American Journal of Biochemistry and Biotechnology

Original Research Paper

Identification of Microbe-Drug Association based on Weighted Profile and Collaborative Matrix Factorization

[1]Lingzhi Zhu, [2]Ying Xiao, [3]Junling Zhang, [3]Chunhua Li and [1]Jun Wang

[1]School of Computer Science and Engineering, Hunan Institute of Technology, Hengyang, China
[2]Hunan Chemical Vocational Technology College, Zhuhou, China
[3]Cooperative Development Office, Hengyang Normal University, Hengyang, China

Article history
Received: 09-07-2021
Revised: 08-11-2021
Accepted: 12-11-2021

Corresponding Author:
Ying Xiao
Hunan Chemical Vocational Technology College, Zhuhou, China
Email: lingzhi0825@yeah.net

Abstract: Previous studies have shown that diseases are associated with microbe. To explore a more effective treatment for these diseases, unknown microbe-drug associations must be identified. However, existing models to identify microbe-drug association are limited. In our article, a predictive model (WPCMF) is presented for identifying microbe-drug associations based on weighted profile and collaborative matrix factorization. In WPCMF, the Gaussian Interaction Profile (GIP) can be used for computing the similarities of microbe and the drug, respectively. Then we use the Canonical SMILES of drugs to compute the chemical structures similarity of drugs. Two drug similarities are fused into an integrated drug similarity matrix. Weighted profile and collaborative matrix factorization are applied for predicting potential microbe-drug associations. Experimental results show that WPCMF achieves the average Area Under the Curve (AUC) values of 0.9096±0.0028, 0.9195±0.0019 and 0.9236 in 5-fold Cross-Validation (5 CV), 10-fold Cross-Validation (10 CV) and Leave-One-Out-Cross-Validation (LOOCV), respectively, which consistently outperforms other related methods (KATZHMDA, WP, CMF and Kron RLS). We think WPCMF is ideal as a supplement in the field of biomedical research.

Keywords: Microbe-Drug Associations, Similarity, Weighted Profile, Collaborative Matrix Factorization, Gaussian Interaction Profile (GIP)

Introduction

Accumulating studies have shown that diseases are associated with microbe (Young, 2017). To explore a more effective treatment for these diseases, unknown microbe-drug associations must be identified (Young, 2017; Ley *et al*., 2005; Larsen *et al*., 2010). For example, a clostridium difficile infectious disease is associated with the function and diversity of microbial communities (Young, 2017). Ley *et al*. (2005) revealed that *Bacteroidetes* is significantly reduced whereas *Firmicutes* is enhanced in obesity. Furthermore, *Firmicutes* obviously decreases in type 2 diabetes (Larsen *et al*., 2010). Based on known microbe-disease associations and symptom-based disease similarity, Zhang *et al*. (2018a) presented a label propagation method to discovery microbe-disease associations. To improve the accuracy of the prediction, a model of graph regularized non-negative matrix factorization was proposed to accurately discovery latent associations between diseases and microbes (Zhang *et al*., 2018b). To cure these diseases, antibiotics can be applied for restoring the function and diversity of microbial communities. However, as the broader application, abusing of antibiotics is becoming increasingly serious problem which followed by dangerous microbial drug resistance. In particular the abusing of antibiotics has been changing so that more sophisticated microbes. More than 70% of bacteria are resistant to at least one class of antibiotics. The finding rate of new antibiotics continues to decline (PCT, 2015). The study of drug combination therapy and drug repurposing looks at measures to prevent development of antibiotic resistance as well as ways to stop its spread. The first step of drug combination therapy and drug repurposing is to identify potential drug-microbe associations.

Microbe communities might take part in mediation of drug activity and drug toxicity (Aarnoudse *et al*., 2008; Haiser *et al*., 2014), such asincreasing 221% in simvastatin AUC for homozygotes (Ong *et al*., 2012; Voora *et al*., 2009; Ramsey *et al*., 2014), altering the activity warfarin (Violi *et al*., 2016), increasing the toxicity of irinotecan (Guthrie *et al*., 2017) and so on. But understanding microbe-drug association mechanisms is limited.

