# SemFunSim: A New Method for Measuring Disease Similarity by Integrating Semantic and Gene Functional Association

**Liang Cheng**[1,9], **Jie Li**[1,9], **Peng Ju**[2], **Jiajie Peng**[1], **Yadong Wang**[1*]

**1** Center for Bioinformatics, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China, **2** School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

## Abstract

***Background:*** Measuring similarity between diseases plays an important role in disease-related molecular function research. Functional associations between disease-related genes and semantic associations between diseases are often used to identify pairs of similar diseases from different perspectives. Currently, it is still a challenge to exploit both of them to calculate disease similarity. Therefore, a new method (SemFunSim) that integrates semantic and functional association is proposed to address the issue.

***Methods:*** SemFunSim is designed as follows. First of all, FunSim (Functional similarity) is proposed to calculate disease similarity using disease-related gene sets in a weighted network of human gene function. Next, SemSim (Semantic Similarity) is devised to calculate disease similarity using the relationship between two diseases from Disease Ontology. Finally, FunSim and SemSim are integrated to measure disease similarity.

***Results:*** The high average AUC (area under the receiver operating characteristic curve) (96.37%) shows that SemFunSim achieves a high true positive rate and a low false positive rate. 79 of the top 100 pairs of similar diseases identified by SemFunSim are annotated in the Comparative Toxicogenomics Database (CTD) as being targeted by the same therapeutic compounds, while other methods we compared could identify 35 or less such pairs among the top 100. Moreover, when using our method on diseases without annotated compounds in CTD, we could confirm many of our predicted candidate compounds from literature. This indicates that SemFunSim is an effective method for drug repositioning.

**Citation:** Cheng L, Li J, Ju P, Peng J, Wang Y (2014) SemFunSim: A New Method for Measuring Disease Similarity by Integrating Semantic and Gene Functional Association. PLoS ONE 9(6): e99415. doi:10.1371/journal.pone.0099415

**Editor:** Ferdinando Di Cunto, University of Turin, Italy

**Received** December 16, 2013; **Accepted** May 14, 2014; **Published** June 16, 2014

**Funding:** This work is partially supported by the China National 863 High-Tech Program (2012AA02A602, 2012AA020404), postdoctoral science research developmental foundation of heilongjiang province no. lbh-q13084 and National natural science foundation of China (61173085). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: ydwang@hit.edu.cn

9 These authors contributed equally to this work.

## Background

The quantitative measurement of similarity between diseases based on qualitative association [1–5] raises more and more attention, because it plays an important role in predicting disease-causing genes [6,7], inferring microRNA function associations [8], and identifying novel drug indications [9]. Currently, there is a critical need to design methods to measure disease similarity.

Methods for calculating disease similarity can be broadly classified as semantic-based [8,10] and function-based [11–13]. Semantic-based methods are widely used for measuring similarity between terms of Gene Ontology (GO) [14,15] and human phenotype ontology (HPO) [16] in the biomedical and bioinformatics domain. Few of them are used for calculating similarity between terms of disease-related ontologies. For computing the similarity of GO terms, Resnik's method [17] has a better performance evaluation result [18] than union-intersection (UI), longest shared path (LP), JC [19] and Lin [20]. Resnik's method has also been used to calculate the similarity between terms of Disease Ontology (DO) [10,21], measuring disease similarity based on the information content (IC) (Figure S1 and File S1) of the most informative common ancestor (MICA) (Figure S1 and File S1) between two terms. In addition, Wang et al.'s method [22] calculates similarity between terms considering multiple common ancestors. It performs very well for computing the semantic similarity between GO terms [22], and has been successfully used for measuring disease similarity between medical subject headings (MeSH) [23] terms and inferring microRNA function network [8].

Function-based methods calculate disease similarity by comparing disease-related gene sets [11–13]. Mathur and Dinakarpandian [11] designed the similarity method based on overlapping gene sets (BOG) between diseases of DO. In comparison to semantic-based methods, the BOG method defines disease similarity from a new perspective. Therefore, it is possible to find unknown relationships [11]. However, it ignores the functional associations between disease-related genes which contribute to disease similarity. In another method, Mathur et al. [13] presented a process-similarity based (PSB) method by involving the associations based on GO [14] terms. PSB outshines BOG, and its performance is better than Resnik [17], Lin [20], LC [24] and JC's [19] methods [13]. Functional associations between genes involve multiple aspects, such as co-expression [25], protein-protein interaction [26], GO terms [27], etc. However, the PSB method only exploits the associations from GO terms. Therefore, the performance would likely be better if multiple associations were considered for calculating disease similarity.

There are many disease-related vocabularies, some of which describe semantic associations between diseases by 'IS_A' relationship (Figure 1), such as MeSH, DO, etc. Among them, DO is an ontology to organize vocabularies around diseases themselves [21]. And it integrates disease and medical vocabularies through extensive cross mapping [21]. Other vocabularies often include not only diseases themselves, but also terms of pathology, anatomical, etc. For example, MeSH is a more comprehensive ontology that has been classified as 16 categories. In these categories, only categories C and F03 define terms around disease. However, not all the terms in these categories are named for diseases themselves, such as pain (D010146). Furthermore, DO has been validated to be suitable for calculating disease similarity [11,13,28]. Therefore, we choose DO as disease terminology to describe disease terms for calculating disease similarity.

Function-based methods calculate disease similarity according to functional associations between genes. Semantic-based methods exploit associations from ontologies and the number of disease-related genes to compute disease similarity. Obviously, not all associations between diseases are represented by the ontology, a part of them are reflected through functional associations among disease-related genes and vice versa. In this paper, a new method (SemFunSim) is proposed, which integrates semantic and gene functional association for measuring similarity between diseases.

## Materials and Methods

### Disease Ontology

DO [21] (Table 1) contains 8,632 disease terms and 7,232 'IS_A' relationships among diseases. The directed acyclic graph (DAG) of DO represents terms linked by 'IS_A' relationship, of which a node represents a DO term and an edge represents an 'IS_A' relationship between diseases. Figure 1 shows a sub-graph of the DAG starting from the specific DO term 'Cutaneous lupus erythematosus (DOID:0050169)' and ending at the root term of DO.

### HumanNet and disease-related gene set

We accessed functional interactions of genes from HumanNet [29], which is an extended gene functional interaction network for Homo sapiens. Multiple distinct lines of evidence, spanning human mRNA co-expression, protein-protein interaction, protein complex, and comparative genomics data sets, in combination with similar lines of evidence from orthologs in yeast, fly and worm are comprehensively analyzed for the network using a probabilistic method [29]. This function network contains 476,399 interactions among 16,243 genes (Table 1).

Disease-related gene sets are from SIDD [30], which integrates five disease-related gene databases: GeneRIF [31], Online Mendelian Inheritance in Man (OMIM) [32], comparative toxicogenomics database (CTD) [33], genetic association database (GAD) [34], and SpliceDisease [35]. In total, 2,817 diseases, 12,063 genes and 117,190 associations between them are involved (Dataset S1). The data sources were downloaded from the web in Jul 2013, and the detailed information is listed in Table 1. Gene names in these sources have been converted to HUGO Gene Nomenclature Committee (HGNC) approved gene symbols [36].

### Disease similarity

Figure 2 gives an overview of SemFunSim. In the figure, $d_1$ and $d_2$ are two diseases from DO, and $d_{MICA}$ is the MICA of $d_1$ and $d_2$. $G_1$, $G_2$ and $G_{MICA}$ are gene sets related to $d_1$, $d_2$ and $d_{MICA}$, respectively. First, a weighted network of human gene function association is used for calculating FunSim (functional similarity) between $G_1$ and $G_2$. Then, semantic associations from DO are used to calculate semantic similarity (SemSim) between diseases. Finally, FunSim and SemSim are integrated into SemFunSim.

**Functional similarity between disease-related gene sets.** Gene function networks are widely used to understand disease [29,37–43]. We accessed the interactions of genes from HumanNet [29], which has been used to understand associations across three GO categories [44]. Each interaction of HumanNet has an associated log likelihood score (LLS) that measures the probability of a functional linkage between genes [29]. We normalized the associated LLS with equation 1.

$$LLS_N(g_i,g_j)=\frac{LLS(g_i,g_j)-LLS_{\min}}{LLS_{\max}-LLS_{\min}} \tag{1}$$

where $g_i$ and $g_j$ indicate the $i$th and $j$th gene, respectively. $LLS_N(g_i,g_j)$ represents LLS between $g_i$ and $g_j$ after normalization. $LLS(g_i,g_j)$ represents LLS between $g_i$ and $g_j$. $LLS_{\min}$ and $LLS_{\max}$ are the minimum LLS and the maximum LLS of HumanNet, respectively.

