

## RESEARCH ARTICLE

# Kernel Bayesian logistic tensor decomposition with automatic rank determination for predicting multiple types of miRNA-disease associations

Yingjun Ma<sup>1\*</sup>, Yuanyuan Ma<sup>2</sup>

**1** School of Mathematics and Statistics, Xiamen University of Technology, Xiamen, China, **2** School of Computer Engineering, Hubei University of Arts and Science, Xiangyang, China

\* [yjma@xmut.edu.cn](mailto:yjma@xmut.edu.cn)

## Abstract

Identifying the association and corresponding types of miRNAs and diseases is crucial for studying the molecular mechanisms of disease-related miRNAs. Compared to traditional biological experiments, computational models can not only save time and reduce costs, but also discover potential associations on a large scale. Although some computational models based on tensor decomposition have been proposed, these models usually require manual specification of numerous hyperparameters, leading to a decrease in computational efficiency and generalization ability. Additionally, these linear models struggle to analyze complex, higher-order nonlinear relationships. Based on this, we propose a novel framework, KBLTDARD, to identify potential multiple types of miRNA–disease associations. Firstly, KBLTDARD extracts information from biological networks and high-order association network, and then fuses them to obtain more precise similarities of miRNAs (diseases). Secondly, we combine logistic tensor decomposition and Bayesian methods to achieve automatic hyperparameter search by introducing sparse-induced priors of multiple latent variables, and incorporate auxiliary information to improve prediction capabilities. Finally, an efficient deterministic Bayesian inference algorithm is developed to ensure computational efficiency. Experimental results on two benchmark datasets show that KBLTDARD has better Top-1 precision, Top-1 recall, and Top-1 F1 for new type predictions, and higher AUPR, AUC, and F1 values for new triplet predictions, compared to other state-of-the-art methods. Furthermore, case studies demonstrate the efficiency of KBLTDARD in predicting multiple types of miRNA-disease associations.

## OPEN ACCESS

**Citation:** Ma Y, Ma Y (2024) Kernel Bayesian logistic tensor decomposition with automatic rank determination for predicting multiple types of miRNA-disease associations. *PLoS Comput Biol* 20(7): e1012287. <https://doi.org/10.1371/journal.pcbi.1012287>

**Editor:** Hugues Richard, Robert Koch Institute; Robert Koch Institut, GERMANY

**Received:** October 19, 2023

**Accepted:** June 27, 2024

**Published:** July 8, 2024

**Copyright:** © 2024 Ma, Ma. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The implemented code and experimental dataset are available online at: <https://github.com/Mayingjun20179/KBLTDARD>.

**Funding:** This work was supported by the Ministry of Education of China project of Humanities and Social Sciences (Grant No. 23YJCZH160 to Yingjun Ma), the Natural Science Foundation of Fujian Province (Grant No. 2021J05260 to Yingjun Ma). The funders had no role in study design, data

## Author summary

Many studies have shown that miRNAs contribute to the onset and development of disease through a variety of potential mechanisms. Therefore, predicting multi-category associations between miRNAs and diseases is helpful for the diagnosis and treatment of complex diseases. Different from existing methods, our KBLTDARD combines the

collection and analysis, decision to publish, or preparation of the manuscript.

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

logistic tensor decomposition and Bayesian methods, realises the automatic search of hyperparameters by introducing sparse induction priors of multiple latent variables, and incorporates auxiliary information to improve the predictive ability of the Bayesian model. Unlike the simple overfitting of point estimation and the slow convergence of MCMC, we developed an efficient deterministic Bayesian inference algorithm to ensure solution efficiency. We performed 5-fold cross-validation to evaluate the performance of KBLTDARD, which can outperform other state-of-the-art methods. Moreover, the average AUC value of KBLTDARD for the prediction of 447 diseases reached 0.8165. For the prediction of miRNAs and association types of four common diseases, 17, 16, 16 and 17 of the top 20 triples were verified from the literature.

## Introduction

MicroRNAs (miRNAs) are a group of small noncoding RNAs that play important roles in many biological processes [1]. They have the ability to inhibit or promote gene expression, thereby affecting protein synthesis. As a result, dysregulation of miRNAs is associated with various biological processes and diseases [2–4]. The identification of disease-related miRNAs is highly significant for studying disease pathogenesis and drug development. In the past, biological experiments were utilized for this purpose, but such methods were not only time-consuming and laborious, but also inadequate for large-scale detection of the miRNA-disease association [5]. With the development of high-throughput sequencing technology, many related databases for miRNA and disease research have been established. Notably, DIANA-TarBase [6], miRTarBase [7], and miRWalk [8] offer a vast collection of miRNA-gene associations. MiRbase [9] and MiREDiBase [10], on the other hand, furnish sequence data and miRNA editing sites, respectively. The Comparative Toxicogenomics Database (CTD) [11] gathers a vast array of biological entities and associated information, including diseases, genes, phenotypes, and chemical compounds. These databases have significantly broadened our comprehension of miRNA functions and their regulatory mechanisms, serving as a foundation for constructing computational models to anticipate potential miRNA-disease associations.

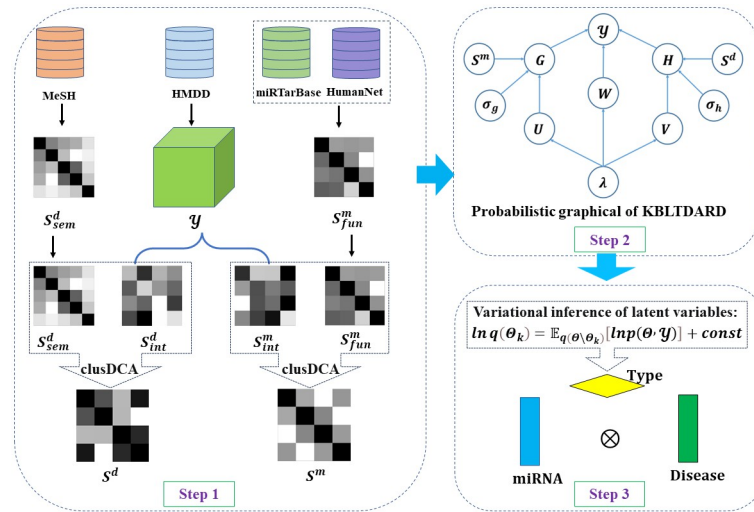
In recent years, numerous models for predicting miRNA-disease associations have been proposed. Tang et al. [12] developed a multi-channel graph convolutional network, which utilized a GCN encoder to capture features under different views, and augmented the learned prediction representation by multi-channel attention. Ma et al. [13] proposed a graph autoencoder model to address the over-smoothing problem of the GNN method, which employed a graph encoder to splice aggregate feature embeddings and self-feature embeddings, and adopted a bilinear decoder for link prediction. In previous research, to obtain higher quality similarity networks, we introduced kernel neighborhood similarity into multi-network bidirectional propagation, which effectively integrates multi-network information to improve prediction performance [14]. Li et al. [5] integrated GCN, CNN and the Squeeze Excitation Network (GCSENet) to devise a novel prediction model for miRNA-disease associations. The model utilizes GCN to gather the features from the miRNA-disease-gene heterogeneous network, performs convolutional operations via CNN, and determines the importance of each feature channel by employing SENet's squeeze and excite blocks.

Most of the aforementioned techniques concentrate on foreseeing a binary association between miRNA and disease. Nevertheless, mounting evidence indicates that the malfunction of miRNAs triggers disease through various conceivable mechanisms [15,16]. On the one hand, only certain types of miRNAs are the cause of disease. For instance, targeted deletion of

heart and muscle-specific miR-1-2 results in defects in cardiac morphogenesis, such as ventricular septal defects, and high mortality before and after birth [17]. On the other hand, the same miRNA can also cause the same disease through a different set of pathways. For example, miR-146a can directly target SMAD4, thereby regulating cell proliferation and apoptosis and playing a role in the onset and development of gastric cancer [18]. The ectopic expression of miR-146a inhibits the migration and invasion of gastric cancer cells, and down-regulates the expression of EGFR and IRAK1 [19]. Therefore, identifying miRNA-disease associations and associated types will enhance our understanding of the pathogenesis of diseases related to miRNA dysregulation on a deeper level.

In recent years, researchers have placed greater emphasis on identifying different types of miRNA-disease associations. Chen et al. [20] first proposed a restricted Boltzmann machine model (RBMMMDA) for predicting associations of miRNAs, diseases, and related types. However, RBMMMDA neglects the use of auxiliary information, which limits its predictive power to some extent. Inspired by the successful application of the existing tensor decomposition method in studying high-order biological relationships [21], Huang et al. [22] initially introduced the CANDECOMP/PARAFAC (CP) decomposition technique to multi-type miRNA-disease association prediction, coupled with biological similarity serving as a constraint to enhance prediction accuracy. Subsequently, Wang et al. [23] developed a NMCMDA method utilizing end-to-end data-driven learning for the prediction of multi-category miRNA-disease associations. To address the challenges of the current tensor decomposition model, which easily falls into local minima and produces false-negative samples, Dong et al. [24] developed a novel multi-type miRNA-disease association prediction model by integrating hypergraph learning and tensor weighting into non-negative tensor decomposition. Due to the intrinsic lack of completeness and noise in miRNA-disease-type datasets, Yu et al. [25] combined tensor decomposition and label propagation, employed robust principal component analysis on tensors to obtain low-rank prediction tensors, and utilized label propagation to transfer information. While the previously developed methods have yielded success in multi-type disease-related miRNAs prediction tasks, there are still limitations. Firstly, calculating miRNA similarities with the help of known miRNA-disease associations will cause the model to rely on the known associations. Secondly, both CP decomposition and non-negative tensor decomposition are linear models, which hinders the ability to identify complex nonlinear relationships among miRNAs, diseases and association types. Finally, the above models include numerous hyperparameters requiring adjustment, affecting both model computational efficiency and generalization ability.

To address the above challenges, we propose a novel computational model called Kernel Bayesian Logistic Tensor Decomposition with Automatic Rank Determination (KBLTDARD) for predicting different types of miRNA-disease associations. Firstly, to reduce dependence on known associations and enhance network precision, we construct the functional similarity of miRNAs from other data sources beyond miRNA-disease-type, and fused multiple similarities to obtain more accurate miRNA(disease) similarity. Secondly, to ensure nonlinear learning ability and avoid tedious hyperparameter debugging, we build hierarchical probability model to formulate logical tensor decomposition, and employ full Bayesian treatment to sparse induction priors of multiple latent variables, enabling automatic search of hyperparameters. Finally, a highly efficient deterministic Bayesian inference algorithm was developed to ensure solution efficiency. Experimental results indicate that LTDSSL outperforms other state-of-the-art methods in predicting multiple types of miRNA-disease associations with high accuracy.



