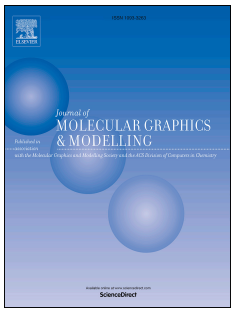
Accepted Manuscript



Computational drug repurposing to predict approved and novel drug-disease associations

Zoya Khalid, Osman Ugur Sezerman

PII: S1093-3263(18)30372-3

DOI: 10.1016/j.jmgm.2018.08.005

Reference: JMG 7212

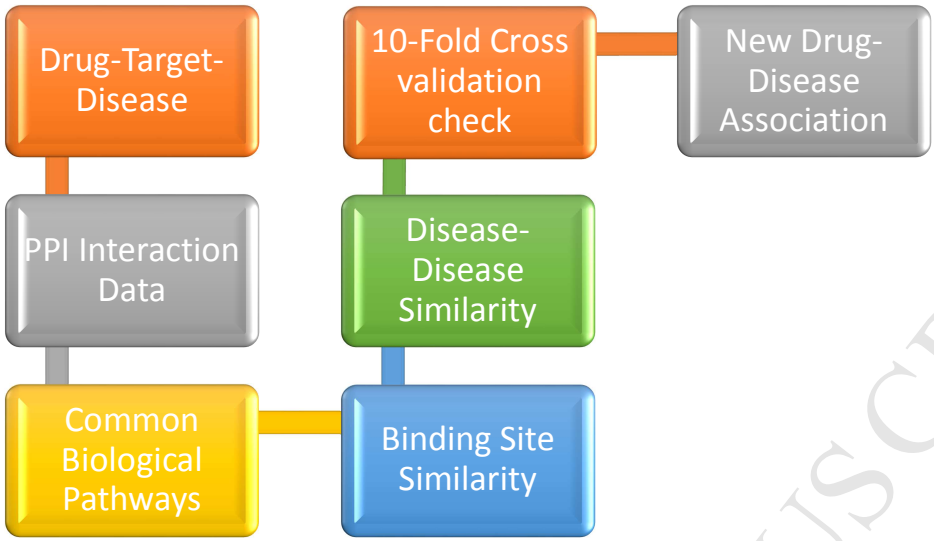
To appear in: *Journal of Molecular Graphics and Modelling*

Received Date: 18 May 2018

Revised Date: 7 August 2018

Accepted Date: 10 August 2018

Please cite this article as: Z. Khalid, O.U. Sezerman, Computational drug repurposing to predict approved and novel drug-disease associations, *Journal of Molecular Graphics and Modelling* (2018), doi: 10.1016/j.jmgm.2018.08.005.



Computational Drug Repurposing to Predict Approved and Novel Drug-Disease Associations

ZOYA KHALID[1,2]

[1]*Sabanci University, Istanbul, Turkey*
zoyakhalid@sabanciuniv.edu

[2]*National University of Medical Sciences, Rawalpindi*

OSMAN UGUR SEZERMAN[3]

[3]*Acibadem University, Istanbul, Turkey*
ugur.sezerman@acibadem.edu.tr

Abstract

The Drug often binds to more than one targets defined as polypharmacology, one application of which is drug repurposing also referred as drug repositioning or therapeutic switching. The traditional drug discovery and development is a high-priced and tedious process, thus making drug repurposing a popular alternate strategy. We proposed an integrative method based on similarity scheme that predicts approved and novel Drug targets with new disease associations. We combined PPI, biological pathways, binding site structural similarities and disease-disease similarity measures. The results showed 94% Accuracy with 0.93 Recall and 0.94 Precision measure in predicting the approved and novel targets surpassing the existing methods. All these parameters help in elucidating the unknown associations between drug and diseases for finding the new uses for old drugs.

Keywords: Drug Repurposing; Binding site similarity; Multiple data sources; Similarity measures; Integrative method ;Common pathways; Drug promiscuity; Drug repositioning.

