RWRDC: Predicting Efficacious Drug Combinations in Cancer Based on Random Walk with Restart

Qi Wang[1,2] and Guiying Yan[1,2]

Affiliations:
1 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China.
2 University of Chinese Academy of Sciences, Beijing, 100190, China.
Email:
Qi Wang: wangqi@amss.ac.cn
Guiying Yan: yangy@amss.ac.cn (corresponding author)

Abstract

(1) Background: Compared with monotherapy, efficacious drug combinations can increase the therapeutic effect, decrease drug resistance of experimental subjects and the side effects of drugs. Therefore, efficacious drug combinations are widely used in the treatment of complex diseases, such as various cancers. However, compared with the mathematical model and computational method, experimental screening efficacious drug combinations is time-consuming, costly, laborious, and inefficient;

(2) Methods: we predicted efficacious drug combinations in cancer based on random walk with restart (RWRDC). An efficacious score can be obtained between any two individual drugs by RWRDC;

(3) Results: As a result, we analyzed the rationality of the efficacious score first. Besides, compared with the other methods by leave-one-out cross-validation, all the Area Under Receiver Operating Characteristic Curves (AUROCs) of RWRDC were higher for data sets of breast cancer, colorectal cancer, and lung cancer. Moreover, the case study of breast cancer showed that RWRDC could discover potential efficacious drug combinations;

(4) Conclusions: These results suggest that RWRDC is a novel way to discover efficacious drug combinations in cancer, which provides new prospects for cancer treatment. Furthermore, RWRDC is a semi-supervised learning framework that can be used to predict combinations of drugs for other complex diseases.

Keywords: **Drug combination, Cancer, Computational method, Random walk with restart**

I. BACKGROUND

A combinatorial drug is a fixed-dose combination, including two or more active drug components in a single dosage form, which is manufactured and distributed in a fixed dose (Collier, 2012). Drug combinations are classified as efficacious or non-efficacious depending on whether they can produce expected improvements in clinical trials or preclinical studies (Liu *et al.*, 2014). Compared with monotherapy, efficacious drug combination can increase the efficacy of the therapeutic effect, and decrease the drug side-effects and drug resistance of experimental subjects. Especially, drug combinations interact with multiple targets in the molecular networks of complex diseases, as a consequence, efficacious drug combinations is targeted treating a large number of complex diseases, such as various cancers (Feliu *et al.*, 2009), human immunodeficiency virus (HIV) as well as fungal (Groll and Tragiannidis, 2009) and bacterial infections (Tamma *et al.*, 2012), etc.

Although efficacious drug combinations are important and meaningful, there are too many plausible combinations of individual drugs, and it is time-consuming and costly to screen efficacious drug combinations by experiment. Mathematical models and computational methods show their advantages in drug screening. For example, Kristina Preuer *et al.* (Preuer *et al.*, 2018) presented a novel deep learning model, DeepSynergy, to predict synergy scores of drug combinations for cancer cell lines. They used chemical and genomic information as model inputs and took into account standardization strategies for heterogeneity of input data. Janizek *et al.* (Janizek *et al.*, 2018) used the XGBoost method combing with a tool named TreeSHAP to predict drug combinations and make prediction results explainable. Zhang *et al.* (Tianyu Zhang *et al*, 2018) proposed a novel deep learning model to predict drug combinations by integrating multi-type of genomics data and chemical structure data. They uncovered potential mechanisms of the synergy of specific drug combinations against given cancer cell lines. Chen *et al.* (Chen *et al.*, 2016) developed a model named Network-based Laplacian regularized Least Square Synergistic drug combination prediction (NLLSS) to predict potential synergistic drug combinations. They also carried out biological experiments, finally, 7 of the 13 antifungal synergistic drug combinations predicted by the NLLSS were verified. (Bai *et al.*, 2018) devised a novel improved naïve Bayesian algorithm to construct classification models to

predict effective drug combinations. (Xu *et al.*, 2017) predict effective drug combinations by integrating biological, chemical and pharmacological information based on a stochastic gradient boosting algorithm. (Li *et al.*, 2020) used neighbor recommender method combined with ensemble learning algorithms to predicting drug combinations, and they predicted seven candidate drug combinations for paclitaxel, and successfully verified the promising effects of two of the predicted combinations.

Most computational models about predicting efficacious drug combinations are using supervised learning methods. In fact, due to the very few known non-efficacious drug combinations, it has been very difficult to obtain all the label information (efficacious or non-efficacious) of the training set. In this paper we used a semi-supervised learning method, called random walk with restart model, to predict efficacious drug combinations in three specific cancers (RWRDC). The random walk is a powerful mathematical model that is evolving rapidly in some important tasks, such as ranking (Minkov and Cohen, 2007; Bubnicki and Orski, 2005), similarity and recommendation (Li *et al.*, 2012; Tayebi *et al.*, 2011), or link prediction (Backstrom and Leskovec, 2011). Then, we analyzed the value of restart probability and the rationality of scoring for RWRDC. Moreover, our model achieved reliable predictions with cross-validation and it obtained higher the Area Under Receiver Operating Characteristic Curves (AUROCs) than other methods for three data sets. A case study of breast cancer demonstrated the effectiveness and potential of RWRDC in identifying potential drug combinations. Finally, we discussed the strengths and limitations of RWRDC and its prospects.

