



# RWHMDA: Random Walk on Hypergraph for Microbe-Disease Association Prediction

Ya-Wei Niu<sup>1</sup>, Cun-Quan Qu<sup>1,2</sup>, Guang-Hui Wang<sup>1,2\*</sup> and Gui-Ying Yan<sup>3</sup>

<sup>1</sup> School of Mathematics, Shandong University, Jinan, China, <sup>2</sup> Data Science Institute, Shandong University, Jinan, China,

<sup>3</sup> Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Based on advancements in deep sequencing technology and microbiology, increasing evidence indicates that microbes inhabiting humans modulate various host physiological phenomena, thus participating in various disease pathogenesises. Owing to increasing availability of biological data, further studies on the establishment of efficient computational models for predicting potential associations are required. In particular, computational approaches can also reduce the discovery cycle of novel microbe-disease associations and further facilitate disease treatment, drug design, and other scientific activities. This study aimed to develop a model based on the random walk on hypergraph for microbe-disease association prediction (RWHMDA). As a class of higher-order data representation, hypergraph could effectively recover information loss occurring in the normal graph methodology, thus exclusively illustrating multiple pair-wise associations. Integrating known microbe-disease associations in the Human Microbe-Disease Association Database (HMDAD) and the Gaussian interaction profile kernel similarity for microbes, random walk was then implemented for the constructed hypergraph. Consequently, RWHMDA performed optimally in predicting the underlying disease-associated microbes. More specifically, our model displayed AUC values of 0.8898 and 0.8524 in global and local leave-one-out cross-validation (LOOCV), respectively. Furthermore, three human diseases (asthma, Crohn's disease, and type 2 diabetes) were studied to further illustrate prediction performance. Moreover, 8, 10, and 8 of the 10 highest ranked microbes were confirmed through recent experimental or clinical studies. In conclusion, RWHMDA is expected to display promising potential to predict disease-microbe associations for follow-up experimental studies and facilitate the prevention, diagnosis, treatment, and prognosis of complex human diseases.

## OPEN ACCESS

### Edited by:

George Tsiamis,  
University of Patras, Greece

### Reviewed by:

Anastasio Chalaris,  
University of Oxford, United Kingdom  
Edwin Wang,  
University of Calgary, Canada

### \*Correspondence:

Guang-Hui Wang  
ghwang@sdu.edu.cn

### Specialty section:

This article was submitted to  
Systems Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 January 2019

**Accepted:** 25 June 2019

**Published:** 10 July 2019

### Citation:

Niu Y-W, Qu C-Q, Wang G-H and  
Yan G-Y (2019) RWHMDA: Random  
Walk on Hypergraph  
for Microbe-Disease Association  
Prediction. *Front. Microbiol.* 10:1578.  
doi: 10.3389/fmicb.2019.01578

**Keywords:** hypergraph, random walk, microbe, human diseases, association prediction

## INTRODUCTION

Microbes exist in almost all habitats of flora and fauna, including humans. Deeper microbiological insights have indicated more compact associations between humans and their microflora (Sommer and Backhed, 2013). Some microbes are harmless and vital for host health in various manners, such as enhancement of host immunity, improvement of host metabolic capability, and protection of the host against pathogens (Eckburg et al., 2003; Ventura et al., 2009). Over the past few decades,

numerous studies have focused on microbes inhabiting humans (Peterson et al., 2009). For instance, the gut flora are a complicated microbial community in the human digestive tract (Sommer and Backhed, 2013). Human gut microbes potentially benefit the host by synthesizing different vitamins, metabolizing bile acids, etc., thus exhibiting a fundamentally mutualistic association between some gut flora and the human host (Clarke et al., 2014). Therefore, microbes may be considered a supplemental “organ” in the host (Bäckhed et al., 2005). Furthermore, the number of microbial cells in the human body is reportedly approximately 10-fold the number of human cells (Rosner, 2014). Therefore, it is essential to systematically analyze associations between microbes and humans. The Human Microbiome Project (HMP) has furthered the current understanding of microbial structure, diversity, and function over the years (Human Microbiome Project Consortium, 2012). However, numerous basic and clinical studies have investigated the association between the human microbiome and human health (Moore and Moore, 1995; Dethlefsen et al., 2007; Zhang et al., 2009; Brown et al., 2011).

It is important to understand microbe-host interactions, which could benefit the prevention, diagnosis, treatment, and prognosis of human diseases (Bao et al., 2017; Zou et al., 2018). Microbial communities could be influenced by not only maternal genetic factors (Khachatryan et al., 2008; Turnbaugh et al., 2009; Goodrich et al., 2014) but also the habitat environments, such as the change of season (Davenport et al., 2014), host diet (David et al., 2014), antibiotic consumption (Donia et al., 2014), host smoking habits (Mason et al., 2015), and residential hygiene of the host (Sommer and Backhed, 2013). Changes in environmental variables may modify microbial communities and alter host-microbe interactions (Ma et al., 2014). In the past decades, with the development of high-throughput sequencing techniques and ensuing computational tools, increasing evidence demonstrates the close association between microbial dysbiosis and various human diseases (Neish, 2009), such as inflammatory bowel disease (IBD) (Frank et al., 2007), diabetes (Brown et al., 2011; Giongo et al., 2011), asthma (Chen and Blaser, 2007), obesity (Ley et al., 2006), and some cancers (Moore and Moore, 1995; Schwabe and Jobin, 2013). For example, through 16S rRNA microarray and parallel clone library-sequencing analysis, Huang et al. (2011) collected bronchial epithelial brushings from 65 asthma patients and compared them with 10 other samples from healthy control subjects, reporting that members of the airway microbiota, such as *Comamonadaceae*, *Sphingomonadaceae*, and *Oxalobacteraceae*, were greater in asthma patients. Hoppe et al. (2011) evaluated the effect of *Oxalobacter formigenes* on primary hyperoxaluria, a rare genetic disease. In particular, the urinary oxalate test and *ad hoc* analysis in their study revealed a reduction in *Oxalobacter formigenes* in patients with kidney stones. Furthermore, to analyze and elucidate the microbiota of colon cancer patients, Sobhani et al. (2011) extracted bacterial DNA from 179 colon cancer patients. Through qPCR and the immunohistochemical analyses, *C. coccoides*, *Bacteroides*, *Lactobacillus* groups, and *Faecalibacterium prausnitzii* species were reportedly increased in colon cancer patients. Moreover, on comparing microbes from 83 healthy control individuals

