 Apply EPN to identify the drug effects on molecules

The firing of transitions and blasting of bombs result in the change of gene marking μ_G. The state of EPN is defined by its gene marking μ_G. The task for EPN is to mediate the state, $\mu_G^{(C)}$, before treatment to that, $\mu_G^{(D)}$, after the treatment of a drug. It is formulated as an optimization problem in (1). The hypothesis is that the final state $\mu_G^{(D)}$ is a reachable state from the control state $\mu_G^{(C)}$. There exists a transition set $\{\mathbf{T}^0, \mathbf{T}', \mathbf{T}'', \ldots, \mathbf{T}^{(k-1)}\}$ and a change function set $\{\delta^0, \delta'', \delta''', \ldots, \delta^{(k-1)}\}$ that satisfies $\delta^0\left(\mu_G^{(C)}, \mathbf{T}^0\right) = \mu'_G, \ldots, \delta^{(k-1)}\left(\mu_G^{(k-1)}, \mathbf{T}^{(k-1)}\right) = \mu_G^{(k)} = \mu_G^{(D)}$. The EPN outputs the simulated drug effects on molecules.

Simulating of drug effects using EPN can be described as follows:

$$
\begin{aligned}
&\mathbf{min} \left| \mu_{\mathrm{G}}^{(\mathrm{k})} - \mu_{\mathrm{G}}^{(\mathrm{D})} \right| \\
&\text{s.t.} \quad \delta^{(i-1)}\left(\mu_G^{(i-1)}, \mathbf{T}^0\right) = \mu_G^{(i)}, \\
&\qquad \mu_G^{(i)} = f^{(i)}(\mathbf{P}), \\
&\qquad \pi^{(j)} = g^{(j)}(\mathbf{T}^j), \\
&\qquad \mu_G^{(i)} \in N^{4n} \\
&\qquad \pi^{(j)} \in N^m \\
&\qquad i = 1, 2, \ldots, k \\
&\qquad j = 1, 2, \ldots, m
\end{aligned}
\tag{1}
$$

Input: $\mu_G^{(0)}$: gene expressions before treatment;
$\mu_G^{(n)}$: gene expressions after treatment;
$\aleph$: downstream signaling network of targets;
Output: $\mu_B^{(n)}$: drug effects on the molecules.

3.4 Transition algorithm

To optimize (1), we proposed a greedy transition algorithm, called Final State Oriented Token Transition. The transitions of drug tokens and gene tokens are implemented by the algorithm. The steps in the algorithm are: (1) shuffling of the interactions in the downstream targets of the signaling network, (2) transition of drug or gene tokens, and (3) decision of termination. The pseudo code is described as following (2)

Step 1. Shuffle all interactions as Π.
Step 2. Token transition

```
for every interaction (G1 → G2) in Π{
G2^fold-change = μ_G^Treatment(G2) / μ_G^δ(G2)
if (G2^fold-change < 1){
    transit drug tokens
    blast effect tokens
    update markings of drug, gene, effect, transportation
    tokens}
if (G2^fold-change > 1){
    transit gene tokens
    update markings of gene tokens}
}
```

(2)

Step 3. Termination

$\Delta^{\delta} = \sum_i |1\text{-}G_i^{\delta}/G_i^{\text{Treatment}}|$, $\Delta^{\text{Control}} = \sum_i |1\text{-}G_i^{\text{Control}}/G_i^{\text{Treatment}}|$
if $(\Delta^{\delta}/\Delta^{\text{Control}} < 10^{-10})$ {Exit}, else go back to *Step* 2.

4 RESULTS

We developed an EPN model to predict synergistic effects of pairwise drug combinations using drug-treated gene expression microarray data. To test the EPN model, we generated a gene expression profile with 16 combinations of two drugs with different dosage pairs. The drug effects on molecules identified by EPN model were used to predict the synergistic effects of the combinations.

4.1 Drug combinations for EPN

Sixteen pairwise drug combinations for gefitinib with doses 0, 5, 10 and 40 μMol and docetaxel with 0, 0.15, 0.6 and 1.2 μMol, are treated on MCF7 breast cancer cell line. We generated the phenotypic response data, that is, the inhibition rates on cell proliferation, and the treatment of microarray data simultaneously. Applying CalcuSyn software (Tallarida, 2001) on the phenotypic response data, we identified that the combinations of (5, 1.2), (20, 0.15), (20, 0.6), (20, 1.2), (40, 0.15), (40, 0.6) and (40, 1.2) exhibit synergistic effects. The results are shown in Table 1. Takabatake *et al.* (2007) also revealed that gefitinib with dosage 19 ± 2.4 μMol and docetaxel with dosage 0.29 ± 0.042 μMol are synergistic.

