**BMC
Systems Biology**

PROCEEDINGS   **Open Access**

Prediction of drugs having opposite effects on disease genes in a directed network

Hasun Yu[1,2], Sungji Choo[1,2], Junseok Park[1,2], Jinmyung Jung[1,2], Yeeok Kang[1,2] and Doheon Lee[1,2*]

Abstract

Background: Developing novel uses of approved drugs, called drug repositioning, can reduce costs and times in traditional drug development. Network-based approaches have presented promising results in this field. However, even though various types of interactions such as activation or inhibition exist in drug-target interactions and molecular pathways, most of previous network-based studies disregarded this information.

Methods: We developed a novel computational method, Prediction of Drugs having Opposite effects on Disease genes (PDOD), for identifying drugs having opposite effects on altered states of disease genes. PDOD utilized drug-drug target interactions with 'effect type', an integrated directed molecular network with 'effect type' and 'effect direction', and disease genes with regulated states in disease patients. With this information, we proposed a scoring function to discover drugs likely to restore altered states of disease genes using the path from a drug to a disease through the drug-drug target interactions, shortest paths from drug targets to disease genes in molecular pathways, and disease gene-disease associations.

Results: We collected drug-drug target interactions, molecular pathways, and disease genes with their regulated states in the diseases. PDOD is applied to 898 drugs with known drug-drug target interactions and nine diseases. We compared performance of PDOD for predicting known therapeutic drug-disease associations with the previous methods. PDOD outperformed other previous approaches which do not exploit directional information in molecular network. In addition, we provide a simple web service that researchers can submit genes of interest with their altered states and will obtain drugs seeming to have opposite effects on altered states of input genes at http://gto.kaist.ac.kr/pdod/index.php/main.

Conclusions: Our results showed that 'effect type' and 'effect direction' information in the network based approaches can be utilized to identify drugs having opposite effects on diseases. Our study can offer a novel insight into the field of network-based drug repositioning.

Keywords: Drug repositioning, Network biology, Directed network

Background

A drug is a chemical compound used to treat diseases and it usually has opposite effects to disease states. The traditional drug development remains expensive and time consuming process with low success rate. Developing a new drug to market takes about 15 years and 1 billion US dollars [1] and more than 85 % of drugs are failed to be released [2]. The identification of new therapeutic indications for existing drugs suggests an alternative for this situation [3]. The repurposing of approved drugs reduces costs and times involved in early phase of drug discovery.

In silico methods have provided candidates for novel uses of existing drugs. With the emerging technologies generating gene expression profiles, one of prevalent computational approaches to identify novel indications for existing drugs is based on drug and disease gene expression signatures from public databases [4–7]. These

* Correspondence: dhlee@biosoft.kaist.ac.kr
[1]Department of Bio and Brain Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
[2]Bio–Synergy Research Center, 291 Daehak–ro, Yuseong–gu, Daejeon 305-701, Republic of Korea

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 18 of 119

methods predicted drug and disease pairs showing therapeutic relationships through comparison of signatures of drugs effect and diseases pathophysiology. They discovered some previously uncharacterized uses of known drugs to diseases. However, these studies may have low precision since gene expression profiles, especially for drugs, can be generated under various conditions such as different tissue, cell line/types, and different doses of drugs [5].

Other recent popular *in silico* approaches in drug repositioning are network-based approaches including: 1. prediction of interactions between drugs and target proteins and using them for drug repositioning [8–11]; 2. prediction of drug effects to disease genes on biological networks [9, 12, 13]. These network-based studies have revealed promising results. However, even though there are diverse types of drug-drug target interactions such as activation or inhibition, the first type of network-based methods using drug-target interactions (DTI) regarded drug-target protein interactions as binary interactions, except for one [11], which did not illustrate how this method could be used in drug repositioning. The second type of network-based methods using molecular biological networks did not fully take into account the way drugs affect to disease genes because these methods did not consider 'effect types', 'effect directions', and 'altered disease genes' in diseases (Fig. 1). Thus, candidate drugs for the treatment of diseases from previous studies could activate over expressed genes in the disease state or inhibit under-expressed genes and could accelerate progression of the disease. In addition, candidate drugs from previous studies could not exert influence on diseases genes due to 'effect direction' (Fig. 1).

Here, we proposed a novel approach to identify novel indications of existing drugs by predicting drugs having opposite effects on altered disease genes. Our hypothesis is that if a drug compensates differentially regulated states of disease genes in patient, then that drug has the potential to be a therapy of the disease. Entire routes from drugs to disease genes are considered with the 'effect types', 'effect directions', and 'altered disease genes' information: interactions between drugs and drug targets; directed pathway between drug targets and disease genes; regulated states of disease genes in patients. This approach is termed as Predicting Drugs having Opposite effects Disease genes (PDOD).

In our PDOD approach, we integrated directed pathways having 'effect type' and 'effect direction', mapped drug-drug target interactions with their 'effect types', and mapped disease genes with their differentially regulated states. Based on this entire path, we predicted a drug whose targets are close to disease genes and activate down-regulated disease genes or inhibit up-regulated disease genes in patients as a therapy for the diseases (Fig. 2 and Additional file 1: Figure S1). We also showed that our method would be able to verify known uses of drugs and predict novel indications for existing drugs.

Methods
Datasets
To identify how drugs influenced on disease genes, we considered the shortest paths from drugs to diseases. We collected directed molecular pathways, drug-target protein interactions, and disease related genes with their regulated states from various data sources. The datasets used in our method are described here.