The functional similarity score between a pair of genes is defined as $FunSim(g_i,g_j)$:

$$FunSim(g_i,g_j)=\begin{cases}1 & i=j\\ LLS_N(g_i,\, g_j) & i\neq j \;\; and \;\; e(i,j)\in\mathbb{E}(HumanNet)\\ 0 & i\neq j \;\; and \;\; e(i,j)\notin\mathbb{E}(HumanNet)\end{cases} \tag{2}$$

In equation 2, $e(i,j)$ represents the interaction edge between gene pair $g_i$ and $g_j$. $\mathbb{E}(HumanNet)$ is a set which includes all the edges of HumanNet.

Then, we define the functional association between a gene $g$ and a gene set $G=\{g_1,g_2,\cdots,g_k\}$ as $F_G(g)$, which is described in equation 3.

$$F_G(g)=\max_{1\leq i\leq k}(FunSim(g,g_i)),\;\; g_i\in G \tag{3}$$

where $k$ indicates the number of genes in $G$, $g_i$ is the $i$th gene of $G$.

Let a pair of gene sets $G_1=\{g_{11},g_{12},\cdots,g_{1m}\}$ and $G_2=\{g_{21},g_{22},\cdots,g_{2n}\}$ be related to diseases $d_1$ and $d_2$, respectively. $m$ is the number of genes in $G_1$, and $n$ is the number of genes in $G_2$. We define FunSim of $d_1$ and $d_2$ in equation 4 as follows.

$$FunSim(G_1,G_2)=\frac{\sum\limits_{1\leq i\leq m}F_{G_2}(g_{1i})+\sum\limits_{1\leq j\leq n}F_{G_1}(g_{2j})}{m+n}, \tag{4}$$
$$g_{1i}\in G_1,\; g_{2j}\in G_2$$

**Semantic similarity based on Disease Ontology.** We define semantic similarity between disease pair $d_1$ and $d_2$ in equation 5.

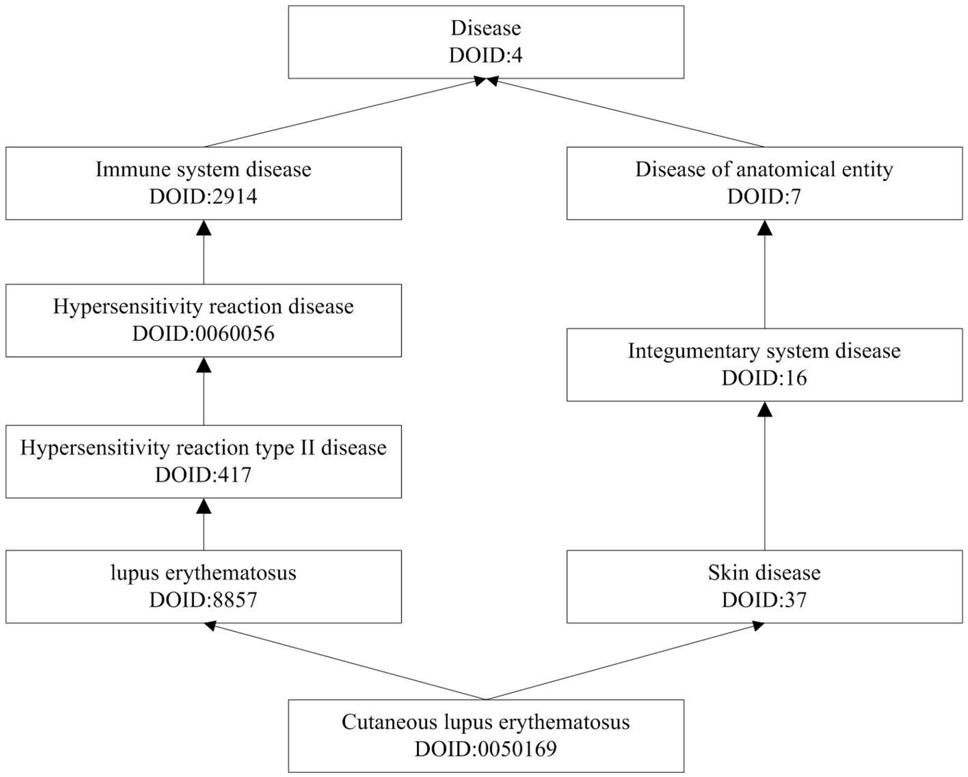
Disease
DOID:4
Immune system disease
DOID:2914
Disease of anatomical entity
DOID:7
Hypersensitivity reaction disease
DOID:0060056
Integumentary system disease
DOID:16
Hypersensitivity reaction type II disease
DOID:417
lupus erythematosus
DOID:8857
Skin disease
DOID:37
Cutaneous lupus erythematosus
DOID:0050169

**Figure 1. A sub-graph of the DAG for DO term 'Cutaneous lupus erythematosus (DOID:0050169)'.** The arrow symbol represents an 'IS_A' link of DO. For example, "Cutaneous lupus erythematosus (DOID:0050169)" is linked to "Skin disease (DOID:37)" by an 'IS_A' relationship.
doi:10.1371/journal.pone.0099415.g001

$$SemSim(d_1,d_2)=\frac{|G_1|}{|G_{MICA}|}\cdot\frac{|G_2|}{|G_{MICA}|} \qquad (5)$$

where $G_1$ and $G_2$ are gene sets related to $d_1$ and $d_2$, respectively. $G_{MICA}$ is gene set related to $d_{MICA}$, which represents the MICA of $d_1$ and $d_2$ in the DAG of DO. $|G_1|$, $|G_2|$, and $|G_{MICA}|$ represent the number of genes in $G_1$, $G_2$ and $G_{MICA}$, respectively.

**Similarity between disease pair by SemFunSim.** The similarity between disease pair $d_1$ and $d_2$ is defined in equation 6.

$$Sim(d_1,d_2)=FunSim(G_1,G_2)\cdot SemSim(d_1,d_2) \qquad (6)$$

where $d_1$ and $d_2$ are two diseases of DO. $G_1$ and $G_2$ are gene sets related to $d_1$ and $d_2$, respectively.

A threshold for significant similarity of the 916 diseases with potential therapeutic chemicals (PTCs) in CTD is defined based on randomized data as follows. First, the 916 disease names in the DAG of DO were randomly shuffled, and the hierarchical structure remained the same as the original DO. Next, gene names in HumanNet were randomly shuffled, and the network topology remained the same as the original HumanNet. Then, the similarity scores for pairs of these 916 diseases were computed by SemFunSim based on the randomized data. The experiment was iterated 1000 times. Finally, we calculate the false discovery rate (FDR) over all pairs according to equation 7.

**Table 1.** Data sources used for measuring disease similarity.

| Data source | Web site (Date of download) |
|---|---|
| DO | https://diseaseontology.svn.sourceforge.net/svnroot/diseaseontology/trunk/ (Apr 2013) |
| SIDD | http://mlg.hit.edu.cn/SIDD (Jul 2013) |
| CTD | http://ctdbase.org/downloads/;jsessionid = 71BC29A1A48AD67BADA2E2C4FC9625F3 (Apr 2013) |
| HumanNet | http://www.functionalnet.org/humannet/download.html (Jul 2013) |
| GO | http://www.geneontology.org/GO.downloads.ontology.shtml (Jul 2013) |
| GOA | http://www.geneontology.org/GO.downloads.annotations.shtml (Jul 2013) |
| MimMiner | http://www.cmbi.ru.nl/MimMiner/suppl.html (Feb 2014) |