and 98 liver cirrhosis patients, Qin et al. (2014) identified several biomarkers associated with liver cirrhosis, reporting that certain groups were reduced (e.g., *Alistipes finegoldii*, *Bacteroides eggerthii*, and *Coprococcus*) while certain others were enriched (e.g., *Fusobacterium*, *Haemophilus parainfluenzae*, and *Phascolarctobacterium*). Therefore, elucidation of the association between microbes and human diseases may facilitate novel drug discovery.

Despite some reported microbe-disease associations, they are not sufficient to completely understand disease pathogenesis, diagnosis, and treatment. Fortunately, Wang et al. (2015) proposed the excellent work about cancer hallmark network framework in the predictive genomics. The cancer hallmark network framework offered great insights on modeling genome sequencing data to predict cancer evolution and associated clinical phenotypes, which provided valuable designment strategies for using the framework in conjunction with genome sequencing data in any other attempt to prediction works on human diseases, drug targets and other fields, microbe included. Indeed, construction of a computationally efficient model from existing associations to predict potential ones is practical, potentially providing novel insights into time-consuming microbiology experiments by elucidating the most promising previously unknown associations (Chen et al., 2017c). Specifically, in determining lncRNA-disease associations (Chen, 2015), studies on drug targets (van Laarhoven et al., 2011; Yamanishi, 2013) and miRNA-disease associations (Wang et al., 2010; Chen et al., 2017d) have yielded various efficient *in silico* models to predict the underlying associations. Recently, based on experimentally verified microbe-disease associations, Ma et al. (2017) constructed the first Human Microbe-Disease Association Database (HMDAD). Thereafter, several computational models have been proposed to further contribute to the HMDAD. For example, Chen et al. (2017a) generated a model based on the KATZ measure, named KATZHMDA. In their model, they first constructed an association network showing pairwise relationships between microbes and human disease. Furthermore, they introduced Gaussian interaction profile kernel similarity for microbes and diseases to predict novel associations. Moreover, Huang Z.A. et al. (2017) developed a model of Path-Based Human Microbe-Disease Association Prediction (PBHMDA), wherein they used a special depth-first search algorithm on the heterogeneous biological network. In particular, they investigated all possible paths between diseases and microbes to infer highly probable associations. Resulting from the idea of collaborative recommendation model, Huang Y.A. et al. (2017) provided a computational model by adopting neighbor-based collaborative filtering and a graph-based scoring approach to calculate the association possibility of unknown microbe-disease pairs. The usage of hybrid approach based on two single recommendation methods contributed much more on their prediction results. Based on the microbe-disease interaction network, Wang et al. (2017) developed the model of Laplacian Regularized Least Squares for Human Microbe-Disease Association (LRLSHMDA). LRLSHMDA is a semi-supervised computational model using the Laplacian regularized least squares classifier.

Recently, Zou et al. (2018) integrated symptom-based disease similarity to predict novel human microbe-disease associations based on network consistency projection (NCPHMDA). In detail, they conducted microbe space projection and disease space projection and combined the projections to design an advisable non-parametric approach. Based on adaptive boosting approach, Peng et al. (2018) developed a model named Adaptive Boosting for Human Microbe-Disease Association prediction (ABHMDA) to reveal the underlying associations between microbes and human diseases by calculating the association probability of concerned disease-microbe pair by grouped weak classifiers to form a stronger classifier for further scoring and sorting samples. Not long time ago, Qu et al. (2019) proposed a computational model on the basis of HMDAD by the methods of matrix decomposition and label propagation, which divided the original adjacency matrix about the relationship between microbes and diseases into a linear combination of itself and a low-rank matrix to predict novel disease-microbe associations.

Herein, we present a Random Walk on Hypergraph for Microbe-Disease Association Prediction (RWHMDA) model to predict underlying microbe-disease associations. In particular, we constructed a higher-order hypergraph model to accurately determine the implicit inherent association between microbes and human diseases. Thereafter, we generalized the well-known random walk process to the hypergraph in a modified manner, wherein vertices (microbes) within a hyperedge (human disease) were differentiated by the walker depending on their features. Finally, we ranked all candidate microbes for every investigated human disease. The merit of this study is the introduction of the concept and method of hypergraph to predict microbe-disease associations. Hypergraph is practical and suitable because it could provide biologically decipherable aspects by placing all disease-associated microbes in one hyperedge. Furthermore, we implemented global and local Leave-one-out cross-validation (LOOCV) to evaluate the predictive performance of RWHMDA.

## MATERIALS AND METHODS

### Human Microbe-Disease Associations

In this study, we utilized microbe-disease associations in HMDAD database (Ma et al., 2017)<sup>1</sup>, containing 483 known microbe-disease associations among 292 microbes inhabiting the human body and 39 human diseases. The associations in HMDAD were obtained from sequencing-based microbiological analyses. In addition, if different data are available for overlapping microbe-disease associations in the database, only one record would be maintained. Finally, we obtained 450 distinct known microbe-disease associations for further prediction. Microbe-disease associations could be stored in an adjacency matrix  $A$ , where element  $A(i, j)$  represented the binary association of disease  $d(i)$  and microbe  $m(j)$ . In other words, we obtained a  $nd \times nm$  matrix  $A$ , where 450 elements were 1 and the others were 0.