1. Introduction

Growth in drug research and development has been dropping down for few years as drug discovery is becoming an expensive job. Pharmaceutical companies are unsuccessful in keeping pace with bringing new drugs to the market, reasons behind this are manifold, broadly the safety and efficacy factors are coupled with the overpricing. To cope with this dilemma, the theory of reusing the existing drugs gained a lot of thoughtfulness among bio-pharmaceutics. The phenomena of using the old drugs with new indications paired with the original indications are termed as repurposing or repositioning. The concept of drug repurposing causing a huge cut down in terms of price in drug discovery domain as it costs almost half the price of a new drug causing an overhaul in drug development. Before carrying out drug repositioning, few aspects should be given a look and these include drug side effects and its relevancy to the disease. Since the approved drugs have already been passed through various validation steps which includes target identification and ADMET (absorption, distribution, metabolism, excretion, and toxicity) characteristics, thus can facilitate in identifying new uses for the same drug [1].

Among various approaches for finding new uses of old drugs, using GWAS data, gene expression data, pathway analysis, and structural features are considered as the most distinguished methods. A drug can bind to several different targets making it promiscuous in nature. This is a well-known phenomenon which is also considered as a significant factor for the efficacy of drugs. Drug promiscuity reveals the drug off-targets, thus making them prime candidates for drug repurposing. Among the in-silico approaches for finding drug targets, binding site structure similarity to structural bioinformatics domain is highly contributing. The proteins sharing similar binding sites tend to bind same ligands hence paving ways for drug repositioning. Previous studies reported that flexibility is also one of the physiochemical properties of ligand contributing in drug promiscuity. With sharing similar binding sites, the ligand flexibility is similarly influencing drug binding to multiple targets, hence these features are important for drug repositioning [2].

In this study, we proposed a method based on **S**imilarity scheme that **P**redicts **A**pproved and **N**ovel Drug **T**argets with new **D**isease associations (SPANTD) by combining PPI neighbors, common pathways, binding site structural similarity and disease-disease similarity measures. Given drug-target as a query, our method developed a scoring scheme that measures the similarity of drug-disease pairs based on the ranking system. Also, we have optimized the scores by applying Genetic algorithm. To obtain the best combination of parameters we have applied GA by tuning its parameter with the number of runs 10. Unlike one drug one target model, our approach integrates multiple source information with emphasizing on binding site structural similarity as this feature turns out to be the most significant one among others. To the best of our knowledge, no study on drug repositioning has used this combination of features to predict drug-disease associations.

2. Related Work

The classical method of drug discovery is one drug one target model which has expected to provide less efficacy with more side effects. This model doesn't consider the biological mechanisms that make a drug to bind with more than one target hence limiting the efficiency of this model [3,4]. To overcome this many studied have worked on analyzing drug associated biological pathways. These are beneficial in exploring the mechanism of action of drugs and also the upstream or downstream genes in a pathway [5]. It is important to look for the pathway associated genes, as there are possibilities if the drug is not directly binding to its target but the target gene is interacting other genes in a pathway hence binding to the ligand. Among drugs retrieved biological pathways, some of the drug targets share common pathways which in turn helps in the revelation of clinical functions along with the drug's mechanism of action. Also, these pathways might have associated with some other diseases than those they were initially used for, hence providing a substitute for drug repurposing [6,7].

Network-based inference (NBI) for developing the method for drug repurposing has been reported in earlier studies. The idea is to construct drug target bipartite graph to infer new uses for old drugs. Unlike target based and drug based similarity measures NBI works better in finding new disease potentials with AUC rate of 0.865±0.009 and 0.849±0.012 [8]. Another reported study uses NBI to develop a method for drug repurposing by targeting mutated cancer genes. The method is based on the hypothesis that if a drug's up/down-regulated genes tend to have somatic mutations, there would have a higher tendency for anticancer indications. The results showed 284 potential

indications connecting 28 cancer types and 48 existing drugs with a 66.7% success rate[9]. In one study the authors used chemical structures of drug and its target starting from this the network will expand linking new indications [10]. The Pairwise similarity to conduct drug repositioning was also proposed [11]. The similarity measures include drug similarity, drug target similarity, and target- target interaction. Furthermore, the literature mining and pathway analysis were also proposed to build drug disease network [12].

