A Bidirectional Label Propagation Based Computational Model for Potential Microbe-Disease Association Prediction

Lei Wang[1,2]*, Yuqi Wang[1], Hao Li[1], Xiang Feng[1,2], Dawei Yuan[3] and Jialiang Yang[3]*

[1] Key Laboratory of Hunan Province for Internet of Things and Information Security, Xiangtan University, Xiangtan, China,
[2] College of Computer Engineering and Applied Mathematics, Changsha University, Changsha, China, [3] Geneis Beijing Co., Ltd., Beijing, China

A growing number of clinical observations have indicated that microbes are involved in a variety of important human diseases. It is obvious that in-depth investigation of correlations between microbes and diseases will benefit the prevention, early diagnosis, and prognosis of diseases greatly. Hence, in this paper, based on known microbe-disease associations, a prediction model called NBLPIHMDA was proposed to infer potential microbe-disease associations. Specifically, two kinds of networks including the disease similarity network and the microbe similarity network were first constructed based on the Gaussian interaction profile kernel similarity. The bidirectional label propagation was then applied on these two kinds of networks to predict potential microbe-disease associations. We applied NBLPIHMDA on Human Microbe-Disease Association database (HMDAD), and compared it with 3 other recent published methods including LRLSHMDA, BiRWMP, and KATZHMDA based on the leave-one-out cross validation and 5-fold cross validation, respectively. As a result, the area under the receiver operating characteristic curves (AUCs) achieved by NBLPIHMDA were 0.8777 and 0.8958 ± 0.0027, respectively, outperforming the compared methods. In addition, in case studies of asthma, colorectal carcinoma, and Chronic obstructive pulmonary disease, simulation results illustrated that there are 10, 10, and 8 out of the top 10 predicted microbes having been confirmed by published documentary evidences, which further demonstrated that NBLPIHMDA is promising in predicting novel associations between diseases and microbes as well.

Keywords: microbe-disease association, bidirectional label propagation, leave-one-out cross validation, 5-fold cross validation, COPD

1. INTRODUCTION

With the development of sequencing technologies, studies on microbes in soils, oceans, human bodies and other places have received increasing attention from the scientific community (Methé et al., 2012). Human microbiota, i.e., the collection of microbes existing in human tissues and biological fluids, includes various species such as archaea, eukaryotes, bacteria, and viruses. It is known that all parts of human body contain microbes, the number of which is more than 10 times the number of cells in a human body (Turnbaugh et al., 2007; Sender et al., 2016). Fortunately, the

vast majority of microbes are harmless to human body, some of which are even indispensable for our metabolism, growth, and development. For example, there are parasitic microbes involving in the physiological mechanisms of the human body and playing a vital role in the process of energy acquisition and storage, salvage of energy, and nutrient, resistance to pathogens and foreign microorganisms, immune responses, and other metabolic processes (Guarner and Malagelada, 2003). As a result, human health will be greatly affected by the human microbiota, and the disorder and imbalance of them will sometimes lead to diseases.

The microbiota has been living and evolving in human body since the emerging of human beings, and they have gradually formed a close symbiotic relationship. The dynamic balance of microbiota and human diseases has become a hot topic recently. For example, a few studies indicate that the host's diet affects the activity and structure of microbiota in the human gut. Long-term high-fat diets can cause changes in the intestinal microbiota, resulting in an increase in intestinal deoxycholic acid concentration (DCA), which may promote the development of liver cancer (David et al., 2013). In addition, researchers also found that smoking creates a microenvironment in which a dangerous microbiota is produced, and reducing the niche saturation of the microbiota will increase the risk of diseases (Mason et al., 2014).

The microbiota can also be affected by genes (Khachatryan et al., 2008; Turnbaugh et al., 2008; Goodrich et al., 2014), seasons (Davenport et al., 2014), hygiene and antibiotics (Donia et al., 2014), and so on. With the advent of high-throughput methods and advanced analytical techniques, the collective genome of microbes (human microbiome) inhabiting the human body has become a key player in human health and diseases. Based on the high-resolution molecular analysis, a few researchers found that the disorder of microbiota in the gastrointestinal tract is closely associated with several idiopathic diseases, such as diabetes, obesity, inflammatory bowel disease, cancer, kidney stones, and neurodegenerative diseases (Shah et al., 2016). For example, according to the study of inflammatory bowel disease, the intestinal flora is found to be able to regulate the inflammatory response by regulating regulatory T cells (Singh et al., 2001). In addition, as pointed out by Christopher et al., microbiota that produces lactic acid and butyl hydrochloric acid such as Akkermansia and Prevotella can induce the synthesis of intestinal mucin, which may contribute to intestinal health (Brown et al., 2011). Furthermore, microbiota has been reported to play an important role in autoimmune diseases such as type 1 diabetes. For example, a population of dysfunctional microbes were found in subjects immunized with type 1 diabetes (Murri et al., 2013).

As disease-associated flora can provide important insights into the understanding of the formulation and development of diseases, a few related large-scale projects including the Human Microbiome Project (HMP) (Turnbaugh et al., 2007) and the Earth Microbiome Project (EMP) (Gilbert et al., 2010) have been launched to entangle the relationship between microbial flora and diseases. In addition, many useful databases have also been developed to curate disease-related microbial information. For example, Ma et al. sorted out the confirmed microbe-disease association from published literatures through large-scale text mining, and established the Human Microbe-Disease Association Database (HMDAD) (Ma et al., 2016). These datasets can be served as valuable sources for predicting novel microbe-disease associations.

Traditional microbial identification are performed mainly by independent culture methods and quantitative methods, which are costly and labor intensive. This presents the need for more effective computational methods to scale down the potential microbe-disease associations for further experimental validation. In fact, similar computational models have been successfully implemented in many other related fields such as drug-target interaction prediction (Chen et al., 2012), gene-disease association prediction (Zou, 2016; Meng et al., 2017; Zeng et al., 2017; Zhu et al., 2018), lncRNA-disease association prediction (Chen et al., 2016c; Yu et al., 2018), miRNA-disease relationship prediction (Zeng et al., 2015, 2018; Tang et al., 2017; You et al., 2017), protein structure prediction, and so on. For the first time, Chen et al. proposed the KATZHMDA model for measuring the human microbe-disease association based on KATZ method (Chen et al., 2016a). KATZHMDA first constructed a heterogeneous network composed of the microbe-disease association network, the disease similarity network and the microbe similarity network, and introduced the concept of variable step number to predict microbe-disease associations, which achieves reliable prediction performance. Subsequently, Shen et al. adopted a random walk with restart algorithm to score each candidate microbe-disease pair on a heterogeneous network composed of the Spearman correlation-based microbe network, the symptom-based disease network and the microbe-disease association network (Shen et al., 2016). Huang et al. proposed a path-based computational model called PBHMDA, which adopts a special depth-first search algorithm to traverse all possible paths between microbes and diseases in the heterogeneous network (Huang et al., 2017b). Wang et al. put forward a semi-supervised model of Laplacian regularized least squares (LRLSHMDA), which utilizes Laplace's regular least squares classification combined with topological information of the known microbe-disease association network to train an optimal classifier (Wang et al., 2017). Based on the Gaussian kernel similarity and symptom-based similarity, Huang et al. presented NGRHMDA combining the neighbor-based collaborative filtering model and the bipartite graph-based prediction model (Huang et al., 2017a). Shen et al. developed Bi-Random Walk based on Multiple Path with different length of path (BiRWMP) to predict microbe-disease associations (Shen et al., 2018).