**Fig 1. The workflow of our proposed KBLTDARD model.**

<https://doi.org/10.1371/journal.pcbi.1012287.g001>

## Materials and methods

### Method review

In this section, we propose a new computational model called KBLTDARD for predicting miRNA-disease-type associations, which mainly consists of three steps (shown in Fig 1). Firstly, multiple similarities of miRNA (disease) are calculated and fused to obtain a more accurate miRNA (disease) similarity network (shown in step 1 of Fig 1). Secondly, a Bayesian framework of logistic tensor decomposition is established, and the auxiliary information of miRNA (or disease) and the prior probability of latent variables are introduced to construct the probabilistic graphical model of KBLTDARD (shown in step 2 of Fig 1). Finally, the Bayesian variational inference framework for KBLTDARD is established to realize the prediction of potential miRNA-disease-type associations (shown in step 3 of Fig 1).

### Dataset

The Human MiRNA Disease Database (HMDD) contains extensive data on experimentally validated human miRNA-disease associations [15,26]. Many computational models utilize HMDD to establish benchmark data sets for execution studies [22–25,27]. To facilitate comparison, we utilize two widely used multi-type miRNA-disease data sets (HMDD v2.0 and HMDD v3.2) established by Huang et al. [22] as benchmark data sets. Specifically, HMDD v2.0 classifies miRNA-disease associations into four types based on evidence from circulation, epigenetics, genetics, and target, containing 1,675 associations for 324 miRNAs and 169 diseases under the four types, with a density of 0.681% in this dataset. HMDD v3.2 contains 16,341 associations of 713 miRNAs and 447 diseases under five types: circulation, epigenetics, genetics, target, and tissue, with a density of 1.025% in the dataset. To obtain additional auxiliary information beyond the associated data, Huang et al. also downloaded disease descriptors from the Medical Subject Headings (MeSH) and calculated the semantic similarity of the diseases [28].

Furthermore, to avoid dependence on known miRNA-disease-type associations, according to previous studies [12, 14], we extracted miRNA-gene associations from miRTarBase Release 8.0 [7] and functional association probabilities of genes from HumanNet [29]. Then, the functional similarity of miRNAs can be obtained by combining the above two kinds of association information.

### Tensor construction

Given a set of miRNAs  $\mathcal{M} = \{m_1, m_2, \dots, m_I\}$ , a set of diseases  $\mathcal{D} = \{d_1, d_2, \dots, d_J\}$ , and a set of association types  $\mathcal{T} = \{t_1, t_2, \dots, t_K\}$ . Then, all associations of miRNAs, diseases and types can be described by the third-order tensor  $\mathcal{Y} \in \{0, 1\}^{I \times J \times K}$ . The  $(i, j, k)$  element in tensor  $\mathcal{Y}$  is recorded as  $\mathcal{Y}_{i,j,k}$ , which represents the relationship between miRNA  $m_i$  and disease  $d_j$  under association type  $t_k$ . When  $m_i$  and  $d_j$  are associated under  $t_k$ ,  $\mathcal{Y}_{i,j,k} = 1$ ; otherwise,  $\mathcal{Y}_{i,j,k} = 0$ . The matrix  $Y_{::k}$  is the  $k$ th frontal slice of  $\mathcal{Y}$ , representing the  $k$ -th type of miRNA-disease association. Previous studies on miRNA-disease associations [5,12,30,31] ignored the impact of association types and were equivalent to research on a certain association type.

Although some associations have been discovered and validated, there still exists many associations that have not been verified, and inferring these potential associations may improve our understanding of the pathogenic mechanisms of different types of miRNAs. To this end, our goal is the prediction of potential miRNA-disease-type triples, which is the tensor completion problem. However, since there are few known associations, the tensor  $\mathcal{Y}$  is very sparse and contains very limited useful information. Therefore, extracting the auxiliary information of the miRNA and the disease is an effective way to improve the performance of prediction.

### MiRNA functional similarity

To prevent dependency on known miRNA-disease associations, we exploited known miRNA-gene associations and gene similarity to calculate miRNA functional similarity. Referring to previous studies [12, 32], let  $LLS(g_i, g_j)$  denote the association log-likelihood score between gene  $g_i$  and  $g_j$  obtained from HumanNet. Then, the similarity  $S(g_i, g_j)$  between  $g_i$  and  $g_j$  is

$$S(g_i, g_j) = \begin{cases} 0 & e(g_i, g_j) \notin \text{HumanNet} \\ 1 & g_i = g_j \\ \frac{LLS(g_i, g_j) - LLS_{min}}{LLS_{max} - LLS_{min}} & e(g_i, g_j) \in \text{HumanNet} \end{cases} \tag{1}$$

where  $e(g_i, g_j)$  is the edge between genes  $g_i$  and  $g_j$ .  $LLS_{min}$  and  $LLS_{max}$  denote the minimum and maximum log-likelihood score in HumanNet, respectively. Let  $G_i$  and  $G_j$  denote the gene sets associated with miRNA  $m_i$  and  $m_j$  respectively, then the functional similarity can be calculated as follows:

$$S(m_i, m_j) = \frac{\sum_{g \in G_i} S(g, G_j) + \sum_{g \in G_j} S(g_i, G_i)}{|G_i| + |G_j|} \tag{2}$$

where  $S(g, G) = \max\{S(g, g_i) \mid g_i \in G\}$ ,  $|G_i|$  and  $|G_j|$  represent the number of genes contained in  $G_i$  and  $G_j$ , respectively.

In summary, for each benchmark dataset, we obtain miRNA-disease-type tensor  $\mathcal{Y}$ , disease semantic similarity  $S_{sem}^d$  and miRNA functional similarity  $S_{fun}^m$ .

### Network integration

The known tensor  $\mathcal{Y}$  also contains important information, which can be an important supplement to the similarity of miRNA (disease) [22,27]. Dong et al. [27] employed  $\mathcal{Y}$  to construct a miRNA-disease association matrix, and then calculated the Gaussian similarity of miRNAs,

which ignored the influence of association type. Therefore, to retain the information of miRNA, disease and association types simultaneously, we extract features directly from  $\mathcal{Y}$  and fuse them to obtain a more accurate similarity network.

Let matrix  $Y^{(1)} \in \mathbb{R}^{I \times K}$  represent the mode-1 matrixization of  $\mathcal{Y}$ , that is, project the disease and type dimensions of  $\mathcal{Y}$  onto the columns of  $Y^{(1)}$ . Then,  $Y_i^{(1)}$  is the  $i$ -th row of  $Y^{(1)}$ , which represents the interaction profile feature of the  $i$ -th miRNA. Similarly, let matrix  $Y^{(2)} \in \mathbb{R}^{J \times K}$  represent the mode-2 matrixization of  $\mathcal{Y}$ , which also represents the interaction profile feature of diseases. Since  $Y^{(1)}$  and  $Y^{(2)}$  are both high-dimensional and sparse association matrices, which inevitably contain noise. Therefore, we first employ non-negative double singular value decomposition (NDSVD) [33] for dimensionality reduction to obtain the non-negative low-dimensional features  $F^m$  and  $F^d$  of miRNA and disease respectively.

Then, we adopt Kernel Soft-neighborhood Similarity (KSNS) to build the similarity network. Referring to previous studies [14,34–39], KSNS hierarchically integrates neighborhood information and mines nonlinear relationships of samples, and has been well applied to the prediction of various types of biological interactions. Therefore, according to  $F^m$  and  $F^d$ , the interaction profile similarities  $S_{int}^m$  and  $S_{int}^d$  of miRNAs and diseases are obtained by KSNS, respectively.

Finally, we obtained two types of miRNA similarity ( $S_{fun}^m$  and  $S_{int}^m$ ) and two types of disease similarity ( $S_{sem}^d$  and  $S_{int}^d$ ), which measured the similarity relationship of miRNA (or diseases) from different perspectives. Referring to previous studies [36,40], we utilized clusDCA to fuse  $S_{fun}^m$  and  $S_{int}^m$  to obtain the integrated similarity  $S^m$  of miRNAs, and fused  $S_{sem}^d$  and  $S_{int}^d$  to obtain the integrated similarity  $S^d$  of diseases.

### KBLTDARD

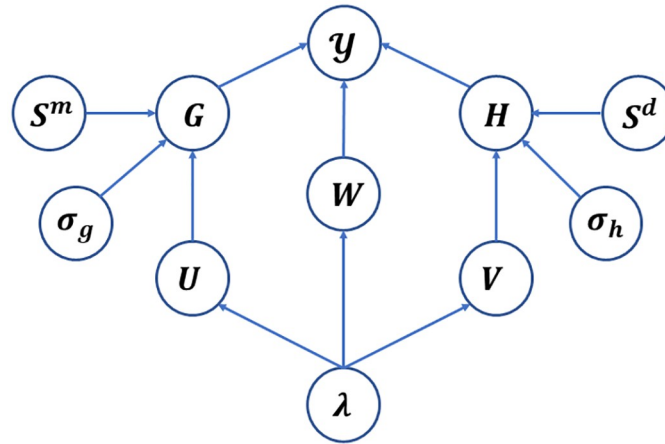
In previous research, we established a new tensor decomposition model (LTDSSL), which introduces logistic functions into tensor decomposition to improve nonlinear learning capabilities, showing strong performance in higher-order relation prediction problems [41]. However, LTDSSL requires manually specifying the rank of the tensor and the values of the hyperparameters, without considering the uncertainty information of potential factors. Based on this, this study combines tensor decomposition and Bayesian inference to establish a new tensor decomposition model.

Let  $G \in \mathbb{R}^{I \times R}$ ,  $W \in \mathbb{R}^{J \times R}$  and  $H \in \mathbb{R}^{K \times R}$  represent the latent factor matrices of miRNAs, diseases and association types respectively. Then, the association probability  $\mathcal{P}_{ijk}$  of the  $i$ th miRNA,  $j$ th disease, and the  $k$ th type is as follows:

$$\mathcal{P}_{ijk} = \frac{\exp(\tilde{\mathcal{Y}}_{ijk})}{1 + \exp(\tilde{\mathcal{Y}}_{ijk})} \tag{3}$$

where  $\tilde{\mathcal{Y}} = \llbracket G, W, H \rrbracket \in \mathbb{R}^{I \times J \times K}$  is reconstructed tensor, and its  $(i, j, k)$ th entry  $\tilde{\mathcal{Y}}_{ijk}$  is  $\sum_{r=1}^R G_{ir} W_{jr} H_{kr}$ . The known miRNA-disease-type triplets are experimentally verified and has higher reliability. Therefore, the weighted logical tensor decomposition model is obtained as follows:

$$P(\mathcal{Y}|G, W, H) = \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \mathcal{P}_{ijk}^{c_{\mathcal{Y}}_{ijk}} (1 - \mathcal{P}_{ijk})^{1 - c_{\mathcal{Y}}_{ijk}} \tag{4}$$



**Fig 2. The probability graph framework of KBLTDARD.**

<https://doi.org/10.1371/journal.pcbi.1012287.g002>

where  $c \geq 1$  represents the importance level parameter. Fig 2 presents the probabilistic graphical model of KBLTDARD with latent variables and corresponding priors.

In Fig 2, the occurrence probability of  $\mathcal{Y}$  is calculated from  $G$ ,  $W$  and  $H$  by (6). The probability distribution of the factor matrix  $G$  is obtained by  $U \in \mathbb{R}^{I \times R}$  combined with the miRNA similarity  $S^m$ , and the probability distribution of the factor matrix  $H$  is obtained by  $V$  combined with the disease similarity  $S^d$ .  $\sigma_g$ ,  $\sigma_h$  and  $\lambda$  are precision parameters. We specify the priors of all latent variables and parameters in this section.