Integrated directed pathways
We constructed a directed backbone network having 'effect type' and 'effect direction' by integrating Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [14]. Although KEGG pathway does not cover entire molecular entries and pathways in human, it provides
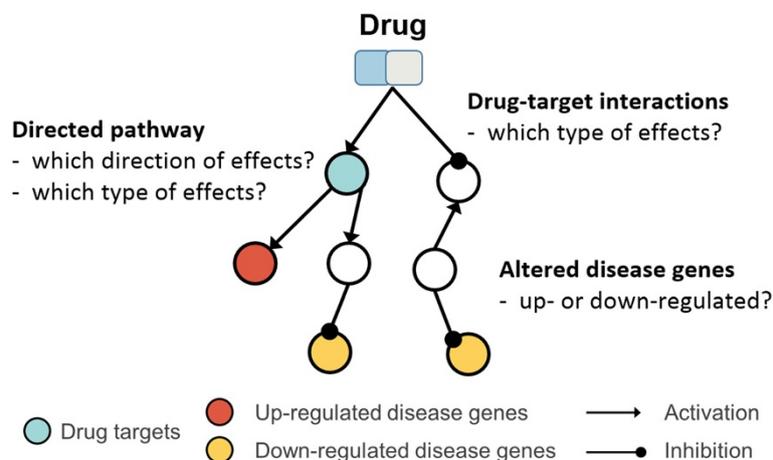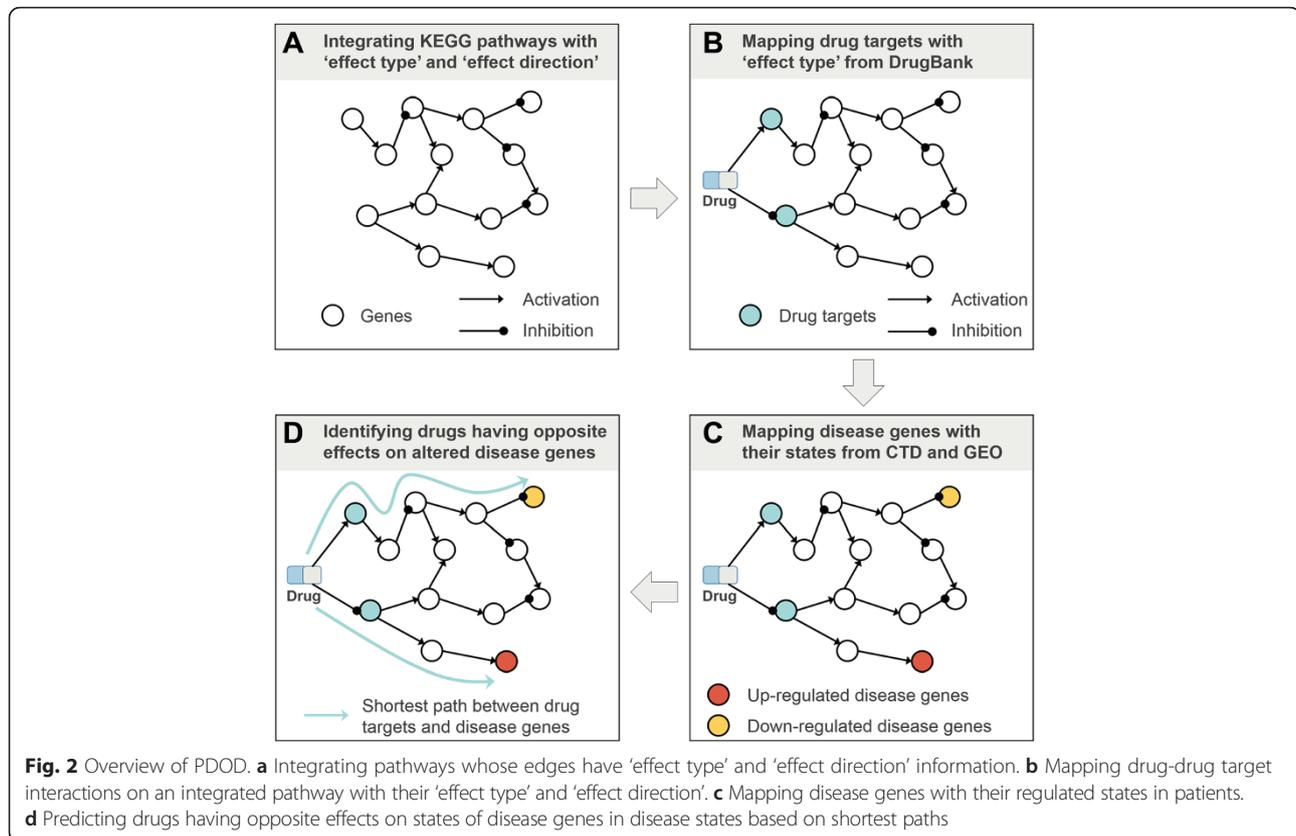

Fig. 1 The way a drug has effects on disease genes. The entire path from a drug to disease genes consists of 'effect type' such as activation or inhibition, 'effect direction', and 'altered states of disease genes' like up- or down-regulated states in disease patients