II. Results

In this section, we first analyzed the value of restart probability and the rationality of scoring for RWRDC. Next, cross-validation evaluated the predictive performance of RWRDC. Then, we conducted a case study to verify the efficiency of RWRDC in discovering novel combinatorial drugs.

A. *Effects of restart probability*

A receiver operating characteristic (ROC) curve is a graphical plot that shows the diagnostic ability of the binary classifier system because its recognition thresholds are different (Fawcett, 2006). AUROC (Area Under Receiver Operating Characteristic Curve) is the area under the ROC curve with a value between 0 and 1. AUROC can intuitively evaluate the quality of classifier, the larger the value, the better.

Leave-one-out cross-validation (LOOCV) is a widely used cross-validation, especially when the data size is not very large. Specifically, one positive sample (a known efficacious drug combination) was taken as the test set in turn, the other positive samples (other known efficacious drug combinations) were taken as the training set, and the other non-positive samples (other possible drug combinations) were taken as the CS. After scoring by RWRDC, every sample of PS and CS was scored and sorted. After taking all positive samples as the test set in turn, the score of the test set was compared with the scores of all candidate samples, then the predictive performance of the RWRDC was evaluated by changing the preset threshold.

To determine the effect of restart probability $1-c$ on the performance of RWRDC, we calculated AUROCs with various values of $1-c$ ranging from 0.05 to 0.95 by LOOCV. **Figure 1** shows the performance of different restart probabilities on the LOOCV in the different data sets. The orange line represented the average AUROC when the restart probability changed. It could be observed that the performance of RWRDC was better when restart probability was about 0.05. Therefore, we set $1-c$ to 0.05.

B. *Rationality Analysis of Scoring*

To check whether the drug-drug pairs with high efficacious scores were more likely to be efficaciously combined, all candidate drug-drug pairs in three data sets were ranked by RWRDC. In this paper, the average efficacious score of PS (efficacious combinations), CS (unknown combinations), and NS (non-efficacious combinations in CS) were calculated respectively. If the average efficacious score of PS was higher than that of CS and also the average efficacious score of CS higher than that of NS on different data sets, the scoring process was reasonable.

In **Table 1**, The average efficacious score of PS, CS, and NS were expressed by *PS score*, *CS score*, and *NS score*, respectively. As shown in **Table 1**, all efficacious scores of NS were close to zero, drug-drug pairs with higher efficacious scores were more likely to be efficacious drug combinations of three data sets.

C. *Cross-validation tests*

In this paper, we used LOOCV for different methods and data sets to test the predictive ability of the model. Briefly speaking, the drug-drug pairs corresponding to the elements of 1 in the adjacent matrix A were proportionally used as the training set and the test set, and the elements of 0 in the adjacent matrix A were used as the CS. Then the efficacious score of each sample of the test set and CS was predicted.

Besides, the model was compared with the random walk model and NLLSS for different data sets. To compare the predictive power of different models more fairly, we used the same data, that is, we only used the data of existing efficacious drug combinations, and we did not use data such as the similarity between drugs. Especially, in the NLLSS model, we regarded each drug as the main drug and the adjuvant drug.

As shown in **Figure 2**, in the case of LOOCV, the blue line represented the ROC curve of RWRDC, the red line and orange line represented the ROC curve of the random walk model and NLLSS, respectively. The graphs show that the prediction result of RWRDC is better than that of the random walk model and NLLSS.

D. Case study

The case study aimed to examine the capability of RWRDC in discovering novel efficacious drug combinations. For breast cancer, we ranked all drug-drug pairs based on their corresponding predicted efficacious scores. Prediction results were verified based on not only the DCDB database but recently published experimental literature. The efficacious drug combinations account for about 5%-9% of the total combinations of each cancer data set (**Table 3**), therefore, we analyzed the top 5% and 10% predictions. **Table 2** listed all the corresponding top 10% ranking results. The "Null" in the table indicates that documentation has not been found.

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related mortality in females worldwide (Zheng *et al.*, 2014), which comprises 22% of all cancers in women (Donahue and Genetos, 2013; Karagoz *et al.*, 2015). We used RWRDC to predict potential efficacious drug combinations. As a result, 100% and 73% of drug-drug pairs which ranked top 5% and 10% in the data set had been validated, respectively. When we took the drug-drug pairs of top 10% the highest predicted efficacious score as the predicted result, all the known efficacious drug combinations contained in the data set would be successfully predicted. Although some predictions have not been validated in databases and literature, there is still a strong possibility that predicted efficacious drug combinations will be available. For example, everolimus-lapatinib was predicted to be 59th by RWRDC and there was no direct evidence that it was an efficacious drug combination. But (Barr *et al.*, 02152016) showed that they were conducting a phase II pilot study, and some HER-2 positive metastatic breast cancer (MBC) were treated with the combination of everolimus and lapatinib for pts. We are confident that the combination of everolimus and lapatinib will perform well in clinical trials. Also, the combination of sorafenib and lapatinib ranked 75th by RWRDC, (Kacan *et al.*, 2014) showed that both sorafenib and lapatinib alone were efficacious in the treatment of breast cancer, also a combination of these two agents may be a promising therapeutic option of the treatment of breast cancer. Besides, sorafenib and docetaxel ranked 81st in the prediction results, although there was no specific literature indicating that they could be combined to treat breast cancer. A phase II trial of sorafenib combined with docetaxel in the treatment of HER2-negative MBC was conducted (Marme *et al.*, 2014). The existing data could not conclude that sorafenib and docetaxel could treat breast cancer, with the development of the trial, there would be clear conclusions soon.