To effectively integrate the auxiliary information, the elements of the factor matrix  $G$  are independent, and the  $(i, r)$ th element  $G_{i,r}$  satisfies the multivariate Gaussian distribution with expectation  $S_i^{(m)} U_{.r}$  and precision  $\sigma_g$

$$P(G|U, S^{(m)}, \sigma_g) = \prod_{i=1}^I \prod_{r=1}^R \mathcal{N}(G_{i,r} | S_i^{(m)} U_{.r}, \sigma_g^{-1}) \tag{5}$$

Similarly, let the elements of  $H$  be independent, and the  $(j, r)$ th element  $H_{j,r}$  satisfies the multivariate Gaussian distribution with expectation  $S_j^{(d)} V_{.r}$  and precision  $\sigma_h$

$$P(H|V, S^{(d)}, \sigma_h) = \prod_{j=1}^J \prod_{r=1}^R \mathcal{N}(H_{j,r} | S_j^{(d)} V_{.r}, \sigma_h^{-1}) \tag{6}$$

Here, the accuracy parameters  $\sigma_g$  and  $\sigma_h$  of the Gaussian distribution satisfy the Jeffreys prior

$$\begin{aligned} P(\sigma_g) &\propto \sigma_g^{-1} \\ P(\sigma_h) &\propto \sigma_h^{-1} \end{aligned} \tag{7}$$

In general, the effective dimension  $R$  of the latent space is the tuning parameter, the selection of which is quite challenging and costly. To both infer the value of  $R$  and avoid overfitting, we introduce automatic rank determination into the priors of  $U$ ,  $V$  and  $W$  [42]. Specifically, let each column of  $U$  be an independent random vector, and whose  $r$ th column satisfy the

multivariate Gaussian distribution with a mean vector  $\mathbf{0}$  and precision matrices  $\lambda_r E_I$

$$P(U|\lambda) = \prod_{r=1}^R \mathcal{N}(U_r|\mathbf{0}, \lambda_r^{-1} E_I) \tag{8}$$

where  $E_I$  represents the identity matrix of size  $I \times I$ , and  $\lambda_r$  controls the  $r$ th column of  $U$ . When  $\lambda_r$  has a large value,  $U_r$  approaches 0, indicating that they make little contribution to  $\mathcal{Y}$  and can be removed from  $U$ . This process realizes the automatic determination of  $R$ . Similarly, the  $r$ th column of  $V$  satisfies the multivariate Gaussian prior with a mean vector  $\mathbf{0}$  and precision matrices  $\lambda_r E_J$

$$P(V|\lambda) = \prod_{r=1}^R \mathcal{N}(V_r|\mathbf{0}, \lambda_r^{-1} E_J) \tag{9}$$

where  $E_J$  represents the identity matrix of size  $J \times J$ . Similar to  $U$  and  $V$ , the prior distribution of  $W$  is as follows

$$P(W|\lambda) = \prod_{r=1}^R \mathcal{N}(W_r|\mathbf{0}, \lambda_r^{-1} E_K) \tag{10}$$

where  $E_K$  represents the identity matrix of size  $K \times K$ .

To achieve joint column sparsity of  $U$ ,  $V$  and  $W$  [43], we further assign the conjugate Gamma prior to  $\lambda$

$$P(\lambda|\alpha, \beta) = \prod_{r=1}^R \text{Gamma}(\lambda_r|\alpha, \beta) = \prod_{r=1}^R \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_r^{\alpha-1} e^{-\beta\lambda_r} \tag{11}$$

where,  $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_R\}$ ,  $\{\alpha, \beta\}$  are the two parameters of the Gamma distribution, and we adopt the uninformative prior, that is,  $\alpha = 1, \beta = 1$  [38].

For simplicity of notation, all unknown latent variables are collected and denoted together by  $\Theta = \{G, H, W, U, V, \lambda, \sigma_g, \sigma_h\}$ . The probabilistic graphical model is shown in Fig 2, from which we can easily write the joint distribution of the model as

$$P(\mathcal{Y}, \Theta) = P(\mathcal{Y}|G, H, W)P(G|U, S^m, \sigma_g)P(H|V, S^d, \sigma_h) \\ P(W|\lambda)P(U|\lambda)P(V|\lambda)P(\lambda)P(\sigma_g)P(\sigma_h) \tag{12}$$

Combining the likelihood in (4), the priors of model parameters  $G$  and  $H$  in (5) and (6), the prior distributions of  $U$ ,  $V$  and  $W$  in (8), (9) and (10), and the hyperpriors in (7) and (11), the logarithmic joint distribution of KBLTDARD can be obtained (see S1 Text of S1 File for details)

$$\ell(\Theta) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[ c\mathcal{Y}_{ijk} \ln(\sigma(\tilde{\mathcal{Y}}_{ijk})) + (1 - \mathcal{Y}_{ijk}) \ln(1 - \sigma(\tilde{\mathcal{Y}}_{ijk})) \right] \\ - \frac{1}{2} \left[ \|\sigma_g G - S^m U\|_F^2 + \sigma_h \|H - S^d V\|_F^2 \right] \\ - \frac{1}{2} \left[ \text{tr}(WAW^T) + \text{tr}(U\Lambda U^T) + \text{tr}(V\Lambda V^T) + 2\beta \text{tr}(\Lambda) \right] \\ + \left( \frac{IR}{2} - 1 \right) \ln(\sigma_g) + \left( \frac{JR}{2} - 1 \right) \ln(\sigma_h) \\ + \left( \frac{I+J+K}{2} + \alpha - 1 \right) \text{tr}(\ln(\Lambda)) + \text{const} \tag{13}$$

where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_R)$ ,  $\ln(\Lambda) = \text{diag}(\ln(\lambda_1), \ln(\lambda_2), \dots, \ln(\lambda_R))$ , and  $\text{diag}(\cdot)$  denotes converting the vector into a diagonal matrix. In (13),  $\text{tr}(\cdot)$  represents the trace of the square matrix,  $\text{const}$  represents a constant independent of  $\Theta$ , and  $\ell(\Theta)$  represents the logarithmic joint distribution, that is,  $\ln p(\Theta, \mathcal{Y})$ . Without losing generality, performing the maximum posterior estimate of  $\Theta$  by maximizing (13) is somewhat equivalent to optimizing the square error function with regularization applied to the logical tensor decomposition and additional constraints on the regularization parameters.

However, unlike point estimation, our goal is to compute the complete posterior distribution of all variables in  $\Theta$  given the data tensor  $\mathcal{Y}$  and the similarities ( $S^m$  and  $S^d$ ), that is,

$$P(\Theta|\mathcal{Y}) = \frac{P(\Theta, \mathcal{Y})}{\int P(\Theta, \mathcal{Y})d\Theta} \tag{14}$$

### Model Inference of KBLTDARD

An exact Bayesian inference in (14) requires integration over all latent variables, which is analytically intractable. Therefore, this study adopts variational inference to calculate the approximate posterior distribution  $q(\Theta)$  of the latent variable [42–44]. The principle of variational inference is to define a parameter distribution group on the latent variable and update the parameters to minimize the Kullback–Leibler (KL) distance between  $P(\Theta|\mathcal{Y})$  and  $q(\Theta)$

$$\begin{aligned} \min_{q(\Theta)} KL(q(\Theta)|P(\Theta|\mathcal{Y})) &= \min_{q(\Theta)} \left\{ \int q(\Theta) \ln \frac{q(\Theta)}{P(\Theta|\mathcal{Y})} d\Theta \right\} \\ &= \ln P(\mathcal{Y}) - \min_{q(\Theta)} \left\{ \int q(\Theta) \ln \left[ \frac{P(\Theta, \mathcal{Y})}{q(\Theta)} \right] d\Theta \right\} \end{aligned} \tag{15}$$

where  $\ln P(\mathcal{Y})$  is a constant, representing model evidence, and its lower bound is defined as  $\mathcal{L}(q) = \int q(\Theta) \ln \left\{ \frac{P(\Theta, \mathcal{Y})}{q(\Theta)} \right\} d\Theta$ . With a mean field approximation,  $q(\Theta)$  is factorized according to the latent variables as

$$q(\Theta) = \prod_k q(\Theta_k) = q(G)q(H)q(W)q(U)q(V)q(\lambda)q(\sigma_g)q(\sigma_h) \tag{16}$$

It is worth noting that (16) is the only assumption for the posterior distribution of the latent variable. When other variables are fixed, the approximate logarithmic posterior distribution of the latent variable  $\Theta_k$  can be accurately obtained

$$\ln q(\Theta_k) = \mathbb{E}_{q(\Theta \setminus \Theta_k)} [\ln p(\Theta, \mathcal{Y})] + \text{const} \tag{17}$$

where  $\mathbb{E}[\cdot]$  represents expectation, and  $\text{const}$  denotes a constant that is not dependent on the current variable.  $\Theta \setminus \Theta_k$  represents the set of all latent variables except  $\Theta_k$ .

1) *Estimate latent variables G and H:* Combined with (3) and (4), the likelihood function  $P(\mathcal{Y}|G, H, W)$  contains exponential forms of  $G$  and  $H$ , resulting in it having no conjugate priors. Therefore, with reference to [45], we adopt the following approximation

$$\begin{aligned} \sigma(z) &\geq \sigma(\xi) \exp \left\{ \frac{z - \xi}{2} - \lambda(\xi)(z^2 - \xi^2) \right\} \\ \lambda(\xi) &= \frac{1}{2\xi} \left[ \sigma(\xi) - \frac{1}{2} \right] \end{aligned} \tag{18}$$

where  $\sigma(x) = 1/(1 + \exp(-x))$  represents the sigmoid function. Combining (3), (4) and (18),

the logarithmic likelihood of  $\mathcal{Y}_{ijk}$  satisfies (see S2 Text of S1 File for details)

$$\begin{aligned} \text{Ln} \left[ P \left( \mathcal{Y}_{ijk} | G, H, W \right) \right] &= \text{Ln} \left[ \mathcal{P}_{ijk}^{c\mathcal{Y}_{ijk}} \left( 1 - \mathcal{P}_{ijk} \right)^{1-\mathcal{Y}_{ijk}} \right] \\ &\geq \text{Ln} \left( h \left( \xi_{ijk}, G, H, W \right) \right) = \left( \frac{c\mathcal{Y}_{ijk} - 1 + \mathcal{Y}_{ijk}}{2} \right) \tilde{\mathcal{Y}}_{ijk} \\ &\quad - \left( c\mathcal{Y}_{ijk} + 1 - \mathcal{Y}_{ijk} \right) \lambda \left( \xi_{ijk} \right) \tilde{\mathcal{Y}}_{ijk}^2 + \left( c\xi_{ijk} + 1 - \xi_{ijk} \right) \\ &\quad \text{Ln} \left[ \sigma \left( \xi_{ijk} \right) \right] - \left( c\mathcal{Y}_{ijk} + 1 - \mathcal{Y}_{ijk} \right) \\ &\quad \frac{\xi_{ijk}}{2} + \left( c\mathcal{Y}_{ijk} + 1 - \mathcal{Y}_{ijk} \right) \lambda \left( \xi_{ijk} \right) \xi_{ijk}^2 \end{aligned} \tag{19}$$

