Mining Biomedical Literature and Ontologies for Drug Repositioning Discovery

Chih-Ping Wei, Kuei-An Chen, and Lien-Chin Chen

Department of Information Management, National Taiwan University, Taipei, Taiwan, R.O.C.
{cpwei,r00725020,lcchen101}@ntu.edu.tw

Abstract. Drug development is time-consuming, costly, and risky. Approximate 80% to 90% of drug development projects fail before they ever get into clinical trials. To reduce the high risk of failure for drug development, pharmaceutical companies are exploring the drug repositioning approach for drug development. Previous studies have shown the feasibility of using computational methods to help extract plausible drug repositioning candidates, but they all encountered some limitations. In this study, we propose a novel drug-repositioning discovery method that takes into account multiple information sources, including more than 18,000,000 biomedical research articles and some existing ontologies that cover detailed relations between drugs, proteins and diseases. We design two experiments to evaluate our proposed drug repositioning discovery method. Overall, our evaluation results demonstrate the capability and superiority of our proposed drug repositioning method for discovering potential, novel drug-disease relationships.

Keywords: Drug repositioning, Drug repurposing, Literature-based discovery, Medical literature mining.

1 Introduction

Drug development is time-consuming, costly, and risky. The development process to bring a new drug to market requires about 10-15 years and costs between 500 million and 2 billion U.S. dollars [1]. However, as the U.S. National Institutes of Health reported, 80 to 90 percent of drug development projects fail before they ever get tested in human [2]. To reduce the high risk of failure for *de novo* drug development, pharmaceutical companies have been evaluating alternative paradigms for drug development, e.g., drug repositioning. The goal of drug repositioning is to find new indications (i.e., treatment for diseases) for existing drugs. Because existing drugs already have their preclinical properties and established safety profiles, many experiments, analyses and tests can therefore be bypassed [3]. Thus, drug repositioning can reduce significant time and cost in the discovering and preclinical stage. Moreover, drug repositioning helps a company exploit its intellectual property portfolio by extending its old or expiring patents, or getting new method-of-use patents [3].

One notable example of repositioned drug is Thalidomide. It was originally marketed as a sedative and antiemetic for pregnant women to treat morning sickness,

V.S. Tseng et al. (Eds.): PAKDD 2014, Part II, LNAI 8444, pp. 373–384, 2014.

but was completely withdrawn from the market after the drug was found responsible for severe birth defects [4]. After Celgene Corporation's repositioning works, FDA approved Thalidomide for use in the treatment of Erythema Nodosum Leprosum (ENL) in 1998 [5]. The company further discovered that Thalidomide is also effective against several other diseases, including multiple myeloma. Accordingly, Celgene gets several utility patents for the repositioned Thalidomide, and it brings in over 300 million U.S. dollars in revenue annually since 2004 [6-8].

Several computational drug repositioning approaches have been developed to help medical researchers sift the most plausible drug-disease pairs from a wide range of combinations. Dudley et al. [9] summarized several computational drug repositioning methods and categorized them as either drug-based, where discovery from the chemical or pharmaceutical perspective, or disease-based, where discovery from the perspective of disease management, symptomatology, or pathology. Another excellent review in [10] highlights computational techniques for systematic analysis of transcriptomics, side effects and genetics data to generate new hypotheses for additional indications. Several network based approaches use various heterogeneous data resources to discovery drug repositioning opportunity [11, 12]. Moreover, Wu et al. [13] summarized 26 different sources of databases related to disease, genes, proteins, and drugs for drug repositioning. In summary, existing drug-repositioning methods can broadly be classified into two approaches: *literature-based* and *ontology-based*. The literature-based approach assumes that if a drug frequently co-occurs with some biomedical concepts (such as enzymes, genes, pathological effects, and proteins) and many of these concepts also frequently co-occur with a disease in biomedical literature (e.g., MEDLINE), it is likely that the disease is a new indication for the focal drug [14]. In contrast, the ontology-based approach relies on existing ontologies (or knowledge bases) to discover hidden relationships between drugs and diseases. These existing methods have shown their feasibility for drug repositioning. However, most existing methods rely only on single information source, i.e., literature or ontologies.