doi:10.1371/journal.pone.0099415.t001

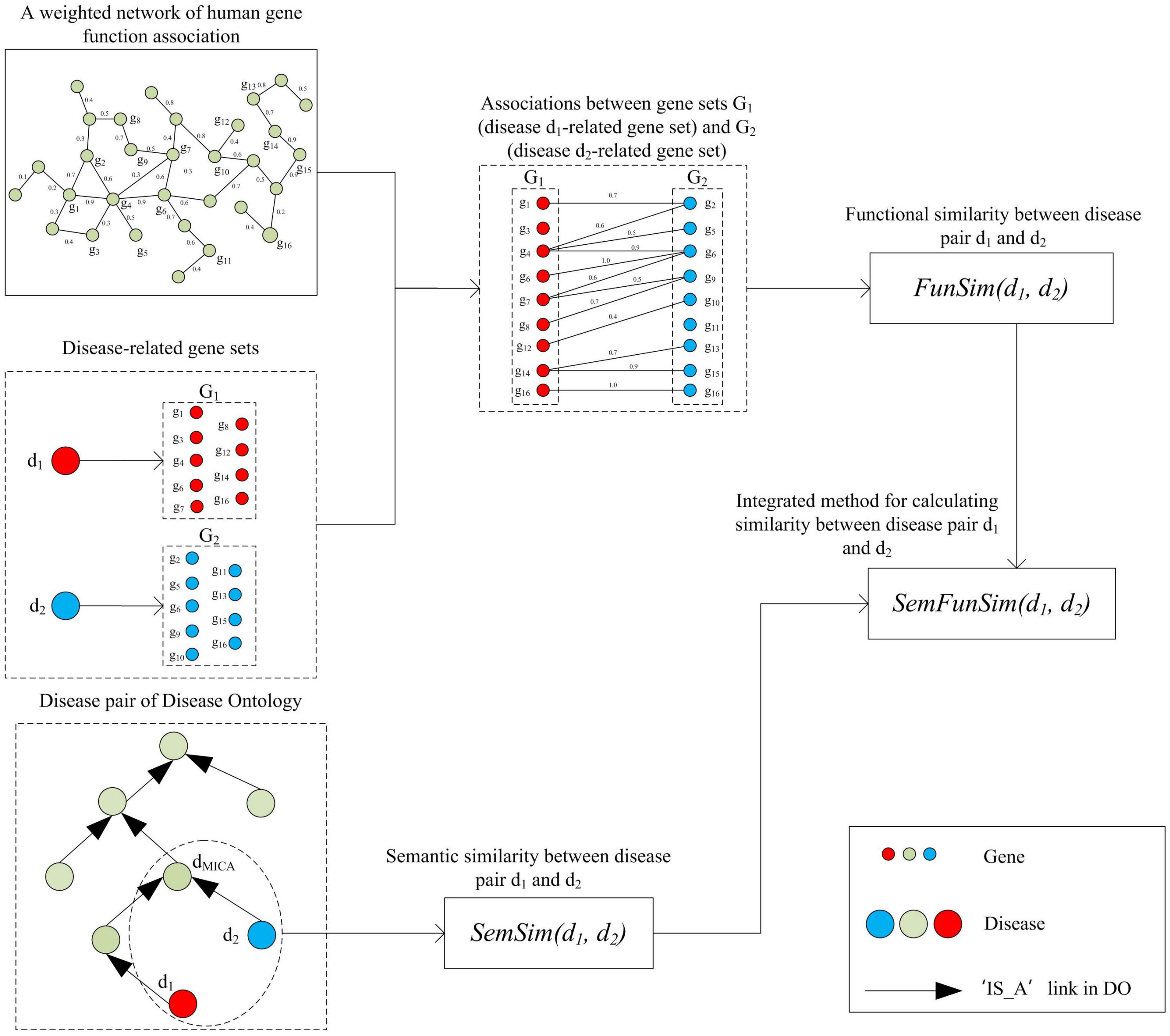
A weighted network of human gene function association
Disease-related gene sets
d1
d2
G1
G2
Disease pair of Disease Ontology
dMICA
Associations between gene sets G1 (disease d1-related gene set) and G2 (disease d2-related gene set)
Functional similarity between disease pair d1 and d2
FunSim(d1, d2)
Integrated method for calculating similarity between disease pair d1 and d2
SemFunSim(d1, d2)
Semantic similarity between disease pair d1 and d2
SemSim(d1, d2)
Gene
Disease
'IS_A' link in DO

**Figure 2. Overview of SemFunSim.** $d_1$, $d_2$ are two diseases, and $d_{MICA}$ is the MICA of $d_1$ and $d_2$. $G_1$, $G_2$ and $G_{MICA}$ represent gene sets related to $d_1$, $d_2$ and $d_{MICA}$, respectively.
doi:10.1371/journal.pone.0099415.g002

$$FDR(Sim_T) = \frac{\sum_{i=1}^{1000} N_i}{1000 \cdot N_T} \tag{7}$$

where $Sim_T$ represents a similarity score, $N_i$ indicates the number of hits in the $i$th permutation with the similarity score $> Sim_T$, and $N_T$ is the number of hits in the real case with the similarity score $\geq Sim_T$.

## Results and Discussion

### Validation of disease similarity methods on benchmark set

We calculated similarities of disease pairs on a benchmark set and another 100 random sets. The performance of SemFunSim was accessed by drawing a receiver operating characteristic (ROC) [45] curve. In Figure 3A, two types of disease pair sets are introduced as input in the validation process. On one hand, two manually checked datasets [12,13,46] of disease pairs with high similarity were integrated into a benchmark set. One dataset was obtained from diseases analyzed in the study by Suthram et al [12]. Disease pairs of the dataset were marked as similar after validation from literature by Mathur et al [13]. The other dataset was derived from the judgment of medical residents for semantic similarity, and pairs of similar diseases were extracted by Pakhomov et al [46]. In total, 47 diseases and 70 pairs of these two disease pair datasets were merged as the benchmark set (Dataset S2). On the other hand, each random set contains 700 disease pairs randomly selected from DO.

In order to further test the performance of the proposed method, SemFunSim was compared with disease similarity methods including Resnik [17], Wang [22], BOG [11], and PSB

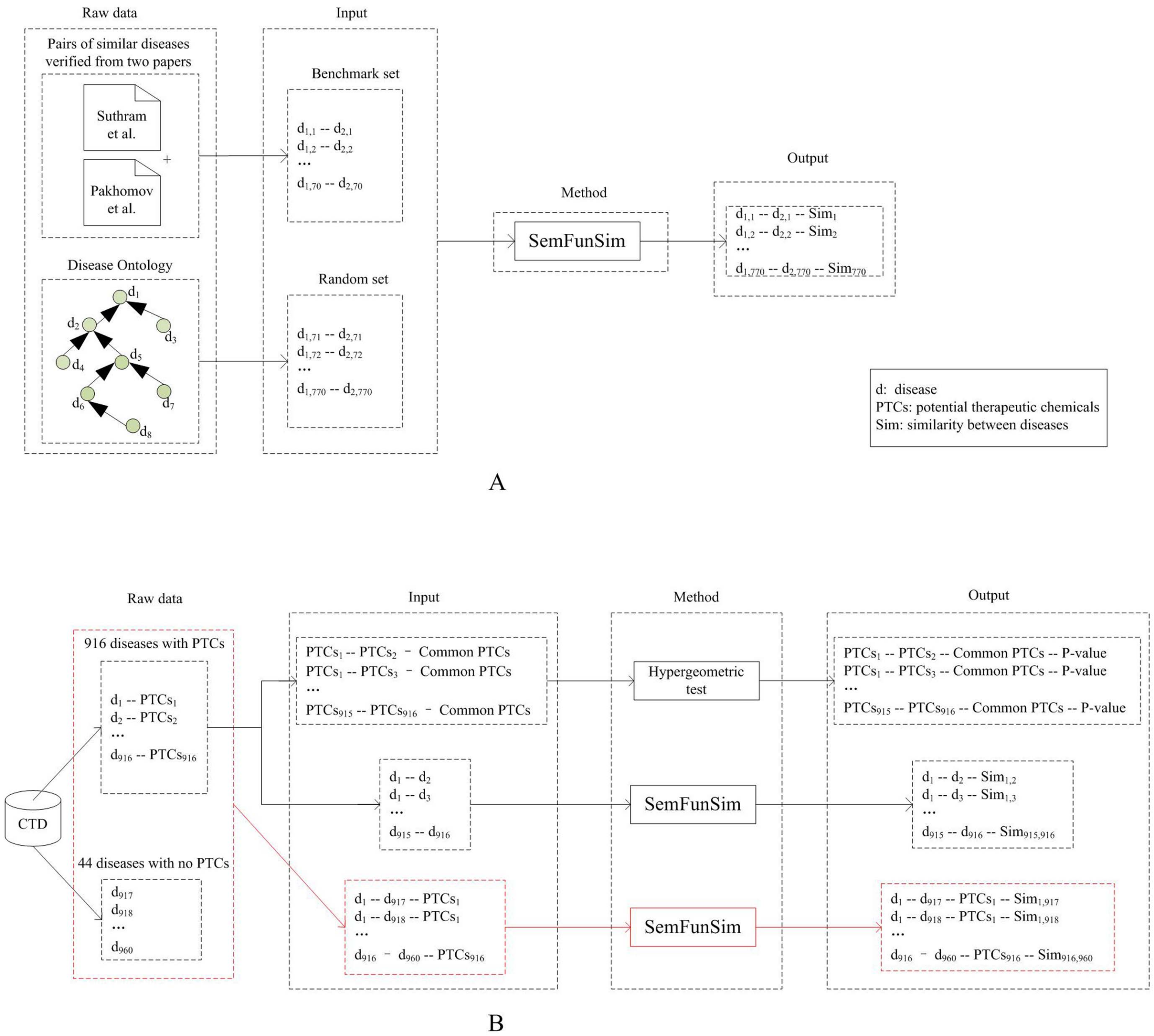
Raw data
Pairs of similar diseases verified from two papers
Suthram et al.
+
Pakhomov et al.
Disease Ontology
$d_1$
$d_2$
$d_3$
$d_4$
$d_5$
$d_6$
$d_7$
$d_8$
Input
Benchmark set
$d_{1,1}$ -- $d_{2,1}$
$d_{1,2}$ -- $d_{2,2}$
...
$d_{1,70}$ -- $d_{2,70}$
Random set
$d_{1,71}$ -- $d_{2,71}$
$d_{1,72}$ -- $d_{2,72}$
...
$d_{1,770}$ -- $d_{2,770}$
Method
SemFunSim
Output
$d_{1,1}$ -- $d_{2,1}$ -- $Sim_1$
$d_{1,2}$ -- $d_{2,2}$ -- $Sim_2$
...
$d_{1,770}$ -- $d_{2,770}$ -- $Sim_{770}$
d: disease
PTCs: potential therapeutic chemicals
Sim: similarity between diseases
A
Raw data
916 diseases with PTCs
$d_1$ -- $PTCs_1$
$d_2$ -- $PTCs_2$
...
$d_{916}$ -- $PTCs_{916}$
CTD
44 diseases with no PTCs
$d_{917}$
$d_{918}$
...
$d_{960}$
Input
$PTCs_1$ -- $PTCs_2$ – Common PTCs
$PTCs_1$ -- $PTCs_3$ – Common PTCs
...
$PTCs_{915}$ -- $PTCs_{916}$ – Common PTCs
$d_1$ -- $d_2$
$d_1$ -- $d_3$
...
$d_{915}$ -- $d_{916}$
$d_1$ -- $d_{917}$ -- $PTCs_1$
$d_1$ -- $d_{918}$ -- $PTCs_1$
...
$d_{916}$ – $d_{960}$ -- $PTCs_{916}$
Method
Hypergeometric test
SemFunSim
SemFunSim
Output
$PTCs_1$ -- $PTCs_2$ -- Common PTCs -- P-value
$PTCs_1$ -- $PTCs_3$ -- Common PTCs -- P-value
...
$PTCs_{915}$ -- $PTCs_{916}$ -- Common PTCs -- P-value
$d_1$ -- $d_2$ -- $Sim_{1,2}$
$d_1$ -- $d_3$ -- $Sim_{1,3}$
...
$d_{915}$ -- $d_{916}$ -- $Sim_{915,916}$
$d_1$ -- $d_{917}$ -- $PTCs_1$ -- $Sim_{1,917}$
$d_1$ -- $d_{918}$ -- $PTCs_1$ -- $Sim_{1,918}$
...
$d_{916}$ – $d_{960}$ -- $PTCs_{916}$ -- $Sim_{916,960}$
B