<sup>1</sup><http://www.cuilab.cn/hmdad>

Meanwhile,  $nd$  was the number of diseases, and  $nm$  was the number of microbes.

### Gaussian Interaction Profile Kernel Similarity for Microbes

Gaussian interaction profile kernel similarity was calculated on the basis of a type of Radial Basis Function (RBF), namely Gaussian kernel function. In this study, we adopted the Gaussian interaction profile kernel similarity to determine the similarity between microbes. In detail, based on the constructed adjacency matrix  $A$ , microbial interaction profiles could be defined as a binary vector  $IP(m(j))$ , representing the absence or presence of the interaction between microbe  $m(j)$  and diseases.  $IP(m(j))$  was the  $j$ -th column of matrix  $A$ . Thereafter, we calculated the Gaussian kernel similarity between microbe  $m(j)$  and microbe  $m(i)$ , using Gaussian kernel function as follows:

$$GM(m(i), m(j)) = \exp(-r_m \|IP(m(i)) - IP(m(j))\|^2) \quad (1)$$

where  $r_m$  was set to balance the kernel bandwidth, and  $GM$  defined the Gaussian interaction profile kernel similarity matrix for microbes. Specially,  $r_m$  was calculated in accordance with a new parameter  $r_m$  and the average known association number per miRNA as follows:

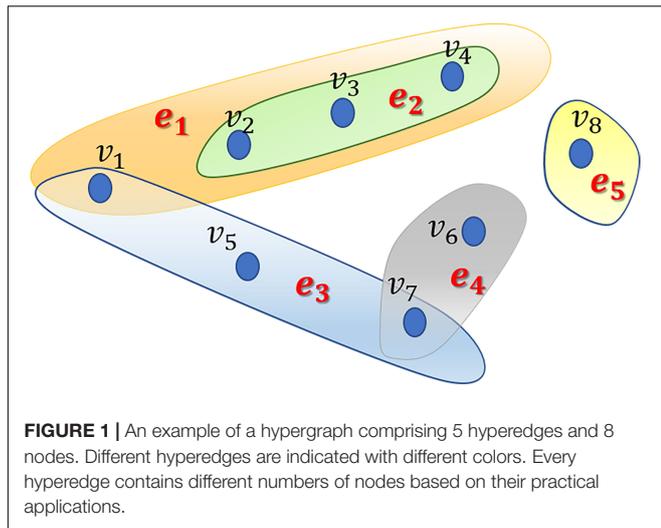
$$r_m = \frac{r'_m}{\left(\frac{1}{nm} \sum_{i=1}^{nm} \|IP(m(i))\|^2\right)} \quad (2)$$

where  $nm$  is the total number of microbes. Technically,  $r_m$  was set as 1 here (Chen et al., 2017b).

### RWHMDA

In this study, we proposed the RWHMDA model from the random walk on hypergraph to predict novel microbe-disease associations. Although Gaussian interaction profile kernel similarity for microbes is also accounted for in this method, RWHMDA is still a graph structure-based model without extra domain information in microbiological studies. Random walks on simple graphs have been investigated extensively in various biological fields. However, random walks on hypergraph have not been reported with respect to the prediction of microbe-disease associations thus far. Hypergraph is a type of higher-order graphical representation of biological data, compensating for information loss in the normal graph method, exclusively describing pair-wise association structures (Figure 1).

Generally, in the present model, we first constructed a hypergraph comprising microbes and diseases, wherein diseases are presented as hyperedges and microbes are presented as nodes. If several microbes have been confirmed to be associated with one disease, they would be presented as nodes in the hyperedge corresponding to the disease. In the hypergraph, hyperedges can join numerous vertices (not limited to two nodes as in simple graph). Specifically, if microbe  $m(i)$  is associated with disease  $d(j)$ , then node  $m(i)$  belongs to hyperedge  $d(j)$ . Obviously, one microbe might belong to different hyperedges. We assessed all known microbe-disease associations and established a hypergraph with 39 hyperedges. Without loss of generalizability,



we defined the hypergraph as  $HG(V, \mathcal{E})$  where  $\mathcal{E}$  is the set of hyperedges, and  $V$  is the set of vertices. Hyperedge  $e \in \mathcal{E}$  is a subset of  $V$ , hyperedge  $e$  is incident with node  $v$  if the node belongs to the hyperedge. Neighborhood relationships among nodes can be defined if  $v \in e, w \in e$ .

After constructing the hypergraph, we implement random walk with restart on it. The random walk on the normal graph is a type of Markov process. The surfer travels between nodes in the graph by starting at a node and shifting to an adjacent node at each discrete time step  $t$ . The transition probability between nodes is completely independent of the time  $t$ . Therefore, we could define the transition probability matrix  $P \in R^{|V| \times |V|}$  for the whole process. Matrix  $P$  represents the transition probabilities of the random internodal movements. Matrix  $P$  is actually a critical factor calculated on the basis of multifarious filed knowledge. Furthermore, we introduced the random walk on the hypergraph. Basically, the surfer shifts between two nodes only if they are neighbors in the currently visited hyperedge. Briefly, this process may be considered a two-step procedure as follows: the surfer randomly selects a hyperedge incident with a currently visited node in step 1; thereafter, the surfer selects a destination neighbor node within the selected hyperedge in step 2 (Figure 2). Thereafter, we would focus on capturing transition matrix  $P$  with respect to the random walk on the hypergraph.