In our article, a Predictive Model (WPCMF) is presented for identifying microbe-drug associations based on weighted profile and collaborative matrix factorization. In WPCMF, the Gaussian Interaction Profile (GIP)can be used for computing the similarities of microbe and the drug, respectively. Then we use the Canonical SMILES of drugs to compute the chemical structures similarity of drugs. Two drug similarities are fused into an integrated drug similarity matrix. Weighted profile and collaborative matrix factorization are applied for predicting potential microbe-drug associations. To validate the capability of WPCMF, we compare WPCMF with four related models, such as HMDAKATZ (Zhu *et al*., 2019), WP (Yamanishi *et al*., 2008), CMF (Shen*et al*., 2017) and Kron RLS (Van Laarhoven *et al*., 2011). 5-fold cross-validation (5 CV), 10-fold Cross-Validation (10 CV) and Leave One Out Cross-Validation (LOOCV) are introduced to confirm whether WPCMF could be more effective in predicting microbe-drug associations. In 5 CV, the AUCs of WPCMF, KATZHMDA, WP, CMF and Kron RLS are 0.9096±0.0028, 0.9010±0.0024, 0.897±0.0024, 0.6918±0.0085 and 0.6809±0.0064, respectively. The predictive performance of WPCMF is better than five related models. In 10CV, WPCMF is also better as AUC of 0.9195±0.0019, compared with four related models above (KATZHMDA: 0.9066±0.0014, WP: 0.903±0.0016, CMF: 0.7201±0.0048 and Kron RLS: 0.6897±0.0051). WPCMF is also better than five related models. In LOOCV, the AUC of WPCMF is 0.9236, while the AUC values of KATZHMDA, WP, CMF and Kron RLS are 0.9116, 0.9086, 0.762 and 0.6936, respectively. WPCMF is also better than three other models. It is obvious that WPCMF is consistently superior to five related models (KATZHMDA, WP, CMF and Kron RLS) in 5 CV, 10 CV and LOOCV.

Materials and Methods

Materials

As a commonly-used databases, MDAD (Sun *et al*., 2018) saves the information of 5,055 known associations of 1,388 drugs and 180 microbes. We download known associations from MDAD. Then we further remove the redundant associations and take the key information of known microbe-drug associations as the benchmark dataset. 1152 known associations are chosen from the data set. The specific information of known microbe-drug associations are shown in the following Table 1.

Methods

Drug Similarity

In our study, some approaches are used to measure the drug similarities, which include the drug GIP similarity and the drug chemical structure similarity. The drug GIP similarity can be computed with the known associations of drugs (Van Laarhoven *et al*., 2011; Zhu *et al*., 2021a; Lan *et al*., 2020; Chen *et al*., 2021; Zhu *et al*., 2021c; Luo *et al*., 2018). Let $M = \{m_1, m_2, m_3,\ldots, m_{nm}\}$ represent a set of *nm* microbes and $D = \{d_1, d_2, d_3,\ldots, d_{nd}\}$ be a set of *nd* drugs. And *Y* is a adjacency matrix of known microbe-drug associations, which include *nd* rows and *nm* columns. If microbe m_i and drug d_j have a known association, y_{ij} has a value of 1, otherwise 0. For drug d_i and drug d_j, the drug GIP similarity can be defined as:

$$D_G\left(d_i, d_j\right) = exp\left(-\gamma_d \| y_{d_i} - y_{d_j} \|^2\right) \quad (1)$$

$$\gamma_d = \gamma_d' \Big/ \left(\frac{1}{nd}\sum_{i=1}^{nd} \| y_{d_i} \|^2\right), \quad (2)$$

in which $y_{d_i} = \{y_{i1}, y_{i2},..., y_{inm}\}$ and $y_{d_j} = \{y_{j1}, y_{j2},..., y_{jnm}\}$ denote the interaction profiles of disease d_i to disease d_j, respectively.

In addition, Chemical Development Kit (Steinbeck *et al*., 2006) is used to compute the drug chemical structure similarity with based on the Canonical SMILES of drugs (Weininger, 1988; Wishart *et al*., 2018). Binary fingerprints of all drugs are computed by Chemical Development Kit. We use the Tanimoto (1958) of their binary fingerprints to measure the drug chemical structure similarity $D_{ch}(d_i, d_j)$.