**Figure 3. The process of validation.** A. The similarities of disease pairs from the benchmark set and 100 random sets were calculated by SemFunSim, FunSim, Resnik, Wang, BOG, and PSB. B. The similarities of all the disease pairs between 916 diseases with PTCs in CTD were measured by SemFunSim, FunSim, Resnik, Wang, BOG, and PSB. In addition, the similarities of all the disease pairs between these 916 diseases with PTCs and 44 diseases without PTCs in CTD were computed by SemFunSim.
doi:10.1371/journal.pone.0099415.g003

[13]. During the experiment, the parameters of these methods are selected according to the original paper.

Similarities of disease pairs of the benchmark set and a random set were calculated by SemFunSim. We examined whether similarities of disease pairs of benchmark set could be prioritized in the top to produce an ROC curve. In Figure 4A, the area under the ROC curve (AUC) of each method is listed as follows, Resnik (63.14%), Wang (68.04%), BOG (78.10%), PSB (89.52%), and SemFunSim (96.36%). FunSim is part of SemFunSim, and has an AUC of 94.37%. The AUC shows that Wang et al.'s method is a little better than Resnik's method. The BOG method has the worst performance among function-based methods. When linking genes based on the GO biological process category [14] by the PSB method, the result has been improved significantly. Although the PSB method shows a very high AUC, FunSim still improves the results of the PSB method by about 5%. After integrating gene functional and semantic association, the SemFunSim method improves the performance further to nearly 100%. This experiment was iterated 100 times by calculating similarities of 100 random sets and the benchmark set. In Figure 4B, the average AUC of the 100 permutations is 0.6345, 0.6784, 0.7657, 0.8984, 0.9415, and 0.9637 for Resnik, Wang, BOG, PSB, FunSim, and SemFunSim, respectively. The result is consistent with Figure 4A.

Currently, functionally relevant gene associations can be defined in multiple ways (e.g. annotations for co-expression [25], protein-protein interaction [26], etc.). However, only one or two types of gene functional associations have been used to calculate the similarity by BOG and PSB [11,13]. FunSim was designed for calculating disease similarity based on a comprehensive weighted gene functional association network. In Figure 4, the AUC of

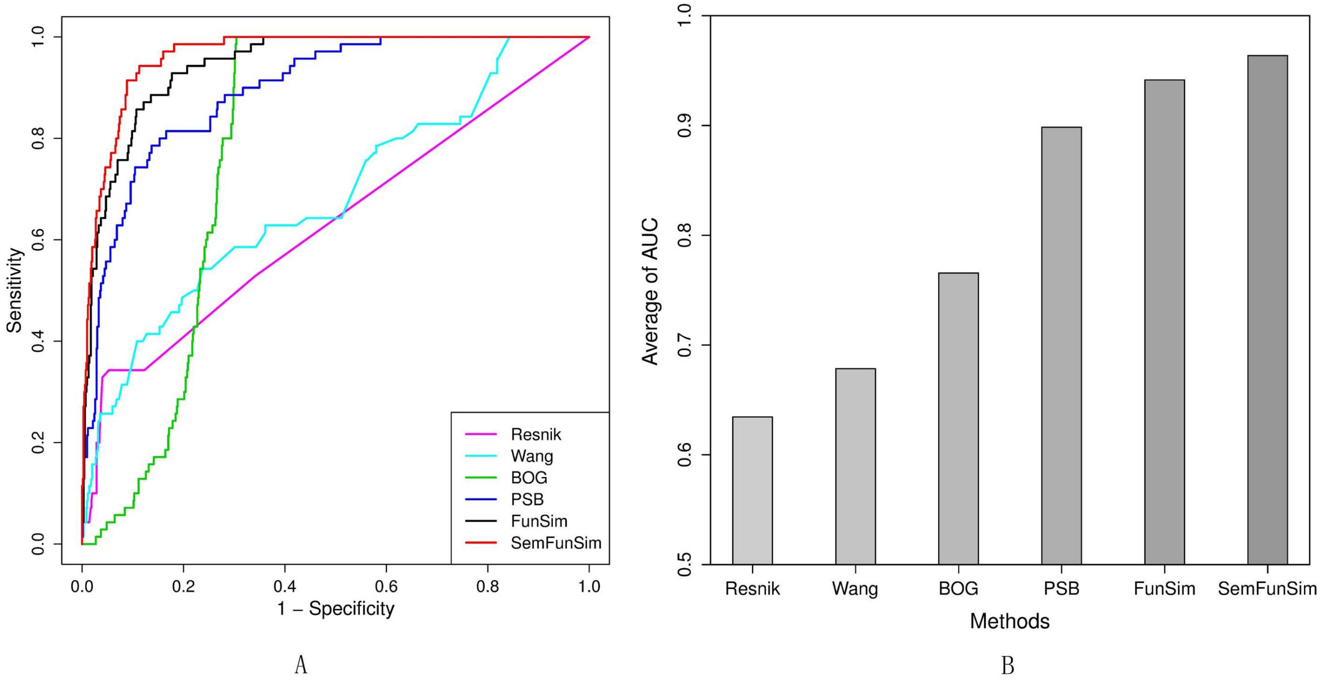

**Figure 4. AUC analysis of the benchmark set and random sets.** A. ROC curves for the experimental results on the benchmark set and a random set. It shows 1-specificity versus sensitivity of each method for calculating the similarities of disease pairs. B. Average of AUC for 100 permutations.
doi:10.1371/journal.pone.0099415.g004

FunSim is higher than BOG and PSB. The results show that comprehensive gene functional association is suitable for calculating disease similarity.