Considering an unweighted hypergraph  $HG(V, \mathcal{E})$ , wherein hyperedges and nodes have no weights, the incidence matrix  $H \in R^{|V| \times |E|}$  was defined as follows:

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (3)$$

$$\delta(e) = |e| \quad (4)$$

$$d(v) = |\mathcal{E}(v)| \quad (5)$$

where  $\delta(e)$  is the degree of hyperedge  $e$ ,  $d(v)$  is the degree of vertex  $v$ ,  $|e|$  indicates the number of nodes within hyperedge  $e$ ,  $\mathcal{E}(v)$  is the set of hyperedges incident with vertex  $v$ . Thereafter,

we obtained the diagonal hyperedge degree matrix  $D_e \in R^{|E| \times |E|}$ , the diagonal vertex degree matrix  $D_v \in R^{|V| \times |V|}$ .

Regarding data on microbe-disease associations, it means the surfer would select a disease known to be associated with the current microbe. We could not unambiguously distinguish the more critical disease associated with the referenced microbe. Therefore, we intend for the surfer to uniformly randomly select a hyperedge at step 1. Furthermore, the surfer would walk to a node within this hyperedge. In our predictive case, although it is potentially difficult to evaluate the features of nodes, we differentiated microbes within a hyperedge of disease in accordance with the Gaussian interaction profile kernel similarity. Technically, in step 2, we intend for the surfer to shift to a node within a hyperedge in accordance with the sum of similarities of the node with all other nodes in the hypergraph. In summary, starting from node  $u$ , the surfer would select hyperedge  $e$  incident with  $u$  proportional to the weight of hyperedge  $w(e)$ . Thereafter, the surfer selects node  $v$  proportional to the weight of  $v$  within the current hyperedge  $e$ , namely  $w(v_e)$ .

Considering the afore-mentioned motivation, we then defined the weighted incident matrix  $W \in R^{|V| \times |E|}$  of hypergraph  $HG(V, \mathcal{E})$  as follows:

$$w(v, e) = \begin{cases} w(v_e), & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (6)$$

where  $w(v_e)$  is the weight of node  $v$  in hyperedge  $e$ . In the present model, we calculated  $w(v_e)$  on the basis of matrix  $GM$ . In this study, the weight of a microbe  $m(i)$  in a hyperedge is the sum of  $i$ th row in  $GM$ . Thereafter, we redefined hyperedge degree  $\delta'(e)$  and hyperedge degree matrix  $D_{ve} \in R^{|E| \times |E|}$  as follows:

$$\delta'(e) = \sum_{v \in e} w(v, e) \quad (7)$$

where  $D_{ve}$  is the diagonal hyperedge degree matrix with element  $\delta'(e)$ .

We then calculated the transition probability from vertex  $u$  to vertex  $v$  as follows:

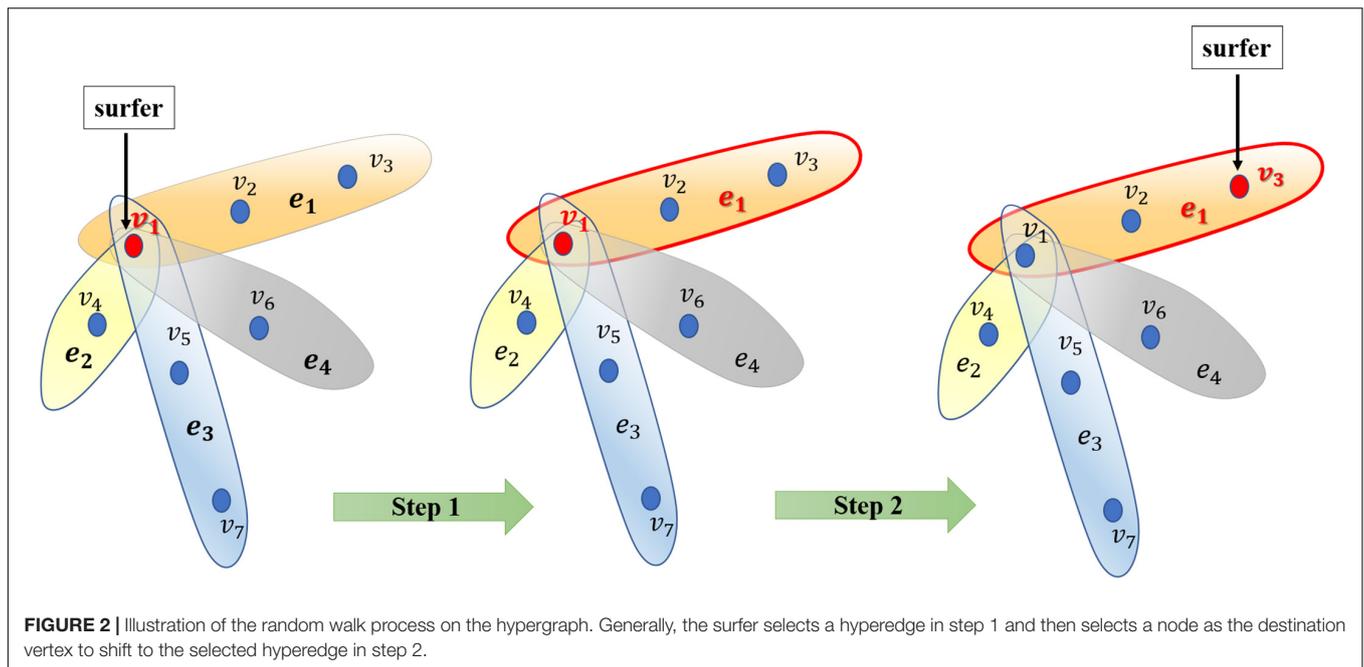
$$P(u, v) = \sum_{e \in E} w(e) \frac{h(u, e)}{\sum_{\hat{e} \in \mathcal{E}} w(\hat{e})} \frac{w(v, e)}{\sum_{\hat{v} \in e} w(\hat{v}, e)} \quad (8)$$

which may also be expressed in matrix form as follows:

$$P = D_v^{-1} H W_e D_{ve}^{-1} W^T \quad (9)$$

where  $W_e \in R^{|E| \times |E|}$  is the diagonal matrix of hyperedge weights, wherein all diseases are considered with equal weightage in accordance with the previously described practical consideration, i.e., all hyperedges are  $1/|E|$ . Naturally, transition matrix  $P$  is stochastic, implying that the sum of every row equals 1.