III. Discussions

The model proposed in this paper is an extension of the classical random walk model in the stochastic process and has a solid mathematical foundation. Mathematically, when the stochastic process reaches a steady-state, we can get the efficacious score s_{ij}, that is, the possibility of connecting the edges between any vertices v_i and v_j. The drug is treated as a vertex in the modeling process, therefore, there is an edge between the two vertices indicating that the two drugs can be effectively combined, so the model in this paper can be used to predict which drug-drug pairs can be efficacious drug combinations. For a given disease, potential efficacious drug combinations can be predicted. Based on the different types of cross-validation and different data sets, all the results showed that RWRDC had a very strong predictive ability.

Despite the effectiveness of RWRDC, it also has several limitations. Firstly, more types of data can be considered in the model, such as chemical structure similarity between drugs, the semantic similarity between diseases, drug-target interactions, and so on. Introducing more data can improve the prediction accuracy. Furthermore, if the model can quantitatively predict the toxicity and side effects of drug combinations, it will be more convincing. Besides, only the efficacious combination data of two drugs are utilized, and efficacious drug combinations of more than two drugs can be considered in the future. Moreover, we think efficacious drug combinations of more than two drugs can be predicted by a more complex random walk model soon afterward.

In brief, RWRDC is expected to be an important and useful resource for identifying potential efficacious drug combinations and providing useful insights into the underlying mechanisms of drug combinations.

IV. Conclusion

In this manuscript, we predicted efficacious drug combinations in cancer based on random walk with restart. An efficacious score can be obtained between any two individual drugs by RWRDC. The analysis results suggest that RWRDC is a novel way to discover efficacious drug combinations in cancer, which provides new prospects for cancer treatment. Furthermore, RWRDC is a semi-supervised learning framework that can be used to predict combinations of drugs for other complex diseases.

V. Methods

A. Data collection and pre-processing

The Drug Combination Database (DCDB) is an available database to collect and organize drug combinations information (Liu *et al.*, 2014), which includes 1363 drug combinations, consisting of 1735 drugs in all. Among these drug combinations, 682 are curated from clinical trials published in the reports, 330 are curated from the FDA orange book and 351 are curated from PubMed articles. There are 1103 efficacious drug combinations in all drug combinations. Besides, DCDB collected 237 non-efficacious drug combinations, which may provide valuable data support for finding efficacious drug combinations. In this paper, we set efficacious drug-drug interactions as positive samples and non-efficacious drug-drug interactions as negative samples.

Because the mortality of breast cancer, colorectal cancer, and lung cancer is very high (Jemal *et al.*, 2010), the drug combinations in DCDB related to these cancers are the data sets studied in this paper. In detail, the data in the database were selected according to the following criteria. ICD10 is an internationally unified disease classification method developed by World Health Organization (WHO) (Quan *et al.*, 2005), each drug combination of the DCDB database had at least one ICD10 code. If a drug combination had an ICD10 code: C18, i.e. malignant neoplasm of colon, we recorded this drug combination as colorectal cancer data. Similarly, the drug combinations of C50 and C34 appearing in those ICD10 codes were recorded as breast cancer and lung cancer data, respectively. Also, due to the experimental results of some drug combinations that were still unclear, we did not use those combinations whose effect type was "Need further study" in the DCDB.

We made a simple analysis of the data in the database, as shown in **Table 3**. Here, the edge represented the efficacious drug combination, the vertex represented the drug involved in the efficacious drug combinations. The number of non-efficacious drug combinations was the total number of non-efficacious drug combinations associated with the drugs involved in the efficacious drug combinations. The non-efficacious drug combinations data could be used to verify whether the efficacious score is reasonable (in section B of the Results section). Density referred to the number of edges divided by the number of possible edges, which meant efficacious drug combinations accounted for the proportion of all possible combinations.

Table 3 shows that the density of efficacious drug combinations used in this paper is very low, that is, only 5% to 9% of all possible drug combinations.

B. Notations and definitions

Let $G(V,E)$ be an undirected graph with the vertex set $V = \{i|\ i\ is\ an\ individual\ drug\}$ and the set of edges $E = \{(i,j)\}$, here i and j are a pair of efficacious drug combinations that already exist in the current database.

Define the set of positive samples be $PS = \{(i,j)|(i,j) \in E\}$, the set of candidate samples be $CS = \{(i,j)|\ (i,j) \notin E\}$, and the set of negative samples $NS = \{(i,j)|\ (i,j)\ is\ a\ non\ efficacious\ drug\ combination\}$. In fact, CS contains possible PS and NS.

