Design the PETS (Pharmacology-Effect-on-Targets Simulator) algorithm

We applied the ‘transmission’ concept in the PETS algorithm, where the drug’s triggering signal transmits between nodes in the network. Here, we modeled the updating mechanism and transmission distribution similarly to PageRank, which has been previously applied and modified in molecular-network-based solutions for diseases biomarker identification[51, 52]. Since regulations in bio-molecular networks include both activation and inhibition, the signal was modeled as either positive or negative values, accordingly. The final drug-disease score, or PETS score, was measured by summarizing the overall effect of the drug to the pathway. The hypothesis is that a therapeutic drug’s molecular signature should reverse molecular expressions from the disease condition to a normal condition on the pathway level [11].

• Annotation

- ***z***: a specific disease or a disease subtype.

**- M**: pathway model. **M***z*: the specific pathway model for disease *z*. When there is only one disease mentioned or the disease is given, we use **M** to simplify the annotation.

- ***p***: molecule.

- **M*z*(*pi*, *pj*)**: the regulation from molecule *i* to molecule *j* in the pathway model of disease ***z***. When there is no ambiguity about the given disease, we use **M**(*i*, *j*) to simplify the annotation

- **Rp*z*(*pi*)**: significance of molecule *i* in disease *z*’s pathway model. Similar to the previous annotation, we use Rp(*i*) to annotate the same content at a given disease..

- **ZP*z*(*pi*)**: the disease expression of molecule *i* (from the pathway model) at disease *z* condition. When there is no disease ambiguity, we use the simplified annotation ZP(*i*).

**-ZP***z*: a vector showing all expression of all molecular in disease z’s pathway model. The simplified annotation is **ZP**.

- ***D***: drug

- **PD(*D*, *pi*)**: HD’s score for the indication from drug *d* to molecule *i*. We use PD(*i*) for the same context when there is no ambiguity about the drug.

- **S*z*(*D*, *pi*)**: predicted drug molecule score using disease *z*’s pathway model, drug *D* and molecule *i*. If there is no ambiguity about the disease, we use the simpler annotation S(*D*, *pi*). If there is no ambiguity about both the drug and the disease, we use the simplest annotation S(*i*) for the same context.

From the iterative nature of the PETS algorithm, given both the disease and the drug, we annotate S(i, k) as the predicted drug molecule score at the kth iteration. S(i) is the final predicted drug molecule score after the iterative process terminates.

- **PETS(*D*, *z*)**: the pharmacological effect on target score. The simplified annotations are PETS(*D*) given the disease, and PETS given both the drug and the disease.

• PETS’s assumption about the disease pathway model

PETS simulates the regulation chain between molecules in the disease pathway model as the transmission of signal in a network. In the signal network, each molecule becomes a node (or station), and each regulation becomes an edge (or channel). Each station stores some signal power, either positive (higher than normal condition), zero (normal condition) or negative (lower than normal condition). Each channel is either an activating channel (positive) or inhibiting channel (negative). Each station can receive signal from other station(s) and send out signals toward its downstream station(s). In the signal network, the drug triggers the initial signals to the drug’s target(s), and this triggering signal remains intact over time.

PETS assumes the following rules about signal transmission:

- The signal of a station at a given iteration only depends on the signal of its upstream station(s) at the previous iteration and the upstream channel.

- The signal at every station converges.

Fig. 12 demonstrates four possible cases of a downstream station, P2, receiving signal from an upstream station, P1. In case a, when the signal at P1 is positive and the channel is activating, the signal at P2 is expected to be increasing, or positive. In case b, when the signal at P1 is negative and the channel is activating, the signal at P2 is expected to be decreasing, or negative. In case c, when the signal at P1 is positive and the channel is negative, the signal at P2 is expected to be decreasing, or negative. And in case d, when the signal at P1 is negative and the channel is negative, the signal at P2 is expected to be increasing, or positive.

• Mathematical framework to simulate signal transmission

To simulate the signal transmission in PETS, we employed the novel iterative updating formula from PageRank:

(4)