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 19 of 119

Fig. 2 Overview of PDOD. **a** Integrating pathways whose edges have 'effect type' and 'effect direction' information. **b** Mapping drug-drug target interactions on an integrated pathway with their 'effect type' and 'effect direction'. **c** Mapping disease genes with their regulated states in patients. **d** Predicting drugs having opposite effects on states of disease genes in disease states based on shortest paths

manually curated pathways with highly reliable relations having 'effect type' and 'effect direction' information. We downloaded KEGG Markup Language (KGML) files of 286 human pathways using KEGGgraph package [15]. We only used gene entries and relations with obvious 'effect direction' like activation, inhibition, expression, and repression. Then, the 166 human pathways including signal transduction and transcription pathways were merged into a Directed KEGG pathways (DKEGG) consisting of 3,307 nodes and 28,808 directed edges. Each node in DKEGG is a gene and each edge represents a relation with 'effect type' and 'effect direction' from gene A to gene B, for example, gene A activates (inhibits) gene B. In DKEGG, if the effect type from gene A to gene B was 'activation-like effect types' like activation or expression, we assigned the edge weight from node A to B as '+1' and if the effect type from node A to B was inhibition or repression, the edge weight from node A to B was set to '-1'.

Drug-drug target interactions

Drug-drug target interactions were obtained from Drug-Bank database (version 4.0) [16]. We collected drugs which possess known 'activation-like' or 'inhibition-like' drug-drug target interactions. We regarded activation, agonist, activator, simulator, and partial agonist interactions as 'activation-like drug-drug target interactions' and inhibition, inhibitor,

antagonist, negative modulator, inverse agonist, suppressor, inhibitor (competitive), partial antagonist, reducer, and blocker as 'inhibition-like drug-drug target interactions'. We filtered out drugs whose targets were not included in DKEGG. Finally, we extracted 2,434 drug-drug target interactions for 364 drug targets (genes) and 898 drugs.

Selection of diseases

We selected diseases satisfying several conditions: 1. Known disease-gene associations with direct evidence are in Comparative Toxicogenomics Database (CTD) [17]; 2. Diseases related tissues are clear; 3. Gene expression profiles of disease cases and controls obtained from related tissues exist in Gene Expression Omnibus (GEO) database [18]. Based on these conditions, we selected original 16 disease. Gene expression data information and related tissues about 16 diseases we firstly chose were shown in Additional file 2: Table S1.

Disease genes with their altered states

Disease-gene associations for 16 diseases were collected from CTD. We only used disease-gene associations having direct evidences. MeSH IDs [19] of diseases and the number of the associated disease genes are described in Additional file 2: Table S1.

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 20 of 119

We assigned altered states of disease genes using gene expression data. We downloaded microarray gene expression datasets for each disease from GEO. The altered states of disease genes were assigned to up-regulated or down-regulated if disease genes are significantly differentially expressed genes between disease and control samples. We identified significantly differentially expressed genes using linear models and empirical Bayes methods (limma) [20] representing several practical advantages compared to other methods [21]. We found altered disease genes of 16 diseases whose FDR corrected *p*-values were less than 0.05. Five diseases, Alzheimer's disease, bipolar disorder, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and major depressive disorder, were excluded because none of their disease genes was significantly differentially expressed in disease states. Then, we checked whether differentially expressed disease genes exist in DKEGG or not. Multiple sclerosis was removed because its differentially expressed disease genes do not exist in DKEGG. Pulmonary hypertension was also filtered out because a differentially expressed disease gene is an isolated node in DKEGG. Finally, nine diseases were used in our further analysis.

Collecting therapeutic drug-disease associations

The chemical-disease association file was downloaded from CTD. MeSH IDs, presented in Additional file 2: Table S1, were utilized to match nine diseases in our study to diseases in the chemical-disease association field. We collected answer drugs for each disease which mean the drugs could be used to treat the diseases. The chemicals were employed as the answer sets in the further analysis according to two criteria: 1. Chemicals have therapeutic associations for diseases; 2. Chemicals have DrugBank ID and drug-drug target interactions with their effect types. The entire list of answer drugs with DrugBank ID for each disease is shown in Additional file 3: Table S2.

Resolving conflicts in DKEGG

We hypothesized drugs whose targets have short distances to disease genes and activate (inhibit) down (up)-regulated genes in disease states can be used to treat the diseases. To calculate distances between drug targets and disease genes, we inferred all possible shortest paths from each drug target to each disease gene using Breadth-first-search algorithm in DKEGG with 'effect type' and 'effect direction'. In our approach, a distance from gene A to gene B can be a negative value if gene A inhibits gene B.

In directed biological pathways, there exist conflicts. Conflict refers to a situation that two or more contradictory relations coexist in directed biological pathways

[22]. Conflicts usually occur in biological pathways due to different biological contexts [23], for example, a dual way of TAK1 working corresponding to different cellular contexts [24]. In addition, conflicts were also observed in shortest paths from node A and node B in the directed pathway. For example, there were cases that gene A activated gene B in one shortest path while gene A inhibited gene B in other shortest paths: gene A activates gene C and gene C activates B; gene A activates gene D and gene D inhibits B.