In this study, we propose a drug repositioning method that exploits multiple information sources to discover hidden relationships between drugs and diseases. Specifically, a comprehensive network of biomedical concepts is first constructed by combining and integrating relations of biomedical concepts extracted from literature and existing ontologies. Subsequently, we follow Swanson's ABC model [14] to obtain links between a focal drug (*A*) and intermediate terms (*B*s) and then between *B*s and diseases (*C*s) from the comprehensive concept network. A novel link weighting method and two target term ranking measures are proposed to effectively rank candidate diseases that are likely to be new indications of the focal drug *A*. To evaluate the proposed method, we collect the literature from MEDLINE and three ontologies (i.e., DrugBank [15], Online Mendelian Inheritance in Man (OMIM) [16], and Comparative Toxicogenomics Database (CTD) [17]), and follow the evaluation procedure proposed in [18] to conduct a series of experiments. According to our empirical evaluation results, our proposed method outperforms the existing method for drug repositioning.

The remainder of this paper is organized as follows: Section 2 reviews existing methods related to this study, and discuss their limitations to justify our research

motivation. In Section 3, we describe the design of our proposed drug repositioning discovery method. Section 4 reports on our evaluation of our proposed method. Finally, we conclude our study in Section 5.

2 Related Work

In this section, we review existing drug repositioning methods, which can be classified into two major approaches: literature-based and ontology-based.

2.1 Literature-Based Approach

Swanson [19] first introduced the idea of discovering hidden relationships from biomedical literatures in the mid-1980s. He examined across disjoint literatures, manually identified plausible new connections, and found fish oil might be beneficial to the treatment of Raynaud's syndrome [19]. Furthermore, Swanson and Smalheiser developed a computational model, namely "*ABC model*" or "*undiscovered public knowledge* (*UPK*) *model*" [14]. The basic assumption of ABC model is that if a biomedical concept *A* relates to intermediate concept *B* and intermediate concept *B* relates to another concept *C*, there is a logically plausible relation between *A* and *C*. The ABC model generally consists of three major phases: *term selection* to extract textual terms (concepts) from the literature, *link weighting* to assess the link strength between two concepts, and *target term ranking* to rank target terms by assigning a score to each target term on the basis of the connections and link weights between the starting term and the target term. This approach is often referred to as the literature-based discovery.

Weeber et al. [20] followed Swanson's idea of co-occurrence analysis and mapped words from titles and abstracts extracted from MEDLINE articles to Unified Medical Language System (UMLS) concepts to filter link candidates with the help of semantic information. Similarly, Wren et al. [21] mapped full text from articles into OMIM concepts. They measured link weights between concepts by mutual information. Lee et al. [22] further combined multiple thesauruses to better translate text into biomedical concepts. These studies employed the full text for concept extraction with the help of thesauri. On the other hand, some other studies used Medical Subject Headings (MeSH) as keywords to annotate each article in MEDLINE [18] [23-24]. They applied tf-idf, association rule, and z-score as the measurement of link weights. All of them reported the metadata-only approach is feasible, though Hristovski et al. [24] noted some shortcoming of using MeSH such as insufficient information of involving genes. Based on the ABC model, several drug repositioning methods [20-21] [25-26] were proposed to find undiscovered relations between drugs and diseases through selecting different semantic groups of intermediate terms such as adverse effects, genes, and proteins.

For evaluating the performance of the literature-based approach, Yetisgen-Yildiz and Pratt [18] developed an evaluation methodology. They used two literature sets collected from separated time spans, and trained systems by using the older set to

predict novel relations in the newer set. They compared the effectiveness of various link weighting methods. According to their study, association rule appears to achieve the best performance over tf-idf, mutual information measure, and z-score. They also compared different target term ranking algorithms and suggested that the use of link term count with average minimum weight can achieve the best effectiveness.