Among the five methods, Resnik's method used the IC of the MICA to calculate similarity between diseases. A few disease pairs of the benchmark set have only one common ancestor node, consequently the similarities of these diseases are zero according to Resnik (File S1). For example, the similarity between disease pair 'diabetes mellitus (DOID:9351)' and 'Alzheimer's Disease (DOID:10652)' is zero (File S1), because the MICA of these two diseases is the root node of DO (Figure S1), and the IC of the root node is zero. To avoid this problem for pairs of similar diseases with only one common ancestor, the IC is not used for measuring disease similarity in SemSim. The ROC curves in Figure 4A show clearly that SemFunSim has the highest AUC, which validates that the integrated semantic association helps to enhance the true positive rate and reduce the false positive rate.

## Assessment of disease similarity by means of common therapeutic compounds

CTD (Table 1) [33] was introduced to compare PTCs for diseases (Figure 3B). CTD not only documents disease-related genes, but also documents disease-related markers and potential therapeutic compounds for diseases. Only potential therapeutic compounds for diseases were extracted as PTCs. In a previous study, disease terms of CTD were integrated with DO [30]. After extracting PTCs for diseases from CTD, 916 diseases, 3,522 chemicals and 11,134 associations were retained (Dataset S3). In addition, 44 diseases without PTCs in CTD were also kept.

In order to illustrate the point that similar diseases can often be treated with similar drugs [9,47–49], PTCs for the top 100 pairs of similar diseases (T100-PSDs) and top 100 pairs of dissimilar diseases (T100-PDDs) (Dataset S4) identified using SemFunSim were compared. We counted the number of pairs with common PTCs and used a hypergeometric test to calculate the P-value for common PTCs for each pair of diseases. The P-value was adjusted by FDR [50]. There are 419,070 pairs between these 916 diseases. 1,251 pairs of them can be linked to each other by an 'IS_A' relationship of DO, which were not compared for avoiding diseases with common PTCs caused by the inclusion relationship. The results of the comparison are shown in Figure 5. 79 pairs of the T100-PSDs can be treated with common PTCs and 43 pairs have an adjusted P-value $<0.05$. In comparison, only 1 pair of the T100-PDDs can be treated with common PTCs and no pair has an adjusted P-value $<0.05$. The results show that the higher the similarity of a pair of diseases, the more likely they can be treated with common PTCs. Therefore, SemFunSim confirms the assumption that similar diseases can often be treated with similar drugs [9,47–49].

We further compared the PTCs for the T100-PSDs identified by the five methods (Dataset S5). The results are shown in Figure 6. 2, 15, 29, 31, 35, 79 pairs of the T100-PSDs identified by BOG, PSB, Resnik, FunSim, Wang and SemFunSim respectively can be treated with common PTCs, and 0, 4, 19, 17, 10, 43 pairs of the T100-PSDs identified by BOG, PSB, Resnik, FunSim, Wang and SemFunSim respectively have an adjusted P-value $<0.05$. FunSim is part of SemFunSim and is designed by considering comprehensive gene functional association. It identifies a higher number of pairs of diseases with common PTCs than BOG and PSB. It shows that disease similarity calculated by comprehensive gene function association is appropriate for taking advantage of the fact that similar diseases can often be treated with similar drugs [9,47–49]. The SemFunSim method identifies more than twice the number of pairs with common PTCs than the other methods. This confirms that SemFunSim is very suitable for the task.

The same test was applied to the top 500 pairs of similar diseases (T500-PSDs) and the top 1000 pairs of similar diseases

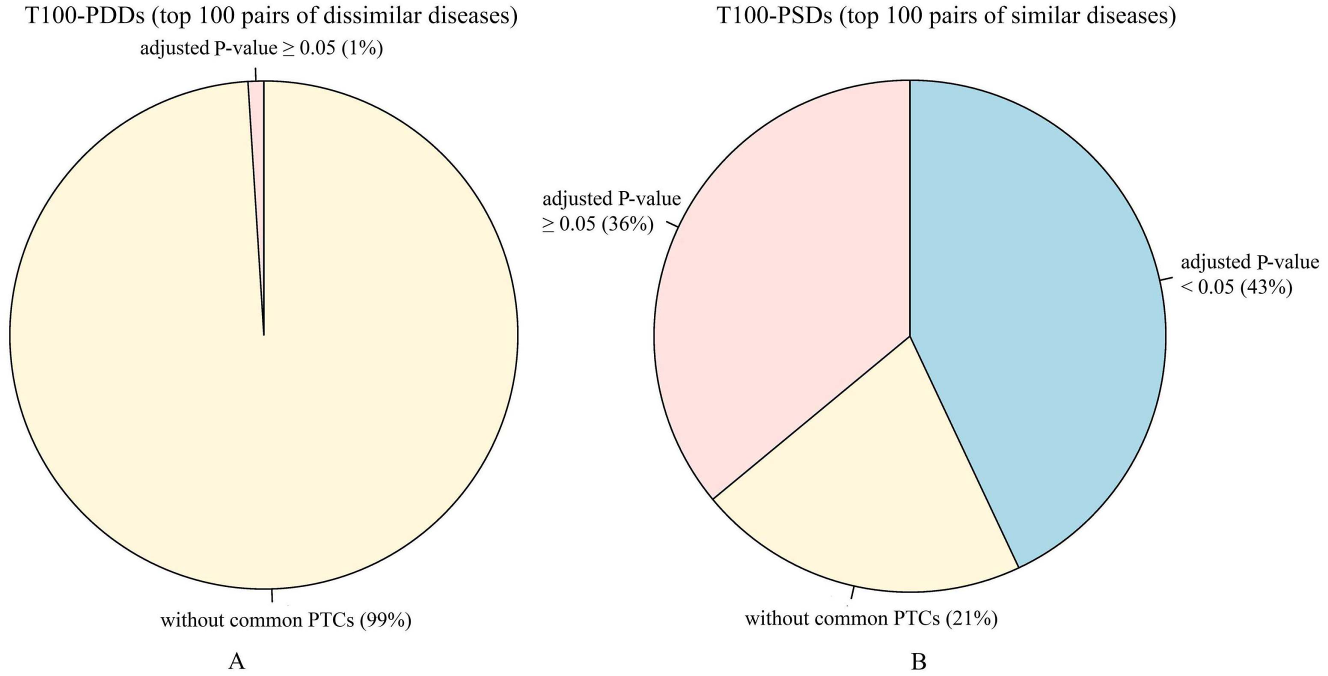

**Figure 5. The number of pairs of diseases identified using SemFunSim with common PTCs.** A. The number of pairs of the T100-PDDs with common PTCs. B. The number of pairs of the T100-PSDs with common PTCs. The yellow area represents the number of pairs without common PTCs. The pink area indicates the number of pairs with common PTCs and adjusted P-value ≥0.05. The light blue area represents the number of pairs with common PTCs and adjusted P-value <0.05.
doi:10.1371/journal.pone.0099415.g005

(T1000-PSDs) identified by the five methods (Dataset S5). The results are shown in Table S1. In the table, 57, 247, 281, 308, 457, and 556 pairs of the T1000-PSDs identified by BOG, Resnik, Wang, PSB, FunSim, and SemFunSim respectively can be treated with common PTCs. And 9, 99, 90, 104, 170, and 237 pairs of the T1000-PSDs identified by BOG, Resnik, Wang, PSB, FunSim, and SemFunSim respectively have an adjusted P-value <0.05. The performance of Resnik, FunSim and Wang appears to be roughly the same in the T100-PSDs. After comparing more pairs of similar diseases (T500-PSDs and T1000-PSDs), FunSim performs better than Resnik and Wang (Table S1). The experimental results in Table S1 show that SemFunSim has an advantage over other compared methods.

Using random permutations of the functional gene network and the 916 diseases with PTCs in CTD, as described in the Methods section, we defined thresholds for significant similarity. We found that 448 pairs of diseases have a similarity score above 0.06060 at an FDR less than 0.05, and 6,981 pairs of diseases have a similarity score above 0.00111 at an FDR less than 0.10. The FDRs for pairs of diseases with the similarity score above 0.00111 are listed in Dataset S6. The threshold can be defined as 0.06060 (FDR <0.05). In addition, researchers can also adjust the threshold to validate more disease pairs, such as 0.00111 (FDR <0.10).

In an early study, van Driel et al. [51] developed a tool (MimMiner), which was extensively used to calculate similarity between phenotype terms from OMIM [52]. We obtained the similarity score between 5,080 OMIM phenotype records from MimMiner (Table 1). As mentioned before, CTD includes 916 diseases with PTCs. 127 common diseases between the 5,080 OMIM phenotype records and these 916 diseases (Dataset S7) were found through DO's extensive cross mapping [21]. Then, SemFunSim and MimMiner were compared on the basis of these 127 diseases.