C. Random walk with restart algorithm

The random walk on the graph is a transition process by moving from a given vertex to a randomly selected neighboring vertex for each step. We regard the vertex set $\{v_1, v_2, \dots, v_n\}$ as a set of states $\{s_1, s_2, \dots, s_n\}$ in a finite Markov chain $\mathcal{M}$. The transition probability of $\mathcal{M}$ is a conditional probability defined as $P(u,v) = Prob(s_{t+1} = v|s_t = u)$ which means the probability that the $\mathcal{M}$ will be at v at time $t+1$ given that it was at u at time t. For any u of V we have $\sum_{v \in V} P(u,v) = 1$, we can define a transition matrix $P \in \mathbb{R}^{|V|\times|V|}$ of $\mathcal{M}$.

The random walk from u to v of graph G is to choose an edge e_{uv} randomly. Define transition probability $P(u,v)$ as follows:

$$P(u,v) = \frac{1}{d(u)}, \tag{1}$$

where $d(u)$ is the degree of vertex u. We denote D_G to be the diagonal matrix containing the vertex degrees of the graph and A to be the adjacency matrix of G. Thus, P can be rewritten in matrix notation as follows:

$$P = D_G^{-1}A. \tag{2}$$

Define $r_t \in \mathbb{R}^{|V|\times 1}$ as a vector in which the i-th element represents the probability of discovering the random walk at vertex i at step t, then the probability r_{t+1} can be calculated iteratively by:

$$r_{t+1} = P^T r_t. \tag{3}$$

For the random walk with restart (RWR) algorithm (Tong *et al.*, 2006), there is an additional restart item compared to the above algorithm. The probability r_{t+1} can be calculated iteratively by the following expression:

$$r_{t+1} = cP^T r_t + (1-c)r_0. \tag{4}$$

Define initial probability $r_0 \in \mathbb{R}^{|V|\times 1}$ as a vector in which the i-th element is equal to one, while other elements are zeros. And $1-c$ is the restart probability ($0 \le c \le 1$).

At the beginning of the RWR, we choose a starting vertex v_i, then it would walk based on transition probability. Suppose we come to vertex v_j, then it has a probability of c to walk based on transition probability and also have a probability of $1-c$ to restart the walk, that is, to go back to vertex v_i. After some steps, the RWR will be stable, that is, the $L1$ norm between r_{t+1} and r_t is less than an arbitrarily small value. Here we took the cutoff as 10^{-6}, and this criterion is often used in some researches (Chen *et al.*, 2012; Chen *et al.*, 2010; Köhler *et al.*, 2008; Li and Patra, 2010). When the RWR is stable, stable probability between vertex i and vertex j is defined as the j-th element of r_t corresponding to the starting vertex is v_i.

D. Efficacious scoring of drug-drug pairs

Each candidate drug-drug pair can be ranked twice according to RWR. Define $\Pi = \{\pi_{ij}\}_{|V|\times|V|}$ be the stable probability matrix where π_{ij} indicates the stable probability between vertex i and vertex j, that is, RWR starts from vertex v_i and the probability of reaching vertex v_j when the process is stable. Briefly speaking, the value of π_{ij} is the j-th element of steady-state probability $\lim_{t\to\infty} r_t$ when the i-th element of r_0 is 1. The significance of π_{ij} in drug combination can be understood as the possibility of efficacious combination of drug j (i.e. for $n-1$ candidate drugs except for drug i) and given drug i; similarly, π_{ji} can be understood as the possibility of an efficacious combination of drug i (i.e. for $n-1$ candidate drugs except for drug j) and given drug j. The goal of this paper is to predict the efficacious combination of two drugs, so each drug cannot be combined with itself, but can be combined

with other $n-1$ drugs. Therefore, define efficacious score $S=\{s_{ij}\}_{|V|\times|V|}$ as follows:

$$s_{ij}=\begin{cases}\pi_{ij}+\pi_{ji} & if\ i\neq j\\ 0 & if\ i=j\end{cases} \quad (5)$$

An efficacious score of any two drugs can be calculated by RWRDC. Drug-drug pairs with higher efficacious scores can be expected to have a greater probability to be combinatorial drugs. Experimenters can prioritize the validation of these drug-drug pairs, therefore, the cost to identify potential drug combinations can be significantly reduced. The flowchart of the RWRDC are shown in **Figure 3**.

The code and dataset of RWRDC are freely available at https://github.com/wangqi27/ERWR

VI. List of Abbreviations

RWRDC: predicting efficacious Drug Combinations in cancer based on Random Walk with Restart
NLLSS: Network-based Laplacian regularized Least Square Synergistic drug combination prediction
AUROCs: Area Under Receiver Operating Characteristic Curves
ROC: Receiver Operating Characteristic
LOOCV: Leave-One-Out Cross-Validation
DCDB: Drug Combination Database
MBC: Metastatic Breast Cancer
WHO: World Health Organization
RWR: Random Walk with Restart

VII. Declarations

- Ethics approval and consent to participate
Not applicable.

- Consent to publish
Not applicable.

- Availability of data and materials
Publicly available datasets were analyzed in this study. This data can be found here: http://www.cls.zju.edu.cn/dcdb/. (Temporary DCDB database address: http://public.synergylab.cn/dcdb/.)
The code and dataset of RWRDC are freely available at https://github.com/wangqi27/ERWR.

- Competing interests
The authors declare that they have no competing interests.