2.2 Ontology-Based Approach

Campillos et al. [27] constructed a network of side-effect driven drug-drug relations from UMLS ontology by measuring side-effect similarity between drugs. Assuming that similar side effects of unrelated drugs may be caused by common targets, they can be used to predict new drug-target interactions. They also experimentally validated their results, and thus reported the feasibility of using phenotypic information to infer unexpected biomedical relations. Yang and Agarwal [28] also based on side effect likelihood between drugs, but they constructed Naïve Bayes models to make predictions. They took PharmGKB and SIDER knowledge bases, rather than phenotype database, as their information sources. Cheng et al. [29] built a bipartite network by extracting known drug-target interaction data from DrugBank, and used the network similarity to predict new targets of drugs. Li and Lu [30] built a network similar to Cheng et al.'s work, but added the similarity of drug chemical structure into consideration.

Qu et al. [31] and Lee et al. [26] both attempted to increase the size and scope of semantic data by constructing integrated network or database of ontologies. However, to our best knowledge, few prior studies, if any, take both ontologies and literature into account, which may be a good way to acquire deeper and broader biomedical knowledge for making predictions of drug-disease relations. Li et al. [32] tried to incorporate more knowledge by using protein-protein interactions extracted from Online Predicted Human Interaction Database (OPHID) to expand disease-related proteins, and built disease-specific drug-protein connectivity maps based on literature mining. His work inspires us to build a network over multiple information sources.

3 The Proposed Method

We propose a drug repositioning discovery method that is based on Swanson's ABC model [14] but takes both biomedical literature and existing ontologies into account. Fig. 1 illustrates our proposed method, which consists of four main phases: *comprehensive concept network construction*, *related concept retrieval*, *link weighting*, and *target term ranking*.

3.1 Phase 1: Comprehensive Concept Network Construction

The goal of this phase is to construct a literature-based concept network from the biomedical literature and an ontology-based concept network from existing ontologies (i.e., DrugBank, OMIM, and CTD in this study) and, subsequently, integrate them into a comprehensive concept network.

For the literature-based concept network construction, we collect the biomedical literature from MEDLINE 2011 baseline. U.S. National Library of Medicine (NLM) indexes the publication type for each article. We follow Yetisgen-Yildiz's preprocessing procedure [18] to remove 18 irrelevant types (e.g., address, bibliography, comment, etc.) from the 61 publication types in MEDLINE 2011 baseline. NLM also indexes several MeSH terms for each biomedical article. As a result, our literature database consists of 18,712,338 biomedical articles. The number of MeSH terms per article ranges from 1 to 97, and its average is 9.44. We further select several MeSH subcategories related to drug repositioning, as shown in Table 1. Next, association rule mining is applied on the collected literature where biomedical articles and MeSH terms are considered as transactions and items, respectively. We follow Yetisgen-Yildiz and Pratt's experiment [18] by setting the minimum support threshold to 2.6 and the minimum confidence threshold to 0.0055 in this study. After filtering, we extract 12,278 MeSH terms and 2,623,222 relations to construct a literature-based concept network.

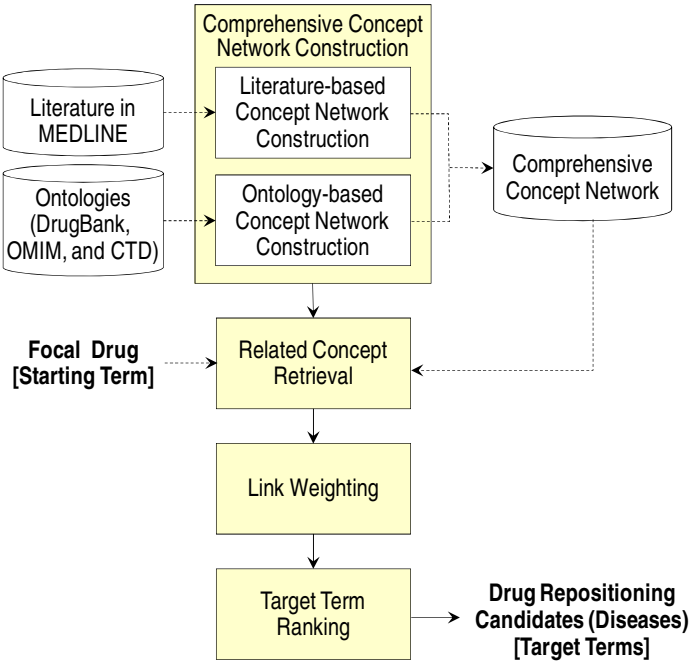

Fig. 1. Overall Process of the Proposed Drug Repositioning Discovery Method

For the ontology-based concept network construction, the ontologies we use in this study include DrugBank [15], OMIM [16], and CTD [17]. DrugBank (http://www.drugbank.ca/) is a richly annotated database which provides extensive information about targets, pathways, indications, adverse effects, and related proteins of various drugs, whereas OMIM (http://www.omim.org/) is a comprehensive and