As shown above, two drug similarity matrices are computed. We combine $D_G(d_i, d_j)$ and $D_{ch}(d_i, d_j)$ into an integrated drug similarity matrix S_d by the linear weighted method:

$$S_d = \frac{D_{ch} + D_G}{2} \quad (3)$$

Microbe Similarity

Similarly, the Gaussian Interaction Profile (GIP) can be used for computing the microbe GIP similarity (Zhu *et al*., 2019; Lan *et al*., 2021; Zhu *et al*., 2021b). The GIP similarity M_G is calculated as below:

$$S_m = M_G\left(m_i, m_j\right) = exp\left(-\gamma_m \| y_{m_i} - y_{m_j} \|^2\right) \quad (4)$$

$$\gamma_m = \gamma_m' \Big/ \left(\frac{1}{nm}\sum_{i=1}^{nm} \| y_{m_i} \|^2\right), \quad (5)$$

in which $y_{d_i} = \{y_{i1}, y_{i2},..., y_{inm}\}$ and $y_{d_j} = \{y_{j1}, y_{j2},..., y_{jnm}\}$ is the interaction profiles of microbe m_i and microbe m_j, respectively. S_m is a similarity matrix of microbes.

Weighted Profile

The weighted profile (Yamanishi *et al*., 2008) model has been successfully applied in the field of bioinformatics. In our study, the weighted profile model adopt similarities to all the other microbe and drugs by a weighted average method. We define the weighted profile model as follows:

$$\hat{Y}(d_i)=\frac{\sum_{j=1}^{nd} S_d(d_i,d_j)\times Y(d_j)}{\sum_{j=1}^{nd} S_d(d_i,d_j)} \tag{6}$$

$$\hat{Y}(m_i)=\frac{\sum_{j=1}^{nm} S_m(m_i,m_j)\times Y(m_j)}{\sum_{j=1}^{nm} S_m(m_i,m_j)} \tag{7}$$

in which $Y(d_i)$ and $Y(m_i)$ are the interaction profiles of drug d_i and microbe m_j, respectively. After running the weighted profile model, we get the average of the predictive results.

WPCMF for Microbe-Drug Association Prediction

As a traditional method, collaborative matrix factorization (Shen *et al*., 2017) have been used for identifying hidden associations, but its prediction performance need improve. In our study, a new method (WPCMF) is presented for identifying microbe-drug associations via weighted profile and collaborative matrix factorization. We describe the detail of WPCMF:

$$\min_{A,B}\left\|Y-AB^T\right\|_F^2+\lambda_l\left(\|A\|_F^2+\|B\|_F^2\right)$$

$$+\lambda_d\left\|S_d-AA^T\right\|_F^2+\lambda_m\left\|S_m-BB^T\right\|_F^2 \tag{8}$$

with:

$$Y\approx AB^T \tag{9}$$

$$S_d\approx AA^T \tag{10}$$

$$S_m\approx BB^T \tag{11}$$

where, λ_l, λ_d and λ_m are the non-negative parameters, ||.|| denote a Fresenius norm and Y is a adjacency matrix. In the above expression, we use the first term to denote the approximate model of Y, a tikhonov regularization term to minimize the norms and the last two terms to find the least-squared-error between $S_m(S_d)$ and $AA^T(BB^T)$. For λ_l, λ_d and λ_m, 5 CV, 10 CV and LOOCV are preformed to find the most appropriate values.

Initialize Matrix A and Matrix B

The first step of collaborative matrix factorization is to initialize the matrix A and matrix B. Here, we use the singular value decomposition method to decompose Y into matrix A and matrix B:

$$[U,S,V]=SVD(Y,k), \tag{12}$$

$$A=\mathrm{US}_k^{1/2} \tag{13}$$

$$B=\mathrm{VS}_k^{1/2} \tag{14}$$

in which diagonal matrix S_k includes k singular values.