4.2 Downstream signaling network of the targets

Before implementation of EPN on the drug-treated microarray data of pairwise drug combinations, we have to identify the downstream signals and related protein–protein interactions (PPIs) of the targets, i.e. epidermal growth factor receptor (EGFR) of gefitinib and B-cell lymphoma 2 (BCL2) of docetaxel. The combined data for signaling pathways were derived from three pathway databases, including KEGG: Kyoto Encyclopedia of Genes and Genomes (Kanehisa *et al.*, 2004), NCI PID: Pathway Interaction Database (Schaefer *et al.*, 2009) and BioCarta (Schaefer *et al.*, 2009). The PPIs used here are the physical interactions gathered from five PPI databases, IntAct (Kerrien *et al.*, 2007), DIP (Xenarios *et al.*, 2002), MINT (Chatr-aryamontri *et al.*, 2007), MIPS (Mewes *et al.*, 2002) and BioGrid

Table 1. The CI identified by phenotypic responses

Doses of gefitinib (μMol)	Doses of docetaxel (μMol)	CI
5	0.15	3.361
5	0.6	5.121
5	1.2	0.894
20	0.15	0.610
20	0.6	0.687
20	1.2	0.664
40	0.15	0.672
40	0.6	0.753
40	1.2	0.619

CI used here is developed by CalcuSyn. If $CI < 0.7$, strong synergism or synergism; if $0.7 < CI < 0.9$, moderate or slight synergism; and if $CI > 0.9$, nearly additive or antagonism.

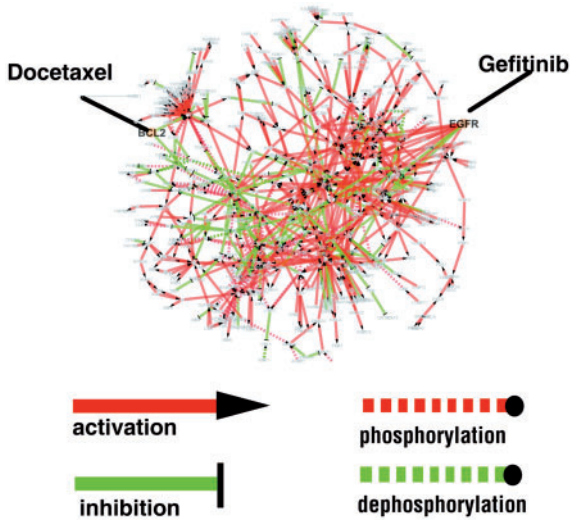

Fig. 2. The downstream signaling network for targets of the pairwise combination of gefitinib and docetaxel. The related PPI data are not shown.

(Breitkreutz *et al.*, 2008). The identified downstream signaling network comprises 575 proteins with 919 signaling interactions, as shown in Figure 2, which interacts with another 433 proteins via 4365 PPIs.

4.3 Apply EPN to the pairwise-combination microarray data

The task of EPN is to figure out those molecules with enhanced effects by comparing the effects of pairwise combinations with those effects of the corresponding single drugs. It involves three steps: (i) initiate the EPN model, (ii) detect the effects on the molecules of the pairwise combinations and of two single drugs, and (iii) identify the molecules with enhanced effects.

4.3.1 Initiate the EPN model To construct an EPN, we have to add transitions and tokens in the identified downstream signaling network of the targets of the combinations. The signaling molecules with their interacting protein molecules in Figure 2 are considered as the gene places in EPN while the marking of the gene places is determined by the expression values in the treatment microarray data. The drugs are considered as drug places, and the marking of their drug tokens is initiated as relative large numbers, which ensures their related transitions are enabled. The transitions for drugs and genes were assigned to the interactions between the places. If the interaction $(p_1 \rightarrow p_2)$ is activation or phosphorylation or drug $\rightarrow$ target, the new interactions were defined as from p_1 to transition, $p_1 \rightarrow T$ and from transition to p_2, $T \rightarrow p_2$. If the interaction $(p_1 \rightarrow p_2)$ is inhibition or dephosphorylation, the new interactions were defined reversely, from p_2 to transition, $p_2 \rightarrow T$ and from transition to p_1, $T \rightarrow p_1$.