- Funding
This work was supported by the National Natural Science Foundation of China under Grant No. 11631014. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

- Authors' Contributions
QW conceived the project, developed the prediction method, designed the experiments, implemented the experiments, analyzed the result, and wrote the paper. GY conceived the project, analyzed the result, and revised the paper.
All authors have read and approved the manuscript.

- Acknowledgements
This work was supported by the National Natural Science Foundation of China. And the authors would like to acknowledge Dr. Yaqi Zhang for technical support. We thank Dr. Ming Chen *et al.* for making the drug combination data complete, published in ref. (Liu *et al.*, 2014), freely available via the DCDB Database.

- Authors' Information
Affiliations
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China.
Qi Wang, Guiying Yan

University of Chinese Academy of Sciences, Beijing, 100190, China.
Qi Wang, Guiying Yan

Corresponding author
Correspondence to Guiying Yan (yangy@amss.ac.cn).

VIII. References

Bachelot, T. *et al.* (2012) Randomized phase II trial of everolimus in combination with tamoxifen in patients with hormone receptor-positive, human epidermal growth factor receptor 2-negative metastatic breast cancer with prior exposure to aromatase inhibitors: a GINECO study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **30**, 2718–2724.

Backstrom, L. and Leskovec, J. (2011) Supervised random walks. In King, I., Nejdl, W. and Li, H. (eds), Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11. New York, New York, USA, ACM Press, 635.

Bai, L.-Y. *et al.* (2018) Prediction of Effective Drug Combinations by an Improved Naïve Bayesian Algorithm. *International journal of molecular sciences*, **19,** 2018: 10.3390/ijms19020467.

Barr, J.A. *et al.* (02152016) Abstract OT3-01-12: Phase II trial of lapatinib and everolimus for HER2 positive metastatic breast cancer. In, Ongoing Clinical Trials. American Association for Cancer Research, OT3-01-12-OT3-01-12.

Bartsch, R. *et al.* (2007) Capecitabine and trastuzumab in heavily pretreated metastatic breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **25**, 3853–3858.

Bubnicki, Z. and Orski, D. (2005) Resource and Task Allocation in Complex of Parallel Operations with Uncertain and Random Parameters. In, 18th International Conference on Systems Engineering (ICSEng'05). IEEE, 499–504.

Canonici, A. *et al.* (2013) Neratinib overcomes trastuzumab resistance in HER2 amplified breast cancer. *Oncotarget*, **4**, 1592–1605.

Chen, H. *et al.* (09042019) Large-scale Analysis of Drug Combinations by Integrating Multiple Heterogeneous Information Networks. In Shi, X., Buck, M., Ma, J. and Veltri, P. (eds), Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. New York, NY, USA, ACM, 67–76.

Chen, X. *et al.* (2010) A novel candidate disease genes prioritization method based on module partition and rank fusion. *Omics : a journal of integrative biology*, **14**, 337–356.

Chen, X. *et al.* (2012) Drug-target interaction prediction by random walk on the heterogeneous network. *Molecular bioSystems*, **8**, 1970–1978.

Chen, X. *et al.* (2016) NLLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning. *PLoS computational biology*, **12**, e1004975.

Collier, R. (2012) Reducing the "pill burden". *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, **184**, E117-8.

Ding, P. *et al.* (2019) Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological, and Network Knowledge. *IEEE journal of biomedical and health informatics*, **23**, 1336–1345.

Donahue, H.J. and Genetos, D.C. (2013) Genomic approaches in breast cancer research. *Briefings in functional genomics*, **12**, 391–396.

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861–874.

Feliu, J. *et al.* (2009) Chemotherapy for colorectal cancer in the elderly: Whom to treat and what to use. *Cancer treatment reviews*, **35**, 246–254.

G. Steger et al. (2005) 455 PUBLICATION Results from a pilot study of goserelin plus fulvestrant in premenopausal women with advanced, hormone-sensitive breast cancer, **3**, 127–128.

Gradishar, W.J. *et al.* (2004) Capecitabine plus paclitaxel as front-line combination therapy for metastatic breast cancer: a multicenter phase II study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **22**, 2321–2327.

Groll, A.H. and Tragiannidis, A. (2009) Recent advances in antifungal prevention and treatment. *Seminars in hematology*, **46**, 212–229.

Janizek, J. D., Celik, S. and Lee, S.-I. (2018) Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine.

Jemal, A. *et al.* (2010) Global patterns of cancer incidence and mortality rates and trends. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, **19**, 1893–1907.

Kacan, T. *et al.* (2014) Investigation of Antitumor Effects of Sorafenib and Lapatinib Alone and in Combination on MCF-7 Breast Cancer Cells. *Asian Pacific Journal of Cancer Prevention*, **15**, 3185–3189.

Karagoz, K. *et al.* (2015) Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. *Omics : a journal of integrative biology*, **19**, 115–130.

Köhler, S. *et al.* (2008) Walking the interactome for prioritization of candidate disease genes. *American journal of human genetics*, **82**, 949–958.

Li, J. *et al.* (2020) A Machine Learning Method for Drug Combination Prediction. *Front. Genet.*, **11,** 2020: 10.3389/fgene.2020.01000.