Table 1. Selected MeSH Subcategories

Semantic Group	MeSH Subcategories
Drugs	D01-D05, D09, D10, D20, D26, D27
Genes, Proteins, and Enzymes	D06, D08, D12, D13, D23
Pathological Effects	G03-G16
Diseases	C01-C23

authoritative knowledge base of human genes and genetic phenotypes. CTD (http://ctdbase.org/) is a database that integrates data from scientific literature to describe chemical interactions with genes and proteins, and diseases and genes/proteins, and others. The collected information includes various types of relations, such as drug-target, gene-disease, gene-gene, protein-protein, chemical-gene, gene-disease relations, etc. And, we also try to translate those terms to MeSH term and select only MeSH subcategories shown in Table 1. The summary of these collected ontology-based relations is shown in Table 2.

Table 2. Summary of Extracted Ontology-based Relations

Data Source	Number of Selected Relations
DrugBank	7,808 drug-target interactions
OMIM	2,404 gene-disease relations
CTD	195,033 chemical-gene interactions and 27,397 gene-disease associations with direct evidences

3.2 Phase 2: Related Concept Retrieval

Given a user-specified focal drug, the goal of this phase is to retrieve related relations from the comprehensive concept network and to construct a subgraph for the focal drug. In other words, given a focal drug, we retrieve concepts related to the given drug in the network as intermediate terms. These intermediate terms may be gene, protein, disease, etc. Then, we extract the disease concepts that related to these intermediate terms but not related to the given drug.

3.3 Phase 3: Link Weighting

Given the extracted subgraph for the user-specified drug, the goal of this phase is to weight each link in the subgraph according to different weighting strategies for literature-based links and ontology-based links. We propose a link weighting method, namely *Extended Normalized MEDLINE Similarity* (ExtNMS), based on the Normalized Google Distance (NGD) [33] to calculate similarity between MeSH term A and B as follows:

$$NMD(A,B) = \frac{\max\{\log|D_A|,\log|D_B|\}-\log|D_A\cap D_B|}{logM-\min\{\log|D_A|,\log|D_B|\}}, \text{NMD} = 1 \text{ if NMD} > 1, \quad (1)$$

$$NMS(A,B) = 1 - NMD(A,B), \quad (2)$$

$$ExtNMS(A,B) = \begin{cases} NMS(A,B), & if(A,B) \in Literature\ and \notin Ontology \\ 1, & if(A,B) \in Ontology \end{cases}, \quad (3)$$

where D_A is the set of articles including MeSH term A, and M denotes the total number of articles in MEDLINE. This weighting measure ranges from 0 to 1, in which 0 being completely unrelated and 1 being credibly related. If link (A,B) is from the literature-based concept network and is not found in the ontology-based concept network, we weight it by calculating its *Normalized MEDLINE Distance* (NMD) and subtracting from 1, called *Normalized MEDLINE Similarity* (NMS); otherwise, if the link appears in the ontology-based concept network, we assign its weight as 1 since this relation should have been validated.

3.4 Phase 4: Target Term Ranking

Yetisgen-Yildiz & Pratt [18] suggested that using *Link Term Count with Average Minimum Weight* (LTC-AMW) can achieve the best effectiveness for target term ranking. LTC-AMW takes the number of intermediate terms between starting term and target term as the primary ranking criteria, i.e. the number of paths. The average minimum weight of paths is used only when two target terms are identical in their number of paths. In this study, we propose two target term ranking measures, *Summation of Minimum Weight* (SumMW) and *Summation of Average Weight* (SumAW), as follows:

$$SumMW(A,C) = \sum_{B \in N(A) \cap N(C)} \min\{Wt(A,B), Wt(B,C)\}, \quad (4)$$

$$SumAW(A,C) = \sum_{B \in N(A) \cap N(C)} \frac{Wt(A,B)+Wt(B,C)}{2}, \quad (5)$$

where $N(A)$ denotes the neighbor concepts of term A, and $Wt(A,B)$ is the weight of link between term A and B. The above measures differentiate the importance of each path according to their minimum or average weight of constituent links, and assign an importance score to each target term according to the cumulative information of all paths between the starting term and the target term. We then order target terms according to their importance scores.