To resolve this problem, we searched all possible shortest paths and set a distance $d_c(r,g)$ from a drug target r to a disease gene g with the consideration of the conflict problem by

$$d_c(r,g) = \frac{n_a + n_i}{n_a - n_i} |d(r,g)| \qquad (1)$$

where n_a is the number of 'activation-like shortest paths', n_i is the number of 'inhibition-like shortest paths', and $|d(\cdot)|$ is an absolute value of distances of shortest paths between a drug target and a disease gene. Whether a shortest path is an activation or inhibition-like path is determined by the sum of the number of inhibition and repression edges. If the sum of the number of inhibition and repression edges is an even number including zero in the shortest path, it is regarded as an activation-like path and if an odd number, an inhibition-like path. The sign of d_C indicates whether the drug target would activate or inhibit the disease gene. When the number of activation-like shortest paths is greater than inhibition-like shortest paths, the sign of d_C is positive, or vice versa. The absolute value of d_C reveals how reliable the effect from a drug target to a disease gene is. As the number of activation and inhibition-like paths are similar, the absolute value of d_C increases than an original absolute value of distances $|d(\cdot)|$ and it represents the drug target would give little influence on the disease gene. If the number of activation and inhibition-like paths are equal or a drug target r and disease gene g are not connected in DKEGG, $d_C(r,g)$ will be a value of infinity. The distribution of d_C between all genes in DKEGG is presented in Additional file 4: Figure S2.

A score between a drug and a disease

For each drug and disease pair, we calculated a score using drug-target interactions, distances from drug targets disease genes with consideration of conflicts, and the regulated states of disease genes. In our PDOD approach, a PDOD score between a drug R having drug targets r_i and a disease G with its associated disease genes g_j was computed as below:

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 21 of 119

$$\mathrm{PDOD\ score}(R, G) = \frac{1}{n_g n_r} \sum_{i=1}^{n_r} \sum_{j=1}^{n_g} sgn\left(r_i\ , g_j\right)$$
$$\times \frac{1}{1 + \left|\dfrac{d_c\left(r_i, g_j\right)}{\alpha}\right|^2} \qquad (2)$$

$$sgn\left(r_i\ , g_j\right) = i(r_i) \times sign\left(d_c\left(r_i, g_j\right)\right) \times s\left(g_j\right) \qquad (3)$$

where α is the parameter of a bell shaped function meaning the half width of the bell shaped function, n_r is the number of drug targets of the drug R, n_g is the number of disease genes of the disease G, $i(r_i)$ is the corresponding value for the predefined 'effect type' of drug-target interactions and it is assigned to '+1' for 'activation-like drug-drug target effect type' and '-1' for 'inhibition-like drug-drug target effect type', $s(g_j)$ is '+1' in case that g_j is a down-regulated gene in disease states and '-1' in case of an up-regulated-gene, and $sign(\cdot)$ is the sign function. Thus, the function $sgn(r_i\ , g_j)$ has a positive value when a drug has on opposite effect on a disease gene g_j through a drug target r_i. We set our scoring function as a bell shaped function since we thought that if the distances from drug targets to a disease genes were longer than the specific distance, α, the possibility that the drug had an effect on the disease genes would decrease rapidly. Thus, an absolute value of our PDOD score is in inverse proportion to the norm of $d_c(r_i, g_j)$. We divided our scores by the number of drug targets of the drug because we thought target proteins which did not affect disease genes might act as 'off-targets' for treatment of diseases and these targets could lead to unexpected effects such as side effects [25]. Our scores were also divided by the number of disease genes for normalization between diseases and thus our scores had values between −1 and 1. Finally, in our scoring function, a drug will receive a high score in case that drug targets of the drug are close to disease genes, drugs targets activate down-regulated disease genes and inhibit up-regulated disease genes in patients, and the number of off-targets is small.

Results and discussions

We applied our PDOD approach to nine diseases, adenocarcinoma of lung, acute myelocytic leukemia, astrocytoma, asthma, glioblastoma, oligodendroglioma, Parkinson disease, schizophrenia, and thyroid carcinoma. For each disease, we estimated the possibility of 898 drugs for the treatment of each disease based on PDOD scores which use drug-drug target interactions with 'effect type' from DrugBank, shortest paths from drug targets to disease genes with the consideration of conflicts in DKEGG, and changed states of disease genes in the disease with disease genes from CTD and expression profiles of GEO. The distribution of PDOD scores between all drugs and nine diseases we used presented in in Additional file 4: Figure S2–S3.

Performance of PDOD and previous methods
Comparison to a network-based method without directional information
To verify the importance of directional information used in PDOD, we make a Predicting Drugs having effects on Disease genes (PDD) score which uses drug-drug target interactions without 'effect type', paths from drug targets to disease genes on DKEGG without 'effect type' and 'effect direction', and disease genes without their regulated states.