As described in transition algorithm of (2), we randomly selected an interaction from the downstream signaling network of the targets, then determine to transit either drug tokens or gene tokens. The number of gene tokens for every transition is a randomized number μ_{Gamma} derived from a Gamma distribution,

$$f(x;k,\vartheta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k \Gamma(k)}, \quad x \geq 0 \text{ and } k,\vartheta > 0 \tag{3}$$

where k, ϑ are two parameters for scale and shape of the Gamma distribution, and $\Gamma(k) = (k-1)!$. Implementing the Final State Oriented Token Transition algorithm of (2), we do not expect the final state (treated state) is achieved immediately by transition of a relatively large number of tokens, which will cause a high deviation for the identified numbers of gene or drugs tokens by running the algorithm for 1000 times. Instead, we selected a large shape parameter ($k = 10$) and a small scale parameter ($\theta = 1$). The large shape parameter enables the Gamma distribution to have a relatively high probability for small numbers while the small-scale parameter determines that only a relatively small range of numbers would have higher probabilities.

The transition of drug tokens is forced by the output place. The number of drug tokens for every transition is determined by two parts, one is the number of drug tokens in the output place that have been transited by its downstream interactions, and another is a randomized number also derived from (3) to indicate how many drug tokens should be transited additionally. The markings for gene tokens, drug tokens, transportation drug tokens and transitions are updated with the transitions.

In contrast to conventional PNs, EPN has not only the action of firing, that is, transitions of tokens, but also the action of blasting. The role of blasting is to consume drug tokens as effect tokens and to reduce the number of gene tokens in the output place simultaneously. The markings for gene tokens, drug tokens and effect tokens are updated with blasting. Here, a blasting parameter b is needed to indicate how many gene tokens can be destroyed by one effect token. We set b as 1. The identified effect tokens by blasting are used to denote the drug effects on the molecules.

4.3.2 Detect the effects on the molecules We implemented EPN on every combination microarray data for 1000 times. A list of effect token markings is generated, $\Lambda = [\mu_{B,1}, \mu_{B,2}, \ldots, \mu_{B,1000}]$. For every place or gene p^i, the drug effect E^{p^i} is defined as

$$E^{p^i} = \frac{1}{1000}\sum_j \mu_{B,\,\mathrm{j}}^{p^i} \tag{4}$$

4.3.3 Find the molecules with enhanced effects The synergistic effects of drug combinations are evaluated by the identified drug

effects on molecules. A pairwise combination (a,b) denotes the treatment achieved by gefitinib with dose a and docetaxel with dose b. If the combination generates a synergistic effect, there exists at least one gene p^i satisfied with $E^{p^i}_{(a,b)} > E^{p^i}_{(a)} + E^{p^i}_{(b)}$, where $E^{p^i}_{(a,b)}$, $E^{p^i}_{(a)}$ and $E^{p^i}_{(b)}$ are the drug effects on gene p^i generated by combination(a,b), gefitinib with dose a, and docetaxel with dose b. We defined

$$S^{p^i} = E^{p^i}_{(a,b)} - (E^{p^i}_{(a)} + E^{p^i}_{(b)}) \quad (5)$$

The genes with $S^{p^i} > 0$ were filtered out as synergistic markers for combination (a, b).

We designed a random process to evaluate to what degree a combination shows synergistic effect. We randomly generated the treatment expression data for the pairwise combinations and single drugs. The expression of each gene was randomly selected from the 16 expression values of this gene in the treatment microarrays with 16 dose pairs. We repeated the random process for a thousand times and generated 1000 randomized microarrays. To evaluate the significance of synergistic effects, we randomly chose three microarray data for pairwise combination (a, b) and single drugs from the generated randomized microarrays so that we could compute the randomized effect on every place. Thus, we found a randomized distribution f for the total synergistic effects $\sum_i S^{p^i}$ of a combination (a, b). The P-value is derived from the complementary cumulative distribution function (ccdf) of f.

We applied EPN model to our generated combination microarrays. Our results are consistent with the prediction results based on phenotypic response data and the published results in (Takabatake *et al.*, 2007). EPN model not only indicates which combination has synergistic effects but also shows which molecules result in the synergistic effects. The prediction results are shown in Table 2.