Li, X. *et al.* (2012) Social network-based recommendation. In Boughida, K. B., Howard, B., Nelson, M. L., van de Sompel, H. and Sølvberg, I. (eds), Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries - JCDL '12. New York, New York, USA, ACM Press, 409.

Li, Y. and Patra, J.C. (2010) Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics*, **26**, 1219–1224.

Liu, Y. *et al.* (2014) DCDB 2.0: a major update of the drug combination database. *Database : the journal of biological databases and curation*, **2014**, bau124.

Mark Verrill (2013) A trial looking at either ixabepilone or paclitaxel with bevacizumab for breast cancer that has come back (CA163115). https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-trial-looking-at-either-ixabepilone-or-paclitaxel-chemotherapy-in-combination-with-bevacizumab-for-breast-cancer-that-has-come-back-locally-or-has-spread#undefined.

Marme, F. *et al.* (2014) Sorafenib (SOR) plus docetaxel (DOC) as first-line therapy in patients with HER2-negative metastatic breast cancer (MBC): A randomized, placebo-controlled phase II trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **32**, 1072.

Minkov, E. and Cohen, W.W. (2007) Learning to rank typed graph walks. In Zhang, H., Mobasher, B., Giles, L., McCallum, A., Nasraoui, O., Spiliopoulou, M., Srivastava, J. and Yen, J. (eds), Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis - WebKDD/SNA-KDD '07. New York, New York, USA, ACM Press, 1–8.

Pan, G.-Q. *et al.* (2009) Expression of semaphorin 5A and its receptor plexin B3 contributes to invasion and metastasis of gastric carcinoma. *World journal of gastroenterology*, **15**, 2800–2804.

Pegram, M.D. and Slamon, D.J. (1999) Combination therapy with trastuzumab (Herceptin) and cisplatin for chemoresistant metastatic breast cancer: evidence for receptor-enhanced chemosensitivity. *Seminars in oncology*, **26**, 89–95.

Preuer, K. *et al.* (2018) DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics (Oxford, England)*, **34**, 1538–1546.

Quan, H. *et al.* (2005) Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, **43**, 1130–1139.

Tamma, P.D. *et al.* (2012) Combination therapy for treatment of infections with gram-negative bacteria. *Clinical microbiology reviews*, **25**, 450–470.

Tas, F. *et al.* (2008) Circulating levels of vascular endothelial growth factor (VEGF), matrix metalloproteinase-3 (MMP-

3), and BCL-2 in malignant melanoma. *Medical oncology (Northwood, London, England)*, **25**, 431–436.

Tayebi, M.A. *et al.* (2011) CrimeWalker. In Mobasher, B., Burke, R., Jannach, D. and Adomavicius, G. (eds), Proceedings of the fifth ACM conference on Recommender systems - RecSys '11. New York, New York, USA, ACM Press, 173.

Thor, A. *et al.* Metformin Enhances Trastuzumab Efficacy and Overcomes Resistance in HER2 Type Breast Cancer Cells. In, Poster Session Abstracts 2009, 1133.

Tianyu Zhang, Liwei Zhang, Philip R.O. Payne, Fuhai Li (2018) Synergistic Drug Combination Prediction by Integrating Multi-omics Data in Deep Learning Models, 2018.

Tong, H. *et al.* (2006 - 2006) Fast Random Walk with Restart and Its Applications. In, Sixth International Conference on Data Mining (ICDM'06). IEEE, 613–622.

Wang, L. (2007) Docetaxel-cisplatin combined chemotherapy in advanced breast cancer, 1102–1103.

Xiaorong Y, Lingru M, Huanxin Y, et al. (2015) Clinical Observation of Pemetrexed Combined with Capecitabine in the Treatment of Recurrent and Metastatic Breast Cancer. *Journal of Basic and Clinical Oncology*, 2015.

Xu, Q. *et al.* (2017) PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm. *Journal of theoretical biology*, **417**, 1–7.

Yang L, Xin T, Rong W U (2014) Efficacy of combined therapy of goserelin and exemestane on the advanced and metastatic breast cancer patients in the premenopausal and perimenopausal. *Practical Pharmacy and Clinical Remedies*, 2014.

Yardley, D.A. *et al.* (2015) A phase II trial of ixabepilone and cyclophosphamide as neoadjuvant therapy for patients with HER2-negative breast cancer: correlation of pathologic complete response with the 21-gene recurrence score. *Breast cancer research and treatment*, **154**, 299–308.

Yi Z, Qingli Z, Ji M (2018) Proliferation and apoptosis effect of combining metformin and docetaxel on breast cancer cells. *Journal of Modern Oncology*, 2018.

Zheng, R.-R. *et al.* (2014) Effects of doxorubicin and gemcitabine on the induction of apoptosis in breast cancer cells. *Oncology reports*, **32**, 2719–2725.