In this paper, we have proposed a novel computational model called NBLPIHMDA based on the bidirectional label propagation to predict potential microbe-disease associations. NBLPIHMDA first calculates the disease similarity matrix and the microbe similarity matrix by introducing the Gaussian interaction profile kernel similarity, based on which a disease similarity network and a microbe similarity network are constructed simultaneously. After that, the edge weights of nodes in these two networks are calculated by using the Gaussian interaction profile kernel similarity. Finally, the bidirectional label propagation is performed on these two weighted networks to obtain the correlation score matrix between diseases and microbes. The

final prediction results are obtained by integrating these two correlation score matrices. The leave-one-out cross validation (LOOCV) and 5-fold cross validation (5-fold CV) are adopted to evaluate the predictive performance. As a result, NBLPIHMDA can achieve the area under the receiver operating characteristic (ROC) curves (AUCs) of 0.8777 and 0.8958 ± 0.0027, respectively. In addition, case studies on asthma, colorectal carcinoma and Chronic obstructive pulmonary disease (COPD) further demonstrate that NBLPIHMDA can be considered as an effective tool to discover reliable pathogenic microbes in the future.

2. MATERIALS AND METHODS

We downloaded known microbe-disease associations from the Human Microbe-Disease Association database (HMDAD on http://www.cuilab.cn/hmdad), which contains 483 microbe-disease associations including 39 diseases and 292 microbes collected from 61 publications (Ma et al., 2016). After removing redundant associations, we finally obtained 450 distinct microbe-disease associations (Chen et al., 2016a). Based on these associations, we constructed a 39×292 dimensional adjacency matrix A as our data source, where $A(i, j)=1$ if and only if there is a known association between the disease i and the microbe j, and $A(i, j) = 0$ otherwise. For convenience, we denote the number of collected diseases by N_D and the number of collected microbes by N_M.

2.1. Diseases Similarity Based on Gaussian Interaction Profile Kernel Similarity

Based on the assumption that diseases related to similar microbes tend to have more functional similarity and share similar interaction and non-interaction patterns with microbes (Chen et al., 2016a), we calculated the Gaussian interaction profile kernel similarity between each pair of diseases by using the Gaussian kernel for the interaction profiles of them. Specifically, for any two given diseases d_i and d_j, their Gaussian interaction profile kernel similarity can be calculated as follows:

$$SD(d_i, d_j) = exp(-\gamma_d \left\| IP(d_i) - IP(d_j) \right\|^2) \quad (1)$$

Where the interaction profile $IP(d_t)$ indicates whether there is an association between disease d_t and each microbe, and it is defined as a binary vector, i.e., the tth row of the adjacency matrix A. $||IP_{d_t}||$ represents the norm of the binary vector $IP(d_t)$. The parameter γ_d is used to control the kernel bandwidth, which needs to be calculated by normalizing a new bandwidth parameter γ'_d according to the average number of associations between each disease and microbes (Chen et al., 2016a). The calculation formula is as follows:

$$\gamma_d = \gamma'_d / (\frac{1}{N_D} \sum_{k=1}^{N_D} \left\| IP(d_k) \right\|^2) \quad (2)$$

Although it may be possible to set the new bandwidth parameter γ'_d to a better value by cross validation experiments, in this paper

we will set γ'_d to 1 for the sake of simplicity. Hence, based on above formulas, we can finally obtain a matrix SD, where $SD(i, j)$ represents the score of the Gaussian interaction profile kernel similarity between diseases d_i and d_j.

2.2. Microbes Similarity Based on Gaussian Interaction Profile Kernel Similarity

Similar to the way presented in above section 3.1, based on the assumption that microbes related to similar diseases tend to show more functional similarity and share similar interaction and non-interaction patterns with diseases (Chen et al., 2016a), the Gaussian interaction profile kernel similarity between each pair of microbes can be computed according to the following formula (3) as well:

$$SM(m_i, m_j) = exp(-\gamma_m \left\| IP(m_i) - IP(m_j) \right\|^2) \quad (3)$$

Where the interaction profile $IP(m_t)$ indicates whether there is an association between microbe m_t and each disease, and it is defined as a binary vector, i.e., the tth column of the adjacency matrix A. The parameter γ_m utilized to control the kernel bandwidth can be calculated by normalizing a new bandwidth parameter γ'_m as follows:

$$\gamma_m = \gamma'_m / (\frac{1}{N_M} \sum_{k=1}^{N_M} \left\| IP(m_k) \right\|^2) \quad (4)$$

And in this paper, the new bandwidth parameter γ'_m will be set to 1 for the sake of simplicity as well. Hence, based on above formulas, we can finally obtain another matrix SM, where $SM(i, j)$ represents the score of the Gaussian interaction profile kernel similarity between microbes m_i and m_j.

2.3. Constructing Weighted Networks for Diseases and Microbes

According to these two kinds of Gaussian interaction profile kernel similarity score matrices SM and SD calculated above, it is obvious that we can construct a microbe similarity network based on SM and a disease similarity network based on SD, respectively. Moreover, considering that the similarity values between any two diseases or microbes calculated by the Gaussian kernel for the interaction profiles will not be zero, therefore it is obvious that both the newly constructed microbe similarity network and disease similarity network will be fully connected networks.

Additionally, while implementing the label propagation method (Zhu and Ghahramani, 2002) on the newly constructed disease similarity network, for any given disease node d_i, we will assign a initial label $A(i, :)$ to d_i first, where $A(i, :)$ represent the ith row of the adjacency matrix A constructed in above section 2. And then, these label information can be propagated between neighboring nodes in the disease similarity network, thereafter each node can update its label information according to the label information received from its neighboring nodes. However, while updating its label information, it is reasonable to the node that its label information should be updated according to these neighboring nodes with high similarity to it rather than all of

its neighboring nodes. And moreover, these neighboring nodes with higher similarity to it should be assigned larger weights in the process of updating as well. Hence, based on above analysis, for any given disease node d_i, let Q_d represent the set of disease nodes other than d_i itself and these K disease nodes with the top K lowest similarity to d_i, then we can a novel matrix SD^* based on above obtained matrix SD as follows:

$$SD^*(d_i, d_j) = \begin{cases} \frac{SD(d_i,d_j)}{\sum_{d_k \in Q_d} SD(d_i,d_k)}, & \text{if } d_j \in Q_d \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Furthermore, in a similar way, while implementing the label propagation method on the newly constructed microbe similarity network, for any given microbe m_i, we will assign a initial label $A(:, i)$ to m_i first, where $A(:, i)$ represent the ith column of the adjacency matrix A constructed in above section 2. And thereafter, let Q_m represent the set of microbe nodes other than m_i itself and these K microbe nodes with the top K lowest similarity to m_i. then we can a novel matrix SM^* based on above obtained matrix SM as follows:

$$SM^*(m_i, m_j) = \begin{cases} \frac{SM(m_i,m_j)}{\sum_{m_k \in Q_m} SM(m_i,m_k)}, & \text{if } m_j \in Q_m \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Thus, according to above formula (5) and formula (6), we can further construct a updated disease similarity network and a updated microbe similarity network based on these two kinds of newly obtained matrices such as SD^* and SM^*. Thereafter, in this way, we have built two novel weighted networks that are adapted to label propagation.