4 Evaluation and Results

In this study, we conduct two experiments to evaluate our proposed method for drug repositioning. The first experiment is to evaluate our proposed comprehensive concept network and link weighting measure (i.e., ExtNMS). The second experiment is to evaluate our proposed target term ranking measure.

4.1 Evaluation Design

We follow the evaluation procedure proposed by Yetisgen-Yildiz and Pratt [18]. Specifically, we describe our experiment procedure step by step in the following.

Given a starting term A (i.e., drug):

1. We set cut-off date as January 1, 2000 and divide MEDLINE 2011 baseline into two datasets:
   (a) *Pre-cut-off* set (S_{t1}) includes the documents prior to 1/1/2000.
   (b) *Post-cut-off* set (S_{t2}) includes the documents on and after 1/1/2000.
2. The documents in the pre-cut-off set are used along with ontologies as the input to construct the comprehensive concept network.
3. We define a gold-standard set G_A, which contains terms that satisfy the following rules:
   (a) Terms are within our specified target semantic group, i.e., disease.
   (b) Terms co-occur with term A in the post-cut-off set, but do not co-occur with term A in the pre-cut-off set. In other words, these terms co-occur with term A in literature only after the cut-off date (i.e., 1/1/2000).
   (c) Terms are not related to term A in the ontology-based concept network.
4. The discovery effectiveness is estimated by using the information retrieval metrics as follows:
   (a) Precision:

$$P_A = \frac{|T_A \cap G_A|}{|T_A|}, \tag{6}$$

   (b) Recall:

$$R_A = \frac{|T_A \cap G_A|}{|G_A|}, \tag{7}$$

where T_A is the set of target terms generated by our discovery method.

Table 3 shows the list of semantic groups used in our experiments. In this study, we randomly select 100 terms from the semantic group of drugs as the starting terms, i.e., the focal drugs.

Table 3. Selected Semantic Groups for Our Experiments

Selected Intermediate Terms	Selected Target Terms
Drugs	Diseases
Genes, Proteins, and Enzymes	
Pathological Effects	
Diseases	

4.2 Exp. 1: Evaluation of the Comprehensive Concept Network and Link Weighting Method

The performance benchmark is the original ABC model over only the literature which uses association rules as link weighting algorithm [14]. We evaluate three versions of our discovery method, i.e., one is over only the literature-based concept network,

another one is over only the ontology-based concept network, and the third one is over the comprehensive concept network. All methods (including the benchmark) under investigation apply LTC-AMW as the target term ranking measure. Table 4 shows the evaluation results, and the Area Under Curve of Precision and Recall (AUC-PR) represents the overall performance. A higher AUC-PR value represents a greater effectiveness. As Table 4 illustrates, our proposed ExtNMS measure outperforms the benchmark link weighting measure (i.e., association rules) when the information source is the literature only or the integrated information sources (i.e., the comprehensive concept network that contains literature and ontologies). Moreover, using both literature and ontologies as information sources improves the overall performance, especially precisions on higher ranks. This would better help researchers sift plausible drug-disease relations for the purpose of drug repositioning.