Figure legends

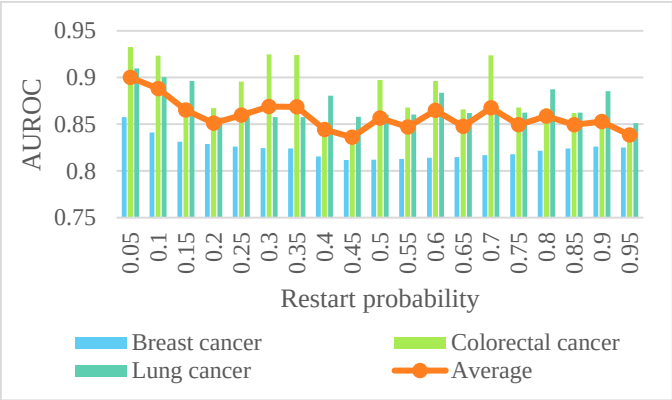


Fig. 1. The performance of restart probability on the LOOCV.

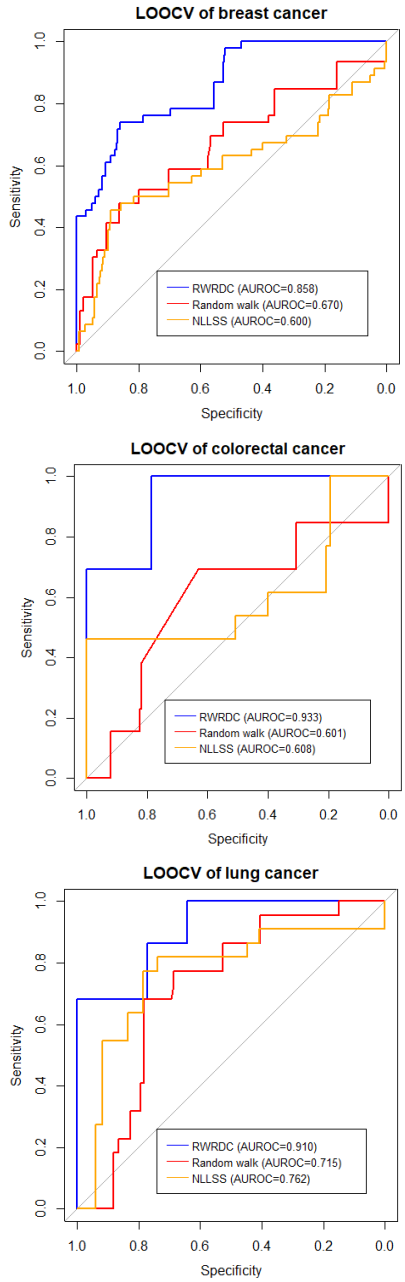
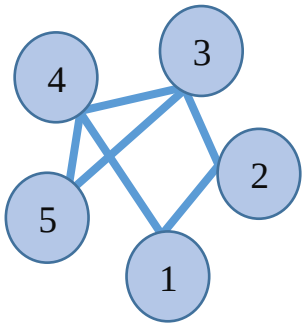


Fig. 2. The ROC curves of breast cancer, colorectal cancer and lung cancer with LOOCV.

Step 1:
Select a disease from DCDB database and construct the drug combination network of it:



Step 2:
Select initial node 1, the calculation process is as follows:

$$P = D_G^{-1}A = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$t = 0, r_0 = [1\ 0\ 0\ 0\ 0]'$$
$$r_1 = cP^T r_0 + (1-c)r_0$$
$$r_2 = cP^T r_1 + (1-c)r_0$$
$$\dots$$
$$r_{t+1} = cP^T r_t + (1-c)r_0$$

When

$$\|r_{t+1} - r_t\| \ll 10^{-6}$$
$$\Pi(:,1) = r_t = [0.2081\ 0.1714\ 02290\ 0.2423\ 0.1492]'$$

Step 3:
Select initial node 2/3/4/5 to compute $\Pi(:,2)$ / $\Pi(:,3)$ / $\Pi(:,4)$/ $\Pi(:,5)$, then we got Π.

$$S = \Pi + \Pi'(The\ diagonal\ element\ of\ S\ is\ 0)$$
$$= \begin{bmatrix} 0 & 0.3428 & 0.3817 & 0.4038 & 0.2985 \\ 0.3428 & 0 & 0.4038 & 0.3817 & 0.2985 \\ 0.3817 & 0.4038 & 0 & 0.4797 & 0.4124 \\ 0.4038 & 0.3817 & 0.4797 & 0 & 0.4124 \\ 0.2985 & 0.2985 & 0.4124 & 0.4124 & 0 \end{bmatrix}$$

Step 4:
Obtain the ranking list of all combinations of drugs:

Rank	Drug 1	Drug 2	Efficacious score	Evidence
1	3	4	0.4797	DCDB
2	3	5	0.4124	DCDB
3	1	4	0.4038	DCDB
4	1	3	0.3817	Potential
5	1	2	0.3428	DCDB
6	2	5	0.2985	Potential
…	…	…	…	…

Fig. 3. Flowchart of the RWRDC.