Furthermore, we implement the random walk with restart on the hypergraph. In particular, assuming that microbes associated with disease  $d(i)$  are to be predicted, all microbes with known associations with  $d(i)$  are considered seed microbes, while the others are considered candidate microbes. Thereafter, we set the initially normalized probability vector  $\vec{v}(0) \in R^{|V| \times 1}$  such that seed microbes are assigned with equal probability and the



non-seed miRNAs are zero. After the first step,  $\vec{v}(1) = P^T \vec{v}(0)$ . Moreover, we set the restart probability at every step at source nodes as  $\alpha$  ( $0 < \alpha < 1$ ),  $\vec{v}(1) = (1 - \alpha) P^T \vec{v}(0) + \alpha \vec{v}(0)$ . Finally, we obtained the random walk with the following formula:

$$\vec{v}(t+1) = (1 - \alpha) P^T \vec{v}(t) + \alpha \vec{v}(0) \quad (10)$$

$\vec{v}(t)$  is defined such that the  $i$ th element means the probability of moving to node  $i$  at step  $t$ . After some steps, the random walk would stabilize, implying that the difference between  $\vec{v}(t+1)$  and  $\vec{v}(t)$  measured by the L1 norm is smaller than the provided threshold. The stable state of the random walk with restart is defined as  $\vec{v}(\infty)$ . Stationary probability in  $\vec{v}(\infty)$  indicates the probable associations between candidate microbes with the currently investigated disease. We conducted the random walk for every disease in the HMDAD database and ranked the underlying microbe-disease associations in accordance with the corresponding  $\vec{v}(\infty)$  of the current disease (Figure 3).

As a supplement, we set the  $\alpha$ -value as 0.2 and set the cutoff value as  $10^{-6}$ .

## RESULTS

### Performance Evaluation

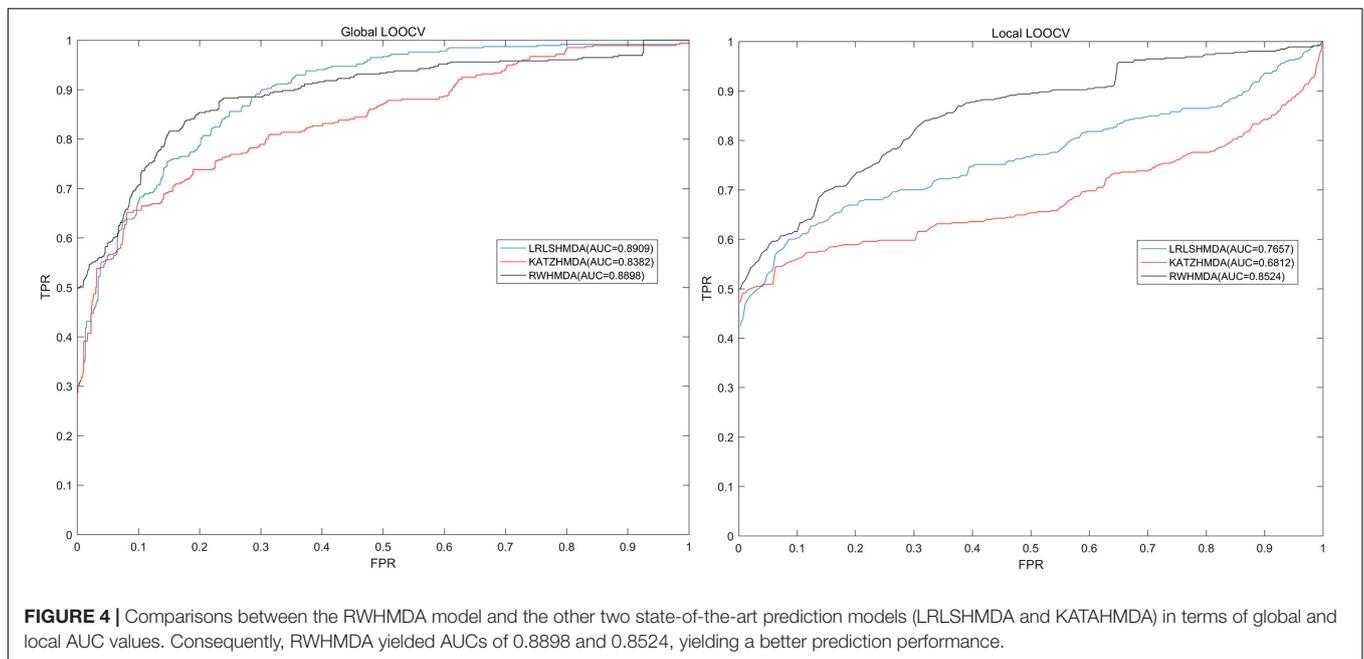
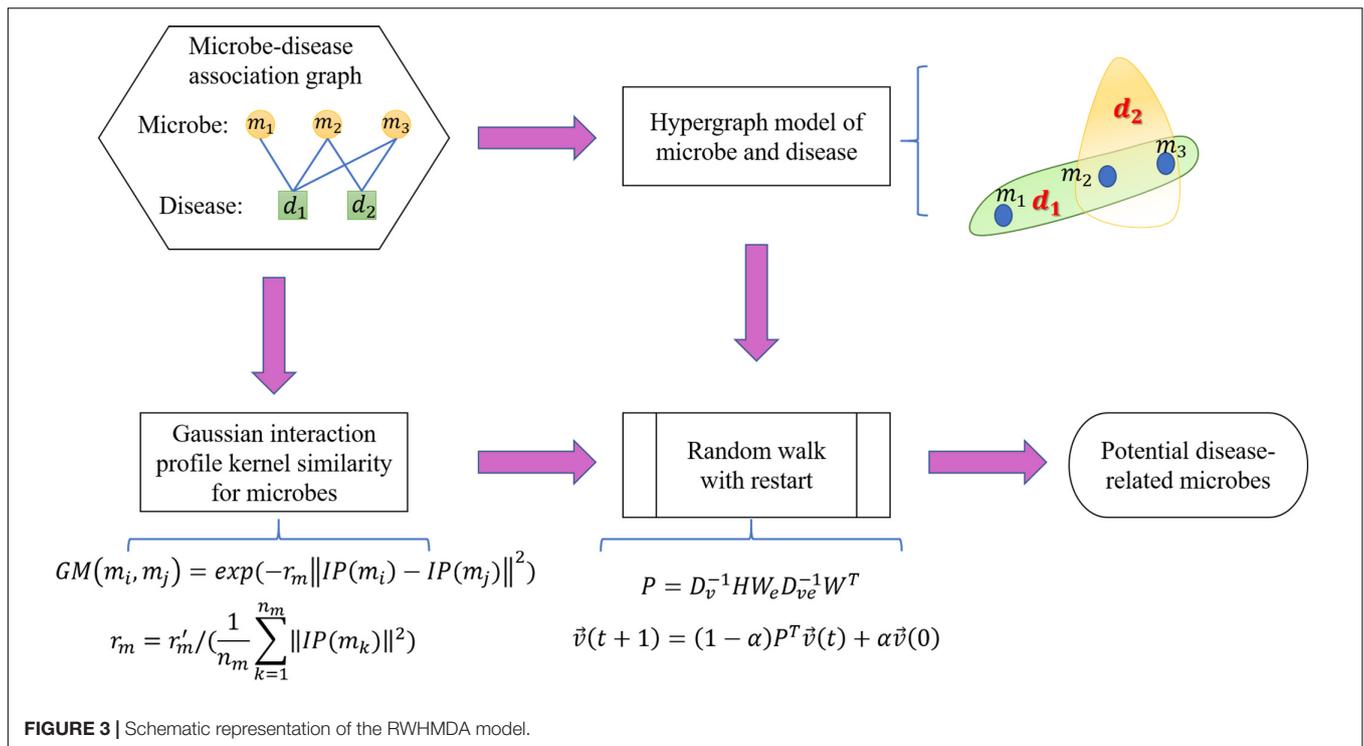
LOOCV was usually implemented to assess the performance of the prediction model. Global and local LOOCV in the present study were both conducted to comprehensively assess the performance of RWHMDA. Specifically, global LOOCV was conducted on the basis of the known microbe-disease associations in the HMDAD database (Ma et al., 2017). Each association was left out in turn as the test sample, while others were set as candidate samples. If the rank of the test sample was higher than that of the candidate samples, the test association