The result of the comparison is shown in Figure 7. 39, 129, and 218 pairs of the T100-PSDs, T500-PSDs, and T1000-PSDs identified by MimMiner respectively can be treated with common PTCs. And 17, 52, and 79 pairs of the T100-PSDs, T500-PSDs, and T1000-PSDs respectively have an adjusted P-value <0.05. In comparison, 74, 271, and 441 pairs of the T100-PSDs, T500-PSDs, and T1000-PSDs identified by SemFunSim respectively can be treated with common PTCs. And 43, 100, and 130 pairs of the T100-PSDs, T500-PSDs, and T1000-PSDs respectively have an adjusted P-value <0.05. Result shows that similar diseases identified using SemFunSim are very likely to be treated with common drugs.

We further compared MimMiner and SemFunSim based on their thresholds. 53 pairs of the 127 common diseases identified by MimMiner have a similarity >0.4 (threshold of MimMiner) (Dataset S7). 23 (43.4%—23/53) of them can be treated with common PTCs, and 9 (17.0%—9/53) have an adjusted P-value < 0.05. In comparison, 107 pairs of the 127 diseases identified by SemFunSim have a similarity >0.00111 (threshold of SemFunSim) (Dataset S7). 78 (72.9%—78/107) of them can be treated with common PTCs, and 44 (41.1%—44/107) have an adjusted P-value <0.05. The experiment results based on these 127 diseases show that SemFunSim's performance in measuring disease similarity is better than MimMiner's.

## Prediction of novel therapeutic applications of known compounds

SemFunSim was used to find PTCs for 44 diseases without PTCs in CTD. First, as shown in Figure 3B, we calculated similarities of 40,304 pairs between these 44 diseases and 916 diseases with PTCs in CTD (Dataset S8). In order to avoid

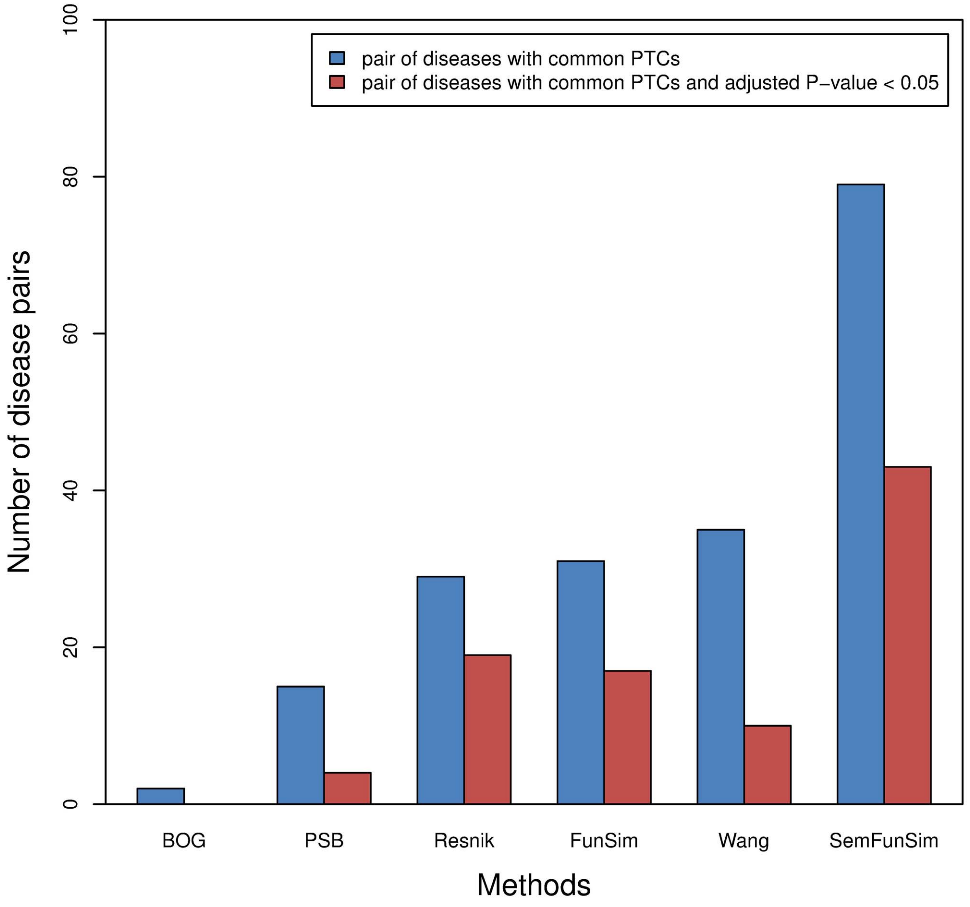

**Figure 6. The number of pairs of similar diseases identified using the five methods with common PTCs.** Blue bar indicates the number of pairs with common PTCs. Red bar represents the number of pairs with common PTCs and adjusted P-value <0.05.
doi:10.1371/journal.pone.0099415.g006

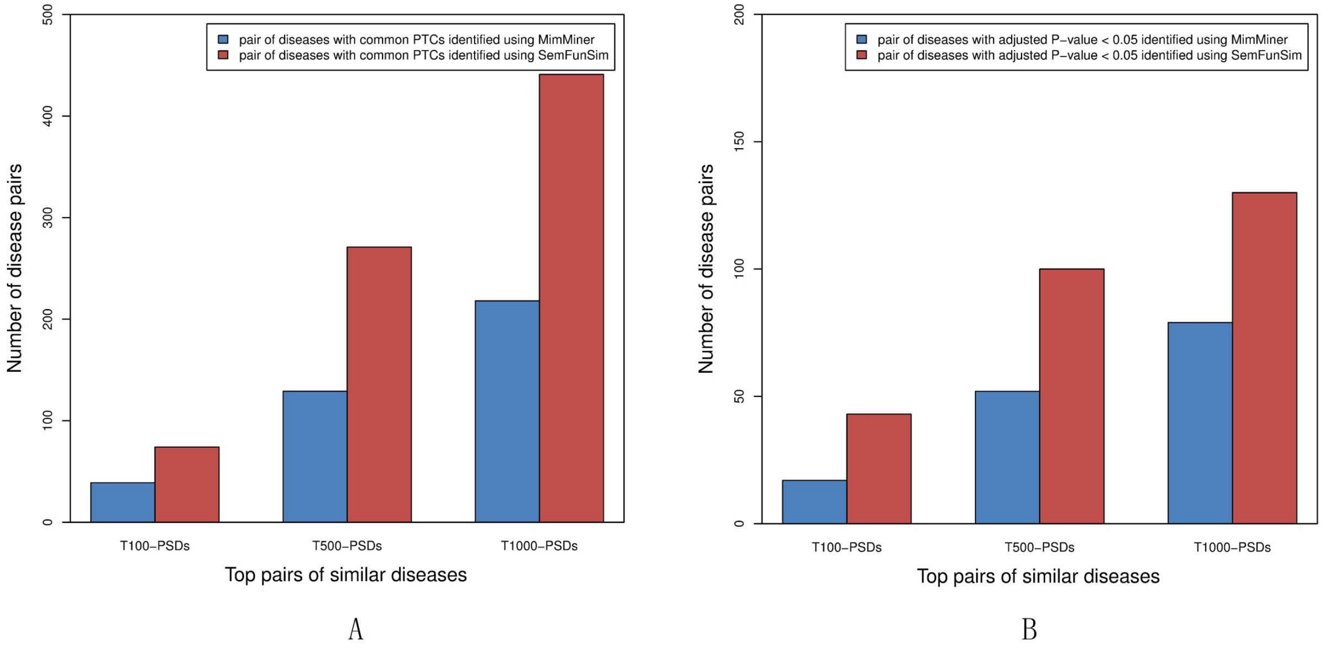

**Figure 7. The number of pairs of similar diseases identified using MimMiner and SemFunSim with common PTCs.** A. The number of pairs of the top pairs of similar diseases with common PTCs. The red bar represents the number of pairs with common PTCs measured by SemFunSim. The blue bar indicates the number of pairs with common PTCs measured by MimMiner. B. The number of pairs of the top pairs of similar diseases with common PTCs and adjusted P-value <0.05. The red bar represents the number of pairs with common PTCs and adjusted P-value <0.05 measured by SemFunSim. The blue bar indicates the number of pairs with common PTCs and adjusted P-value <0.05 measured by MimMiner.
doi:10.1371/journal.pone.0099415.g007