We calculated PDOD scores and PDD scores between all 898 drugs and each disease respectively and compared performance of PDOD to PDD for predicting known drugs of each disease from CTD. Fig. 3 shows the area under receiver operating characteristic curves (AUC) values of PDOD and PDD approaches for nine diseases. As shown in Fig. 3, PDOD shows better performance than PDD in the prediction of answer drugs for diseases except for Parkinson disease. This illustrates the importance of directional information used in PDOD. We set a value
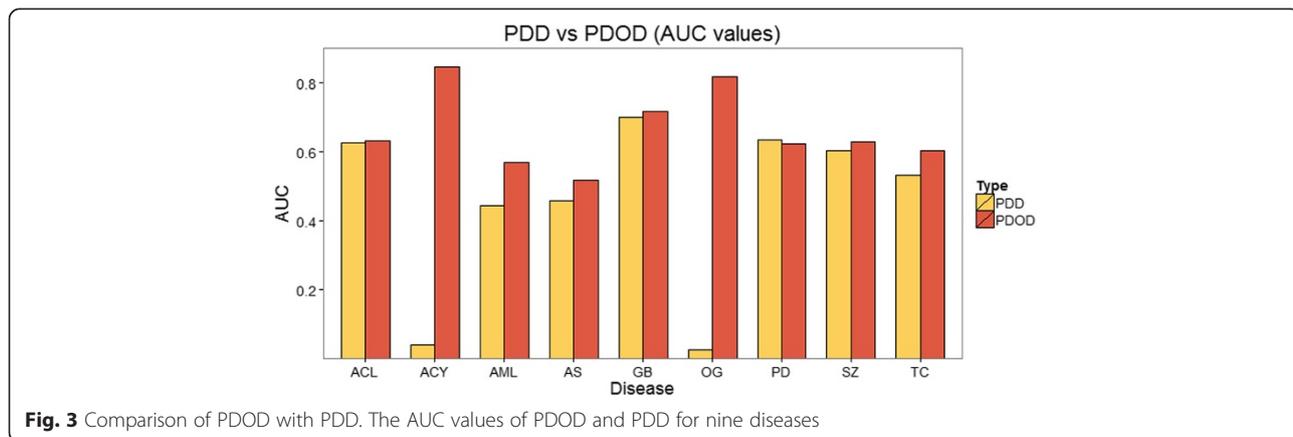

Fig. 3 Comparison of PDOD with PDD. The AUC values of PDOD and PDD for nine diseases

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 22 of 119

of parameter α to three for AUC values of both PDOD and PDD in Fig. 3. AUC values through various α values are given in Additional file 5: Figure S4–S12. As shown in figures of Additional file 5: Figure S4–S12, some diseases like adenocarcinoma of lung and schizophrenia revealed slightly different AUC values according to values of parameter α and they have high AUC values when α is bigger than three. It may be related to results from the previous work [26] which shows that the averages of distances between drug targets and disease genes are dependent on the types of diseases.

Comparison to a previous method without directed network

Our approach was also compared with a previous method using gene expression profiles of both drugs and diseases [6, 7]. These previous approaches are analogous to PDOD in that they intended to identify drugs likely to have complementary effects on a disease state as therapies of the disease. Thus, we compared performance of PDOD, which is a network–based approach, to the previous method using only gene expression profiles for prediction of known therapeutic relationship between drugs and diseases.

Disease gene expression signatures of the previous method were obtained identical datasets from GEO (Additional file 3: Table S1). We assigned that the numbers of up- and down-regulated disease gene expression signatures for each disease are equal to the numbers of up- and down-regulated disease genes used in PDOD. With disease gene expression signatures, the enrichment score (ES) introduced in Connectivity Map [6] was used for determining the potentiality of a drug for the treatment of the disease. A drug having high negative enrichment score was expected to repress the disease. Since some of answer drugs did not have drug exposure expression data in Connectivity Map and some drugs existing in Connectivity Map did not have drug-target interactions

used in our approach. Thus, the intersection of two drug sets, 317 drugs, was used in this evaluation. In this process, the answer drug of astrocytoma and oligodendroglioma are filtered out. Thus, we compared performances of PDOD and the previous method for seven diseases (Fig. 4). Figure 4 indicates PDOD using directed network usually outperformed a previous approach using only gene expression signatures.

Potentiality of high ranked drugs in PDOD as therapies

Top 10 ranked drugs receiving high scores in our PDOD for whole diseases are revealed in Table 1. All of the ten drugs getting high scores are not included in answer drugs for each disease. We investigated published literature which can support the possibility of uses of these drugs for the treatment of the disease. Six drugs have supporting evidence in literature for their expected usages [27–36]. These results show the potential of our PDOD approach for identification of novel drug indications.

Specially, tetracycline [DrugBank:DB00759], which is the drug receiving the highest PDOD score and is not contained in our answer drug sets for oligodendroglioma, has already a supporting literature evidence for the treatment of oligodendroglioma [27]. This drug inhibits its target [Entrez ID:5621]. The distance d_c from its target [Entrez ID:5621] to up-regulated disease gene [Entrez ID:3280] is '+3' which is obtained by two 'activation like paths' from DKEGG and thus its target activates up-regulated disease genes. Therefore, this drug may inhibit an up-regulated disease gene [Entrez ID:3280] of oligodendroglioma not directly but through its target [Entrez ID:5621] and DKEGG.