Here, *N* is the total number of nodes in the network; *k* denotes the *k*th iteration, *i* and *j* denote different nodes; out\_deg(*i*) is the out-degree. In other words, the number of downstream connectors from *i*; *cj* is the initial value of S(*j*). Damping factor *d* controls how much the new signal S(*j*, *k*) is updated from other nodes in the network. Hence, setting *d* = 0 means that there is no update at *j*, implying that PETS is not applied, which leads to the PETS(-) framework. Setting *d* = 1 implies that we totally update *j* value from the network. The first term in (4) refers to a constant factor applied to node *j*’s value at every time *j* is updated, which reflects the PETS’ assumption about the constant drug’s triggering signal(s) to the drug’s target(s). The triggering signal of molecule *j* (*cj*) is 0 if *j* is not a drug’s target and is nonzero otherwise. The second term in (4) implements PageRank’s uniform delivery value idea, in which a node sends the node’s value uniformly to all downstream node(s). PETS adopts this idea if all regulations in M are uniformly weighted. Otherwise, we replaced the denominator out\_deg(*i*) by . The product matches with the expectation in Fig. 12. Different from PageRank in which **M**(*i*, *j*) is always non-negative, **M**(*i*, *j*) in PETS could be negative if the regulation from molecule *i* to molecule *j* is an inhibition reaction.

To meet more biological assumptions, we introduced several modifications in (4). First, since each biological interaction has a boosting factor, we add a factor *b* standing for boosting factor into the second term of (4). *b* receives a constant value for default, but it could be regulation-dependent based on some domain knowledge. For constant *b*, we design *b* > 1 favoring the indication of the longer chain of regulation and vice versa. Second, due to the design of *b*, we introduced the notion of layer to decide when *b* is applied. From the network breadth-first search point of view, the *k*th layer occurs only nodes visited at the *k*th iteration. Thus, when updating signal S(*j*) by at *k*th iteration (currentLayer), *b* is applied if and only if station *i* is visited at the *k*-1th iteration (previousLayer). Overall, we adjusted PageRank’s updating mechanism into

(5)

• Rp score adjustment

Estimating missing *Rp* scores is necessary before applying PETS algorithm since we included proteins outside C2MAP database in the pathway model. Let *i* be the protein in the pathway with the minimum C2MAP *Rp* score, and deg(*i*) be the degree of protein *i* in the pathway. For any protein *j* without *Rp* score, we estimate *Rp*(*j*) as following:

After estimating missing Rp score, we selected the top 5% proteins having the highest Rp score and set the *Rp* score of these proteins to the 95% largest *Rp* score.

PETS pseudo code and pseudo code explanation

Based on the assumptions in section 7, we design PETS as follows to complete the following tasks:

- Predicting the drug-molecular indication based on known drug-target indications and disease pathway models. From this task, we were able to suggest more drug-molecular indications which were not covered by HD or any other databases.

- Compute the drug therapeutic score based on the predicted drug-molecular indications and *ZP* score. We complete this task by applying Lamb et al novel statement: a therapeutic drug should reverse the molecular expression in the disease condition [11].

• Initialization

Given the drug-target of drug *D* and disease *z*’s pathway model **M**, we trigger the signal at a molecule based on the molecule’s degree and the drug’s direct effector information:

if *pi* is the designed target of the drug

if *pi* is not the designed target of the drug (off-target)

Where *D*(*i*) = 0 if *pi* is not the direct effector of the drug. *D*(*i*) = 1 if *pi* is the direct effector of the drug and receives an activation signal from the drug. *D*(*i*) = -1 if *pi* is the direct effector of the drug and receives an inhibition signal from the drug. This initialization assumes that the drug sends stronger signals towards its designed targets. We set the trigger signal as to ensure that the target still receives some signal even if the target does not have any downstream connection.

• Updating the s array for drug-effector prediction

For all molecules *i* that are not in the *currentLayer*, their signals are maintained:

S(*i*, *k*) = S(*i*, *k*-1)**,** Otherwise, we update the signal for all proteins *i* in *currentLayer* based on (5).

In theory, step 2 is completed at the time step *k*con when the signal power of every node converges. We chose the stopping condition as the inequality holds for all molecular *i*. In practice, for the breast cancer pathway model, it is completed when *k*con = 150, and may be different with different models. At the convergent point, we check the sequence *s*(*i*, 1*k*con) is converging to 0 by examining the ratio . If the ratio is less than 0.05, the sequence *s*(*i*, 1*k*con) is considered converging to 0.

PETS completes task 1 by calculating *s*(*i*, *k*con) as the final effect the molecule receives. If *s*(*i*, *k*con) > 0, the effect is predicted as activation; if *s*(*i*, *k*con) < 0, the effect was predicted as inhibition. Otherwise, the effect is unknown and not considered in evaluation. The score *s*(*i*, *k*con) shows the algorithm’s prediction of the drug-molecular indication associated with *pi*.

• PETS score evaluation

The drug is scored using a weight-averaging technique, extending Lamb et al’s idea [13], based on the following formula:

PETS-score = (6)

Drug having high PETS score is predicted as therapeutic, and vice versa.