2.4. NBLPIHMDA

As illustrated in the following **Figure 1**, the implementation process of our prediction model NBLPIHMDA can be divided into the following major steps.

Step 1: Firstly, we will implement the label propagation on the updated disease similarity network, for each given disease node d_i, supposing that its label information is updated by absorbing the labels from its neighboring nodes with probability α and retaining its previous label with probability $1-\alpha$, and in addition, let $L_{d_i}^0 = A(i, :)$ denote the initial label of the node d_i, $L_{d_i}^k$ represent the label of d_i after k rounds of updating, then $L_{d_i}^k$ can be calculated as follows:

$$L_{d_i}^k = \alpha \cdot SD^*(d_i, :) \cdot A + (1 - \alpha) \cdot L_{d_i}^{k-1} \quad (7)$$

Moreover, as for all disease nodes, supposing that their label vectors be $L_{d_1}^k, L_{d_2}^k, \cdots, L_{d_{N_D}}^k$ after k rounds of updating in the updated disease similarity network, and for convenience, let $LD^k = (L_{d_1}^k; L_{d_2}^k; \cdots; L_{d_{N_D}}^k)$, then, we can rewrite above formula (7) into the following form of matrix:

$$LD^k = \alpha \cdot SD^* \cdot A + (1 - \alpha) \cdot LD^{k-1} \quad (8)$$

According to above (8), the labels of each disease node in the updated disease similarity network can be updated iteratively. And during the process of iterative updating, we will consider the iteration to be convergent and stop the process of iterative updating while the change between the updated label matrix LD^k and the former label matrix LD^{k-1} measured by the absolute loss function is less than a predetermined threshold P. Thus, supposing that the process of iterative updating stopped after $n1$ rounds of iterations, we can obtain a microbe-disease association score matrix W_{LPD} as follows:

$$W_{LPD} = LD^0 + LD^1 + ... + LD^{n1} \quad (9)$$

Step 2: Next, we will implement the label propagation on the updated microbe similarity network, for each given microbe node m_i, supposing that its label information is updated by absorbing the labels from its neighboring nodes with probability α and retaining its previous label with probability $1-\alpha$, and in addition, let $L_{m_i}^0 = A(:, i)$ denote the initial label of the node m_i, $L_{m_i}^k$ represent the label of m_i after k rounds of updating, then $L_{m_i}^k$ can be calculated as follows:

$$L_{m_i}^k = (\alpha \cdot SM^*(m_i, :) \cdot A^T)^T + (1 - \alpha) \cdot L_{m_i}^{k-1} \quad (10)$$

Moreover, as for all microbe nodes, supposing that their label vectors be $L_{m_1}^k, L_{m_2}^k, \cdots, L_{m_{N_M}}^k$ after k rounds of updating in the updated disease similarity network, and for convenience, let $LM^k = (L_{m_1}^k, L_{m_2}^k, \cdots, L_{m_{N_M}}^k)$, then, we can rewrite above formula (10) into the following form of matrix:

$$LM^k = (\alpha \cdot SM^* \cdot A^T)^T + (1 - \alpha) \cdot LM^{k-1} \quad (11)$$

Thus, supposing that the process of iterative updating stopped after $n2$ rounds of iterations, then in a similar way, we can obtain another microbe-disease association score matrix W_{LPM} as follows:

$$W_{LPM} = LM^0 + LM^1 + ... + LM^{n2} \quad (12)$$

Finally, through combining these two kinds of score matrices W_{LPD} and W_{LPM} obtained by bidirectional label propagation, we can ultimately obtain a final microbe-disease association score matrix W as follows:
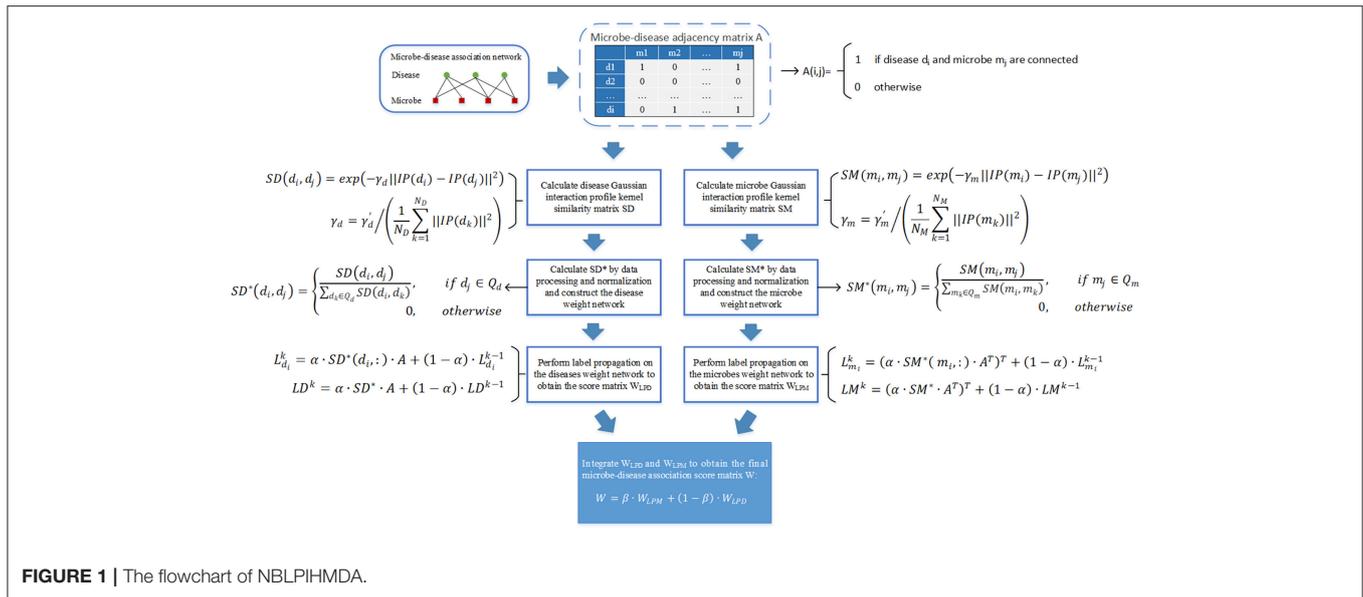
$$W = \beta \cdot W_{LPM} + (1 - \beta) \cdot W_{LPD} \quad (13)$$

Here, β is a parameter with value between 0 and 1 for controlling the weights of W_{LPD} and W_{LPM}.