The Optimization Process

In this study, the least square method is still a useful and often vital, part of the optimization process. In the optimization process, we must take into account the least square method to update matrix A and matrix B until converge. Let I_k denote the identity matrix of $k \times k$ and $\partial S/\partial A$be 0 and A can be represented as follows:

$$A=\left(YB+\lambda_d S_d A\right)\left(B^TB+\lambda_l I_k+\lambda_d A^TA\right)$$

$$\partial S/\partial A=0 \tag{15}$$

Similarly, we also use the same approach to get the representation of A as below:

$$B=(Y^TA+\lambda_m S_m A)(A^TA+\lambda_l I_k+\lambda_m B^TB)$$

$$\partial S/\partial B=0 \tag{16}$$

Results

Performance Evaluation

In our article, 5 CV, 10 CV and LOOCV are used for evaluating the ability of WPCMF. In 5CV, we divide 1152 known associations $\mathbb{S}^+$ into five parts as follows:

$$\mathbb{S}^+=\mathbb{S}_1^+\cup\mathbb{S}_2^+\cup\ldots\cup\mathbb{S}_5^+ \tag{17}$$

With:

$$\varnothing=\mathbb{S}_1^+\cap\mathbb{S}_2^+\cap\ldots\cap\mathbb{S}_5^+ \tag{18}$$

$$\left|\mathbb{S}_1^+\right|\approx\left|\mathbb{S}_2^+\right|\approx\ldots\approx\left|\mathbb{S}_5^+\right| \tag{19}$$

where $\cup$ denotes an union symbol, $\cap$ denotes an intersection symbol and $\varnothing$ denotes an empty set symbol. $\mathbb{S}_i^+$ is an exclusive subset and *i* ranges from 1 to 5. We select each part, in turn, as a testing data and other data as training data. 5 CV can be performed 100 times.

Similarly, $\mathbb{S}^+$ is divided into 10 parts continually. The process is ended until each part has approximately the same amount of microbe-drug associations as below:

$$\mathbb{S}^+ = \mathbb{S}_1^+ \cup \mathbb{S}_2^+ \cup ... \cup \mathbb{S}_{10}^+ \tag{20}$$

With:

$$\varnothing = \mathbb{S}_1^+ \cap \mathbb{S}_2^+ \cap ... \cap \mathbb{S}_{10}^+ \tag{21}$$

$$\left|\mathbb{S}_1^+\right| \approx \left|\mathbb{S}_2^+\right| \approx ... \approx \left|\mathbb{S}_{10}^+\right| \tag{22}$$

in which we select each part, in turn, as a testing data and other data as training data. 10 CV can be performed 100 times.

In LOOCV, we select each association is selected as a testing data and other data as training data. Each known microbe-drug association is ranked relative to the candidate associations.

According to the combination of the real category and the predicted category of the learner, it can be divided into True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), as listed in the following Table 2.

Comparison with other Models

To confirm the ability WPCMF, it is compared with the other four methods, such as HMDAKATZ (Zhu *et al*., 2019), WP (Yamanishi *et al*., 2008), CMF (Shen *et al*., 2017) and Kron RLS (Van Laarhoven *et al*., 2011). HMDAKATZ is a model to identify associations between drugs and targets. As a recommendation model, WP is based on the similarity of microbes and drugs to predict microbe-drug associations. Kron RLS uses the Kronecker product kernel to calculate the prediction scores and is based on regularised least squares to predict associations between microbes and drugs.

Then, we choose 5 CV, 10 CV and LOOCV to verify the ability of WPCMF and other models. In 5 CV, Fig. 1 shows that the AUCs of WPCMF, KATZHMDA, WP, CMF and Kron RLS are 0.9096±0.0028, 0.9010±0.0024, 0.897±0.0024, 0.6918±0.0085 and 0.6809±0.0064, respectively. The predictive performance of WPCMF is better than five related models.

As the Fig. 2 shows, WPCMF can achieve the AUC value of0.9195±0.0019, compared with four related models above (KATZHMDA: 0.9066±0.0014, WP: 0.903±0.0016, CMF: 0.7201±0.0048 and Kron RLS: 0.6897±0.0051). WPCMF is also better than five related models.

We can see from Fig. 3 that the AUC of WPCMF is 0.9236 in LOOCV, while the AUC values of KATZHMDA, WP, CMF and Kron RLS is 0.9116, 0.9086, 0.762 and 0.6936, respectively.