**Table 2.** Top 20 pairs of similar diseases.

| Order | Diseases with PTCs in CTD | Diseases without PTCs in CTD | Similarities |
|---|---|---|---|
| 1 | Liver Cirrhosis | Hepatopulmonary syndrome | 0.03460 |
| 2 | agranulocytosis | lymphopenia | 0.01665 |
| 3 | neutropenia | lymphopenia | 0.01566 |
| 4 | macroglobulinemia | alpha 1-antitrypsin deficiency | 0.01424 |
| 5 | hepatitis | hepatopulmonary syndrome | 0.00887 |
| 6 | wilson disease | hemochromatosis | 0.00862 |
| 7 | systemic scleroderma | polymyalgia rheumatica | 0.00717 |
| 8 | drug-induced hepatitis | hepatopulmonary syndrome | 0.00710 |
| 9 | myasthenia gravis | lambert-eaton myasthenic syndrome | 0.00644 |
| 10 | dilated cardiomyopathy | restrictive cardiomyopathy | 0.00643 |
| 11 | sarcoidosis | cryoglobulinemia | 0.00607 |
| 12 | berylliosis | asbestosis | 0.00600 |
| 13 | berylliosis | extrinsic allergic alveolitis | 0.00575 |
| 14 | intestinal disease | hepatopulmonary syndrome | 0.00564 |
| 15 | placenta disease | bacterial vaginosis | 0.00499 |
| 16 | hyperthyroidism | congenital hypothyroidism | 0.00461 |
| 17 | biliary tract disease | hepatopulmonary syndrome | 0.00454 |
| 18 | bile duct disease | hepatopulmonary syndrome | 0.00452 |
| 19 | inflammatory bowel disease | hepatopulmonary syndrome | 0.00447 |
| 20 | primary biliary cirrhosis | hepatopulmonary syndrome | 0.00421 |

The first column is the descending order number of similarity between diseases. The second column represents diseases with PTCs in CTD. The third column indicates diseases without PTCs in CTD. The fourth column represents the similarities between pairs of diseases in the second and third columns.
doi:10.1371/journal.pone.0099415.t002

diseases with common PTCs caused by the inclusion relationship, 64 of the 40,304 pairs which can be linked with each other by an 'IS_A' relationship of DO were not included. Each pair of the 40,240 pairs includes one disease without PTCs in CTD and one disease with PTCs in CTD. The top 20 pairs of similar diseases (T20-PSDs) (Table 2) contain 12 diseases without PTCs in CTD. Then, we searched PubMed to find PTCs for these 12 diseases. According to the idea that similar diseases can often be treated with similar drugs, the PTCs for one disease in pair of similar diseases can be used as a reference for the other without PTCs. For example, 'systemic scleroderma' is similar with 'polymyalgia rheumatica', 11 PTCs for the former are docu-

**Table 3.** Associations between PTCs and diseases retrieved from PubMed.

| Order | PTCs | Diseases with PTC in CTD | Diseases without PTC in CTD | Similarities | PMIDs |
|---|---|---|---|---|---|
| 1 | Pentoxifylline | Liver Cirrhosis | Hepatopulmonary syndrome | 0.03460 | 23002364 [57] |
| 5 | Acetylcysteine | hepatitis | hepatopulmonary syndrome | 0.00887 | 18341514 [58] |
| 7 | Azathioprine | systemic scleroderma | polymyalgia rheumatica | 0.00717 | 2750226 [53] |
| 7 | Methylprednisolone | systemic scleroderma | polymyalgia rheumatica | 0.00717 | 1768166 [54] |
| 7 | Prednisolone | systemic scleroderma | polymyalgia rheumatica | 0.00717 | 8523341 [55] |
| 7 | Prednisone | systemic scleroderma | polymyalgia rheumatica | 0.00717 | 15466766 [56] |
| 8 | Pentoxifylline | drug-induced hepatitis | hepatopulmonary syndrome | 0.00710 | 23002364 [57] |
| 9 | Prednisolone | myasthenia gravis | lambert-eaton myasthenic syndrome | 0.00644 | 10555101 [59] |
| 11 | Methylprednisolone | sarcoidosis | cryoglobulinemia | 0.00607 | 6851261 [60] |
| 13 | Prednisone | berylliosis | extrinsic allergic alveolitis | 0.00575 | 9489437 [61] |
| 16 | Methimazole | hyperthyroidism | congenital hypothyroidism | 0.00461 | 22672871 [62] |
| 19 | Acetylcysteine | inflammatory bowel disease | hepatopulmonary syndrome | 0.00447 | 18341514 [58] |

The first column is the descending order number of similarity between diseases. PTCs (in the second column) for diseases (in the third column) are documented in CTD. The fourth column represents diseases without PTCs in CTD. The fifth column indicates the similarities between pairs of diseases in the third and fourth column. The sixth column is the PubMed IDs that record the associations between PTCs (in the second column) and diseases (in the fourth column).
doi:10.1371/journal.pone.0099415.t003

mented in CTD. We searched from PubMed for finding associations between these 11 PTCs and 'polymyalgia rheumatica'. And we found that four of them were also PTCs for 'polymyalgia rheumatica', such as azathioprine [53], Methylprednisolone [54], Prednisolone [55] and Prednisone [56]. Finally, 6 of these 12 diseases from the T20-PSDs can be treated with PTCs confirmed by literature. The detailed results are listed in Table 3, which indicate that SemFunSim is an effective method to find PTCs for diseases.

## Conclusions

In this article, we devise an algorithm (SemFunSim) to measure disease similarity by integrating FunSim and SemSim effectively. Experimental evaluation was performed on the benchmark set and 100 random sets from DO. The high average AUC (96.37%) shows that SemFunSim achieves a high true positive rate and a low false positive rate.

SemFunSim is in agreement with the notion that similar diseases can often be treated with similar drugs [9,47–49]. SemFunSim not only helps to understand associations between diseases, but also provides an effective way to predict PTCs for diseases. We found associations between diseases and PTCs that were not documented in CTD using SemFunSim (Table 3).

## Supporting Information

**File S1 Description of IC and MICA.**
(DOCX)

**Figure S1 A sub-graph of the DAG for DO term 'pick's disease (DOID:11870)', 'Alzheimer's Disease (DOID:10652)' and 'Diabetes mellitus (DOID:9351)'.** The arrow symbol represents an 'IS_A' link of DO. For example, "Alzheimer's Disease (DOID:10652)" is linked to "Dementia (DOID:1307)" by an 'IS_A' relationship.
(DOCX)

**Table S1 The number of pairs of the T500-PSDs and T1000-PSDs measured by the five methods with common PTCs.**
(DOCX)

**Dataset S1 Associations between diseases and genes.**
(XLSX)

**Dataset S2 Benchmark set.**
(XLSX)

**Dataset S3 Associations between diseases and PTCs.**
(XLSX)

**Dataset S4 The T100-PSDs and T100-PDDs measured by SemFunSim.** 916 diseases can be treated with PTCs in CTD. The T100-PSDs and T100-PDDs between these 916 diseases were identified by SemFunSim. The disease pairs, the number of common PTCs between diseases, and adjusted P-values are listed.
(XLSX)

**Dataset S5 The T1000-PSDs measured by the five methods.** 916 diseases can be treated with PTCs in CTD. The T1000-PSDs between these 916 diseases were accessed by SemFunSim, FunSim, Wang, Resnik, PSB, and BOG. The disease pairs, the number of common PTCs between diseases, and adjusted P-values are listed.
(XLSX)

**Dataset S6 FDRs for the similarity scores.** The FDRs for pairs of diseases with the similarity score above 0.00111 are listed.
(XLSX)

**Dataset S7 The T1000-PSDs measured by MimMiner and SemFunSim.** The T1000-PSDs were identified using MimMiner and SemFunSim. The disease pairs, the number of common PTCs between diseases, and adjusted P-values are listed. In addition, 127 DO terms and their extensive cross-referenced OMIM phenotype records are listed.
(XLSX)

**Dataset S8 Similarities between 44 diseases without PTCs and 916 diseases with PTCs in CTD measured by SemFunSim.** We calculated the similarities between 44 diseases (without PTCs in CTD) and 916 diseases (with PTCs in CTD) based on SemFunSim. These 44 diseases without PTCs in CTD and the similarities are listed.
(XLSX)

## Acknowledgments

The authors thank Yanshuo Chu, Zhenxing Wang, Yang Hu, and Kang Xu at Harbin Institute of Technology for informative suggestions. We also thank the reviewers for their constructive comments.

## Author Contributions

Conceived and designed the experiments: YW LC JL. Performed the experiments: LC. Analyzed the data: YW LC JL PJ. Contributed reagents/materials/analysis tools: LC. Wrote the paper: YW LC JL PJ JP.