Comparing with the prediction results based on the phenotypic response data, we confirmed that the pairwise combinations (5, 1.2), (20, 0.6), (20, 1.2) and (40, 1.2) have synergistic effects, in which (20, 1.2) shows the strongest synergistic effect. The other three combinations, (20, 0.15), (40, 0.15) and (40, 0.6) are abandoned by the randomized analysis although they have the genes with enhanced effects ($0 < S^{p^i} < 50.1$). Our results are consistent with the published results in (Takabatake *et al.*, 2007), which indicate the efficacious dose $19 \pm 2.4\mu$Mol of gefitinib and $0.29 \pm 0.042\mu$Mol of docetaxel. We also predict the combination (5, 0.6) has synergistic effect.

Table 2. The predicted effects on molecules for combinations

Combs(a,b)	P-value	Genes with significantly higher synergistic effects	
		Gene P^i	Synergistic effect S^{p^i}
(5, 0.15)	0.85		
(5, 0.6)	$<\mathbf{10}^{-2}$	NFKBIA	646.15
(5, 1.2)	$<\mathbf{10}^{-3}$	NFKBIA	1020.46
		PXN	51.39
(20, 0.15)	0.32		
(20, 0.6)	**0.03**	NFKBIA	280.39
(20, 1.2)	$<\mathbf{10}^{-10}$	KRT8	3248.89
		KRT18	2229.01
		ACTB	1563.20
		STMN1	862.19
		HMGN2	628.15
		CDH1	482.05
		SREBF1	254.53
		CDC2	223.16
		RAC1	163.42
		PCNA	113.11
		CCND1	107.30
		SYN1	89.33
		RRAS	89.12
		GYS1	80.93
		RHOA	65.68
		TRADD	56.81
		HLA-A	54.14
(40, 0.15)	0.44		
(40, 0.6)	0.053		
(40, 1.2)	$<\mathbf{10}^{-2}$	NFKBIA	691.81

The application of EPN to randomized gene expression microarray data indicates a cutoff for S^{p^i}, i.e. 50.1. The genes with $S^{p^i} > 50.1$ are shown. The bold values in the P-value column indicates the P-value is significant ($P < 0.01$).

4.4 Further understanding of synergism

The conventional prediction methods based on phenotypic response data only indicates whether two drugs are synergistic, but cannot explain how the two drugs derive the synergistic effect. EPN fills this void.

4.4.1 How is synergism generated? Conventional studies on synergism aim to find the combination whose combined dosage is lower than dosages of individual drugs. The CI was defined as (Chou and Talalay, 1983):

$$\mathrm{CI} = C_{a,r}/\mathrm{IC}_{X,a} + C_{b,r}/\mathrm{IC}_{X,b} \quad (6)$$

where in case of an inhibitory drug, X refers to a specific percent inhibition level (e.g. 50%), $C_{a,r}$ and $C_{b,r}$ are the concentrations of drugs A and B given in a combination of the two drugs, and $\mathrm{IC}_{X,a}$ and $\mathrm{IC}_{X,b}$ are the concentrations of drugs A and B yielding the same effect level, when treated alone, as the combination.

If CI<1 and $R = \mathrm{IC}_{X,a}/\mathrm{IC}_{X,\mathrm{b}}$ is defined as the relative potency of the two drugs, (6) is reformatted as

$$C_{a,r} + R \cdot C_{b,r} < \mathrm{IC}_{X,a} \text{ or } C_{a,r}/R + C_{b,r} < \mathrm{IC}_{X,b} \quad (7)$$

From (6) and (7), we can see that the explanation of synergism is based on the doses used for achievement of a specific effect level in the treatments of combinations and single drugs. It is still difficult to explain how the synergistic effects are exactly generated.