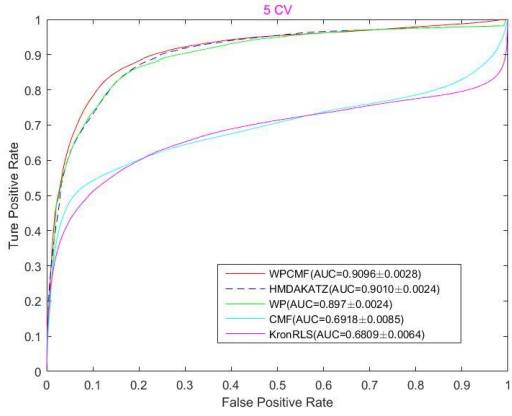

Fig. 1: The AUC curves of five models in 5 CV

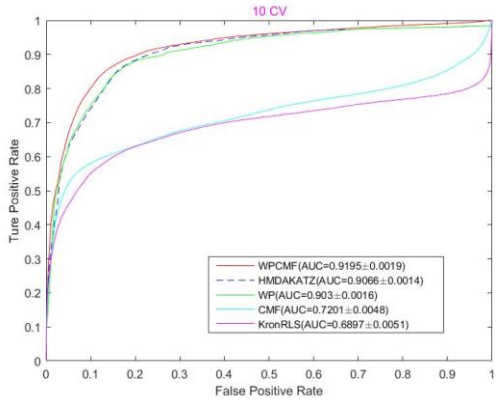

Fig. 2: The AUC curves of five models in 10 CV

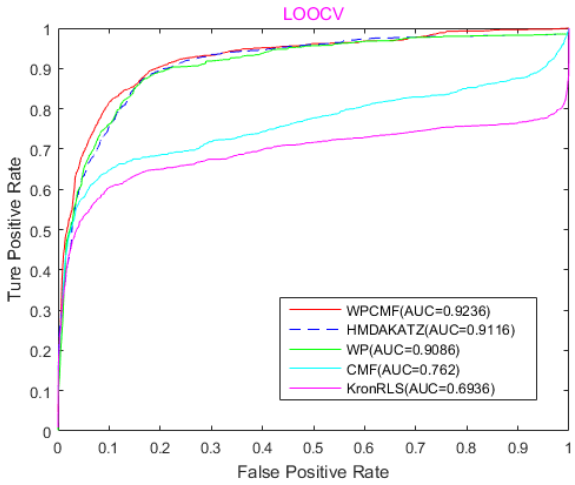

Fig. 3: The AUC curves of five models in LOOCV

Table 1: Microbe, drug and associations in the benchmark

Dataset	The benchmark dataset
Microbes	142
Drugs	627
Microbe-drug associations	1152

Table 2: Confusion matrix

	Predictive results	
Real results	Positive	Negative
Positive	*TP*	*FN*
Negative	*FP*	*TN*

It is obvious that WPCMF is consistently superior to five related models (KATZHMDA, WP, CMF and Kron RLS) in 5 CV, 10 CV and LOOCV.

Conclusion

Accumulating studies have shown that diseases are associated with microbe. To explore a more effective treatment for these disease, unknown microbe-drug associations must be identified. However, existing models to identify microbe-drug association are limited. In our study, a predictive model (WPCMF) presented for identifying microbe-drug associations based on weighted profile and collaborative matrix factorization, which can efficiently fuse multiple drug information and microbe information. In WPCMF, the Gaussian Interaction Profile (GIP) can be used for computing the similarities of microbe and the drug, respectively. Then we use the Canonical SMILES of drugs to compute the chemical structures similarity of drugs. Two drug similarities are fused into an integrated drug similarity matrix. Weighted profile and collaborative matrix factorization are applied for predicting potential microbe-drug associations. To validate the capability of WPCMF, we compare WPCMF with four related models. 5 CV, 10 CV and LOOCV are introduced to confirm that WPCMF is consistently superior to five related models (KATZHMDA, WP, CMF and Kron RLS) in 5 CV, 10 CV and LOOCV. Some multiple kernel boosting algorithm methods (Liu *et al*., 2016) and traditional machine learning (Cheng *et al*., 2020) should also be considered in the future.

Acknowledgement

This study is supported in part by the Scientific Research Foundation of Hunan Provincial Education Department (17C0229, 20C0561), the Research Foundation of Hunan Chemical Vocational Technology College (HNHY2020001), the Teaching Reform Research Foundation of Hunan Provincial Education Department (Xiang Jiao Tong [2019] 291-1023) the Hengyang City Science and Technology Program (2019jh010657, 202010031491).