3. RESULTS

3.1. Performance Evaluation

We adopted LOOCV and 5-fold CV to evaluate the performance of NBLPIHMDA. In the framework of LOOCV, each known microbe-disease association was taken as the test sample in
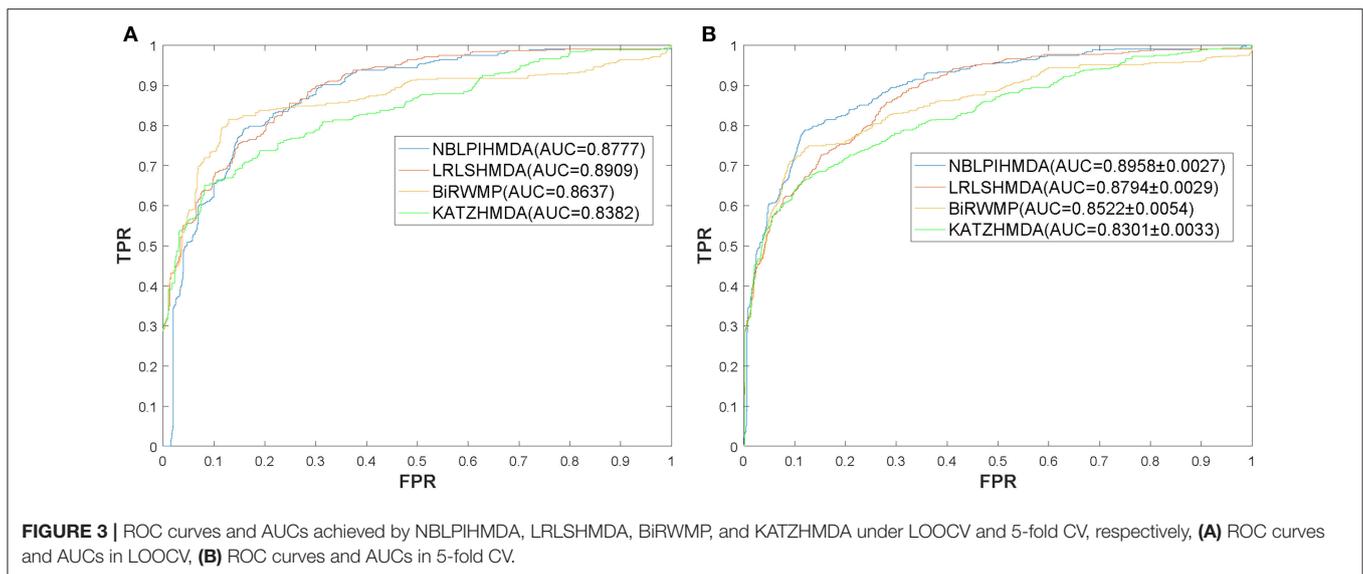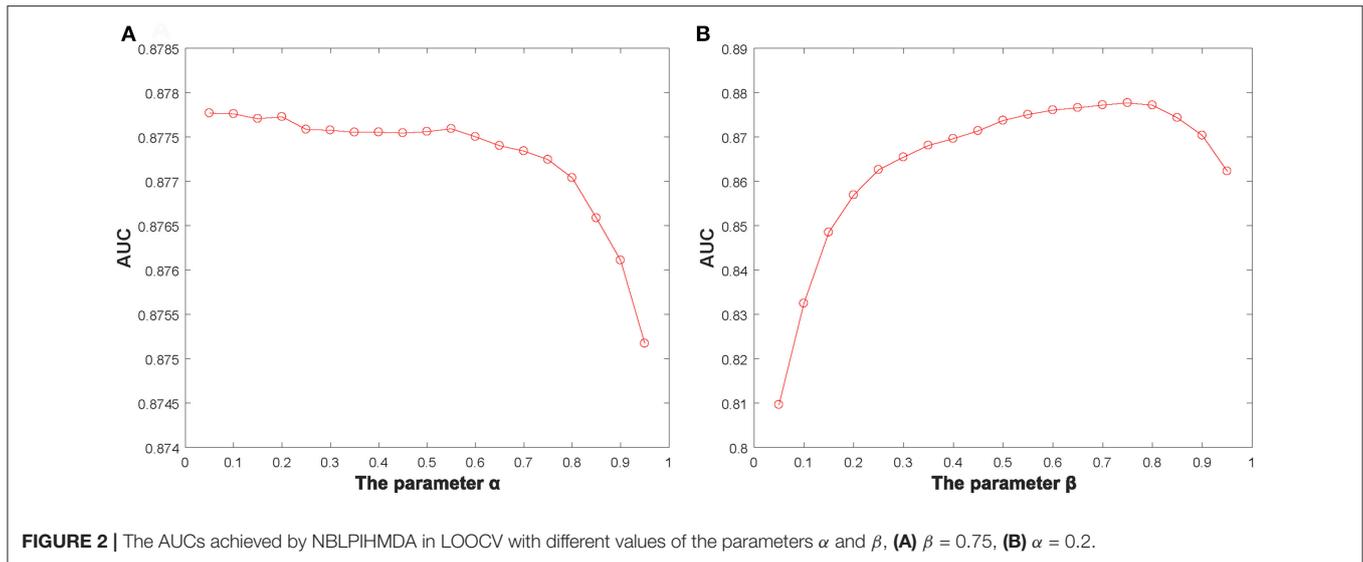
FIGURE 1 | The flowchart of NBLPIHMDA.

turn, while the remaining known associations were taken as the training set. In addition, all microbe-disease pairs without known associations would be considered negative samples. Thereafter, based on the predicted scores, each test sample would be ranked with all microbe-disease pairs that were not confirmed to be associated. Samples with rankings above the given threshold were predicted to be positive, whereas samples with rankings below the given threshold were predicted to be negative. In addition, test samples with rankings above the given threshold were considered to be successful samples. Next, in the framework of 5-fold CV, known microbe-disease associations were randomly divided into five groups, and each group was selected as a test sample in turn, while the remaining four groups were used as training samples. In order to reduce the deviation caused by random grouping, this process would be performed 100 times. Under the setting of different thresholds, the ROC curve could be further plotted by calculating the corresponding true and false positive rates. In our experiments, sensitivity referred to the percentage of positive samples with rankings above the given threshold, and the specificity was the percentage of negative samples with rankings below the given threshold. Subsequently, the AUC would be further calculated to evaluate the performance. Obviously, the AUC value of 1 represented a perfect prediction, while the AUC value of 0.5 represented a random prediction. Thereafter, simulation results show that NBLPIHMDA can achieve reliable AUCs of 0.8777 and 0.8958 ± 0.0027 under the frameworks of LOOCV and 5-fold CV, respectively, which indicated that NBLPIHMDA has satisfactory prediction performances.