where  $\xi_{ijk}$  represents the local variation parameter. From (19),  $\text{Ln}(h(\xi_{ijk}, G, H, W))$  is quadratic functions of  $G$  and  $H$ , which is the lower bounds of log likelihood. Replace  $P(\mathcal{Y}|G, H, W)$  in (12) with  $h(\xi, G, H, W)$ , combine (5), and substitute them into (17), it is found that the approximate posterior density of the  $i$ th row  $G_i$ . of  $G$  obeys the multivariate Gaussian distribution with expectation  $\tilde{G}_i$  and covariance matrix  $\Sigma(G_i)$ , that is,  $q(G_i) = \mathcal{N}(G_i | \tilde{G}_i, \Sigma(G_i))$  (see S3 Text of S1 File for details)

$$\begin{aligned} \Sigma(G_i) &= \left( 2 \sum_{j=1}^J \sum_{k=1}^K \mathcal{B}_{ijk} \left[ \left( H_j^T \widetilde{H}_j \right) \odot \left( W_k^T \widetilde{W}_k \right) \right] + \tilde{\sigma}_g E_R \right)^{-1} \\ \tilde{G}_i &= \left\{ \Sigma(G_i) \left[ \left( \tilde{W} \odot \tilde{H} \right)^T A_i^{(1)T} + \tilde{\sigma}_g \tilde{U}^T S_i^T \right] \right\}^T \end{aligned} \tag{20}$$

where  $\mathcal{A}_{ijk} = \frac{c\mathcal{Y}_{ijk} - 1 + \mathcal{Y}_{ijk}}{2}$ ,  $\mathcal{B}_{ijk} = \left( c\mathcal{Y}_{ijk} + 1 - \mathcal{Y}_{ijk} \right) \lambda \left( \xi_{ijk} \right)$ .  $A^{(1)}$  represents the mode-1 matrixization of tensor  $\mathcal{A}$ .  $\odot$  represents the Hadamard product, and  $\odot$  represents the Khatri-Rao product.  $E_R$  represents the identity matrix of size  $R \times R$ .  $H_j^T \widetilde{H}_j$  and  $W_k^T \widetilde{W}_k$  denote the expectation of  $H_j^T H_j$  and  $W_k^T W_k$  respectively as follows:

$$\begin{aligned} H_j^T \widetilde{H}_j &= \tilde{H}_j^T \tilde{H}_j + \Sigma(H_j) \\ W_k^T \widetilde{W}_k &= \tilde{W}_k^T \tilde{W}_k + \Sigma(W_k) \end{aligned} \tag{21}$$

Similarly, the posterior density of the  $j$ th row  $H_j$ . of  $H$  obeys the multivariate Gaussian distribution with expectation  $\tilde{H}_j$  and covariance matrix  $\Sigma(H_j)$ , that is,

$$\begin{aligned} q(H_j) &= \mathcal{N} \left( H_j | \tilde{H}_j, \Sigma(H_j) \right) \\ \Sigma(H_j) &= \left( 2 \sum_{i=1}^I \sum_{k=1}^K \mathcal{B}_{ijk} \left[ \left( W_k^T \widetilde{W}_k \right) \odot \left( G_i^T \widetilde{G}_i \right) \right] + \tilde{\sigma}_h E_R \right)^{-1} \\ \tilde{H}_j &= \left\{ \Sigma(H_j) \left[ \left( \tilde{W} \odot \tilde{G} \right)^T A_j^{(2)T} + \tilde{\sigma}_h \tilde{V}^T S_j^T \right] \right\}^T \end{aligned} \tag{22}$$

where  $G_i^T \widetilde{G}_i = \tilde{G}_i^T \tilde{G}_i + \Sigma(G_i)$ , and  $A^{(2)}$  represents the mode-2 matrixization of tensor  $\mathcal{A}$ .

2) *Estimate the latent variable  $W$* : Similar to the solution of  $G$  and  $H$ , replace  $P(\mathcal{Y}_{ijk}|G, H, W)$  in (19) with  $h(\xi_{ijk}, G, H, W)$ , combine with (10), and substitute them into (17), it can be found that the approximate posterior density of the  $k$ th row  $W_k$ . of  $W$  obeys the

multivariate Gaussian distribution with expectation  $\tilde{W}_k$  and covariance matrix  $\Sigma(W_k)$ , that is,  $q(W_k) = \mathcal{N}(W_k | \tilde{W}_k, \Sigma(W_k))$  (see S4 Text of S1 File for details)

$$\Sigma(W_k) = \left[ 2 \sum_{i=1}^I \sum_{j=1}^J \mathcal{B}_{ijk} \left[ \left( \widetilde{G}_i^T \widetilde{G}_i \right) \odot \left( \widetilde{H}_j^T \widetilde{H}_j \right) \right] + \text{diag}(\tilde{\lambda}) \right]^{-1}$$

$$\tilde{W}_k = \left( \Sigma(W_k) (\tilde{H} \odot \tilde{G})^T A_k^{(3)T} \right)^T$$
(23)

where  $\text{diag}(\cdot)$  represents the conversion of a vector to diagonal matrix form, and  $A^{(3)}$  represents the mode-3 matrixization of tensor  $\mathcal{A}$ .

3) *Estimate latent variables U and V:* Substituting the priors of  $U$  and  $G$  into (17), the logarithmic posterior approximation of the  $r$ th column  $U_r$  of  $U$  satisfies (see S5 Text of S1 File for details)

$$\begin{aligned} \text{Ln}q(U_r) &= \mathbb{E}_{q(\Theta \setminus U_r)} \left[ \text{Ln} \left\{ P(G|U, S^m, \sigma_g) P(U|\lambda) \right\} \right] + \text{const} = \\ &\mathbb{E}_{q(\Theta \setminus U_r)} \left[ - \frac{U_r^T \left[ \sigma_g S^m T S^m + \lambda_r E_I \right] U_r - 2 \sigma_g U_r^T S^m T G_r}{2} \right] + \text{const} = \\ &- \frac{U_r^T \left[ \tilde{\sigma}_g S^m T S^m + \tilde{\lambda}_r E_I \right] U_r - 2 \tilde{\sigma}_g U_r^T S^m T \tilde{G}_r}{2} + \text{const} \end{aligned}$$
(24)

From (24), the posterior approximation  $U_r$  obeys the multivariate Gaussian distribution with expectation  $\tilde{U}_r$  and covariance matrix  $\Sigma(U_r)$ , that is,  $q(U_r) = \mathcal{N}(U_r | \tilde{U}_r, \Sigma(U_r))$

$$\Sigma(U_r) = \left[ \tilde{\sigma}_g S^m T S^m + \tilde{\lambda}_r E_I \right]^{-1}$$

$$\tilde{U}_r = \tilde{\sigma}_g \Sigma(U_r) S^m T \tilde{G}_r$$
(25)

Apparently, the posterior approximation  $V_r$  also obeys the multivariate Gaussian distribution  $q(V_r) = \mathcal{N}(V_r | \tilde{V}_r, \Sigma(\tilde{V}_r))$ , as follows:

$$\Sigma(V_r) = \left[ \sigma_h S^v T S^v + \lambda_r E_J \right]^{-1}$$

$$\tilde{V}_r = \sigma_h \Sigma(V_r) S^v T \tilde{H}_r$$
(26)

4) *Estimate the latent variable λ:* Substituting the priors of  $U$ ,  $V$ ,  $W$  and  $\lambda$  in (8), (9), (10) and (11) into (17), the logarithmic posterior approximately of  $\lambda$  satisfies (see S6 Text of S1 File for details)

$$\begin{aligned} \text{Ln}q(\lambda) &= \mathbb{E}_{q(\Theta \setminus \lambda)} \left[ \text{Ln} \{ P(U|\lambda) P(V|\lambda) P(W|\lambda) P(\lambda|\alpha, \beta) \} \right] \\ &+ \text{const} = \sum_{r=1}^R \left[ \left( \frac{(I+J+K)}{2} + \alpha - 1 \right) \text{Ln}(\lambda_r) \right. \\ &\left. - \left( \frac{U_r^T U_r + V_r^T V_r + W_r^T W_r}{2} + \beta \right) \lambda_r \right] + \text{const} \end{aligned}$$
(27)

From (27), the approximate posterior of  $\lambda_r$  follows a Gamma distribution with parameters  $\tilde{\alpha}_r$  and  $\tilde{\beta}_r$ , that is,  $q(\lambda_r) = \text{Gamma}(\lambda_r | \tilde{\alpha}_r, \tilde{\beta}_r)$

$$\begin{aligned} \tilde{\alpha}_r &= \frac{(I + J + K)}{2} + \alpha \\ \tilde{\beta}_r &= \left( \frac{U_r^T U_r + V_r^T V_r + W_r^T W_r}{2} + \beta \right) \\ \tilde{\lambda}_r &= \frac{\tilde{\alpha}_r}{\tilde{\beta}_r} = \frac{I + J + K + 2\alpha}{U_r^T U_r + V_r^T V_r + W_r^T W_r + 2\beta} \end{aligned} \tag{28}$$

where  $\tilde{\lambda}_r$  represents the expectation of  $\lambda_r$ . In (28),  $U_r^T U_r$ ,  $V_r^T V_r$ , and  $W_r^T W_r$  are the expectations of  $U_r^T U_r$ ,  $V_r^T V_r$ , and  $W_r^T W_r$ , respectively, as follows:

$$\begin{aligned} U_r^T U_r &= \tilde{U}_r^T \tilde{U}_r + \text{tr}(\Sigma(U_r)) \\ V_r^T V_r &= \tilde{V}_r^T \tilde{V}_r + \text{tr}(\Sigma(V_r)) \\ W_r^T W_r &= \tilde{W}_r^T \tilde{W}_r + \sum_{k=1}^K \sigma^2(W_{kr}) \end{aligned} \tag{29}$$

where  $\text{tr}(\cdot)$  represents the trace of the square matrix, and  $\sigma^2(\cdot)$  represents the variance.