One study used the microarray gene expression data for finding drug-disease interactions. Their network contains disease-disease, disease-drug, and drug-drug associations that provides insights about drug repositioning. The methodology was based on a scoring system that calculates the similarity scores among the drug and disease pairs. Using this method, the authors have discovered many new indications for the approved drugs [13]. Another method used gene expression profiles to check the effect of a drug to various treatments. The network contains those nodes that either have a similar mechanism of action or targeting same biological pathway. This network was developed on consensus transcriptional response which shows the transcriptional activity of a drug towards drug treatments. This approach helps in capturing the similarities and differences in drug responses, hence is useful for drug repositioning [14]. For creating a disease-drug, disease-disease and drug-drug network Guanghui and Agarwal used gene expression profiles, they have used two approaches correlation and enrichment. Correlation takes in profile-profile similarity while enrichment measures signature-profile similarity. This helps in identifying novel relations among drug and disease hence can be used for repositioning [15,16].

A drug-drug similarity network was proposed by [17] where the network has gene expression profiles as features. Many studies have been proposed lately that used the drug and disease expression profiles to provide a plausible candidate for drug repositioning [18, 19]. Another study performed drug repositioning without using the gene signatures. A scoring function was formulated to compute drug-gene-disease network, which takes in both the contribution of a gene and effect of a drug on a gene. Drug-disease association can be computed by measuring the similarity and dissimilarity of their gene expression profiles [20]. In one study authors used both efficacy and side effects measure for drug repurposing. The gene regulation has been observed before and after drug treatment to measure drug efficacy and the number of essential genes and correlated genes were measured for side effects. Based on this a scoring scheme was developed to align drug-disease association for repurposing [21]. Machine learning is contributing much to building drug repositioning strategies. It combines multiple information including how similar their chemical structures are, closeness in a PPI network of drug targets and correlation among the gene expression patterns [22]. In addition to the computational approaches, few studies are also devoted to manually analyze the drug associated pathways for drug repositioning [23,24]. For instance, bexarotene which was used for cancer treatment can also be used for Alzheimer's disease [25]. The manual curation has performed which is based on drug target, target associated pathways, transcriptional responses of the pathway and the gene-based analysis for understanding mechanism of disease.

3. Methods

We have proposed a drug prioritization algorithm SPANTD to reposition drugs by using two different benchmarked datasets. We developed a ranking algorithm to find diseases that a drug can be repurposed for. We trained our

method on 150 FDA approved drugs with their associated diseases referenced in [26,27]. Further for testing the proposed model, we used independent test set of 50 drugs with the same attributes.

3.1 Computing Drug-Disease Network

Our aim is to start with the drug and its target in the old disease indication as a seed value and ends up finding the new/repurposed disease label. We extracted drug targets from DrugBank, comparative Toxicogenomics Database (CTD) [28] and Therapeutic Target Database (TTD) for the disease in query [29]. For those drug-disease association in which direct target is not reported, we looked for the genes highly associated with a drug-disease pair. To look for common pathways among targets we used GenesLikeMe from GeneCards. It works by finding the shared pathways with the query gene by assigning a weighted score. For finding binding site structural similarity we used online tool PROBIS. This server takes in PDB structures as query proteins and compares it with 42270 structures available in the database which shares similar binding sites. The threshold of 1.0 was selected which filter out the significant similarity scores with the non-significant ones. The tool is freely available at http://probis.cmm.ki.si/[30]. Binding sites can be similar in two proteins, a ligand binding to one protein can bind to another protein sharing the similar binding site that was not binding with this ligand at first place. Lastly, Gene associated diseases were retrieved from DisGenet and further disease-disease similarity measure was computed using DisGenet and CTD based on the number of shared genes between the two. Figure S1 provides a flowchart of complete methodology.