Author's Contributions

Lingzhi Zhu and Ying Xiao: Designed and performed the experiments and wrote the paper.

Junling Zhang: Participated to collect the materials related to the experiment.

Chunhua Li: Participated to collect the materials related to the experiment and revised the manuscript.

Jun Wang: Wrote the paper and checked the experiment.

Ethics

The authors declare their responsibility for any ethical issues that may arise after the publication of this manuscript.

Conflict of Interest

The authors declare that they have no competing interests. The corresponding author affirms that all of the authors have read and approved the manuscript.

References

Aarnoudse, A. J. L., Dieleman, J. P., Visser, L. E., Arp, P. P., van der Heiden, I. P., van Schaik, R. H., ... & Stricker, B. H. C. (2008). Common ATP-binding cassette B1 variants are associated with increased digoxin serum concentration. Pharmacogenetics and genomics, 18(4), 299-305. doi.org/10.1097/FPC.0b013e3282f70458

Chen, Q., Lai, D., Lan, W., Wu, X., Chen, B., Chen, Y. P. P., & Wang, J. (2021). ILDMSF: Inferring associations between long non-coding RNA and disease based on multi-similarity fusion. IEEE/ACM transactions on computational biology and bioinformatics., 18(3), 1106-1112, doi.org/10.1109/TCBB.2019.2936476

Cheng, J., Liu, J., Yue, H., Bai, H., Pan, Y., & Wang, J. (2020). Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric MRI images. IEEE/ACM Transactions on Computational Biology and Bioinformatics. doi.org/10.1109/TCBB.2020.3033538.

Guthrie, L., Gupta, S., Daily, J., & Kelly, L. (2017). Human microbiome signatures of differential colorectal cancer drug metabolism. NPJ biofilms and microbiomes, 3(1), 1-8. doi.org/10.1038/s41522-017-0034-1

Haiser, H. J., Seim, K. L., Balskus, E. P., & Turnbaugh, P. J. (2014). Mechanistic insight into digoxin inactivation by Eggerthella lenta augments our understanding of its pharmacokinetics. Gut microbes, 5(2), 233-238. doi.org/10.4161/gmic.27915

Lan, W., Dong, Y., Chen, Q., Liu, J., Wang, J., Chen, Y. P. P., & Pan, S. (2021). IGNSCDA: Predicting CircRNA-Disease Associations Based on Improved Graph Convolutional Network and Negative Sampling. IEEE/ACM Transactions on Computational Biology and Bioinformatics. doi.org/10.1109/TCBB.2021.3111607

Lan, W., Lai, D., Chen, Q., Wu, X., Chen, B., Liu, J., ... & Chen, Y. P. P. (2020). LDICDL: LncRNA-disease association identification based on collaborative deep learning. IEEE/ACM Transactions on Computational Biology and Bioinformatics. doi.org/10.1109/TCBB.2020.3034910.

Larsen, N., Vogensen, F. K., Van Den Berg, F. W., Nielsen, D. S. andreasen, A. S., Pedersen, B. K., ... & Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PloS one, 5(2), e9085. doi.org/10.1371/journal.pone. 0009085

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. Proceedings of the national academy of sciences, 102(31), 11070-11075. doi.org/10.1073/pnas.0504978102

Liu, J., Li, M., Lan, W., Wu, F. X., Pan, Y., & Wang, J. (2016). Classification of Alzheimer's disease using whole brain hierarchical network. IEEE/ACM transactions on computational biology and bioinformatics, 15(2), 624-632. doi.org/10.1109/TCBB.2016.2635144

Luo, H., Lan, W., Chen, Q., Wang, Z., Liu, Z., Yue, X., & Zhu, L. (2018). Inferring microRNA-environmental factor interactions based on multiple biological information fusion. Molecules, 23(10), 2439. doi.org/10.3390/molecules23102439

Ong, F. S., Deignan, J. L., Kuo, J. Z., Bernstein, K. E., Rotter, J. I., Grody, W. W., & Das, K. (2012). Clinical utility of pharmacogenetic biomarkers in cardiovascular therapeutics: A challenge for clinical implementation. Pharmacogenomics, 13(4), 465-475. doi.org/10.2217/pgs.12.2

PCT. (2015). Antibiotics currently in clinical development.