Five drugs, fulvestrant [DrugBank:DB00947], cetuximab [DrugBank:DB00002], gefitinib [DrugBank:DB00317], erlotinib [DrugBank:DB00530], and panitumumab [DrugBank:DB01269], do not exist in our answer drug sets but they also have a potential to treat thyroid carcinoma
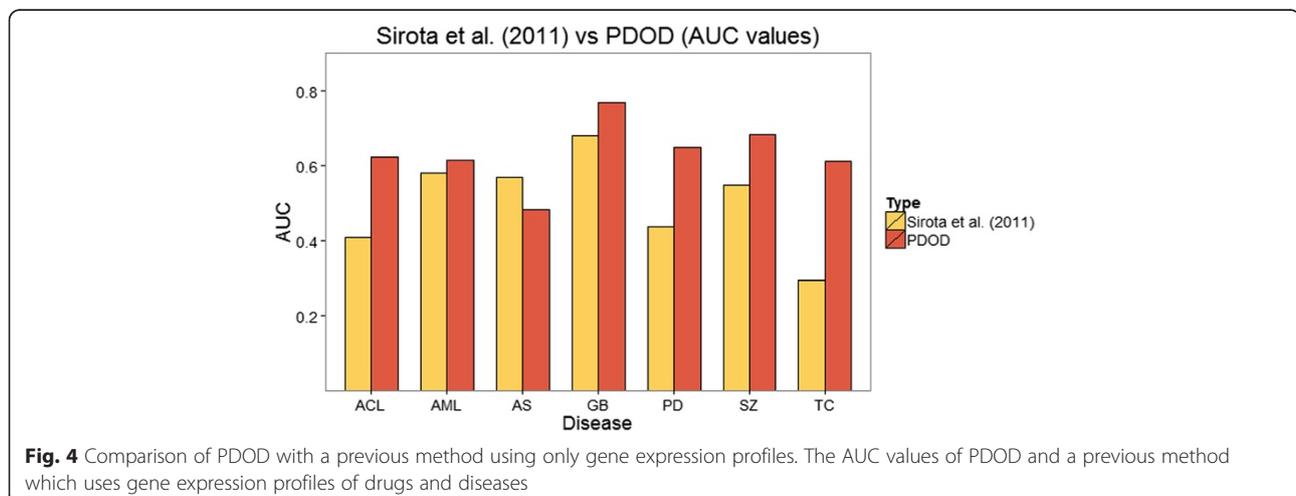

Fig. 4 Comparison of PDOD with a previous method using only gene expression profiles. The AUC values of PDOD and a previous method which uses gene expression profiles of drugs and diseases

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 23 of 119

Table 1 Top 10 ranked drugs in PDD and their potentiality for the therapy of diseases

Disease (MeSH ID)	Drug (DrugBank ID)	Description in DrugBank	Supporting evidence from literature
OG (D009837)	Tetracycline (DB00759)	Bacterial infections	[27]
TC (D013964)	Clomifene (DB00882)	Female infertility due to anovulation	n/a
TC (D013964)	Fulvestrant (DB00947)	Metastatic breast cancer	[28]
TC (D013964)	Ospemifene (DB04938)	Dyspareunia	n/a
TC (D013964)	Cetuximab (DB00002)	Metastatic colorectal cancer	[29, 30]
TC (D013964)	Gefitinib (DB00317)	Certain types of cancer	[31–33]
TC (D013964)	Erlotinib (DB00530)	Non-small cell lung cancer	[34–36]
TC (D013964)	Panitumumab (DB01269)	Antineoplastic agent	[36]
AML (D015470)	Mecasermin (DB01277)	Primary IGF 1 deficiency	n/a
TC (D013964)	Purvalanol (DB02733)	n/a	n/a

according to literature. All of five drugs also inhibit an up-regulated disease gene [Entrez ID:595] and activate a down-regulated disease gene [Entrez ID:83439] through their targets [Entrez ID:1956, 2099] and DKEGG. For example, fulvestrant have an antagonistic effect on its target [Entrez ID:2099]. The distance from its target [Entrez ID:2099] to up-regulated disease gene [Entrez ID:595] is '+1' by one 'activation like path'. The distance from its target [Entrez ID:2099] to down-regulated disease gene [Entrez ID:83439] is '-3' by one 'inhibition like path'. Thus, fulvestrant would inhibit an up-regulated disease genes and activate a down-regulated disease gene. For another example, gefitinib have an antagonistic effect on its target [Entrez ID:1956]. The distance from its target [Entrez ID:1956] to up-regulated disease gene [Entrez ID:595] is about '2.57' by eight 'activation like paths' and one 'inhibition like path'. The distance from its target [Entrez ID:2099] to down-regulated disease gene [Entrez ID:83439] is '-2' by one 'inhibition like path'. Therefore, gefitinib also would inhibit an up-regulated disease genes and activate a down-regulated disease gene and have an opposite effects on altered states of disease genes through drug-drug target interactions and paths from drug target to disease genes in DKEGG. These results not only are consistent with the previous study [26] but also indicate the importance of the consideration of biological networks in predicting indications of drugs.

Implementation in online

For researchers to infer drugs having opposite effects on altered genes they interested, we implemented PDOD in online (http://gto.kaist.ac.kr/pdod/index.php/main). Users can submit a gene list with their Entrez gene IDs [37] and altered states which they gained from their experiments. Users can select and submit alpha values according to the disease of interest [26] or the distribution of d_C in Additional file 4: Figure S2. Then, users will get a drug list likely to have reverse effects on their genes if more

than one gene existed in DKEGG. In addition, users can simply access our web page via mobile.

Conclusions

In this paper, we developed a novel network-based method, PDOD, to identify novel indications of existing drugs. We inferred drugs having reverse effects on altered states of disease genes with the consideration of 'effect type' and 'effect direction' in directed network. For a given disease gene set with altered states, we predicted a drug as a medicine of the disease when the drug satisfies two conditions: 1. Target proteins of a drug are close to disease genes; 2. A drug restores altered states of disease genes through paths from a drug to disease genes. We applied our method to infer candidate therapies for nine diseases. Our PDOD is compared to two previous approaches, the network-based method without 'effect type', 'effect direction', and 'altered states of disease gene' information and the method which did not utilize any biological network. This comparison showed the importance of this kind of information. In addition, we represented the potentialities of some of high ranked drugs in PDOD as novel treatments of the diseases.