3.2 Learning Transition Weights and constructing a Feature set

Our scoring system is based on similarity measures. We prioritize genes based on the ranking algorithm which works as follows

1. Starting with drug, target and old disease indication, look in the BIOGRID PPI network for the proteins that are at a distance of ≥4, ≥5, ≥6 from the drug target. Scores S (pi, $pi^{/}$) is the computed similarity between two proteins which is based on the distance with the target, closest one is scored highest.
2. The second task is to check for common pathways among the targets and the proteins retrieved in the previous step. Scores S (p, $p^{/}$) are assigned by using the scoring scheme of GenesLikeMe.
3. The third step is to retrieve the PDB ids of these proteins and look for pairwise binding site structural similarity. PROBIS scoring system was used to score this feature. The scores are stored in S (dr, $dr^{/}$).
4. The final step is to extract diseases associated with the proteins and further check for disease-disease similarity between a query disease and extracted diseases. The score was calculated by using the following formula

$$S\left(di, di^{'}\right) = \left(\frac{X\text{-}Min}{Max\text{-}Min}\right) \times 5$$

Where S (di, $di^{/}$) computed disease-disease similarity value. X is the number of shared genes between the two diseases in a query. Among all possible new disease indication for the query drug, Min is the Minimum number of

shared genes between the old disease and possible new disease. For the query drug, Max is the Max number of shared genes between the old disease and possible new disease. The value obtained was multiplied by 5 in order to normalize the scores. All scores were added together to compute one final score for the drug-disease association which in turn creates a feature set for classification purpose.

$$\boldsymbol{S(dr, di') = S(pi, pi') + S(p, p') + S(sr, sr') + S(di, di')}$$

Further, all these scores are optimized by applying Genetic algorithm using GA toolbox in MATLAB. Genetic algorithm optimization is inspired by evolutionary rules. GA assess many possible solutions simultaneously. To obtain the best combination of parameters we have applied GA by tuning its parameter as follows :

Population Size :50 ; Crossover Rate : 0.95 ; Mutation Rate: 0.05 ; No of generations :100; No of Runs: 10

4. Results

We have developed an algorithm SPANTD to integrate PPI interactions data, common pathway analysis, binding site structure similarity and disease-disease similarity measures to score the relevance of each component in predicting new diseases for which a drug can be repurposed for. From human PPI network, the first task was to find the optimal distance in a PPI network with the target protein. For this, we tried 3 different distance filters $\leq 4, \leq 5, \leq 6$ form the drug target to PPI. We compared the AUC scores for the three and found a distance of ≤5 as an optimal distance so rest of the experiments were carried out using this distance. This gives us a maximum of 60,000 proteins varying from target to target. We picked these proteins and looked for common pathways among them. This filters the proteins from 60,000 to almost 100 for most of the cases. In the third step, we have extracted the PDB structures of the candidate protein and looked for binding site structural similarity between the drug target and these filtered protein list obtained from the previous measure. This further shortens the candidate list to at most 4 proteins. The final step was to look for disease-disease similarity associated with the drug target and the new targets. The final output (combined score) of our method shows that the high scoring candidates are always the target of the new disease indications which was already reported as a repurposed disease of that particular drug.

4.1 10-Fold Cross Validation for validity check

To evaluate SPANTD algorithm we applied 10-fold cross validation check by hiding 10% of the drugs associations in each iteration. The dataset was divided randomly into 10 parts with each part turning into a test set for once. We performed 50 independent runs in each of which a different random partition of the training set to 10 parts were used. We computed average AUC scores with AUPR and MPR (mean percentile ranking) measures. As there is unequal distribution of known drug-disease associations (positive labels) and unknown drug-disease associations (negative labels) AUPR (Area under precision-recall curve) and MPR were the most suitable model evaluation measures. Figure 2 shows the ROC curve with the average AUC of 0.97± 0.007. Further, the results showed 0.923± 0.004 AUPR and 0.453± 0.013 MPR measures. Furthermore, we also checked the performance of our method on an independent test set of 50 drugs, the results are tabulated in Table1. The results also predict some novel targets and diseases which in future can be verified experimentally. The novel targets and their associated diseases are tabulated in Table 2 and Table 3 from the two datasets as referenced above. For testing the importance of binding site

structural similarity score we have tested SPANTD with and without this feature and observed the accuracy measures. We have used python for implementing logistic regression classifier for the prediction task.