Ramsey, L. B., Johnson, S. G., Caudle, K. E., Haidar, C. E., Voora, D., Wilke, R. A., ... & Niemi, M. (2014). The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. Clinical Pharmacology & Therapeutics, 96(4), 423-428. doi.org/10.1038/clpt.2014.125

Shen, Z., Zhang, Y. H., Han, K., Nandi, A. K., Honig, B., & Huang, D. S. (2017). miRNA-disease association prediction with collaborative matrix factorization. Complexity, 2017. doi.org/10.1155/2017/2498957

Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R., & Willighagen, E. L. (2006). Recent developments of the chemistry development kit (CDK)-an open-source java library for chemo-and bioinformatics. Current Pharmaceutical Design, 12(17), 2111-2120. doi.org/10.1021/ci025584y

Sun, Y. Z., Zhang, D. H., Cai, S. B., Ming, Z., Li, J. Q., & Chen, X. (2018). MDAD: A special resource for microbe-drug associations. Frontiers in cellular and infection microbiology, 8, 424. doi.org/10.3389/fcimb.2018.00424

Tanimoto, T. T. (1958). Elementary mathematical theory of classification and prediction. https://agris.fao.org/agris-search/search.do?recordID=US201300372414

Van Laarhoven, T., Nabuurs, S. B., & Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics, 27(21), 3036-3043. doi.org/10.1093/bioinformatics/btr500

Violi, F., Lip, G. Y., Pignatelli, P., & Pastori, D. (2016). Interaction between dietary vitamin K intake and anticoagulation by vitamin K antagonists: Is it really true?: A systematic review. Medicine, 95(10). doi.org/10.1097/MD.0000000000002895

Voora, D., Shah, S. H., Spasojevic, I., Ali, S., Reed, C. R., Salisbury, B. A., & Ginsburg, G. S. (2009). The SLCO1B1* 5 genetic variant is associated with statin-induced side effects. Journal of the American College of Cardiology, 54(17), 1609-1616. doi.org/10.1016/j.jacc.2009.04.053

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of chemical information and computer sciences, 28(1), 31-36. doi.org/10.1021/ci00057a005

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... & Wilson, M. (2018). Drug Bank 5.0: A major update to the Drug Bank database for 2018. Nucleic acids research, 46(D1), D1074-D1082. doi.org/10.1093/nar/gkx1037

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics, 24(13), i232-i240. doi.org/10.1093/bioinfor-matics/btn162

Young, V. B. (2017). The role of the microbiome in human health and disease: An introduction for clinicians. Bmj, 356. doi.org/10.1136/bmj.j831

Zhang, W., Lu, X., Yang, W., Huang, F., Wang, B., Wang, A., & Zhao, Q. (2018a, December). HNGRNMF: Heterogeneous Network-based Graph Regularized Nonnegative Matrix Factorization for predicting events of microbe-disease associations. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 803-807). IEEE. doi.org/10.1109/BIBM.2018.8621085

Zhang, W., Yang, W., Lu, X., Huang, F., & Luo, F. (2018b). The bi-direction similarity integration method for predicting microbe-disease associations. IEEE Access, 6, 38052-38061. doi.org/10.1109/ACCESS.2018.2851751

Zhu, L. Z., Yan, C., & Duan, G. H. (2021b). Prediction of Virus-Receptor Interactions Based on Improving Similarities. Journal of Computational Biology. 2021, 28(07): 650-659. doi.org/10.1089/cmb.2020.0544.

Zhu, L., Duan, G., Yan, C., & Wang, J. (2019, November). Prediction of microbe-drug associations based on Katz measure. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 183-187). IEEE. doi.org/10.1109/BIBM47256.2019.8983209

Zhu, L., Duan, G., Yan, C., & Wang, J. (2021a). Prediction of Microbe-drug Associations Based on Chemical Structures and the KATZ Measure. Current Bioinformatics, 16(6), 807-819. doi.org/10.2174/1574893616666210204144721

Zhu, L., Duan, G., Yan, C., & Wang, J. (2021c, November). Prediction of Virus-Receptor Interactions Based on Similarity and Matrix Completion. In International Symposium on Bioinformatics Research and Applications (pp. 584-595). Springer, Cham. doi.org/10.1007/978-3-030-91415-8_49