Additionally, we identified several important parameters in NBLPIHMDA, such as the propagation probability α and the weighting factor β and so on. Hence, it is necessary to evaluate the impacts of these important parameters to the prediction performance of NBLPIHMDA. And as for evaluating the effects of the parameter α, we calculated the AUCs in framework of LOOCV with α varying from 0.05 to 0.95. The simulation results were shown in the following **Figure 2**, and as a result, it is obvious

that NBLPIHMDA can achieve the highest AUC of 0.8777 while $\alpha = 0.2$, which implies that while updating its label information, a node should retain more of its previous label information. Next, as for evaluating the effects of the parameter β, we calculated the AUCs in framework of LOOCV with β varying from 0.05 to 0.95. The simulation results were shown in the following **Figure 2**, and as a result, it is obvious that NBLPIHMDA can achieve the best prediction performance while $\beta = 0.75$, which means that the weight assigned to W_{LPD} should be greater than that of W_{LPM}. Through analysis, the reason that why the weight assigned to W_{LPD} should be greater than that of W_{LPM} may be that the number of collected microbes is much larger than the number of diseases. Therefore, in NBLPIHMDA, we would set $\alpha = 0.2$ and $\beta = 0.75$. And besides, according to our simulation results, the other two parameters K and P would be set to 5 and 10^{-12}, respectively in this paper.

3.2. Comparison With Other Methods
In this section, in order to attest the superior performance of our computational model NBLPIHMDA, we compared NBLPIHMDA with some state-of-the-art prediction methods such as KATZHMDA (Chen et al., 2016a), LRLSHMDA (Wang et al., 2017), and BiRWMP (Shen et al., 2018) under the frameworks of LOOCV and 5-fold CV, respectively. Here, KATZHMDA adopted the KATZ method and Gaussian interaction profile kernel similarity to infer potential microbe-disease associations by introducing the variable step number into a heterogeneous network. LRLSHMDA combined the Laplace's regular least squares classification with topological information of known microbe-disease association network to infer novel microbe-diseases associations. BiRWMP was a random walk based computational model for potential microbe-disease association prediction, and while comparing NBLPIHMDA with BiRWMP, in order to maintain consistency and accurate contrast, we replaced the Spearman correlations and symptom-based similarity of diseases adopted in BiRWMP

FIGURE 2 | The AUCs achieved by NBLPIHMDA in LOOCV with different values of the parameters α and β, **(A)** $\beta = 0.75$, **(B)** $\alpha = 0.2$.

FIGURE 3 | ROC curves and AUCs achieved by NBLPIHMDA, LRLSHMDA, BiRWMP, and KATZHMDA under LOOCV and 5-fold CV, respectively, **(A)** ROC curves and AUCs in LOOCV, **(B)** ROC curves and AUCs in 5-fold CV.

with the Gaussian interaction profile kernel similarity. And as illustrated in the following **Figure 3** and **Table 1**, simulation results show that NBLPIHMDA can achieve reliable AUCs of 0.8777 and 0.8958 ± 0.0027 in LOOCV and 5-fold CV, respectively, which are not only superior to the AUCs of 0.8382 and 0.8637 achieved by KATZHMDA and BiRWMP in LOOCV, but also superior to the AUCs of 0.8301 ± 0.0033 and 0.8522 ± 0.0054 achieved by KATZHMDA and BiRWMP in 5-fold CV simultaneously. While comparing with LRLSHMDA, although the AUC of 0.8777 achieved by NBLPIHMDA in LOOCV is not as good as the AUC of 0.8909 achieved by LRLSHMDA in LOOCV, however, the AUC of 0.8958 ± 0.0027 achieved by NBLPIHMDA in 5-fold CV is much better than the AUC of 0.8794 ± 0.0029 achieved by LRLSHMDA in 5-fold CV. Hence, it is obvious that the prediction performance of NBLPIHMDA outperforms that of these state-of-the-art prediction models mentioned above.

TABLE 1 | The AUCs achieved by NBLPIHMDA, LRLSHMDA, BiRWMP, and KATZHMDA under the framework of LOOCV and 5-fold CV.

Method	LOOCV	5-fold CV
NBLPIHMDA	0.8777	0.8958 ± 0.0027
LRLSHMDA	0.8909	0.8794 ± 0.0029
BiRWMP	0.8637	0.8522 ± 0.0054
KATZHMDA	0.8382	0.8301 ± 0.0033

3.3. Case Study

We selected three diseases of widespread concern such as colorectal carcinoma, asthma, and COPD for case studies to investigate their pathogenic mechanism from the perspective of microbes. Interestingly, there are 10, 10, and 8 out of the top 10 predicted microbes could be validated, respectively

for the three diseases by literature mining. The rankings of all potential microbe-disease pairs and the top ten related microbes of all diseases predicted by NBLPIHMDA are listed in **Supplementary Tables 1, 5**, respectively.

Asthma, a common chronic inflammatory pulmonary tracheal disease, affects more than 300 million people worldwide. Asthma usually occurs in childhood and is accompanied by a recurrent cough, wheezing, chest tightness, and dyspnea. Currently, there is no cure for asthma. Asthma is generally thought to be caused by a combination of genetic and environmental factors. In recent years, more and more studies have shown that microbes play an important role in the pathogenesis of asthma. Consequently, we conducted a case study of Asthma on our calculation model. And as illustrated in the following **Table 2** and **Supplementary Table 2**, all of these top 10 predicted microbes interrelated with Asthma were verified to be correlative. For example, the colonization of *Clostridium difficile* (Ranking fist in the prediction list) at 1 month of neonatal age was closely related to asthma and eczema that occurred at 6–7 years of age (van Nimwegen et al., 2011). Compared with non-asthmatic healthy people, the increase of Firmicutes (Ranking second in the prediction list) mainly composed of streptococcal OTUs in patients with severe asthma is abnormally significant, and Actinobacteria (Ranking fifth in the prediction list) was found to have a lower proportion in asthmatic patients (Zhang et al., 2016). *Clostridium coccoides* (Ranking third in the prediction list) was confirmed to be significantly associated with a positive Asthma Predictive Index (API), and early fecal colonization with a *Clostridium coccoides* subcluster XIVa species could be an early indicator of asthma later in life (Vael et al., 2011). An increase in sensitivity to *Staphylococcus aureus* (Ranking fourth in the prediction list) enterotoxins in smokers with asthma could be considered as a marker of eosinophilic inflammation and exacerbation of asthma (Nagasaki et al., 2017). Long-term asthma patients have lower levels of Bifidobacteria (Ranking eighth in the prediction list) compared with patients who have recently been diagnosed with asthma (Hevia et al., 2016). The colonization of Bacteroides (Ranking tenth in the prediction list) at three weeks of infants was proved to be positively correlated with the positive asthma predictive index at 3 years of age, which may serve as the early indicator of asthma (Vael et al., 2008).