Table 4. Comparative Evaluation Results of the Link Weighting Measures Under Different Concept Networks

Recall	Precision			
	Association Rules (Literature)	ExtNMS (Literature)	ExtNMS (Ontology)	ExtNMS (Comprehensive Concept Network)
0%	62.61%	57.72%	39.01%	59.33%
10%	29.72%	29.93%	20.16%	30.54%
20%	22.07%	23.75%	15.96%	23.89%
30%	17.80%	18.95%	14.22%	19.01%
40%	15.13%	16.27%	12.58%	16.26%
50%	11.73%	13.80%	12.25%	13.62%
60%	9.52%	11.69%	13.77%	11.53%
70%	7.61%	9.66%	0%	9.61%
80%	7.17%	7.76%	0%	7.69%
90%	2.31%	6.01%	0%	5.97%
100%	0.60%	3.80%	0%	3.84%
AUC-PR	**15.47%**	**16.86%**	**10.84%**	**16.97%**

4.3 Exp. 2: Evaluation of Target Term Ranking Measure

In this experiment, we evaluate the proposed target term ranking measures, *Summation of Minimum Weight* (SumMW) and *Summation of Average Weight* (SumAW). Based on the experiment 1, we apply ExtNMS as the link weighting measure in this experiment. In this experiment, LTC-AMW is used as the benchmark ranking measure. Table 8 shows the comparative evaluation results of the three target term ranking measures using the comprehensive concept network as the information sources. Both SumMW and SumAW outperform the benchmark measure, LTC-AMW. These results show that our link weighting measure, ExtNMS, is a more effective measure to weight links, and considering both the number and weights of paths between the starting term and target terms can improve the discovery effectiveness.

Table 5. Comparative Evaluation Results of Target Term Ranking Measures (Using the Comprehensive Concept Network)

Recall	Precision		
	LTC-AMW	SumMW	SumAW
0%	59.33%	59.14%	61.70%
10%	30.54%	33.46%	33.16%
20%	23.89%	24.86%	24.52%
30%	19.01%	20.91%	20.69%
40%	16.26%	17.32%	17.01%
50%	13.62%	14.74%	14.56%
60%	11.53%	12.42%	12.17%
70%	9.61%	10.50%	10.29%
80%	7.69%	8.41%	8.22%
90%	5.97%	6.35%	6.22%
100%	3.84%	3.84%	3.84%
AUC-PR	**16.97%**	**18.05%**	**17.96%**

5 Conclusions

In this study, we develop a drug repositioning discovery method that uses both biomedical literature and ontologies as information sources for constructing a comprehensive network of biomedical concepts. We also develop a link weighting method (i.e., ExtNMS) and two target term ranking measures. We experimentally evaluate our proposed method and show that taking both literature and ontologies into account and using our ExtNMS measure can improve the effectiveness of predicting novel drug-disease relationships. Besides, our proposed target term ranking measures can better infer plausible drug-disease relations. Overall, our proposed drug repositioning discovery method can help researchers sift most plausible unknown drug-disease relationships, i.e., potential drug repositioning candidates.

References

1. Adams, C.P., Brantner, V.V.: Estimating the cost of new drug development: Is it really $802 million? Health Aff. 25(2), 420–428 (2006)
2. National Institutes of Health. NIH Announces New Program to Develop Therapeutics for Rare and Neglected Diseases (May 20, 2009), http://rarediseases.info.nih.gov/files/TRND%20Press%20Release.pdf
3. Ashburn, T.T., Thor, K.B.: Drug repositioning: identifying and developing new uses for existing drugs. Nat. Rev. Drug Discov. 3(8), 673–683 (2004)
4. McBride, W.G.: Thalidomide and congenital abnormalities. Lancet 278(7216), 1358 (1961)
5. Stephens, T.D., Brynner, R.: Dark Remedy: The Impact of Thalidomide and Its Revival as a Vital Medicine. Perseus Publishing, Cambridge (2001)