5) *Estimate latent variables  $\sigma_g$  and  $\sigma_h$* : Substituting (5) and (7) into (17), the logarithmic posterior approximately of  $\sigma_g$  satisfies

$$\begin{aligned} \ln q(\sigma_g) &= \mathbb{E} \left[ \ln \left\{ P(G | U, S^u, \sigma_g) P(\sigma_g) \right\} \right] + \text{const} = \\ &= \left( \frac{IR}{2} - 1 \right) \ln(\sigma_g) - \sigma_g \left( \frac{\mathbb{E}[\|G - S^m U\|^2]}{2} \right) + \text{const} \end{aligned} \tag{30}$$

Therefore, the posterior distribution of  $\sigma_g$  is a Gamma distribution with mean

$$\tilde{\sigma}_g = \frac{\tilde{a}^g}{\tilde{b}^g} = \frac{IR}{\mathbb{E}[\|G - S^m U\|^2]} \tag{31}$$

Referring to Theorem 1 (see S7 Text of S1 File for details),  $[\|G - S^m U\|^2]$  is transformed into

$$\mathbb{E}[\|G - S^m U\|^2] = \|\tilde{G} - S^m \tilde{U}\|^2 + \sum_{i=1}^I \text{tr}(\Sigma(G_i)) + \text{tr}\left(S^m \sum_{r=1}^R \Sigma(U_r)(S^m)^T\right) \tag{32}$$

Similarly, the posterior approximation of  $\sigma_h$  follows the Gamma distribution, whose expectation is

$$\tilde{\sigma}_h = \frac{\tilde{a}^h}{\tilde{b}^h} = \frac{JR}{\mathbb{E}[\|H - S^d V\|^2]} \tag{33}$$

6) *Estimate the local variation parameter  $\xi_{ijk}$* : From (19), take the derivative of  $\ln(h(\xi_{ijk}, G, H, W))$  with respect to  $\xi_{ijk}$ , set the derivative equal to 0, and obtain  $\xi_{ijk}$  that satisfies (see S8 Text of S1 File for details)

$$\xi_{ijk}^2 = \mathbb{E}(\tilde{\mathcal{Y}}_{ijk}^2) = \mathbb{E}\left(\langle G_i, H_j, W_k \rangle^2\right) = \langle U_i^T U_i, V_j^T V_j, W_k^T W_k \rangle \tag{34}$$

where  $\langle \cdot \rangle$  is the generalized inner product, which means the product of corresponding

elements, and then summed [42]. In (34),  $\widetilde{U}_i^T U_i$ ,  $\widetilde{V}_j^T V_j$ , and  $\widetilde{W}_k^T W_k$  are the expectations of  $U_i^T U_i$ ,  $V_j^T V_j$ , and  $W_k^T W_k$ , respectively, as follows:

$$\begin{aligned} \widetilde{U}_i^T U_i &= \widetilde{U}_i^T \widetilde{U}_i + \text{diag}(\sigma^2(U_{i1}), \sigma^2(U_{i2}), \dots, \sigma^2(U_{iR})) \\ \widetilde{V}_j^T V_j &= \widetilde{V}_j^T \widetilde{V}_j + \text{diag}(\sigma^2(V_{j1}), \sigma^2(V_{j2}), \dots, \sigma^2(V_{jR})) \\ \widetilde{W}_k^T W_k &= \widetilde{W}_k^T \widetilde{W}_k + \Sigma(W_k) \end{aligned} \tag{35}$$

In summary, the optimization algorithm for solving KBLTDARD is presented in Algorithm 1.

Algorithm 1. The Algorithm of KBLTDARD

Input: Known miRNA-disease-type tensor  $\mathcal{Y}$ , miRNA similarity  $S^m$ , disease similarity  $S^d$ .

Output:  $G$ ,  $W$ ,  $H$ , and  $\mathcal{P}$ .

Initialization:

Initialize  $\widetilde{G}$ ,  $\widetilde{H}$ ,  $\widetilde{W}$ ,  $\widetilde{U}$  and  $\widetilde{V}$  with multivariate Gaussian random numbers. Initialize  $\Sigma(G)$ ,  $\Sigma(H)$ ,  $\Sigma(W)$ ,  $\Sigma(U)$  and  $\Sigma(V)$  with unit tensors of corresponding size. Let local variational parameters  $\xi_{ijk} = 1$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ ,  $k = 1, \dots, K$ .

repeat

Update the posterior expectations  $\tilde{\sigma}_g$  and  $\tilde{\sigma}_h$  of  $\sigma_g$  and  $\sigma_h$  via (31) and (33), respectively.

Update the posterior expectations and variances of  $G$ ,  $H$  and  $W$  via (20), (22) and (23), respectively.

Update the posterior expectations and variances of  $U$  and  $V$  via (25) and (26), respectively.

for each  $r$  ( $1 \leq r \leq R$ )

Update the posterior  $q(\lambda_r)$  via (28) and (29).

end for

Update the local variational parameter  $\xi$  via (34) and (35).

Until convergence.

$$\mathcal{P} = \frac{1}{1 + \exp(-\langle G, W, H \rangle)}$$

Return  $\mathcal{P}$ .

Output: Factor matrices  $G$ ,  $W$ ,  $H$ , and association probability  $\mathcal{P}$ .

### Complexity analysis

In algorithm 1, the time complexity is primarily attributed to the updating of the posterior expectation or variance of the latent variable and the iteration of the local variational parameters. Therefore, we combine the updated formula of latent variables to analyze one by one.

In (31) and (33), the computation cost of precision parameters  $\sigma_g$  and  $\sigma_h$  are  $\mathcal{O}(I^3R)$  and  $\mathcal{O}(J^3R)$  respectively. The computation cost of factor matrix  $W$  in (23) is  $\mathcal{O}(IJKR^2 + KR^3)$ , while the computation cost of factor matrix  $G$  in (20) and  $H$  in (22) are  $\mathcal{O}(IJKR^2 + I^2R + IR^3)$  and  $\mathcal{O}(IJKR^2 + J^2R + JR^3)$ , respectively. The computation cost of latent variable  $U$  in (25) and  $V$  in (26) are  $\mathcal{O}(I^3R)$  and  $\mathcal{O}(J^3R)$ , respectively. The computation cost of the rank determination parameter  $\lambda$  in (28) is  $\mathcal{O}(IR + JR + KR)$ , while the computation cost of the local variational parameter  $\xi$  in (34) is  $\mathcal{O}(IJKR^2)$ .

In summary, since the maximum number of iterations is fixed, the total computational cost of an update in Algorithm 1 is  $\mathcal{O}(IJKR^2 + I^3R + J^3R + IR^3 + JR^3 + KR^3)$ , where the potential space dimension  $R \ll \min\{I, J\}$ .

## Results

### Experimental settings

To extensively assess the predictive performance of multi-type miRNA-disease associations, referring to previous studies [22–25,27], we adopt two types of 5-fold cross validation.

1.  $CV_{type}$ : Evaluation of the accuracy of model predictions for types. We randomly divided the miRNA-disease pairs with at least one type association into 5 disjoint equal parts. In each experiment, one subset was alternately selected as the test set and the rest as the training set.
2.  $CV_{triple}$ : Evaluation of the accuracy of model predictions for triples. We randomly divided all miRNA-disease--type triples into 5 disjoint equal parts. In each experiment, one subset was alternately selected as the test set and the rest as the training set.

Regarding  $CV_{type}$ , we are interested in the type with the maximum score in the test set. Therefore, referring to previous studies [22–25,27,41], we sorted the types of miRNA-disease pairs according to prediction scores, selected the type with the highest score, and applied the average Top-1 precision, average Top-1 recall, and average Top-1 F1 as evaluation indicators. For  $CV_{triple}$ , we choose commonly used overall evaluation indicators, namely the area under the precision-recall curve (AUPR), the area under the ROC curve (AUC), and the F1 value [22,24,27]. In this study, the Matlab tensor toolbox "tensor\_toolbox" is used to perform tensor calculations [46].

### Comparison experiments

Several computational models have been developed and applied for the prediction of multiple types of miRNA-disease associations. To comprehensively evaluate the performance of KBLTDARD, we select six state-of-the-art tensor decomposition models as benchmarks.

TFAI [22,47]: TFAI introduces graph Laplacian regularization based on the CP decomposition to keep the information about the local structure of the data.

FBCPARD [42]: FBCPARD is a standard Bayesian tensor decomposition model, which introduces the Bayesian framework into CP decomposition and utilizes automatic rank determination to achieve adaptive inference of CP rank.

TDRC [22]: TDRC established a new way of relation constraint and integrated auxiliary information of miRNAs and diseases into CP decomposition.

WeightTDAIGN [27]: WeightTDAIGN introduces a positive sample weighting strategy based on CP decomposition to improve prediction performance, utilizes the  $L_{2,1}$  norm constraint projection matrix to reduce the impact of redundant information, and employs graph regularization to preserve local structural information.

TFLP [25]: TFLP combines tensor robust principal component analysis and label propagation, introduces multiple similarity information of miRNAs (diseases), and achieves prediction through iteration of label information.

SPLDHyperAWNTF [24]: SPLDHyperAWNTF integrates hypergraph learning and tensor weighting with non-negative tensor decomposition to achieve miRNA disease-type triple prediction.

Except for FBCPARD, the above five methods are applied to multiple types of miRNA-disease association prediction. Therefore, we adopt the optimal parameters recommended by the above methods to perform experiments. In the original literature, WeightTDAIGN, TFLP, SPLDHyperAWNTFTDRC, and TFAI all construct miRNA functional similarity through miRNA-disease associations, which results in bias towards known miRNA-disease associations. Therefore, different from the original method, this paper adopts the functional similarity

**Table 1. The predictive performance of all models under  $CV_{type}$ .**

Method	HMDD v2.0			HMDD v3.2		
	Top-1 precision	Top-1 recall	Top-1 F1	Top-1 precision	Top-1 recall	Top-1 F1
TFAI	0.5530	0.4928	0.5078	0.5827	0.4470	0.4801
FBCPARD	0.5349	0.4765	0.4903	0.5387	0.4132	0.4445
TDRC	0.5195	0.463	0.4727	0.5472	0.4201	0.4455
WeightTDAIGN	0.5134	0.4571	0.4697	0.5716	0.4385	0.4714
TFLP	0.3659	0.3260	0.3334	0.3133	0.2402	0.2488
SPLDHyperAWNTF	0.5952	0.5305	0.5507	0.6056	0.4647	0.5024
KBLTDARD	<b>0.6361</b>	<b>0.5665</b>	<b>0.5869</b>	<b>0.6236</b>	<b>0.4783</b>	<b>0.5197</b>

Note: Bold numbers indicate the optimal value under the corresponding indicator.

<https://doi.org/10.1371/journal.pcbi.1012287.t001>

of miRNAs described in Section 2.2.2 to perform the above model. In addition, when performing  $CV_{triplet}$ , the random selection of negative samples may have an impact on the evaluation metrics of the model. Therefore, referring to the suggestion of Huang et al. [22], for each experiment, we perform negative sample selection under 20 different seeds and calculate the mean as the final evaluation index.

According to Table 1, under the  $CV_{type}$ , KBLTDARD outperforms other methods on both HMDD v2.0 and HMDD v3.2. Specifically, on HMDD v2.0, KBLTDARD's Top-1 precision, Top-1 recall and Top-1 F1 reached 0.6361, 0.5665 and 0.5869 respectively, which are 6.87%, 6.79% and 6.57% higher than the second-ranked SPLDHyperAWNTF. On HMDD v3.2, the Top-1 precision, Top-1 recall and Top-1 F1 of KBLTDARD are 0.6236, 0.4783 and 0.5197, respectively, which are substantially better than other methods.