## References

1. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. (2007) The human disease network. Proc Natl Acad Sci U S A 104: 8685–8690.
2. Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. PLoS One 4: e6536.
3. Zhang X, Zhang R, Jiang Y, Sun P, Tang G, et al. (2011) The expanded human disease network combining protein-protein interaction information. Eur J Hum Genet 19: 783–788.
4. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, et al. (2008) The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A 105: 9880–9885.
5. Li Y, Agarwal P (2009) A pathway-based view of human diseases and disease relationships. PLoS One 4: e4346.
6. Lage K, Karlberg EO, Storling ZM, Olason PI, Pedersen AG, et al. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 25: 309–316.
7. Wu X, Liu Q, Jiang R (2009) Align human interactome with phenome to identify causative genes and networks underlying disease families. Bioinformatics 25: 98–104.
8. Wang D, Wang J, Lu M, Song F, Cui Q (2010) Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics 26: 1644–1650.
9. Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7: 496.
10. Li J, Gong B, Chen X, Liu T, Wu C, et al. (2011) DOSim: an R package for similarity between diseases based on Disease Ontology. BMC Bioinformatics 12: 266.
11. Mathur S, Dinakarpandian D (2010) Automated ontological gene annotation for computing disease similarity. AMIA Summits Transl Sci Proc 2010: 12–16.
12. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, et al. (2010) Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 6: e1000662.
13. Mathur S, Dinakarpandian D (2012) Finding disease similarity based on implicit semantic similarity. J Biomed Inform 45: 363–371.
14. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.
15. Pesquita C, Faria D, Falcao AO, Lord P, Couto FM (2009) Semantic similarity in biomedical ontologies. PLoS Comput Biol 5: e1000443.
16. Robinson PN, Mundlos S (2010) The human phenotype ontology. Clin Genet 77: 525–534.

17. Resnik P (1995) Using information content to evaluate semantic similarity in a taxonomy. Proceedings of the 14th international joint conference on artificial intelligence. Morgan Kaufmann Publishers Inc. pp. 448–453.
18. Guo X, Liu R, Shriver CD, Hu H, Liebman MN (2006) Assessing semantic similarity measures for the characterization of human regulatory pathways. Bioinformatics 22: 967–973.
19. Jiang JJ, Conrath DW (1997) Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint cmp-lg/9709008.
20. Lin D (1998) An information-theoretic definition of similarity. Proceedings of the 15th international conference on Machine Learning. San Francisco, CA: Morgan Kaufmann. pp. 296–304.
21. Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, et al. (2012) Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Res 40: D940–946.
22. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF (2007) A new method to measure the semantic similarity of GO terms. Bioinformatics 23: 1274–1281.
23. Lowe HJ, Barnett GO (1994) Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA: the journal of the American Medical Association 271: 1103–1108.
24. Leacock C, Chodorow M (1998) Combining local context and WordNet similarity for word sense identification. WordNet: An electronic lexical database 49: 265–283.
25. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302: 249–255.
26. Ortutay C, Vihinen M (2009) Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies. Nucleic Acids Res 37: 622–628.
27. Schlicker A, Lengauer T, Albrecht M (2010) Improving disease gene prioritization using the semantic similarity of Gene Ontology terms. Bioinformatics 26: i561–i567.
28. Xu W, Wang H, Cheng W, Fu D, Xia T, et al. (2012) A framework for annotating human genome in disease context. PLoS One 7: e49686.
29. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21: 1109–1121.
30. Cheng L, Wang G, Li J, Zhang T, Xu P, et al. (2013) SIDD: A Semantically Integrated Database towards a Global View of Human Disease. PLoS One 8: e75504.
31. Mitchell JA, Aronson AR, Mork JG, Folk LC, Humphrey SM, et al. (2003) Gene indexing: characterization and analysis of NLM's GeneRIFs. AMIA Annu Symp Proc: 460–464.
32. Amberger J, Bocchini C, Hamosh A (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). Hum Mutat 32: 564–567.
33. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, et al. (2013) The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res 41: D1104–1114.
34. Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. Nat Genet 36: 431–432.
35. Wang J, Zhang J, Li K, Zhao W, Cui Q (2012) SpliceDisease database: linking RNA splicing and disease. Nucleic Acids Res 40: D1055–1059.
36. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, et al. (2013) Genenames.org: the HGNC resources in 2013. Nucleic Acids Res 41: D545–552.
37. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. Cell 144: 986–998.
38. Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. Nature 461: 218–223.
39. Navlakha S, Kingsford C (2010) The power of protein interaction networks for associating genes with diseases. Bioinformatics 26: 1057–1063.
40. Linghu B, Snitkin ES, Hu Z, Xia Y, Delisi C (2009) Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. Genome Biol 10: R91.
41. Kann MG (2007) Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform 8: 333–346.
42. Chen J, Xu H, Aronow BJ, Jegga AG (2007) Improved human disease candidate gene prioritization using mouse phenotype. BMC bioinformatics 8: 392.
43. Baudot A, Gomez-Lopez G, Valencia A (2009) Translational disease interpretation with molecular networks. Genome Biol 10: 221.
44. Peng J, Chen J, Wang Y (2013) Identifying cross-category relations in gene ontology and constructing genome-specific term association networks. BMC Bioinformatics 14 Suppl 2: S15.
45. Heagerty PJ, Zheng Y (2005) Survival model predictive accuracy and ROC curves. Biometrics 61: 92–105.
46. Pakhomov S, McInnes B, Adam T, Liu Y, Pedersen T, et al. (2010) Semantic similarity and relatedness between clinical terms: an experimental study. American Medical Informatics Association. pp. 572.
47. Dakshanamurthy S, Issa NT, Assefnia S, Seshasayee A, Peters OJ, et al. (2012) Predicting new indications for approved drugs using a proteochemometric method. Journal of Medicinal Chemistry 55: 6832–6848.
48. Chan SY, Loscalzo J (2012) The emerging paradigm of network medicine in the study of human disease. Circ Res 111: 359–374.
49. Li J, Lu Z (2012) A new method for computational drug repositioning using drug pairwise similarity. Proceedings of The IEEE International Conference on Bioinformatics and Biomedicine. pp. 1–4.
50. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological): 289–300.
51. van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA (2006) A text-mining analysis of the human phenome. Eur J Hum Genet 14: 535–542.
52. Piro RM (2012) Network medicine: linking disorders. Hum Genet 131: 1811–1820.
53. Reinhardt U (1989) [Azathioprine therapy in polymyalgia rheumatica]. Z Arztl Fortbild (Jena) 83: 489–490.
54. Dasgupta B, Gray J, Fernandes L, Olliff C (1991) Treatment of polymyalgia rheumatica with intramuscular injections of depot methylprednisolone. Ann Rheum Dis 50: 942–945.
55. Krogsgaard MR, Lund B, Johnsson B (1995) A longterm prospective study of the equipotency between deflazacort and prednisolone in the treatment of patients with polymyalgia rheumatica. J Rheumatol 22: 1660–1662.
56. Caporali R, Cimmino MA, Ferraccioli G, Gerli R, Klersy C, et al. (2004) Prednisone plus methotrexate for polymyalgia rheumatica: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 141: 493–500.
57. Kianifar HR, Khalesi M, Mahmoodi E, Afzal Aghaei M (2012) Pentoxifylline in hepatopulmonary syndrome. World J Gastroenterol 18: 4912–4916.
58. Vercelino R, Tieppo J, Dias AS, Marroni CA, Garcia E, et al. (2008) N-acetylcysteine effects on genotoxic and oxidative stress parameters in cirrhotic rats with hepatopulmonary syndrome. Basic Clin Pharmacol Toxicol 102: 370–376.
59. Takata T, Koide S, Ogata K, Motomura M, Yoshimura T, et al. (1999) Lambert-Eaton myasthenic syndrome associated with idiopathic thrombocytopenic purpura and diffuse panbronchiolitis: long-term remission after a course of intravenous immunoglobulin combined with low-dose prednisolone. Am J Med Sci 318: 353–355.
60. De Vecchi A, Montagnino G, Pozzi C, Tarantino A, Locatelli F, et al. (1983) Intravenous methylprednisolone pulse therapy in essential mixed cryoglobulinemia nephropathy. Clin Nephrol 19: 221–227.
61. Ziora D, Szurkowska J, Oklek K (1997) [Two year observation of patients with chronic extrinsic allergic alveolitis during steroid treatment]. Pneumonol Alergol Pol 65: 635–642.
62. Kurtoglu S, Akin MA, Daar G, Akin L, Memur S, et al. (2012) Congenital hypothyroidism due to maternal radioactive iodine exposure during pregnancy. J Clin Res Pediatr Endocrinol 4: 111–113.