was considered to have been correctly predicted. Furthermore, local LOOCV was somewhat different from global LOOCV, and it was implemented as follows: first, for an investigated disease, based on the association records in the HMDAD (Ma et al., 2017) database, each known disease-associated microbe was excluded in turn as the test sample and the others were used as seed samples. Thereafter, the predicted association probability of the current test sample would be ranked with the probability of candidate samples. If the test sample was ranked beyond the threshold, the model successfully predicted this microbe-disease association. Further, we plotted a receiver operating characteristics (ROC) curve. The area under the ROC curve (AUC) was determined to assess the prediction performance of RWHMDA. Specifically,  $AUC = 1$  implied an excellent performance, and  $AUC = 0.5$  indicated a random performance. Consequently, RWHMDA yielded a global AUC value of 0.8898 and local AUC value of 0.8524, which were higher than some previously reported computational models, such as LRLSHMDA (0.8959, 0.7657) (Bao et al., 2017) and KATZHMDA (0.8382, 0.6812) (Chen et al., 2017a; Figure 4).

### Case Studies

To further assess the performance of the proposed model, we conducted case studies of asthma, Crohn's disease (CD), and type 2 diabetes by assessing the 10 highest probable microbes ranked by RWHMDA.

It is unambiguous that the human microflora play an important role in asthma pathogenesis (Li N. et al., 2017). Morbidity rates among asthma patients have significantly increased since the 1960s (Anandan et al., 2010). Asthma caused approximately 400 thousand deaths worldwide in 2015. More recently, on evaluating data regarding the association between *Helicobacter pylori* status with the history of asthma from 7663



adults in the Third National Health and Nutrition Examination Survey, childhood acquisition of *H. pylori* is associated with a reduced risk of asthma (Chen and Blaser, 2007). We implemented RWHMDA for the asthma case study. Consequently, 9 of the 10 most highly ranked asthma-related microbes were confirmed from the literature (Table 1). For example, the study reporting the presence of *Propionibacterium acnes* (1st ranked in the prediction list of RWHMDA) in asthma patients helped diagnose asthma

(Romero-Espinoza et al., 2018). *Pseudomonas*, ranked 3rd by our model, was confirmed to be more prevalent in the sputum of asthma patients (Jung et al., 2016). Moreover, as the 10th predicted asthma-related microbe, *Streptococcus* are associated with asthma, potentially contributing to its pathophysiology (Zhang et al., 2016).