According to Table 2, under the  $CV_{triplet}$ , KBLTDARD also performs better prediction performance on HMDD v2.0 and HMDD v3.2. Specifically, on HMDD v2.0, the AUPR of KBLTDARD is 0.8966, which is 5.00%, 16.29%, 4.23%, 9.74%, 9.08% and 1.69% higher than that of TFAI (0.8539), FBCPARD (0.7710), TDRC (0.8602), WeightTDAIGN (0.8170), TFLP (0.8220) and SPLDHyperAWNTF (0.8817), respectively. The AUC and F1 of HMDD v2.0 are 0.8893 and 0.8218 respectively, which are substantially better than other methods. Furthermore, on HMDD v3.2, the AUPR, AUC and F1 of KBLTDARD are 0.9452, 0.9445 and 0.8775 respectively, which is better than other methods. The comparison of KBLTDARD with other models under 20 random seeds is detailed in S9 Text of S1 File.

**Table 2. The predictive performance of all models under  $CV_{triplet}$ .**

Method	HMDD v2.0			HMDD v3.2		
	AUPR	AUC	F1	AUPR	AUC	F1
TFAI	0.8539	0.8194	0.7742	0.9211	0.9046	0.8498
FBCPARD	0.7710	0.6764	0.6746	0.8741	0.8239	0.8027
TDRC	0.8602	0.8264	0.7754	0.9243	0.9121	0.8517
WeightTDAIGN	0.8170	0.7434	0.7229	0.9099	0.8841	0.8405
TFLP	0.8220	0.7639	0.7771	0.8193	0.7447	0.7774
SPLDHyperAWNTF	0.8817	0.8646	0.8025	0.9294	0.9214	0.8569
KBLTDARD	<b>0.8966</b>	<b>0.8835</b>	<b>0.8185</b>	<b>0.9452</b>	<b>0.9445</b>	<b>0.8775</b>

Note: Bold numbers indicate the optimal value under the corresponding indicator.

<https://doi.org/10.1371/journal.pcbi.1012287.t002>

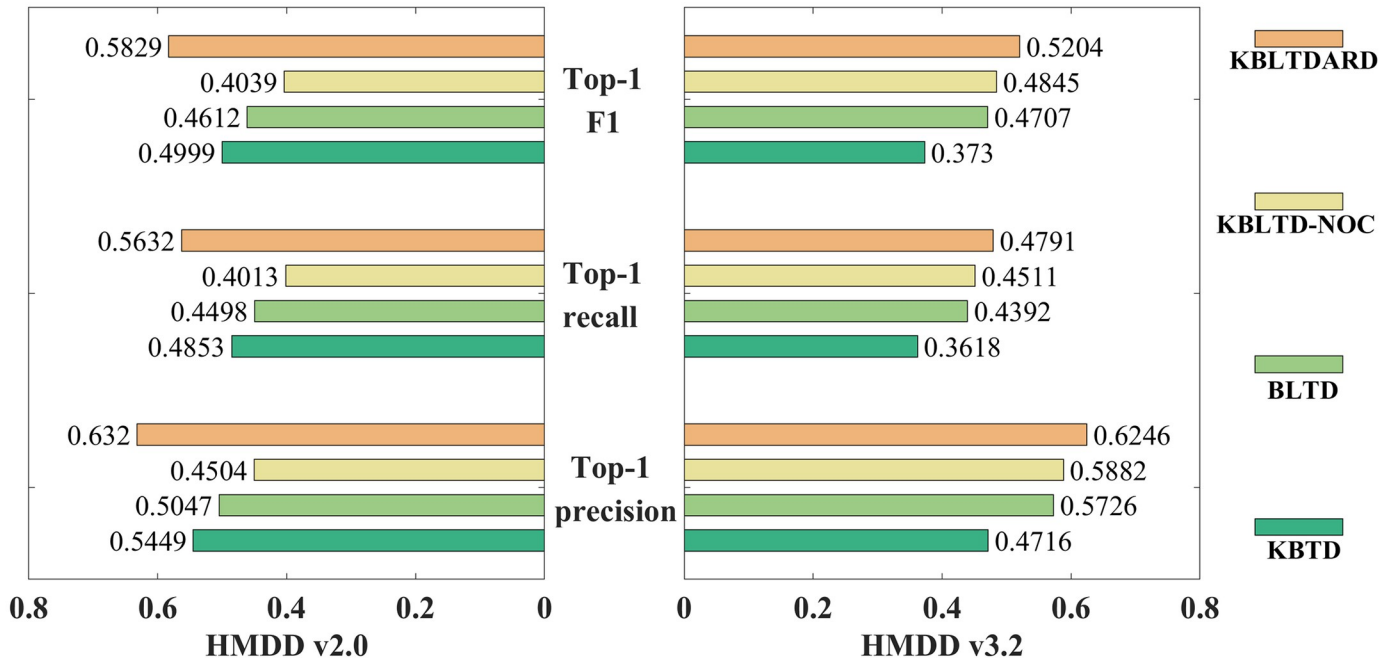


Fig 3. The prediction performance of ablation experiment.

<https://doi.org/10.1371/journal.pcbi.1012287.g003>

In summary, KBLTDARD achieved the most optimal prediction performance, followed by SPLDHyperAWNTF. FBCPARD’s prediction ability was limited to some extent as it solely relied on miRNA-disease-type associations for prediction. Furthermore, models other than KBLTDARD contain many hyperparameters that often demand cumbersome debugging before conducting predictions, greatly affecting their computational efficiency and generalization capabilities. In contrast, KBLTDARD, with the help of Bayesian framework, takes low-dimensional features and model hyperparameters as latent variables, and realizes model solution by inferring the posterior distribution of latent variables, avoiding complex parameter debugging and enhancing generalization ability.

### Ablation studies

Compared with the traditional Bayesian tensor decomposition methods [42,44], the improvement of KBLTDARD is manifested in three aspects: the introduction of logical functions, the addition of auxiliary information, and the set of importance levels. For a better understanding of these contributions, we created comparison models by removing the logistic function, auxiliary information, and importance level from KBLTDARD, respectively. Specifically, KBLTD represents the model acquired by removing the logistic function from KBLTDARD, BLTD represents the model acquired by removing the auxiliary information from KBLTDARD, and KBLTD-NOC represents the model acquired by eliminating the importance level from KBLTDARD.

Fig 3 shows the prediction performance of ablation experiments evaluated by 5-fold cross-validation under HMDD v2.0 and HMDD v3.2. For KBLTD-NOC and KBLTDARD, after setting the importance level, the predictive ability of KBLTDARD is substantially superior to that of KBLTD-NOC, especially on the sparse data set (HMDD v2.0). This result indicates that increasing the importance of known associations can effectively mitigate the influence of false-

negative samples on model performance. For BLTD and KBLTDARD, after combining auxiliary information, KBLTDARD achieves higher prediction performance compared with BLTD. This result shows that the introduction of auxiliary information corrects the iteration direction of logical tensor decomposition and improves the model's prediction ability for isolated samples. For KBTD and KBLTDARD, after the introduction of logistic functions, the prediction performance of KBLTDARD has been significantly improved. The result of this experiment demonstrates that the introduction of logical functions significantly improves nonlinear learning capabilities. To sum up, the addition of logical functions, auxiliary information, and importance levels can improve the predictive ability of the model to a certain extent.

### Case study

To further evaluate the actual prediction performance of KBLTDARD, we conduct two types of case studies. The first strategy evaluates KBLTDARD from a global perspective, that is, testing the model's predictive ability for all diseases. Therefore, we employ KBLTDARD to predict 447 diseases in HMDD v3.2 one by one. The second strategy predicts four common diseases ('Gastric Neoplasms', 'Myocardial Infarction', 'Prostate Neoplasms' and 'Pancreatic Neoplasms') and checks the latest literature to test the prediction results.

In Case Study 1, for each disease from HMDD v3.2, all its associations with all miRNAs and types are removed, and the prediction score for each disease is evaluated. We adopt AUC to evaluate the overall predictive performance with respect to each disease and perform statistics on all AUC values. In addition, for each disease, researchers are more likely to focus on the fraction of known associations included in the top-ranked associations [38,48], that is, the hit rate, as follows:

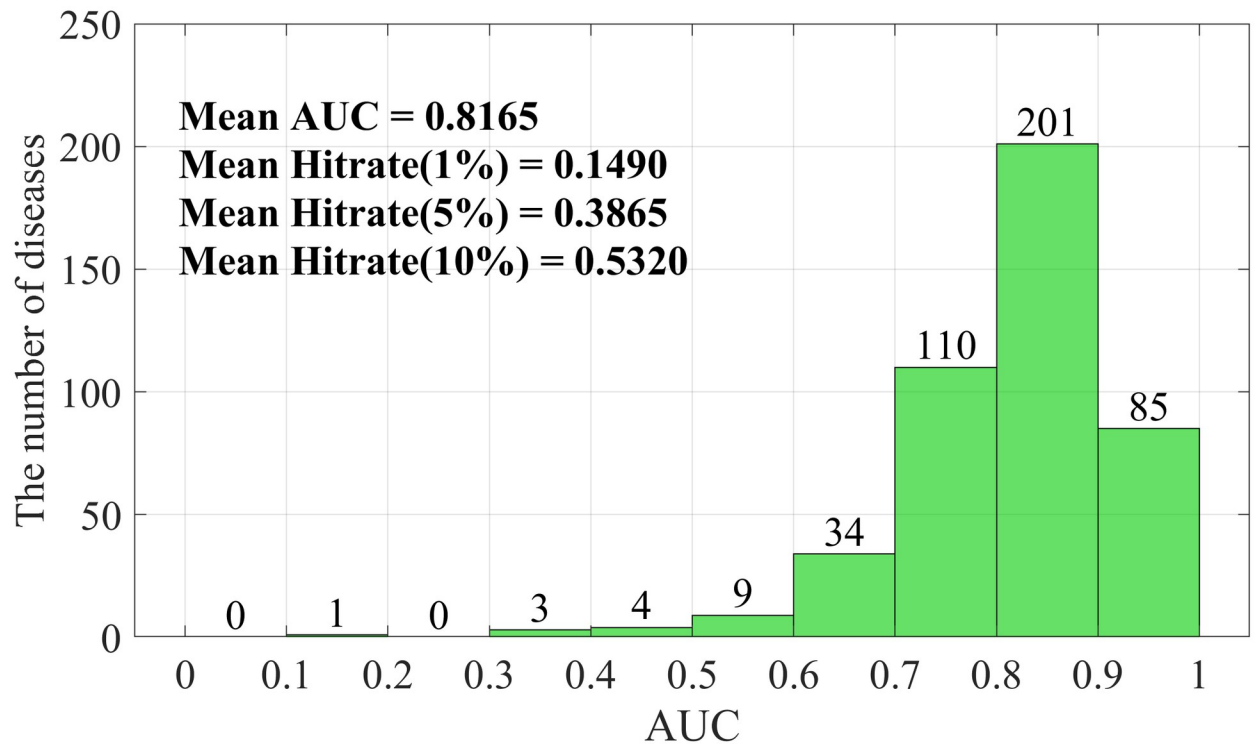
$$\text{Hitrate}(\rho) = \frac{|S_{\text{cand}}([\rho \cdot N]) \cap S_{\text{Test}}|}{|S_{\text{Test}}|} \quad (36)$$

where  $N$  is the total number of triples in the test set,  $\rho$  represents the scaling factor, which in this study is selected as {1%, 5%, 10%}, and  $[\cdot]$  indicates an integer operation.  $S_{\text{cand}}([\rho \cdot N])$  denotes the set of highest scoring top  $[\rho \cdot N]$  triples, and  $S_{\text{Test}}$  denotes the real triple set in the test set.