Colorectal carcinoma, also known as colon cancer, is the third most common cancer, whose symptoms may include blood in the stool, abdominal pain, diarrhea, weight loss, prolonged fatigue, and flagging spirit. The prevalence of colorectal cancer is increasing year by year, about 4 to 5%. The risk is mainly caused by age, lifestyle, and genetic history. The gut microbiota also plays a related role, and disorder of the intestinal flora may affect the chronic inflammation mechanism and induce colon cancer. Hence we conducted a case study of Colorectal Carcinoma on our calculation model. And as illustrated in the following **Table 3** and **Supplementary Table 3**, all of these top 10 predicted microbes interrelated with Colorectal Carcinoma were verified to be correlative. For example, infection with *Helicobacter pylori* (Ranking third in the prediction list), particularly CagA-positive flora, increased the risk of Colorectal cancer and gastric adenocarcinoma (Shmuely, 2001). By 16S rRNA gene denaturing

TABLE 2 | The top 10 predicted microbes associated with asthma.

Rank	Microbe	Evidence
1	*Clostridium difficile*	PMID: 21872915
2	Firmicutes	PMID: 23265859
3	*Clostridium coccoides*	PMID: 21477358
4	*Staphylococcus aureus*	PMID: 17950502
5	Actinobacteria	PMID: 23265859
6	Clostridia	PMID: 21477358
7	Lachnospiraceae	Ciaccio et al. (2014)
8	Bifidobacterium	PMID: 26840903,PMID: 24735374
9	Lactobacillus	Gutkowski et al. (2011), PMID: 20592920
10	Bacteroides	PMID: 18822123, PMID: 29161087

TABLE 3 | The top 10 predicted microbes associated with colorectal carcinoma.

Rank	Microbe	Evidence
1	Proteobacteria	PMID: 24603888
2	*Clostridium difficile*	PMID: 19807912
3	*Helicobacter pylori*	PMID: 11774957
4	*Clostridium coccoides*	PMID: 19807912
5	*Staphylococcus aureus*	PMID: 7074582
6	Actinobacteria	PMID: 24316595
7	Lachnospiraceae	PMID: 29985435
8	Bifidobacterium	PMID: 9111222
9	Haemophilus	PMID: 22761885
10	Lactobacillus	PMID: 15828052

gradient gel electrophoresis and ribosomal gene interval analysis (RISA), it was found that the abundance of *Clostridium difficile* (Ranking second in the prediction list) and *C. coccoides* (Ranking fourth in the prediction list) in colorectal cancer patients was significantly increased compared with healthy controls, which means that *C. cerevisiae* might have a potential effect on the induction of colorectal cancer (Leu et al., 2009). Protein a-containing *Staphylococcus aureus* (Ranking fifth in the prediction list) was used as an immunosorbent to perform *in vitro* immunoadsorption therapy on the plasma of patients with metastatic colorectal carcinoma and has produced good effects (Ishikawa et al., 2005). *Bifidobacterium lactis* (Ranking eighth in the prediction list) could prevent colorectal carcinoma by up-regulating the apoptosis reaction of colon carcinogens (Ray et al., 1982). The ingestion of *Lactobacillus casei* (Ranking tenth in the prediction list) had been shown in a large randomized clinical trial to reduce the incidence of moderate or severe heterogeneity tumors and to inhibit the atypia of colorectal tumors as a preventive measure (Scanlan et al., 2008).

COPD is a progressive obstructive pulmonary disease, which means it will get worse over time. It usually occurs in people over the age of 40, and the risk is the same for men and women. Smoking is the most common habit in patients with COPD. Besides, factors such as air pollution and genes will also increase the risk of COPD. Although treatment can be

TABLE 4 | The top 10 predicted microbes associated with COPD.

Rank	Microbe	Evidence
1	*Clostridium difficile*	PMID: 15655746
2	*Helicobacter pylori*	PMID: 15733502
3	Firmicutes	PMID: 24591822
4	*Clostridium coccoides*	unconfirmed
5	*Staphylococcus aureus*	PMID: 7074582
6	Actinobacteria	PMID: 26852737
7	Clostridia	PMID: 26852737
8	Lachnospiraceae	unconfirmed
9	Bifidobacterium	PMID: 26317628
10	Bacteroides	PMID: 29709671

used to slow down the progression of the disease, there is no cure yet, and it is even more necessary to step up research on the pathogenesis of COPD. Therefore, in this section, COPD is selected as a case for study. And as shown in following **Table 4** and **Supplementary Table 4**, 8 out of the top 10 predicted microbes associated with COPD were confirmed to be relevant. For instance, through the study, it was found that the serum *Helicobacter pylori*-specific IgG in COPD patients was significantly higher than that in the healthy control group, which implies that *Helicobacter pylori* (Ranking second in the prediction list) infection is closely related to COPD (Mammen and Sethi, 2016). In addition, there is evidence that IgE antibodies to *Staphylococcus aureus* (Ranking fifth in the prediction list) enterotoxin in patients with COPD are significantly higher than that in healthy subjects, suggesting that an immune response to this superantigen, *S. aureus* enterotoxin, may be a potential cause of chronic inflammation in COPD (Gencer et al., 2005). The increase of Actinomyces (Ranking sixth in the prediction list) and Proteobacteria was found to aggravate the deterioration of the pathogenesis of COPD (Rohde et al., 2004). *Bifidobacterium breve* (Ranking ninth in the prediction list) can serve as an anti-inflammatory agent to inhibit the expression and release of inflammatory mediators in COPD by inhibiting the activity of NF-kB induced by cigarette smoke which is the main cause of COPD (Mortaz et al., 2015).

4. DISCUSSION

There are numerous microbial communities inhabited in the human body, which is critical to human health. The relationship between human microbiome and diseases received much attention from both medical and bioinformatics community recently. However, traditional methods to detect their association is costly and labor-intensive. Thus, we proposed here a new computational model called NBLPIHMDA to infer potential microbe-disease associations. NBLPIHMDA first combined known microbe-disease associations in HMDAD and the Gaussian interaction profile kernel similarity to construct disease similarity network and microbe similarity network. It then conducted tag transmission on these two networks to obtain the predicted score of each microbe-disease pair. Under the

framework of LOOCV and 5-fold CV, the AUCs reached 0.8777 and 0.8958 ± 0.0027, respectively, In addition, the case studies of asthma, colorectal carcinoma and COPD further demonstrated that NBLPIHMDA could provide valuable insights into the pathogenesis research.

It worth's noting that NBLPIHMDA has certain limitations. First of all, the HMDAD only curated hundreds of known associations between 39 diseases and 292 microbes, which is relatively small. The problem will be partially solved in the future when more associations between diseases and microbes are discovered. In addition, the Gaussian interaction profile kernel similarity of diseases and microbes is calculated based on the known microbe-disease associations, which will bias toward diseases with more known associations and microbes with more known associations. We believe that the bias can be reduced by integrating other effective similarity methods, such as disease semantic similarity, symptom-based disease similarity, and microbe functional similarity. The advancement of association prediction research in various fields of computational biology would also provide valuable insights into the development of microbe-disease association prediction, such as miRNA-disease association prediction (Chen and Huang, 2017; Chen et al., 2018a,b,c), lncRNA-disease association prediction (Chen and Yan, 2013), drug-target interaction prediction (Chen et al., 2015), and synergistic drug combinations (Chen et al., 2016b). Therefore, we will introduce some reliable technologies and optimization strategies in the future work to further improve the quality of the heterogeneous network and the prediction performance of NBLPIHMDA.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00684/full#supplementary-material

Supplementary Table 1 | Final prediction result.