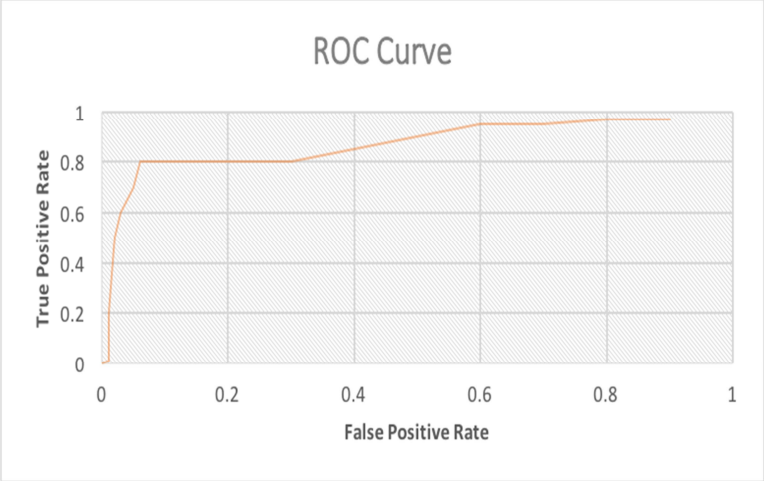


Fig. 2. ROC Curve obtained for SPANTD

Table 1. True and False Predictions

Total Drugs	**True Predictions**	**After Pathway Analysis**	**False Predictions**	**Accuracy**	**Precision**	**Recall**
150 (Train Set)	145	147	3	98%	0.96	0.95
50 (Test Set)	45	47	3	94%	0.94	0.93

Table 2. Novel Indications for Drug Repurposing from Dataset1

Drug	**Old Indication**	**Target**	**New Targets**	**New Indications**	**Reported New Indications**
Zidovudine	Cancer	TP53	MDM2 BCL2	Breast Neoplasms Hypertension	AIDS
Methotrexate	Cancer	DHFR	MYC RB1	Liver Neoplasms Breast Neoplasms	Rheumatoid arthritis
Memantine	Parkinson Disease	GRIN1	RAF1 RAC1	NOONAN SYNDROME Heart Failure	Alzheimer's disease

Thalidomide	Sedative	TNF	IL6 TNFRSF1A	Diabetes Mellitus Liver Syndrome	Multiple Myeloma
Raloxifene	Osteoposrosis	ESR1	FOS TGFR	Hypertensive Disorders Breast Neoplasms	Breast Neoplasms

Table 3. Novel Indications for Drug Repurposing from Dataset2

Drug	**Old Indication**	**Old Target**	**New Target**	**New Indication**
Amantadine	Parkinson Disease	BDNF	PAK1	Malignant Neoplasm of Breast
Capecitabine	Colorectal cancer	TYMS	TOP2A	Mammary Neoplasms
Cyclosporine	Allograft rejection	THBS1	EGFR	Non-small cell lung carcinoma
Itraconazole	Fungal infections	STAT1	MAPK3	Breast Carcinoma
Galantamine	Alzheimer's disease (ad)	ACHE	ALDH2	Alcoholic Intoxication, Chronic
Formoterol	Asthma	ADRB2	CACNA1D	Adenoma

4.2 Low Scoring Candidates

From training set, the computed scores of the two drug targets were really low. Further scanning of these two cases showed that they do not have any binding site structure similarity between the two targets. Similarly, from the test set, we have found 3 similar cases. We then picked these less scoring genes and find the pathways they are associated with and further compare the downstream and upstream genes to check binding site structure similarity. The associated pathways were scanned up to 2 levels for downstream and upstream genes which are either activating, phosphorylating or interacting with the target genes. The idea was to uncover if the pathways proteins are actually binding to the drug and blocking its activity. These genes were looked again for structure similarity, all of the drug targets shared the similar binding site with the pathway genes, the results are tabulated in Table 4.

5. Comparison with Other Methods

We compared our approach with three methods reported in literature namely PREDICT [17], SLAMS [31] and DSESI [32]. We have downloaded the gold standard datasets available online for these three methods and applied our SPANTD algorithm to compare the performance. Our method is successful in finding approved and novel drug-disease associations with the higher AUC score of 0.97 as tabulated in table 5. Also, Figure 3 summarizes the comparison of AUC scores with SPANTD. Unlike the above referenced methods, we have introduced a novel combination of similarity measures by adding binding site structural similarity as the most contributing feature for drug repurposing. All these measures have been used separately for predicting new uses for old drugs, we believed using them all together will provide better results.