TABLE I

COMPARING THE EFFICACIOUS SCORES ON DIFFERENT DATA SETS

Data sets	*PS score*	*CS score*	*NS score*
Breast cancer	0.0425	1.5589e-4	6.9571e-5
Colorectal cancer	0.0659	2.2667e-4	5.9617e-4
Lung cancer	0.0519	3.1624e-4	1.8237e-7

TABLE II

CASE STUDY OF BREAST CANCER

Rank	Drug1	Drug2	Evidence
1	Arzoxifene	LG100268	DCDB
2	Decitabine	Depsipeptide	DCDB
3	Taxane	Gefitinib	DCDB
4	PD98059	Gefitinib	DCDB
5	Fulvestrant	Tipifarnib	DCDB
6	Olaparib	Cediranib	DCDB
7	Vinorelbine	Neratinib	DCDB
8	Cisplatin	Gemcitabine	DCDB
9	Doxorubicin	Mitomycin	DCDB
10	Doxorubicin	Metformin	DCDB
11	Methylseleninic acid	Tamoxifen	DCDB
12	Cyclophosphamide	Pemetrexed	DCDB
13	Sorafenib	Capecitabine	DCDB
14	Ixabepilone	Capecitabine	DCDB
15	Paclitaxel	Tubacin	DCDB
16	Paclitaxel	Trabectedin	DCDB
17	Paclitaxel	Toremifene	DCDB
18	Epirubicin	Trastuzumab	DCDB
19	Trastuzumab	Bortezomib	DCDB
20	Carboplatin	Trastuzumab	DCDB
21	Fulvestrant	Anastrozole	DCDB
22	Anastrozole	Goserelin acetate	DCDB
23	Everolimus	Exemestane	DCDB
24	Everolimus	Letrozole	DCDB
25	Tamoxifen	Exemestane	DCDB
26	Tamoxifen	Goserelin acetate	DCDB
27	Lapatinib	Letrozole	DCDB
28	Paclitaxel	Olaparib	DCDB
29	Vinorelbine	Trastuzumab	DCDB
30	Docetaxel	Gemcitabine	DCDB
31	Bevacizumab	Lapatinib	DCDB
32	Doxorubicin	Docetaxel	DCDB
33	Gemcitabine	Trastuzumab	DCDB
34	Lapatinib	Capecitabine	DCDB
35	Bevacizumab	Capecitabine	DCDB
36	Cyclophosphamide	Capecitabine	DCDB
37	Docetaxel	Capecitabine	DCDB
38	Paclitaxel	Bevacizumab	DCDB
39	Paclitaxel	Lapatinib	DCDB
40	Paclitaxel	Cyclophosphamide	DCDB
41	Docetaxel	Trastuzumab	DCDB
42	Doxorubicin	Trastuzumab	DCDB
43	Cyclophosphamide	Trastuzumab	DCDB
44	Bevacizumab	Trastuzumab	DCDB
45	Trastuzumab	Tamoxifen	DCDB
46	Paclitaxel	Trastuzumab	DCDB
47	PD98059	Taxane	Null
48	Tipifarnib	Anastrozole	Null
49	Paclitaxel	Cediranib	Null
50	Neratinib	Trastuzumab	(Canonici *et al.*, 2013)
51	Fulvestrant	Goserelin acetate	(G. Steger et al., 2005)
52	Letrozole	Exemestane	Null
53	Metformin	Mitomycin	Null
54	Cisplatin	Docetaxel	(Wang, 2007)
55	Everolimus	Tamoxifen	(Bachelot *et al.*, 2012)
56	Anastrozole	Tamoxifen	Null
57	Methylseleninic acid	Exemestane	Null
58	Methylseleninic acid	Goserelin acetate	Null
59	Everolimus	Lapatinib	Null
60	Cisplatin	Trastuzumab	(Pegram and Slamon, 1999)
61	Ixabepilone	Sorafenib	Null
62	Docetaxel	Mitomycin	Null
63	Metformin	Docetaxel	(Yi Z, Qingli Z, Ji M, 2018)
64	Pemetrexed	Capecitabine	(Xiaorong Y, Lingru M, Huanxin Y, et al., 2015)
65	Paclitaxel	Pemetrexed	(Tas *et al.*, 2008)
66	Trastuzumab	Mitomycin	Null
67	Metformin	Trastuzumab	(Thor *et al.*)
68	Trastuzumab	Pemetrexed	Null
69	Methylseleninic acid	Trastuzumab	Null
70	Goserelin acetate	Exemestane	(Yang L, Xin T, Rong W U, 2014)
71	Trabectedin	Tubacin	Null
72	Toremifene	Tubacin	Null
73	Toremifene	Trabectedin	Null
74	Paclitaxel	Capecitabine	(Gradishar *et al.*, 2004)
75	Sorafenib	Lapatinib	Null
76	Ixabepilone	Lapatinib	Null
77	Bevacizumab	Sorafenib	Null
78	Ixabepilone	Bevacizumab	(Mark Verrill, 2013)
79	Ixabepilone	Cyclophosphamide	(Yardley *et al.*, 2015)
80	Cyclophosphamide	Sorafenib	Null
81	Sorafenib	Docetaxel	Null
82	Ixabepilone	Docetaxel	Null
83	Doxorubicin	Gemcitabine	(Pan *et al.*, 2009)
84	Trastuzumab	Capecitabine	(Bartsch *et al.*, 2007)

TABLE III

GLOBAL CHARACTERISTICS OF THE DRUG COMBINATION NETWORK

Data sets	No. of vertices	No. of edges	No. of non-efficacious drug combinations	Density
Breast cancer	41	46	4	0.0561
Colorectal cancer	18	13	1	0.0850
Lung cancer	24	22	2	0.0797