We constructed a web based tool implementing PDOD. Our online tool may help researchers to predict drugs having opposite effects on disease genes and generate a hypothesis easily by using PDOD. Since average values of distances between drug targets and disease genes are various according to categories of diseases [26], users can set parameters depending on the disease of interest. In addition, even though we identified altered states of disease genes using expression data, researchers can input genes with changed states from any kind of experiment.

In spite of our promising results for a case study, our works have several drawbacks. One of the major drawbacks in our approach is that the number of 'effect types' used in our method is limited to four interactions. Relations having ambiguous 'effect type' and 'effect direction',

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 24 of 119

for example, binding, association, and phosphorylation were excluded at this present; thus, the scale of backbone network is relatively small compared to a real molecular network. We did not include these relations because they can bring out false positives in prediction of drugs; however, it causes that some diseases like multiple sclerosis cannot be applied PDOD to since genes related to these diseases do not exist in our constructed network. Another drawback of our approach is that we considered only the shortest paths although there would be longer paths from drug targets to disease genes. Even though we only used the shortest paths based on the assumption that closer genes may be more interactive, which is commonly used in other papers using biological network [38, 39], the consideration of longer paths could give us more comprehensive results.

In further study, we plan to enlarge scales of the backbone network not only by including other interactions like phosphorylation, methylation, binding/association, and metabolic reaction but also by combining other databases containing 'effect type' and 'effect direction information' such as pathway interaction database (PID) and small molecule pathway database (SMPDB). We will apply our method to more disease cases and validate promising drugs from our PDOD with wet lab experiment through collaboration. In addition, we expect that a negative score between a drug and disease in PDOD may explain the drug that is likely to have same effects on altered states of disease genes. In other words, the disease may be regarded as side effects of the drug. Based on this idea, the proposed approach may be feasible to predict side effects of drugs and drug-drug interactions.

Homeostasis is one of the most fundamental features of mankind. If a disease occurred and homeostasis could not be maintained, a drug having opposite effects on the disease could restore the altered states in patients. Our approach would be appropriate to predict these kinds of drugs.

Additional files

Additional file 1: Figure S1. Flow diagram of PDOD. The figure that illustrates how the databases are used to generate the inputs and expected outputs. (PDF 166 kb)

Additional file 2: Table S1. 16 diseases considered in this work. Statistics and information of data for 16 diseases are described in this file: Disease name; MeSH ID of diseases; GSE numbers we used; platforms of the GSE; tissues where GSE are generated; the number of disease genes from CTD; the number of differentially expressed disease genes; the number of differentially expressed disease genes in DKEGG. (XLSX 12 kb)

Additional file 3: Table S2. Answer drugs used in this work for nine diseases. Information about answer drugs used in this work are described in this file: Disease name; MeSH ID of diseases; the number of answer drugs for each disease; the list of answer drugs (DrugBank ID). (XLSX 11 kb)

Additional file 4: Figure S2–S3. The distribution of dc and PDOD scores. Each figure presents the distribution of dc between all genes in

DKEGG and PDOD scores between all drugs and nine diseases we used. (PDF 236 kb)

Additional file 5: Figure S4–S12. AUC values for nine diseases with different values of a. AUC values of PDD and PDOD for nine diseases according to different values of parameter a are enclosed in this file. (PDF 131 kb)

Abbreviations
PDOD: Prediction of Drugs having opposite effects on Disease genes; DKEGG: Directed KEGG pathways, which we constructed, containing only directed relations such as 'activation' and 'inhibition'; PDD: Prediction of Drugs having effects on Disease genes; ACL: Adenocarcinoma of lung; AML: Acute myelocytic leukemia; ACY: Astrocytoma; AS: Asthma; GB: Glioblastoma; OG: Oligodendroglioma; PD: Parkinson disease; PH: Pulmonary hypertension; SZ: Schizophrenia; TC: Thyroid carcinoma.

Competing interests
All authors declare that they have no conflict of interest to the publication of this article, and approved the manuscript and this submission.

Authors' contributions
HY suggested the main idea and DH supervised this work. HY and SC not only resolved conflict problems in directed network but also extracted drug-drug target interaction and drug-disease association information. SC and JP developed a web service. HY and JJ integrated directed pathways. SC and YK found literature evidences of high scored drugs in PDOD. All authors read, wrote, and approved the manuscript.

References
1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ. 2003;22(2):151–85.
2. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014;32(1):40–51.
3. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(8):673–83.
4. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med. 2011;3(96):96ra76.
5. Hu G, Agarwal P. Human disease-drug network based on genomic expression profiles. PLoS One. 2009;4(8):e6536.
6. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–35.
7. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011;3(96):96ra77.
8. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol. 2012;8(5):e1002503.
9. Emig D, Ivliev A, Pustovalova O, Lancashire L, Bureeva S, Nikolsky Y, et al. Drug target prediction and repositioning using an integrated network-based approach. PLoS One. 2013;8(4):e60618.
10. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. Bioinformatics. 2014;30(20):2923–30.

Yu *et al. BMC Systems Biology* 2016, **10**(Suppl 1):2

Page 25 of 119

11. Wang Y, Zeng J. Predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics. 2013;29(13):i126–i34.

12. Qabaja A, Alshalalfa M, Alanazi E, Alhajj R. Prediction of novel drug indications using network driven biological data prioritization and integration. J Cheminformatics. 2014;6(1):1.