Table 4. Pathway Analysis for finding Drug-Disease associations

Drug	**Target**	**Old Indication**	**New Indication**	**Target**	**Pathway**	**Target**
(Train Set) Mifepristone	Pregnancy Termination	FAS	Cushing Syndrome	NR3C1	P53 Signaling Pathway	P53
(Train Set) Heparin	Anticoagulant	SERPINC1	Cystic Fibrosis	TGFB1	Complement and coagulation cascades	F9

Fig. 3. Comparison of AUC scores of four methods with SPANTD

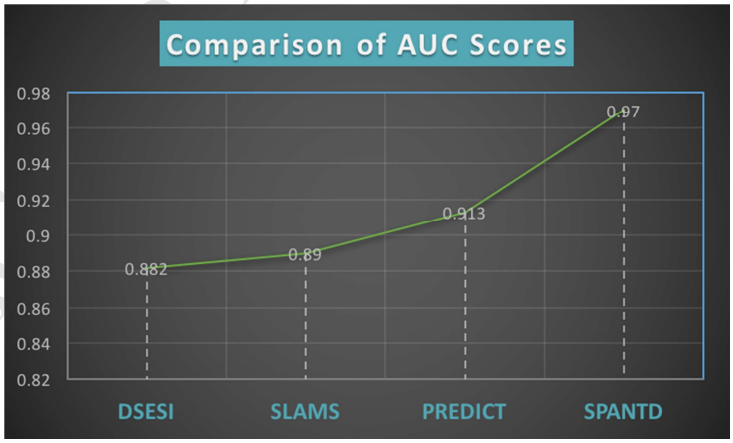


Table 5: Comparison of existing methods on drug repurposing

Method	Similarity Measures	Reported Accuracy Measure
PREDICT	Drug-Drug similarity Disease-Disease similarity	AUC :0.913±0.002
DSESI	Chemical, side effect, therapeutic space	AUC: 0.882 ± 0.011
SLAMS	Chemical biological and phenotypic	AUC: 0.8949
SPANTD	PPI, Pathways, Binding site structural similarity , disease-disease similarity	AUC :0.97± 0.007

6. Discussion

The Drug often binds to more than one target defined as polypharmacology one application of which is drug repurposing also referred as drug repositioning or therapeutic switching. Drug polypharmacology can act both ways, on one hand, it is beneficial for drug repurposing while at the same time is highly unwanted in drug discovery domain because of probable side effects. Two reasons behind this promiscuous phenomenon of drugs are the binding site similarity among the drug targets and the flexibility of drugs making them binding to multiple targets. There has been a number of approaches proposed lately for drug repurposing including side effects, gene expression profiles, and structure similarity. Our study has focused on binding site similarity since it's been in a limelight for drug repositioning as it unravels the drug mode of action.

No study reported before considered the binding site structural similarity for performing drug repurposing hence it is necessary to revise drug repositioning strategies by introducing the important features. SPANTD integrates four different types of similarity measures between drug and disease. We developed a ground basis for evaluating drug disease relationship by prioritizing the candidates for drug repurposing. All these parameters help in elucidating the unknown associations between drug and diseases for finding the novel targets to reiterate old drugs.

To generate a better prediction, researchers should not only focus on building new algorithms but also in identifying a good feature set. Closeness in a PPI network of a drug target is an important measure in identifying new targets that can serve as candidates for drug repurposing. Similarly, understanding the mechanism of action (MOA) of drugs is important for drug development. A biological pathway connecting a drug and its targets may describe the biological function or the disease associated with the drug. Therefore, drug targets sharing common

pathways might be sharing same mechanisms of action hence can serve as new drug targets for drug repurposing. Also, our disease-disease similarity measure is based on shared genes. This feature indicates that the similar diseases can be treated by same drug hence providing a similarity measure to be added for carrying out drug repurposing.