The drug tokens in drug places can be used to explain how the synergistic effects are generated. In EPN, one type of tokens, that is, drug transportation tokens, record how many drug tokens have entered into the downstream signaling network of the targets. By comparing the transportation tokens of μ_T in drug places for combinations and their corresponding single drugs, we found that the synergistic effects are dominated by gefitinib, while docetaxel assists gefitinib to generate the enhanced drug effects (Table 3). The analysis indicates that two pairwise combinations (5, 1.2) and (20, 1.2) are the most confident combinations with synergistic effects. These two combinations were also confirmed by the analysis on phenotypic response data (Table 1). The combination (5, 1.2) is also comparable with the results in (Takabatake *et al.*, 2007),

Table 3. The predicted effects for gefitinib and docetaxel in single and combination treatments

Doses (μMol)		Effects			
		Single drug		Combination	
Gefitinib	Docetaxel	Gefitinib	Docetaxel	Gefitinib	Docetaxel
5	0.15	22 819	150 841	112 439	64 102
5	0.6	22 819	91 527	59 973	39 987
5	**1.2**	22 819	105 586	90 328	50 503
20	0.15	50 728	150 841	89 474	49 301
20	0.6	50 728	88 472	75 844	34 687
20	**1.2**	50 728	104 450	136 433	77 516
40	0.15	118 677	150 841	117 948	67 537
40	0.6	118 677	88 472	97 426	54 183
40	1.2	118 677	104 450	87 338	54 436

The bold value indicates that (i) the dose-pair has a significant *P*-value in Table 2 and (ii) total of the predicted effects of the two drugs in combinations are higher than that in single drugs.

the combination (19 ± 2.4, 0.29 ± 0.042). The combination (5, 1.2) shows that docetaxel assists gefitinib with 5 μMol to generate a higher effect, 90 328, than that of 20 μMol, 50 728, by using a higher dose of 1.2 μMol, instead of 0.3 μMol.

4.4.2 How does the combination achieve the synergism? The combination (20, 1.2) is used to illustrate the molecular mechanisms for the synergistic effects. In EPN, we defined a marking π_T of the drug transition $\mathbf{T}_D$, which records how many drug tokens are transited between gene places by drug transitions. The transition marking, π_T, helps to find the pathways by which the drug tokens are transited from drug places to the synergistic molecules as shown in Table 2.

We chose KRT8 (CK8, keratin 8) from the list of synergistic molecules as an example to show how docetaxel assists gefitinib to achieve the synergistic effects. KRT8 is a member of the type II keratin family clustered on the long arm of chromosome 12. Type I and type II keratins heteropolymerize to form intermediate-sized filaments in the cytoplasm of epithelial cells. The product of this gene typically dimerizes with keratin 18 to form an intermediate filament in simple single-layered epithelial cells. This protein plays a role in maintaining cellular structural integrity and also functions in signal transduction and cellular differentiation. KRT8 with another synergistic molecule KRT18 has been revealed as prognostic biomarkers for the patients of triple negative breast cancers (Williams *et al.*, 2009) and invasive breast cancer (Takei *et al.*, 1995).

Using the transition marking, π_T, we filtered out the transitions T' whose $\pi_{T'} > \mathbf{0}$. These identified transitions facilitate us to chase the drug effects from the two combined drugs to the interested synergistic molecule, KRT8. The transitions are also helpful to understand the mechanisms of the synergistic effects. We showed the identified transitions in Figure 3.

The synergistic effect of (20, 1.2) is elucidated by the identified transitions of the treatment pathway in Figure 3. The reason for the synergism is that the two drugs have many common output places that are, for example, KRT8, the locations for the two drugs to exchange their effects and generate the synergistic effects. Besides

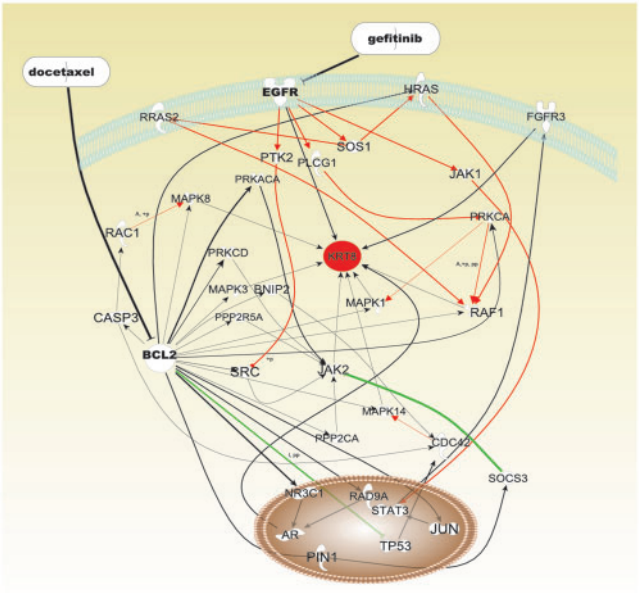

Fig. 3. The pathway for getifinib and docetaxel to generate the synergistic effects on KRT8. The red interactions are for activation or phosphorylation between signaling molecules, while the green ones denote inhibition or dephosphorylation between signaling molecules. And the black interactions are for drug–target relationships and the simulated pathway interactions from PPIs. The pathway was generated using IPA software (Ingenuity Systems, Inc, Redwood City, CA, USA).