6. Celgene Corporation: 2005 Annual Report. Celgene Corporation, Summit, NJ (2006)
7. Celgene Corporation: 2008 Annual Report on Form 10-K. Celgene Corporation, Summit, NJ (2009)
8. Celgene Corporation: 2012 Annual Report on Form 10-K. Celgene Corporation, Summit, NJ (2013)
9. Dudley, J.T., Deshpande, T., Butte, A.J.: Exploiting drug-disease relationships for computational drug repositioning. Brief. Bioinformatics 12, 303–311 (2011)
10. Hurle, M.R., Yang, L., Xie, Q., Rajpal, D.K., Sanseau, P., Agarwal, P.: Computational drug repositioning: From data to therapeutics. Clin. Pharmacol. Ther. 93(4), 335–341 (2013)
11. Emig, D., Ivliev, A., Pustovalova, O., Lancashire, L., Bureeva, S., Nikolsky, Y., Bessarabova, M.: Drug target prediction and repositioning using an integrated network-based approach. PLoS One 8(4), e60618 (2013)
12. Kim, S., Jin, D., Lee, H.: Predicting Drug-Target Interactions Using Drug-Drug Interactions. PLoS One 8(11), e80129 (2013)
13. Wu, Z., Wang, Y., Chen, L.: Network-based drug repositioning. Mol. Biosyst. 9, 1268–1281 (2013)
14. Swanson, D.R., Smalheiser, N.R.: An interactive system for finding complementary literatures: A stimulus to scientific discovery. Artif. Intell. 91(2), 183–203 (1997)
15. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Wishart, D.S.: DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 39(suppl. 1), D1035–D1041 (2011)
16. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A.: Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33(suppl. 1), D514–D517 (2005)
17. Davis, A., King, B.L., Mockus, S., Murphy, C.G., Saraceni-Richards, C., Rosenstein, M., Mattingly, C.J.: The Comparative Toxicogenomics Database: Update 2013. Nucleic Acids Res. 39(suppl. 1), D1067–D1072 (2013)
18. Yetisgen-Yildiz, M., Pratt, W.: A new evaluation methodology for literature-based discovery systems. J. Biomed. Inform. 42(4), 633–643 (2009)
19. Swanson, D.R.: Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspect. Biol. Med. 30(1), 7–18 (1986)
20. Weeber, M., Klein, H., de Jong-van den Berg, L.T., Vos, R.: Using concepts in literature-based discovery: Simulating Swanson's Raynaud–fish oil and migraine–magnesium discoveries. J. Am. Soc. Inf. Sci. Technol. 52(7), 548–557 (2001)
21. Wren, J.D., Bekeredjian, R., Stewart, J.A., Shohet, R.V., Garner, H.R.: Knowledge discovery by automated identification and ranking for implicit relationships. Bioinformatics 20(3), 389–398 (2004)
22. Lee, S., Choi, J., Park, K., Song, M., Lee, D.: Discovering context-specific relationships from biological literature by using multi-level context terms. BMC Medical Informatics and Decision Making (BMC Med. Inform. Decis. Mak.) 12(suppl. 1), S1 (2012)
23. Srinivasan, P.: Text mining: Generating hypotheses from MEDLINE. J. Am. Soc. Inf. Sci. Technol. 55(5), 396–413 (2004)
24. Hristovski, D., Peterlin, B., Mitchell, J.A., Humphrey, S.M.: Using literature-based discovery to identify disease candidate genes. Int. J. Med. Inform. 74, 289–298 (2005)
25. Frijters, R., van Vugt, M., Smeets, R., van Schaik, R., de Vlieg, J., Alkema, W.: Literature mining for the discovery of hidden connections between drugs, genes and diseases. PLoS Comput. Biol. 6(9), e1000943 (2010)

26. Lee, H.S., Bae, T., Lee, J.-H., Kim, D., Oh, Y., Jang, Y., Kim, S.: Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. BMC Syst. Biol. 6(1), 80 (2012)
27. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L.J., Bork, P.: Drug target identification using side-effect similarity. Science 321(5886), 263–266 (2008)
28. Yang, L., Agarwal, P.: Systematic drug repositioning based on clinical side-effects. PLOS ONE 6(12), e28025 (2011)
29. Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., Tang, Y.: Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput. Biol. 8(5), e1002503 (2012)
30. Li, J., Lu, Z.: A new method for computational drug repositioning. In: IEEE International Conference on Bioinformatics and Biomedicine, pp. 1–4. IEEE Press, Philadelphia (2012)
31. Qu, X.A., Gudivada, R.C., Jegga, A.G., Neumann, E.K., Aronow, B.J.: Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. BMC Bioinformatics 10(suppl. 5), S4 (2009)
32. Li, J., Zhu, X., Chen, J.Y.: Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. PLoS Comput. Biol. 5(7), e1000450 (2009)
33. Cilibrasi, R.L., Vitányi, P.M.: The Google similarity distance. IEEE Trans. Knowl. Data Eng. 19(3), 370–383 (2007)