13. Fukuoka Y, Takei D, Ogawa H. A two-step drug repositioning method based on a protein-protein interaction network of genes shared by two diseases and the similarity of drugs. Bioinformation. 2013;9(2):89–93. doi:10.6026/97320630009089.

14. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

15. Zhang JD, Wiemann S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics. 2009;25(11):1470–1.

16. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34 suppl 1:D668–D72.

17. Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegers TC, Mattingly CJ. Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical–gene–disease networks. Nucleic Acids Res. 2009;37 suppl 1:D786–D92.

18. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.

19. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. Jama. 1994;271(14):1103–8.

20. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer; 2005. p. 397–420.

21. Jeanmougin M, De Reynies A, Marisa L, Paccard C, Nuel G, Guedj M. Should we abandon the *t*-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. PLoS One. 2010;5(9):e12336.

22. Stobbe MD, Swertz MA, Thiele I, Rengaw T, Van Kampen AH, Moerland PD. Consensus and conflict cards for metabolic pathway databases. BMC Syst Biol. 2013;7(1):50.

23. Yoon S, Jung J, Yu H, Kwon M, Choo S, Park K et al. Context-based resolution of semantic conflicts in biological pathways. BMC medical informatics and decision making. 2015;15 Suppl 1:S3. doi:10.1186/1472-6947-15-S1-S3.

24. Ajibade AA, Wang HY, Wang R-F. Cell type-specific function of TAK1 in innate immune signaling. Trends Immunol. 2013;34(7):307–16.

25. Scheiber J, Chen B, Milik M, Sukuru SCK, Bender A, Mikhailov D, et al. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. J Chem Inf Model. 2009;49(2):308–17.

26. Yıldırım MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M. Drug—target network. Nat Biotechnol. 2007;25(10):1119–26.

27. Domercq M, Matute C. Neuroprotection by tetracyclines. Trends Pharmacol Sci. 2004;25(12):609–12. doi:10.1016/j.tips.2004.10.001.

28. Rajoria S, Suriano R, George A, Shanmugam A, Schantz SP, Geliebter J, et al. Estrogen induced metastatic modulators MMP-2 and MMP-9 are targets of 3,3′-diindolylmethane in thyroid cancer. PLoS One. 2011;6(1):e15879. doi:10.1371/journal.pone.0015879.

29. Prichard CN, Kim S, Yazici YD, Doan DD, Jasser SA, Mandal M, et al. Concurrent cetuximab and bevacizumab therapy in a murine orthotopic model of anaplastic thyroid carcinoma. Laryngoscope. 2007;117(4):674–9. doi:10.1097/MLG.0b013e318031055e.

30. Kim S, Prichard CN, Younes MN, Yazici YD, Jasser SA, Bekele BN, et al. Cetuximab and irinotecan interact synergistically to inhibit the growth of orthotopic anaplastic thyroid carcinoma xenografts in nude mice. Clin Cancer Res. 2006;12(2):600–7. doi:10.1158/1078-0432.CCR-05-1325.

31. Pennell NA, Daniels GH, Haddad RI, Ross DS, Evans T, Wirth LJ, et al. A phase II study of gefitinib in patients with advanced thyroid cancer. Thyroid. 2008; 18(3):317–23. doi:10.1089/thy.2007.0120.

32. Frasca F, Vella V, Nicolosi ML, Messina RL, Giani F, Lotta S, et al. Thyroid cancer cell resistance to gefitinib depends on the constitutive oncogenic activation of the ERK pathway. J Clin Endocrinol Metab. 2013;98(6):2502–12. doi:10.1210/jc.2012-3623.

33. Kurebayashi J, Okubo S, Yamamoto Y, Ikeda M, Tanaka K, Otsuki T, et al. Additive antitumor effects of gefitinib and imatinib on anaplastic thyroid cancer cells. Cancer Chemother Pharmacol. 2006;58(4):460–70. doi:10.1007/s00280-006-0185-x.

34. Hogan T, Jing Jie Y, Williams HJ, Altaha R, Xiaobing L, Qi H. Oncocytic, focally anaplastic, thyroid cancer responding to erlotinib. J Oncol Pharm Pract. 2009;15(2):111–7. doi:10.1177/1078155208101212.

35. Masago K, Miura M, Toyama Y, Togashi Y, Mishima M. Good clinical response to erlotinib in a patient with anaplastic thyroid carcinoma harboring an epidermal growth factor somatic mutation, L858R, in exon 21. J Clin Oncol. 2011;29(16):e465–7. doi:10.1200/JCO.2010.34.0216.

36. Landriscina M, Maddalena F, Fabiano A, Piscazzi A, La Macchia O, Cignarelli M. Erlotinib enhances the proapoptotic activity of cytotoxic agents and synergizes with paclitaxel in poorly-differentiated thyroid carcinoma cells. Anticancer Res. 2010;30(2):473–80.

37. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2005;33 suppl 1:D54–D8.

38. Li S, Zhang B, Zhang N. Network target for screening synergistic drug combinations with application to traditional Chinese medicine. BMC Syst Biol. 2011;5 Suppl 1:S10. doi:10.1186/1752-0509-5-S1-S10.

39. Zou J, Ji P, Zhao YL, Li LL, Wei YQ, Chen YZ, et al. Neighbor communities in drug combination networks characterize synergistic effect. Mol BioSyst. 2012;8(12):3185–96. doi:10.1039/c2mb25267h.