this, the transitions in Figure 3 also explain why docetaxel assists gefitinib to generate the synergistic effect. The target of gefitinib, EGFR, has a direct effect on KRT8 through the transition, $EGFR \rightarrow KRT8$, whereas the target of docetaxel, BCL2, does not. In addition, gefitinib has a number of positive feed-forward loops through RAS pathway, for instance, $EGFR \rightarrow SOS1 \rightarrow RRAS2 \rightarrow RAF1 \rightarrow KRT8$, $EGFR \rightarrow SOS1 \rightarrow HRAS2 \rightarrow RAF1 \rightarrow KRT8$, and $EGFR \rightarrow PLCG1 \rightarrow PRKCA \rightarrow RAF1 \rightarrow KRT8$. The feed-forward loops in accelerating transductions of signals have been extensively studied (Milo *et al.*, 2002; Shen-Orr *et al.*, 2002). They are, here, considered as the important paths for accelerating gefitinib to have an effect on KRT8. In contrast, besides positive feedback loops, docetaxel has many negative feedback loops to KRT8 caused by two inhibitory interactions, $BCL2 - |TP53$ and $SOCS3 - |JAK2$. The positive feed-forward loops for gefitinib and the negative feedback loops for docetaxel together help us to understand why gefitinib dominates the synergistic effects and the role of docetaxel is to assist gefitinib to generate such effects.

5 DISCUSSION

In this article, we developed a new model, EPN, to predict the synergistic effects of pairwise drug combinations using drug-treated gene expression microarray data. The effects of drugs on molecule expressions and the associated pathways are simulated by the defined tokens and transitions in different colors in the model. Synergism is predicted using enhanced synergistic molecules recognized by EPN. The activated drug transitions help to interpret the mechanisms of the synergistic effects of the combinations.

Table 4. The phenotypic response data for the pairwise drug combinations

Docetaxel (μMol)	Gefitinib (μMol)					
	0	2.5	5	10	20	40
0	100.00	103.16	94.19	92.36	81.37	64.57
0.15	87.94	108.54	91.92	87.33	67.75	55.39
0.3	83.98	105.46	88.21	84.14	63.61	55.03
0.6	72.68	95.56	73.31	67.25	60.87	44.29
1.2	74.48	90.14	68.35	64.73	58.09	47.83
2.4	75.39	96.33	70.41	68.72	62.78	42.08

The application of EPN to the microarray profile for gefitinib and docetaxel not only predicts the synergistic dose pairs but also illustrates the mechanism for the synergistic effects. We identified that gefitinib dominates the synergistic effects while docetaxel assists gefitinib to achieve that. The results are also confirmed by the phenotypic response data as shown in Table 4. We tested another two dosages, 2.5 and 10 μMol, for genfitinib and another two dosages, 0.3 and 2.4 μMol, for docetaxel in the response data. We could see that the proliferation inhibition rates are significantly decreased along with the increase of dosage of gefitinib, and the tendency is, however, not with docetaxel.

The molecule, KRT8, helps us to make it clear how the two drugs generate the synergistic effect. While the drugs treat the cells separately, gefitinib is known to inhibit the Ras pathway by its target, EGFR, so that the drug prevents the proliferation of cells. On the other hand, docetaxel binds to another target, TUBB1, instead of BCL2, which stabilizes microtubules and prevents depolymerization from calcium ions, decreased temperature and dilution, preferentially at the plus end of the microtubule. For the mechanisms of the synergistic combination of two drugs, it remains unclear. Our analysis indicates that gefitinib still uses the target EGFR and its downstream Ras pathway, while docetaxel alternatively imposes its effects on the target, BCL2, so that it helps gefitinib generate the synergistic effects on KRT8. KRT8 is also known as cytokeratin-8 (CK-8) or keratin-8 (K8) is a keratin protein that in human is encoded by the *KRT8* gene. It is overexpressed in the patients of breast cancer. So it has been a prognostic biomarker for the patients of triple negative breast cancers (Williams *et al*., 2009) and invasive breast cancer (Takei *et al*., 1995). Therefore, it is the synergistic molecules, such as KRT8, to help gefitinib and docetaxel to generate the synergistic effects.