The worldwide prevalence of diabetes mellitus has increased continuously over the past few decades

**TABLE 1** | A case study on predicted potential asthma-related microbes.

Rank	Microbe	Evidence
1	<i>Propionibacterium acnes</i>	PMID:29447223
2	<i>Propionibacterium</i>	PMID:29447223
3	<i>Pseudomonas</i>	PMID:27433177
4	<i>Burkholderia</i>	PMID:24451910
5	<i>Enterobacter aerogenes</i>	PMID:18790035
6	<i>Enterobacter hormaechei</i>	Unconfirmed
7	<i>Klebsiella pneumoniae</i>	PMID:26220531
8	<i>Shigella dysenteriae</i>	Unconfirmed
9	<i>Actinobacteria</i>	PMID:23265859
10	<i>Streptococcus</i>	PMID:27078029

Among the 10 highest ranked potential asthma-related microbes, eight were confirmed from the literature.

(Tadic and Cuspidi, 2015). Type 2 diabetes mellitus is a subclass of diabetes mellitus, accounting for approximately 90% of all the diabetes mellitus cases. The traditional view holds that the pathogenesis of type 2 diabetes is associated with both genetic and lifestyle-related factors. Recent evidence suggests that the pathomechanism and pathogenesis of type 2 diabetes mellitus are also associated with an unbalance in microbial communities (Ripsin et al., 2009; Furet et al., 2010). Larsen et al. (2010) assessed the differences in the composition of the intestinal microbiota in individuals without and those with type 2 diabetes via high-throughput 16S rDNA gene pyrosequencing, reporting an increase in *Bacilli*, *Bacteroidetes*, and *Betaproteobacteria* and reductions in *Clostridia*, *Clostridium*, *Firmicutes*, etc. Among the 10 highest ranked microbes by probability, 8 were confirmed through recent evidence (Table 2). For example, *Fusobacterium nucleatum* was ranked first and confirmed to be significantly higher in type 2 diabetes mellitus patients than in those without type 2 diabetes mellitus (Miranda et al., 2017). *Pseudomonas*, abundant in the subgingival plaque, ranked 2nd by our model and was markedly different between individuals with and those without diabetes (Zhou et al., 2013). Furthermore, *Aerococcus* and *Atopobium* were

**TABLE 2** | RWHMDA used to predict candidate microbes associated with type 2 diabetes.

Rank	Microbe	Evidence
1	<i>Fusobacterium nucleatum</i>	PMID:28198980
2	<i>Pseudomonas</i>	PMID:23613868
3	<i>Aerococcus</i>	PMID:28786059
4	<i>Atopobium</i>	PMID:28177125
5	<i>Atopobium vaginae</i>	unconfirmed
6	<i>Candidate division TM7</i>	unconfirmed
7	<i>Eggerthella</i>	PMID:26046242
8	<i>Gardnerella</i>	PMID:28316574
9	<i>Gardnerella vaginalis</i>	PMID:2131794
10	<i>Lactobacillus crispatus</i>	PMID:28608654

Consequently, 8 of the 10 most probable microbes were experimentally confirmed through the relevant literature.

associated with the risk of type 2 diabetes (Li H. et al., 2017; Long et al., 2017).

Crohn’s disease (CD) is a type of IBD. Although the etiology of CD is generally believed to be associated with the combination of immune, environmental, and bacterial factors, however, the precise etiology of CD is still unclear (Dessein et al., 2008; Stefanelli et al., 2008; Cho and Brant, 2011). In fact, no surgical treatment or pharmacotherapeutic methods have been reported to cure Crohn’s disease (Baumgart and Sandborn, 2012). Studies have increasingly investigated the bacterial factors associated with the etiology of CD. Gevers et al. (2014) reported that the increased abundance of *Fusobacteriaceae*, *Enterobacteriaceae*, *Pasteurellaceae*, and *Veillonellaceae* and the decreased abundance of *Clostridiales Erysipelotrichales*, and *Bacteroidales* are closely correlated with Crohn’s disease. A case study on Crohn’s disease revealed that the 10 most probable microbes were confirmed through recent researches (Table 3). For example, the two most promising microbes predicted by our model were *Clostridium difficile* and *Bacteroides fragilis*, both confirmed to be present at high levels in CD patients compared to healthy individuals (Cojocariu et al., 2014; Zhou et al., 2016). Moreover, studies evaluating the association between disease status and gut microbiota in CD patients revealed that *Clostridium coccoides* (3rd place in the ranking list) was abundant in febrile patients presenting with remission in comparison with patients with active CD (Prossberg et al., 2016).

## DISCUSSION AND CONCLUSION

Accumulating falsifiable evidence indicates that microbial involvement is associated with disease pathogenesis in some cases. In this study, with data from microbiological studies, hypergraph theory, and other research areas, we introduced an *in silico* model named RWHMDA to predict underlying microbe-disease associations. Many previous computational models performed pairwise comparisons and illustrated microbe-disease associations as a normal graph. RWHMDA has been developed primarily on the basis of a hypergraph, thus compensating for the information loss issue by normal graph. Known microbe-disease associations in the HMDAD database

**TABLE 3** | A case study on Crohn’s disease verifying all 10 of the 10 most probable candidates of Crohn’s disease-related microbes.

Rank	Microbe	Evidence
1	<i>Clostridium difficile</i>	PMID:25599768
2	<i>Bacteroides fragilis</i>	PMID:27684872
3	<i>Clostridium coccoides</i>	PMID:27687331
4	<i>Bacilli</i>	PMID:29559804
5	<i>Betaproteobacteria</i>	PMID:27833911
6	<i>Lachnospiraceae</i>	PMID:26628508
7	<i>Clostridium</i>	PMID:29722832
8	<i>Prevotella</i>	PMID:28852861
9	<i>Alistipes finegoldii</i>	PMID:28877044
10	<i>Alistipes putredinis</i>	PMID:29311644

and the Gaussian interaction profile kernel similarity for microbes were utilized to design a weighted hypergraph comprising microbes and diseases. Random walk with restart was implemented on the hypergraph for every disease to identify the potential disease-associated microbes. Both cross-validation and case studies on asthma, type 2 diabetes, and Crohn's disease revealed the reliability of RWHMDA. In addition, the predicted microbes for all diseases were publicly released for further validation through biological assays (**Supplementary Table S1**).