As presented in Fig 4, the average AUC of KBLTDARD is 0.8165. Among 447 diseases predicted by KBLTDARD, the AUC of 286 diseases exceeded 0.8, amounting to 63.98%. Only 17 diseases had AUC less than 0.6, amounting to less than 4%. The average hit rates of the top 1%, top 5%, and top 10% are 0.1490, 0.3865, and 0.5320 respectively, which are 14 times, 7 times, and 5 times more than the random hit rates (0.01, 0.05, and 0.10) respectively. The above results indicate that the associations with top prediction scores contain the vast majority of known associations.

Then we focus on these four common diseases: Gastric Neoplasms, Myocardial Infarction, Prostate Neoplasms and Pancreatic Neoplasms for further analysis. Table 3 shows the top 20 predictions and related evidence for the disease Gastric Neoplasms, and S1, S2, and S3 Tables show the top 20 predictions and related evidence for the three diseases Myocardial Infarction, Prostate Neoplasms, and Pancreatic Neoplasms, respectively.

Gastric cancer is the fifth most prevalent cancer in the world and the third major cause of cancer-related deaths worldwide [49]. There are more than 1 million new cases of gastric cancer worldwide every year, and the number of gastric cancer-related deaths exceeds 780,000 [50]. Tchernitsa et al. [51] studied the differential expression of miRNA in adjacent normal and tumor samples from gastric cancer patients. The results found that miR-146a was significantly different in lymph node-positive and node-negative gastric cancer, and its changes may



**Fig 4. The prediction results of KBLTDARD for all diseases.**

<https://doi.org/10.1371/journal.pcbi.1012287.g004>

affect local tumor growth and lymph node spread. Zhang et al. [52] found that hsa-mir-21 is up-regulated in gastric cancer tissues and is significantly related to the degree of differentiation, local invasion and lymph node metastasis of tumor tissues. As shown in Table 3, among the top 20 miRNA and type associations predicted by KBLTDARD, 17 have been verified by relevant literature. Furthermore, in S1, S2, and S3 Tables, among the top 20 associations predicted by KBLTDARD for Myocardial Infarction, Prostate Neoplasms, and Pancreatic Neoplasms, 16, 16, and 17 were confirmed, respectively.

## Discussion and conclusion

In this paper, we proposed a new KBLTDARD model to predict higher-order relationships between miRNAs, diseases and types. This model utilizes miRNA-gene association and gene

**Table 3. Gastric Neoplasms-related miRNAs and types predicted by KBLTDARD.**

Rank	MiRNA	Type	PMID	Rank	MiRNA	Type	PMID
1	hsa-mir-146a	tissue	20726036	11	hsa-mir-206	target	23348698
2	hsa-mir-21	target	21081469	12	hsa-mir-122	tissue	22112324
3	hsa-mir-21	tissue	16461460	13	hsa-mir-223	tissue	16461460
4	hsa-mir-155	tissue	22112324	14	hsa-mir-210	tissue	Unconfirmed
5	hsa-mir-29a	tissue	25889078	15	hsa-mir-483	target	Unconfirmed
6	hsa-mir-146a	target	22020746	16	hsa-mir-483	circulation	Unconfirmed
7	hsa-mir-21	circulation	22860003	17	hsa-mir-146a	circulation	23806809
8	hsa-mir-206	tissue	22112324	18	hsa-mir-125a	tissue	21987613
9	hsa-mir-155	target	26056431	19	hsa-mir-21	genetics	25230738
10	hsa-mir-146a	genetics	21632853	20	hsa-mir-34b	tissue	19148490

<https://doi.org/10.1371/journal.pcbi.1012287.t003>

similarity to calculate the functional similarity of miRNAs, avoiding the re-use of known miRNA-disease associations. Then, we combine logistic tensor decomposition and Bayesian inference, introduce auxiliary information, and build a probabilistic graphical model to describe the dependence between latent variables. In addition, with regard to KBLTDARD, we developed an efficient deterministic Bayesian inference algorithm to ensure the efficiency of the model solution. Under the 5-CV framework, the top-1 precision of KBLTDARD for new type prediction reached 0.6320 and 0.6246, and the AUC values for new triplet prediction reached 0.8834 and 0.9445, respectively, on HMDD v2.0 and HMDD v3.2 datasets. The results show that the performance of KBLTDARD is significantly improved compared to the previous methods. Case studies of 'gastric neoplasia', 'myocardial infarction', 'prostate neoplasia' and 'pancreatic neoplasia' also demonstrated the predictive power of KBLTDARD. Taken together, these results suggest that KBLTDARD can effectively observe multiple types of miRNA-disease associations.

It should be noted that the following factors can contribute to the reliable performance of KBLTDARD. First, we extract important information from the related database of miRNA and disease, and mine key features from the trained tensors to ensure the richness of information about miRNAs and diseases. In addition, we combined logical tensor decomposition and Bayesian inference to realise the automatic search of hyperparameters, which improves the nonlinear learning ability and generalisation ability of the model.

However, there are some limitations that may affect the performance of KBLTDARD. First, to facilitate model inference, we selected Gaussian and Gamma distributions with conjugation properties to represent prior distributions of latent variables, which may not be optimal for model representations. In future research, we will further explore Bayesian theory and try more advanced prior distributions. Second, although our model shows stronger learning ability than some deep learning, in future studies we will try to combine some advanced deep learning methods (such as hypergraph neural networks, etc.) to improve the nonlinear representation ability of the model.

## Supporting information

**S1 File. Supporting information.**  
(DOCX)

**S1 Fig. The boxplot of prediction results of KBLTDARD under  $CV_{triplet}$ .**  
(TIF)

**S1 Table. Myocardial Infarction-related miRNAs and association types predicted by KBLTDARD.**  
(DOCX)

**S2 Table. Prostate Neoplasms-related miRNAs and association types predicted by KBLTDARD.**  
(DOCX)

**S3 Table. Pancreatic Neoplasms-related miRNAs and association types predicted by KBLTDARD.**  
(DOCX)

**S4 Table. Statistics of predictions for all models under  $CV_{triplet}$ .**  
(DOCX)

**S5 Table. The P-value of the Wilcoxon rank sum test for pairing KBLTDARD with other models.**  
(DOCX)

## Author Contributions

**Conceptualization:** Yingjun Ma.

**Data curation:** Yingjun Ma.

**Formal analysis:** Yingjun Ma.

**Funding acquisition:** Yingjun Ma, Yuanyuan Ma.

**Investigation:** Yingjun Ma.

**Methodology:** Yingjun Ma.

**Project administration:** Yingjun Ma.

**Resources:** Yingjun Ma.

**Software:** Yingjun Ma.

**Supervision:** Yingjun Ma.

**Validation:** Yingjun Ma.

**Visualization:** Yingjun Ma.

**Writing – original draft:** Yingjun Ma.

**Writing – review & editing:** Yingjun Ma, Yuanyuan Ma.

## References

1. Hussen BM, Hidayat HJ, Salihi A, Sabir DK, Taheri M, Ghafouri-Fard S. MicroRNA: A signature for cancer progression. *Biomedicine & Pharmacotherapy*. 2021; 138:111528. Epub 2021/03/27. <https://doi.org/10.1016/j.biopha.2021.111528> PMID: 33770669.
2. Katsu M, Hama Y, Utsumi J, Takashina K, Yasumatsu H, Mori F, et al. MicroRNA expression profiles of neuron-derived extracellular vesicles in plasma from patients with amyotrophic lateral sclerosis. *Neurosci Lett*. 2019; 708:134176. Epub 2019/06/08. <https://doi.org/10.1016/j.neulet.2019.03.048> PMID: 31173847.
3. Mahtal N, Lenoir O, Tinel C, Anglicheau D, Tharaux PL. MicroRNAs in kidney injury and disease. *Nat Rev Nephrol*. 2022; 18(10):643–62. Epub 2022/08/17. <https://doi.org/10.1038/s41581-022-00608-6> PMID: 35974169.
4. Roy B, Lee E, Li T, Rampersaud M. Role of miRNAs in Neurodegeneration: From Disease Cause to Tools of Biomarker Discovery and Therapeutics. *Genes (Basel)*. 2022; 13(3). Epub 2022/03/26. <https://doi.org/10.3390/genes13030425> PMID: 35327979.
5. Li Z, Jiang K, Qin S, Zhong Y, Elofsson A. GCSENet: A GCN, CNN and SENet ensemble model for microRNA-disease association prediction. *PLoS Comput Biol*. 2021; 17(6):e1009048. Epub 2021/06/04. <https://doi.org/10.1371/journal.pcbi.1009048> PMID: 34081706.
6. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic acids research*. 2018; 46(D1):D239–d45. Epub 2017/11/21. <https://doi.org/10.1093/nar/gkx1141> PMID: 29156006.
7. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic acids research*. 2020; 48(D1):D148–d54. Epub 2019/10/28. <https://doi.org/10.1093/nar/gkz896> PMID: 31647101.
8. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. 2015; 12(8):697. Epub 2015/08/01. <https://doi.org/10.1038/nmeth.3485> PMID: 26226356.

9. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic acids research*. 2019; 47(D1):D155–D62. Epub 2018/11/14. <https://doi.org/10.1093/nar/gky1141> PMID: 30423142.
10. Marceca GP, Distefano R, Tomasello L, Lagana A, Russo F, Calore F, et al. MiREDiBase, a manually curated database of validated and putative editing events in microRNAs. *Sci Data*. 2021; 8(1):199. Epub 2021/08/06. <https://doi.org/10.1038/s41597-021-00979-8> PMID: 34349127.
11. Davis AP, Wieggers TC, Wieggers J, Wyatt B, Johnson RJ, Sciaky D, et al. CTD Tetramers: a new online tool that computationally links curated chemicals, genes, phenotypes, and diseases to inform molecular mechanisms for environmental health. *Toxicological sciences: an official journal of the Society of Toxicology*. 2023. Epub 2023/07/24. <https://doi.org/10.1093/toxsci/kfad069> PMID: 37486259.
12. Tang X, Luo J, Shen C, Lai Z. Multi-view Multichannel Attention Graph Convolutional Network for miRNA-disease association prediction. *Brief Bioinform*. 2021; 22(6). Epub 2021/05/09. <https://doi.org/10.1093/bib/bbab174> PMID: 33963829.
13. Ma M, Na S, Zhang X, Chen C, Xu J. SFGAE: a self-feature-based graph autoencoder model for miRNA-disease associations prediction. *Brief Bioinform*. 2022; 23(5). Epub 2022/08/30. <https://doi.org/10.1093/bib/bbac340> PMID: 36037084.
14. Ma Y, He T, Ge L, Zhang C, Jiang X. MiRNA-disease interaction prediction based on kernel neighborhood similarity and multi-network bidirectional propagation. *BMC Medical Genomics*. 2019; 12(10):1–14. <https://doi.org/10.1186/s12920-019-0622-4> PMID: 31865912.
15. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, et al. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic acids research*. 2019; 47(D1):D1013–d7. Epub 2018/10/27. <https://doi.org/10.1093/nar/gky1010> PMID: 30364956.
16. Goh JN, Loo SY, Datta A, Siveen KS, Yap WN, Cai W, et al. microRNAs in breast cancer: regulatory roles governing the hallmarks of cancer. *Biological reviews of the Cambridge Philosophical Society*. 2016; 91(2):409–28. Epub 2015/01/30. <https://doi.org/10.1111/brv.12176> PMID: 25631495.
17. Kawahara Y. Human diseases caused by germline and somatic abnormalities in microRNA and microRNA-related genes. *Congenital anomalies*. 2014; 54(1):12–21. Epub 2013/12/18. <https://doi.org/10.1111/cga.12043> PMID: 24330020.
18. Xiao B, Zhu ED, Li N, Lu DS, Li W, Li BS, et al. Increased miR-146a in gastric cancer directly targets SMAD4 and is involved in modulating cell proliferation and apoptosis. *Oncol Rep*. 2012; 27(2):559–66. Epub 2011/10/25. <https://doi.org/10.3892/or.2011.1514> PMID: 22020746.
19. Kogo R, Mimori K, Tanaka F, Komune S, Mori M. Clinical significance of miR-146a in gastric cancer cases. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011; 17(13):4277–84. Epub 2011/06/03. <https://doi.org/10.1158/1078-0432.CCR-10-2866> PMID: 21632853.
20. Chen X, Clarence Yan C, Zhang X, Li Z, Deng L, Zhang Y, et al. RBMMMDA: predicting multiple types of disease-microRNA associations. *Scientific Reports*. 2015; 5(1). Epub 2015/9/8. <https://doi.org/10.1038/srep13877> PMID: 26347258.
21. Wang R, Li S, Cheng L, Wong MH, Leung KS. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. *BMC bioinformatics*. 2019; 20(Suppl 26):628. Epub 2019/12/17. <https://doi.org/10.1186/s12859-019-3283-6> PMID: 31839008.
22. Huang F, Yue X, Xiong Z, Yu Z, Liu S, Zhang W. Tensor decomposition with relational constraints for predicting multiple types of microRNA-disease associations. *Brief Bioinform*. 2021; 22(3). Epub 2020/07/30. <https://doi.org/10.1093/bib/bbaa140> PMID: 32725161.
23. Wang J, Li J, Yue K, Wang L, Ma Y, Li Q. NMCMDA: neural multicategory MiRNA-disease association prediction. *Brief Bioinform*. 2021; 22(5). Epub 2021/03/30. <https://doi.org/10.1093/bib/bbab074> PMID: 33778850.
24. Ouyang D, Liang Y, Wang J, Liu X, Xie S, Miao R, et al. Predicting multiple types of miRNA-disease associations using adaptive weighted nonnegative tensor factorization with self-paced learning and hypergraph regularization. *Brief Bioinform*. 2022; 23(6). Epub 2022/09/29. <https://doi.org/10.1093/bib/bbac390> PMID: 36168938.
25. Yu N, Liu Z-P, Gao R. Predicting multiple types of MicroRNA-disease associations based on tensor factorization and label propagation. *Computers in Biology and Medicine*. 2022; 146. Epub 2022/4/27. <https://doi.org/10.1016/j.compbiomed.2022.105558> PMID: 35525071.
26. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic acids research*. 2014; 42(Database issue): D1070–4. Epub 2013/11/07. <https://doi.org/10.1093/nar/gkt1023> PMID: 24194601.
27. Ouyang D, Miao R, Wang J, Liu X, Xie S, Ai N, et al. Predicting Multiple Types of Associations Between miRNAs and Diseases Based on Graph Regularized Weighted Tensor Decomposition. *Frontiers in*

- Bioengineering and Biotechnology. 2022; 10:859. Epub 2022/7/4. <https://doi.org/10.3389/fbioe.2022.911769> PMID: 35910021.
28. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*. 2010; 26(13):1644–50. Epub 2010/05/05. <https://doi.org/10.1093/bioinformatics/btq241> PMID: 20439255.
  29. Hwang S, Kim CY, Yang S, Kim E, Hart T, Marcotte EM, et al. HumanNet v2: human gene networks for disease research. *Nucleic acids research*. 2019; 47(D1):D573–d80. Epub 2018/11/13. <https://doi.org/10.1093/nar/gky1126> PMID: 30418591.
  30. Hu H, Zhao H, Zhong T, Dong X, Wang L, Han P, et al. Adaptive deep propagation graph neural network for predicting miRNA-disease associations. *Briefings in functional genomics*. 2023. Epub 2023/04/20. <https://doi.org/10.1093/bfpg/elad010> PMID: 37078739.
  31. Zhang H, Fang J, Sun Y, Xie G, Lin Z, Gu G. Predicting miRNA-Disease Associations via Node-Level Attention Graph Auto-Encoder. *IEEE/ACM Trans Comput Biol Bioinform*. 2023; 20(2):1308–18. Epub 2022/05/04. <https://doi.org/10.1109/TCBB.2022.3170843> PMID: 35503834.
  32. Xiao Q, Luo J, Liang C, Cai J, Ding P. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics*. 2017. Epub 2018/1/15. <https://doi.org/10.1093/bioinformatics/btx545> PMID: 28968779.
  33. Boutsidis C, Gallopoulos E. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*. 2008; 41(4):1350–62. <https://doi.org/10.1016/j.patcog.2007.09.010>
  34. Ma Y, Ge L, Ma Y, Jiang X, He T, Hu X. Kernel Soft-neighborhood Network Fusion for MiRNA-Disease Interaction Prediction. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Madrid, Spain: IEEE; 2018.
  35. Ma Y, He T, Jiang X. Projection-Based Neighborhood Non-Negative Matrix Factorization for lncRNA-Protein Interaction Prediction. *Frontiers in Genetics*. 2019; 10:1148. Epub 2019/11/20. <https://doi.org/10.3389/fgene.2019.01148> PMID: 31824563.
  36. Ma Y, He T, Jiang X. Multi-network logistic matrix factorization for metabolite-disease interaction prediction. *FEBS Letters*. 2020; 594(11):1675–84. Epub 2020/4/26. <https://doi.org/10.1002/1873-3468.13782> PMID: 32246474.
  37. Ma Y, He T, Tan Y, Jiang X. Seq-BEL: Sequence-Based Ensemble Learning for Predicting Virus-Human Protein-Protein Interaction. *IEEE/ACM Trans Comput Biol Bioinform*. 2022; 19(3):1322–33. Epub 2020/08/06. <https://doi.org/10.1109/TCBB.2020.3008157> PMID: 32750886.
  38. Ma Y, Liu Q. Generalized matrix factorization based on weighted hypergraph learning for microbe-drug association prediction. *Comput Biol Med*. 2022; 145:105503. Epub 2022/04/16. <https://doi.org/10.1016/j.compbimed.2022.105503> PMID: 35427986.
  39. Ma Y, Ma Y. Hypergraph-based Logistic Matrix Factorization for Metabolite-disease Interaction Prediction. *Bioinformatics*. 2021; 38(2):435–43. Epub 2021/09/10. <https://doi.org/10.1093/bioinformatics/btab652> PMID: 34499104.
  40. Wang S, Cho H, Zhai C, Berger B, Peng J. Exploiting ontology graph for predicting sparsely annotated gene function. *Bioinformatics*. 2015; 31(12):i357–64. Epub 2015/06/15. <https://doi.org/10.1093/bioinformatics/btv260> PMID: 26072504.
  41. Ma Y, Zhong J. Logistic tensor decomposition with sparse subspace learning for prediction of multiple disease types of human-virus protein-protein interactions. *Brief Bioinform*. 2023; 24(1). Epub 2022/12/28. <https://doi.org/10.1093/bib/bbac604> PMID: 36573486.
  42. Zhao Q, Zhang L, Cichocki A. Bayesian CP Factorization of Incomplete Tensors with Automatic Rank Determination. *IEEE Trans Pattern Anal Mach Intell*. 2015; 37(9):1751–63. Epub 2015/09/10. <https://doi.org/10.1109/TPAMI.2015.2392756> PMID: 26353124.
  43. Li C, Xie HB, Fan X, Xu RYD, Van Huffel S, Mengersen K. Kernelized Sparse Bayesian Matrix Factorization. *IEEE Trans Neural Netw Learn Syst*. 2021; 32(1):391–404. Epub 2020/03/24. <https://doi.org/10.1109/TNNLS.2020.2978761> PMID: 32203037.
  44. Takayama H, Zhao Q, Hontani H, Yokota T. Bayesian Tensor Completion and Decomposition with Automatic CP Rank Determination Using MGP Shrinkage Prior. *SN Computer Science*. 2022; 3(3). <https://doi.org/10.1007/s42979-022-01119-8>
  45. Zhang C-X, Xu S, Zhang J-S. A novel variational Bayesian method for variable selection in logistic regression models. *Computational Statistics & Data Analysis*. 2019; 133:1–19. <https://doi.org/10.1016/j.csda.2018.08.025>
  46. Bader BW, Kolda TG. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software (TOMS)*. 2006; 32(4):635–53.
  47. Narita A, Hayashi Kohei, Tomioka R, Kashima H. Tensor Factorization Using Auxiliary Information. *Data Min Knowl Disc*. 2012;( 25):298–324.

48. Ma Y. DeepMNE: Deep Multi-Network Embedding for lncRNA-Disease Association Prediction. *IEEE J Biomed Health Inform.* 2022; 26(7):3539–49. Epub 2022/7/1. <https://doi.org/10.1109/JBHI.2022.3152619> PMID: 35180094.
49. Sexton RE, Al Hallak MN, Diab M, Azmi AS. Gastric cancer: a comprehensive review of current and future treatment strategies. *Cancer Metastasis Rev.* 2020; 39(4):1179–203. Epub 2020/09/08. <https://doi.org/10.1007/s10555-020-09925-3> PMID: 32894370.
50. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68(6):394–424. Epub 2018/09/13. <https://doi.org/10.3322/caac.21492> PMID: 30207593.
51. Tchernitsa O, Kasajima A, Schäfer R, Kuban RJ, Ungethüm U, Györfy B, et al. Systematic evaluation of the miRNA-ome and its downstream effects on mRNA expression identifies gastric cancer progression. *The Journal of pathology.* 2010; 222(3):310–9. Epub 2010/08/21. <https://doi.org/10.1002/path.2759> PMID: 20726036.
52. Zhang BG, Li JF, Yu BQ, Zhu ZG, Liu BY, Yan M. microRNA-21 promotes tumor proliferation and invasion in gastric cancer by targeting PTEN. *Oncol Rep.* 2012; 27(4):1019–26. Epub 2012/01/24. <https://doi.org/10.3892/or.2012.1645> PMID: 22267008.