Supplementary Table 2 | The top 10 predicted microbes associated with asthma.

Supplementary Table 3 | The top 10 predicted microbes associated with colorectal carcinoma.

Supplementary Table 4 | The top 10 predicted microbes associated with COPD.

Supplementary Table 5 | The top ten related microbes of all diseases.

REFERENCES

Brown, C. T., Davis-Richardson, A. G., Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., et al. (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS ONE* 6:e25792. doi: 10.1371/journal.pone.0025792

Chen, X., and Huang, L. (2017). LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction. *PLoS Comp. Biol.* 13:e1005912. doi: 10.1371/journal.pcbi.1005912

Chen, X., Huang, Y.-A., You, Z.-H., Yan, G.-Y., and Wang, X.-S. (2016a). A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* 33, 733–739. doi: 10.1093/bioinformatics/btw715

Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). Drug–target interaction prediction by random walk on the heterogeneous network. *Mol. BioSyst.* 8:1970. doi: 10.1039/c2mb00002d

Chen, X., Ren, B., Chen, M., Wang, Q., Zhang, L., and Yan, G. (2016b). NLLSS: Predicting synergistic drug combinations based on semi-supervised learning. *PLoS Comp. Biol.* 12:e1004975. doi: 10.1371/journal.pcbi.1004975

Chen, X., Wang, L., Qu, J., Guan, N.-N., and Li, J.-Q. (2018a). Predicting miRNA–disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265. doi: 10.1093/bioinformatics/bty503

Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z.-H., and Liu, H. (2018b). BNPMDA: Bipartite network projection for MiRNA–disease association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333

Chen, X., Yan, C. C., Zhang, X., and You, Z.-H. (2016c). Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform.* 18, 558–576. doi: 10.1093/bib/bbw060

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., et al. (2015). Drug–target interaction prediction: databases, web servers and computational models. *Brief. Bioinform.* 17, 696–712. doi: 10.1093/bib/bbv066

Chen, X. and Yan, G.-Y. (2013). Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426

Chen, X., Yin, J., Qu, J., and Huang, L. (2018c). MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comp. Biol.* 14:e1006418. doi: 10.1371/journal.pcbi.1006418

Ciaccio, C. E., Kennedy, K., Barnes, C. S., Portnoy, J. M., and Rosenwasser, L. J. (2014). The home microbiome and childhood asthma. *J. Allergy Clin. Immunol.* 133:AB70. doi: 10.1016/j.jaci.2013.12.274

Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C., and Gilad, Y. (2014). Seasonal variation in human gut microbiome composition. *PLoS ONE* 9:e90731. doi: 10.1371/journal.pone.0090731

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2013). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820

Donia, M. S., Cimermancic, P., Schulze, C. J., Brown, L. C. W., Martin, J., Mitreva, M., et al. (2014). A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 158, 1402–1414. doi: 10.1016/j.cell.2014.08.032

Gencer, M., Ceylan, E., Zeyrek, F. Y., and Aksoy, N. (2005). *Helicobacter pylori* seroprevalence in patients with chronic obstructive pulmonary disease and its relation to pulmonary function tests. *Respiration* 74, 170–175. doi: 10.1159/000090158

Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., et al. (2010). Meeting report: The terabase metagenomics workshop and the vision of an earth microbiome project. *Stand. Genom. Sci.* 3, 243–248. doi: 10.4056/sigs.1433550

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., et al. (2014). Human genetics shape the gut microbiome. *Cell* 159, 789–799. doi: 10.1016/j.cell.2014.09.053

Guarner, F., and Malagelada, J.-R. (2003). Gut flora in health and disease. *Lancet* 361, 512–519. doi: 10.1016/S0140-6736(03)12489-0

Gutkowski, P., Madaliński, K., Grek, M., Dmeńska, H., Syczewska, M., and Michałkiewicz, J. (2011). Clinical immunology effect of orally administered probiotic strains lactobacillus and bifidobacterium in children with atopic asthma. *Central Eur. J. Immunol.* 35, 233–238. Available online at: https://www.termedia.pl/Clinical-immunology-Effect-of-orallyadministered-probiotic-strains-Lactobacillus-and-Bifidobacteriumin-children-with-atopic-asthma,10,16116,0,1.html

Hevia, A., Milani, C., López, P., Donado, C. D., Cuervo, A., González, S., et al. (2016). Allergic patients with long-term asthma display low levels of bifidobacterium adolescentis. *PLoS ONE* 11:e0147809. doi: 10.1371/journal.pone.0147809

Huang, Y.-A., You, Z.-H., Chen, X., Huang, Z.-A., Zhang, S., and Yan, G.-Y. (2017a). Prediction of microbe–disease association from the integration of neighbor and graph with collaborative recommendation model. *J. Trans. Med.* 15:209. doi: 10.1186/s12967-017-1304-7

Huang, Z.-A., Chen, X., Zhu, Z., Liu, H., Yan, G.-Y., You, Z.-H., et al. (2017b). PBHMDA: Path-based human microbe-disease association prediction. *Front. Microbiol.* 8:233. doi: 10.3389/fmicb.2017.00233

Ishikawa, H., Akedo, I., Otani, T., Suzuki, T., Nakamura, T., Takeyama, I., et al. (2005). Randomized trial of dietary fiber andLactobacillus casei administration for prevention of colorectal tumors. *Int. J. Cancer* 116, 762–767. doi: 10.1002/ijc.21115

Khachatryan, Z. A., Ktsoyan, Z. A., Manukyan, G. P., Kelly, D., Ghazaryan, K. A., and Aminov, R. I. (2008). Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE* 3:e3064. doi: 10.1371/journal.pone.0003064

Leu, R. K. L., Hu, Y., Brown, I. L., Woodman, R. J., and Young, G. P. (2009). Synbiotic intervention of *Bifidobacterium lactis* and resistant starch protects against colorectal cancer development in rats. *Carcinogenesis* 31, 246–251. doi: 10.1093/carcin/bgp197

Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2016). An analysis of human microbe–disease associations. *Brief. Bioinform.* 18, 85–97. doi: 10.1093/bib/bbw005

Mammen, M. J., and Sethi, S. (2016). COPD and the microbiome. *Respirology* 21, 590–599. doi: 10.1111/resp.12732

Mason, M. R., Preshaw, P. M., Nagaraja, H. N., Dabdoub, S. M., Rahman, A., and Kumar, P. S. (2014). The subgingival microbiome of clinically healthy current and never smokers. *ISME J.* 9, 268–272. doi: 10.1038/ismej.2014.114