The mainstream computational methods for drug repurposing reported earlier are based on chemical structure, protein target interactions, side effects based similarity and analyzing gene expression profiles. Few studies have focused separately on different drug features and few used combinatorial approach for analyzing drug-like activities hence give biased predictions. One study reported a method named PREDICT which proposed a strategy of drug repositioning that can be applied to personalized medicine as well. They used five different levels of computing drug-drug and disease-disease similarities. Their measure includes chemical-based, sequence-based, genetic-based, closeness in a PPI, side effect and phenotypic based on disease-disease similarity. They have achieved high rates of accuracy measures AUC=0.9 [15]. Our SPANTD is different from the previous methods as we incorporate some additional features like binding site structure similarity and common pathways altogether. These features are the best description for the drug promiscuity hence providing key aspects for carrying out drug repurposing. We used multiple evaluation measures (precision, recall, F-measure) unlike using only AUC to evaluate the performance. Therefore, our model is expected to outperform previously reported algorithms as it helps in providing a better prediction of novel indications, hence can be applied to a large scale for conducting drug repurposing.

7. Conclusion

We have proposed a drug repurposing strategy that serves as a ray of hope in battling drug resistance. Our method SPANTD is based on a similarity scheme that can handle both approved and novel targets for the drug-disease association. Our model integrates protein-protein interaction data, biological pathway, binding site similarity and disease-disease similarity unlike one drug one target models. The algorithm tests the relevance of each parameter and scores accordingly. Results showed that our method is successful in predicting already reported new indications of a drug and along with that some novel indications were also found. The novel targets can serve as leads that require further experimental validation. Repurposed drugs provide the best alternative for treating drug resistance. As a future work, we would like to develop a user interface so that novel predictions can be made given user input.

Data Availability

The dataset analyzed during the current study is accessible at https://www.biomed-data.eu/article/benchmark-dataset-computational-drug-repositioning

REFERENCES

1. Ashburn, Ted T., and Karl B. Thor. "Drug repositioning: identifying and developing new uses for existing drugs." Nature reviews Drug discovery 3.8 (2004): 673-683.
2. Haupt, V. Joachim, Simone Daminelli, and Michael Schroeder. "Drug promiscuity in PDB: protein binding site similarity is key." PLoS one 8.6 (2013): e65894.
3. Dudley, Joel T., et al. "Drug discovery in a multidimensional world: systems, patterns, and networks." Journal of cardiovascular translational research 3.5 (2010): 438-447.
4. Schadt, Eric E., Stephen H. Friend, and David A. Shaywitz. "A network view of disease and compound screening." Nature

reviews Drug discovery 8.4 (2009): 286-295.

5. Li, Jiao, and Zhiyong Lu. "Pathway-based drug repositioning using causal inference." BMC bioinformatics 14.16 (2013): S3.
6. Campillos, Monica, et al. "Drug target identification using side-effect similarity." Science 321.5886 (2008): 263-266
7. Vitali, Francesca, et al. "A Network-Based Data Integration Approach to Support Drug Repurposing and Multi-Target Therapies in Triple Negative Breast Cancer." PloS one 11.9 (2016): e0162407.
8. Cheng, Feixiong, et al. "Prediction of drug-target interactions and drug repositioning via network-based inference." PLoS computational biology 8.5 (2012): e1002503.
9. Cheng, Feixiong, et al. "A network-based drug repositioning infrastructure for precision cancer medicine through targeting significantly mutated genes in the human cancer genomes." Journal of the American Medical Informatics Association 23.4 (2016): 681-691.
10. Keiser, Michael J., et al. "Predicting new molecular targets for known drugs." Nature 462.7270 (2009): 175-181.
11. Li, Jiao, and Zhiyong Lu. "A new method for computational drug repositioning using drug pairwise similarity." Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference On. IEEE, 2012.
12. Iorio, Francesco, et al. "Discovery of drug mode of action and drug repositioning from transcriptional responses." Proceedings of the National Academy of Sciences 107.33 (2010): 14621-14626.
13. Sirota, Marina, et al. "Discovery and preclinical validation of drug indications using compendia of public gene expression data." Science translational medicine 3.96 (2011): 96ra77-96ra77.
14. Shigemizu, Daichi, et al. "Using functional signatures to identify repositioned drugs for breast, myelogenous leukemia and prostate cancer." PLoS Comput Biol 8.2 (2012): e1002347.
15. Li, Jiao, Xiaoyan Zhu, and Jake Yue Chen. "Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts." PLoS Comput Biol 5.7 (2009): e1000450.
16. Li, Yong, and Pankaj Agarwal. "A pathway-based view of human diseases and disease relationships." PloS one 4.2 (2009): e4346
17. Gottlieb, Assaf, et al. "PREDICT: a method for inferring novel drug indications with application to personalized medicine." Molecular systems biology 7.1 (2011): 496.
18. Iorio, Francesco, et al. "Transcriptional data: a new gateway to drug repositioning?." Drug discovery today 18.7 (2013): 350-357.
19. Emig, Dorothea, et al. "Drug target prediction and repositioning using an integrated network-based approach." PLoS One 8.4 (2013): e60618.
20. Hu, Guanghui, and Pankaj Agarwal. "Human disease-drug network based on genomic expression profiles." PloS one 4.8 (2009): e6536.
21. Pan, Yongmei, et al. "Pathway analysis for drug repositioning based on public database mining." Journal of chemical information and modeling 54.2 (2014): 407-418.
22. Napolitano, Francesco, et al. "Drug repositioning: a machine-learning approach through data integration." Journal of cheminformatics 5.1 (2013): 30.
23. Strittmatter, Warren J. "Old drug, new hope for Alzheimer's disease." Science 335.6075 (2012): 1447-1448.