The proposed method of EPN is suitable for studying the target-therapy drugs. The model works best if the targets with their downstream signaling pathways are already known. Otherwise, one has to ask for assistance from other types of computational methods to simulate the downstream signaling network for the targets, for example, drug–target identification or pathway simulation based on Bayesian network, Boolean network or other gene interaction network using the microarray or protein array data from the patients tissues or disease cell lines.

ACKNOWLEDGEMENTS

G.J. and X.Z. proposed the idea for EPN model. G.J. developed the codes for EPN model. G.J. and Z.H. designed the combination experiments and Z.H. conducted the wet lab experiments. S.W. initiated and supervised this work. G.J. and S.W. wrote the paper.

Funding: National Institutes of Health (U54CA149196); John S Dunn Research Foundation.

Conflict of Interest: none declared.

REFERENCES

Berenbaum,M.C. (1989) What is synergy? *Pharmacol. Rev.*, **41**, 93-141.

Bliss,C.I. (1939) The toxicity of poisons applied jointly. *Ann. Appl. Biol.*, **26**, 585-615.

Breitkreutz,B.J. *et al.* (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637-D640.

Chatr-aryamontri,A. *et al.* (2007) MINT: the Molecular INTeraction database, *Nucleic Acids Res*, **35**, D572-574.

Chou,T.-C. and Talalay,P. (1983) Analysis of combined drug effects: a new look at a very old problem. *Trends Pharmacol. Sci.*, **4**, 450-454.

Grunwald,S. *et al.* (2008) Petri net modelling of gene regulation of the Duchenne muscular dystrophy. *Biosystems*, **92**, 189-205.

Hardy,S. and Robillard,P.N. (2008) Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, **24**, 209-217.

Kanehisa,M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277-D280.

Keiser,M.J. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature*, **462**, 175–181.

Kerrien,S. *et al.* (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.

Koch,I. *et al.* (2011) *Modeling in Systems Biology: The Petri net Approach (Computational biology)*. Springer, New York.

Lamb,J. (2007) The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer*, **7**, 54–60.

Lamb,J. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929-1935.

Lehar,J. *et al.* (2009) Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat. Biotechnol.*, **27**, 659-666.

Lehar,J. *et al.* (2007) Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.*, **3**, 80.

MacDonald,M.L. *et al.* (2006) Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat. Chem. Biol.*, **2**, 329-337.

Mewes,H.W. *et al.* (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.

Milo,R. *et al.* (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824-827.

Peterson,J. (1981) *Petri Net Theory and the Modeling of Systems*. Prentice Hall, Englewood Cliffs.

Reddy,V.N. *et al.* (1996) Qualitative analysis of biochemical reaction systems. *Comput. Biol. Med.*, **26**, 9-24.

Ruths,D. *et al.* (2008) The signaling petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks. *PLoS Comput. Biol.*, **4**, e1000005.

Schaefer,C.F. *et al.* (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.*, **37**, D674-D679.

Shen-Orr,S.S. *et al.* (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, **31**, 64-68.

Steggles,L.J. *et al.* (2007) Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics*, **23**, 336-343.

Takabatake,D. *et al.* (2007) Tumor inhibitory effect of gefitinib (ZD1839, Iressa) and taxane combination therapy in EGFR-overexpressing breast cancer cell lines (MCF7/ADR, MDA-MB-231). *Int. J. Cancer*, **120**, 181-188.

Takei,H. *et al.* (1995) Immunohistochemical analysis of cytokeratin #8 as a prognostic factor in invasive breast carcinoma. *Anticancer Res.*, **15**, 1101-1105.

Tallarida,R.J. (2001) Drug synergism: its detection and applications. *J. Pharmacol. Exp. Ther.*, **298**, 865-872.

Voss,K. *et al.* (2003) Steady state analysis of metabolic pathways using Petri nets. *In Silico Biol.*, **3**, 367-387.

Williams,D.J. *et al.* (2009) Triple-negative breast carcinoma in women from Vietnam and the United States: characterization of differential marker expression by tissue microarray. *Hum Pathol.*, **40**, 1176-1181.

Xenarios,I. *et al.* (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303–305.