Meng, X., Zou, Q., Rodriguez-Paton, A., and Zeng, X. (2017). "Iteratively collective prediction of disease-gene associations through the incomplete network," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (IEEE), 1324–1330. doi: 10.1109/BIBM.2017.8217854

Methé B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., et al. (2012). A framework for human microbiome research. *Nature* 486, 215–221. doi: 10.1038/nature11209

Mortaz, E., Adcock, I. M., Ricciardolo, F. L. M., Varahram, M., Jamaati, H., Velayati, A. A., et al. (2015). Anti-inflammatory effects of lactobacillus rahmnosus and *Bifidobacterium breve* on cigarette smoke activated human macrophages. *PLoS ONE* 10:e0136455. doi: 10.1371/journal.pone.0136455

Murri, M., Leiva, I., Gomez-Zumaquero, J. M., Tinahones, F. J., Cardona, F., Soriguer, F., et al. (2013). Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Med.* 11:46. doi: 10.1186/1741-7015-11-46

Nagasaki, T., Matsumoto, H., Oguma, T., Ito, I., Inoue, H., Iwata, T., et al. (2017). Sensitization to *Staphylococcus aureus* enterotoxins in smokers with asthma. *Ann. Allergy Asthma Immunol.* 119, 408–414. doi: 10.1016/j.anai.2017.08.001

Ray, P., Idiculla, A., Mark, R., Rhoads Jr, J., Thomas, H., Bassett, J., et al. (1982). Extracorporeal immunoadsorption of plasma from a metastatic colon carcinoma patient by protein a-containing nonviable *Staphylococcus aureus*. clinical, biochemical, serologic, and histologic evaluation of the patient's response. *Cancer* 49, 1800–1809. doi: 10.1002/1097-0142(19820501)49:9<1800::AID-CNCR2820490912>3.0.CO;2-6

Rohde, G., Gevaert, P., Holtappels, G., Borg, I., Wiethege, A., Arinir, U., et al. (2004). Increased IgE-antibodies to *Staphylococcus aureus* enterotoxins in patients with COPD. *Respir. Med.* 98, 858–864. doi: 10.1016/j.rmed.2004.02.012

Scanlan, P. D., Fergus, S., Yvonne, C., Collins, J. K., O'Sullivan, G. C., Micheal, O., et al. (2008). Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ. Microbiol.* 10, 1382–1382. doi: 10.1111/j.1462-2920.2008.01622.x

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 14:e1002533. doi: 10.1371/journal.pbio.1002533

Shah, P., Fritz, J. V., Glaab, E., Desai, M. S., Greenhalgh, K., Frachet, A., et al. (2016). A microfluidics-based *in vitro* model of the gastrointestinal human–microbe interface. *Nat. Commun.* 7:11535. doi: 10.1038/ncomms11535

Shen, X., Chen, Y., Jiang, X., Hu, X., He, T., and Yang, J. (2016). "Predicting disease-microbe association by random walking on the heterogeneous network," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (IEEE), 771–774. doi: 10.1109/BIBM.2016.7822619

Shen, X., Zhu, H., Jiang, X., Hu, X., and Yang, J. (2018). "A novel approach based on bi-random walk to predict microbe-disease associations," in *Intelligent Computing Methodologies*, (Springer International Publishing), 746–752. doi: 10.1007/978-3-319-95957-3_78

Shmuely, H. (2001). Relationship between *Helicobacter pylori* CagA status and colorectal cancer. *Am. J. Gastroenterol.* 96, 3406–3410. doi: 10.1111/j.1572-0241.2001.05342.x

Singh, B., Read, S., Asseman, C., Malmstrom, V., Mottet, C., Stephens, L. A., et al. (2001). Control of intestinal inflammation by regulatory t cells. *Immunol. Rev.* 182, 190–200. doi: 10.1034/j.1600-065X.2001.1820115.x

Tang, W., Wan, S., Yang, Z., Teschendorff, A. E., and Zou, Q. (2017). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 34, 398–406. doi: 10.1093/bioinformatics/btx622

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2008). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi: 10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244

Vael, C., Nelen, V., Verhulst, S. L., Goossens, H., and Desager, K. N. (2008). Early intestinal bacteroides fragilis colonisation and development of asthma. *BMC Pulmon. Med.* 8:19. doi: 10.1186/1471-2466-8-19

Vael, C., Vanheirstraeten, L., Desager, K. N., and Goossens, H. (2011). Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to

the development of asthma. *BMC Microbiol.* 11:68. doi: 10.1186/1471-21 80-11-68

van Nimwegen, F. A., Penders, J., Stobberingh, E. E., Postma, D. S., Koppelman, G. H., Kerkhof, M., et al. (2011). Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. *J. Allergy Clin. Immunol.* 128, 948–955. doi: 10.1016/j.jaci.2011.07.027

Wang, F., Huang, Z.-A., Chen, X., Zhu, Z., Wen, Z., Zhao, J., et al. (2017). LRLSHMDA: laplacian regularized least squares for human microbe–disease association prediction. *Sci. Rep.* 7:7601. doi: 10.1038/s41598-017-08127-2

You, Z.-H., Huang, Z.-A., Zhu, Z., Yan, G.-Y., Li, Z.-W., Wen, Z., et al. (2017). PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comp. Biol.* 13:e1005455. doi: 10.1371/journal.pcbi.1005455

Yu, J., Ping, P., Wang, L., Kuang, L., Li, X., and Wu, Z. (2018). A novel probability model for LncRNA–disease association prediction based on the naïve bayesian classifier. *Genes* 9:345. doi: 10.3390/genes9070345

Zeng, X., Ding, N., Rodríguez-Patón, A., and Zou, Q. (2017). Probability-based collaborative filtering model for predicting gene–disease associations. *BMC Med. Genom.* 10:76. doi: 10.1186/s12920-017-0313-y

Zeng, X., Liu, L., Lü, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34, 2425–2432. doi: 10.1093/bioinformatics/bty112

Zeng, X., Zhang, X., and Zou, Q. (2015). Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* 17, 193–203. doi: 10.1093/bib/bbv033

Zhang, Q., Cox, M., Liang, Z., Brinkmann, F., Cardenas, P. A., Duff, R., et al. (2016). Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. *PLoS ONE* 11:e0152724. doi: 10.1371/journal.pone.0152724

Zhu, L., Su, F., Xu, Y., and Zou, Q. (2018). Network-based method for mining novel HPV infection related genes using random walk with restart algorithm. *Biochim. Biophys. Acta* 1864, 2376–2383. doi: 10.1016/j.bbadis.2017.11.021

Zhu, X., and Ghahramani, Z. (2002). *Learning From Labeled and Unlabeled Data With Label Propagation.* Technical report, Citeseer.

Zou, Q. (2016). Editorial (thematic issue: machine learning techniques for protein structure, genomics function analysis and disease prediction). *Current Proteomics* 13, 77–78. doi: 10.2174/157016461302160513235846

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.