24. Sivachenko, Andrey, Andrey Kalinin, and Anton Yuryev. "Pathway analysis for design of promiscuous drugs and selective drug mixtures." Current drug discovery technologies 3.4 (2006): 269-277.
25. Cramer, Paige E., et al. "ApoE-directed therapeutics rapidly clear β-amyloid and reverse deficits in AD mouse models." science 335.6075 (2012): 1503-1506.
26. Kissa, Maria, and George Tsatsaronis. "A Benchmark Dataset for Computational Drug Repositioning." Biomed Data J 1.2 (2015): 10-12.
27. Wu, Chao, et al. "Computational drug repositioning through heterogeneous network clustering." BMC systems biology 7.5 (2013): S6.
28. Mattingly, Carolyn J., et al. "The Comparative Toxicogenomics Database (CTD)." Environmental health perspectives 111.6 (2003): 793.
29. Chen, Xin, Zhi Liang Ji, and Yu Zong Chen. "TTD: therapeutic target database." Nucleic acids research 30.1 (2002): 412-415.
30. Konc, Janez, and Dušanka Janežič. "ProBiS: a web server for detection of structurally similar protein binding sites." Nucleic acids research 38. suppl 2 (2010): W436-W440.
31. Zhang, Ping, Pankaj Agarwal, and Zoran Obradovic. "Computational drug repositioning by ranking and integrating multiple data sources." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2013.
32. Cheng, Feixiong, et al. "Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space." Journal of chemical information and modeling 53.4 (2013): 753-762.

Zoya Khalid received her Bachelors and Master Degrees in Bioinformatics from International Islamic University, Islamabad, Pakistan. She joined Sabanci University, Istanbul, Turkey for her PhD studies in 2013. Currently she is working in department of biological sciences and Bioengineering. Her project focuses on Biomedical text mining and to find computational ways for better prediction of drug resistance also improving drug repurposing strategies.

Osman Ugur Sezerman received his M.Sc. degree in Biomedical Engineering from Bosphorus University, Istanbul, Turkey, and his Ph.D. degree in Biomedical Engineering from Boston University, USA, in 1987 and 1993, respectively. He worked in Sabanci University from September 1999 to 31st March 2015. Currently he is working as a full professor in department of biostatistics and Medical informatics, Acibadem University, Istanbul, Turkey.

Highlights

- Building a computational model to propose a drug repurposing strategy that predicts approved and novel drug disease associations.
- We combined PPI, biological pathways, binding site structural similarities and disease-disease similarity measures together
- Features like binding site structural similarity was analyzed in particular.
- The model showed 94% Accuracy with 0.93 Recall and 0.94 Precision measure in predicting the approved and novel targets surpassing the existing methods