Generally, RWHMDA performed reliably, thus revealing several important factors. First, as a representation of a higher-order structure, hypergraphs adequately illustrate and present data on microbe-disease associations without information loss. In particular, the practice of setting disease as a hyperedge and microbe as a node was reasonable and biologically decipherable, thereby naturally benefiting the prediction of potential associations. Second, owing to the valid and updated data on disease-microbe associations through numerous biological analyses, RWHMDA had a greater prediction accuracy with greater probability. Third, random walk process is a widespread and significant physical dynamic process, used extensively in numerous studies. RWHMDA was developed on the basis of the random walk with restart process, following a seemingly iterative 2-step walking strategy to investigate the potential association probability between any pair of microbe and disease.

However, the present RWHMDA model has some limitations. Hypergraph and Gaussian interaction profile kernel were both constructed largely on the basis of known associations. Therefore, the model may have a bias toward those well-known diseases and microbes. Furthermore, some other similarity measures of diseases could also be meticulously integrated into the RWHMDA model, such as symptom-based disease similarity and disease semantic similarity. Finally, the RWHMDA model could not be implemented for new diseases without known

associations with microbes, being an inherent limitation of the graph-based model.

In conclusion, RWHMDA is expected to display promising potential to predict disease-microbe associations for follow-up experimental studies and facilitate the prevention, diagnosis, treatment, and prognosis of complex human diseases.

## AUTHOR CONTRIBUTIONS

Y-WN developed the prediction method, conducted the experiments, analyzed the result, and wrote the manuscript. C-QQ, G-YY, and G-HW conceived the project, analyzed the result, and wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

Y-WN, C-QQ, and G-HW were supported by the National Natural Science Foundation of China under Grant Nos. 11631014 and 11871311. G-YY was supported by the National Natural Science Foundation of China under Grant No. 11631014.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.01578/full#supplementary-material>

**TABLE S1** | We prioritized candidate microbes for all the investigated human diseases in the HMDAD database. The prediction results for each disease were publicly released for further validation. The relatively high ranked disease-microbe associations were anticipated to be confirmed by biological experiments or future clinical observation.

## REFERENCES

- Anandan, C., Nurmatov, U., van Schayck, O. C., and Sheikh, A. (2010). Is the prevalence of asthma declining? Systematic review of epidemiological studies. *Allergy* 65, 152–167. doi: 10.1111/j.1398-9995.2009.02244.x
- Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307:1915. doi: 10.1126/science.1104816
- Bao, W., Jiang, Z., and Huang, D. S. (2017). Novel human microbe-disease association prediction using network consistency projection. *BMC Bioinform.* 18(Suppl. 16):543. doi: 10.1186/s12859-017-1968-2
- Baumgart, D. C., and Sandborn, W. J. (2012). Crohn's disease. *Lancet* 380, 1590–1605. doi: 10.1016/s0140-6736(12)60026-9
- 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. (2015). KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci. Rep.* 5:16840. doi: 10.1038/srep16840
- Chen, X., Huang, Y. A., You, Z. H., Yan, G. Y., and Wang, X. S. (2017a). 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., Niu, Y. W., Wang, G. H., and Yan, G. Y. (2017b). HAMDA: Hybrid approach for MiRNA-disease association prediction. *J. Biomed. Inform.* 76, 50–58. doi: 10.1016/j.jbi.2017.10.014
- Chen, X., Niu, Y. W., Wang, G. H., and Yan, G. Y. (2017c). MKRMDA: multiple kernel learning-based Kronecker regularized least squares for MiRNA-disease association prediction. *J. Transl. Med.* 15:251. doi: 10.1186/s12967-017-1340-3
- Chen, X., Wu, Q. F., and Yan, G. Y. (2017d). RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol.* 14, 952–962. doi: 10.1080/15476286.2017.1312226
- Chen, Y., and Blaser, M. J. (2007). Inverse associations of *Helicobacter pylori* with asthma and allergy. *Arch. Intern. Med.* 167, 821–827. doi: 10.1001/archinte.167.8.821
- Cho, J. H., and Brant, S. R. (2011). Recent insights into the genetics of inflammatory bowel disease. *Gastroenterology* 140, 1704–1712. doi: 10.1053/j.gastro.2011.02.046
- Clarke, G., Stilling, R. M., Kennedy, P. J., Stanton, C., Cryan, J. F., and Dinan, T. G. (2014). Minireview: gut microbiota: the neglected endocrine organ. *Mol. Endocrinol.* 28, 1221–1238. doi: 10.1210/me.2014-1108
- Cojocariu, C., Stanciu, C., Stoica, O., Singeap, A. M., Sfarti, C., Girleanu, I., et al. (2014). Clostridium difficile infection and inflammatory bowel disease. *Turk. J. Gastroenterol.* 25, 603–610. doi: 10.5152/tjg.2014.14054
- Davenport, E. R., Mizrahi-Man, O., Michellini, 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. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820
- Dessein, R., Chamaillard, M., and Danese, S. (2008). Innate immunity in Crohn's disease: the reverse side of the medal. *J. Clin. Gastroenterol.* 42(Suppl. 3 Pt 1), S144–S147. doi: 10.1097/MCG.0b013e3181662c90
- Dethlefsen, L., McFall-Ngai, M., and Relman, D. A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449, 811–818. doi: 10.1038/nature06245
- Donia, M. S., Cimermancic, P., Schulze, C. J., Wieland Brown, L. C., 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
- Eckburg, P. B., Lepp, P. W., and Relman, D. A. (2003). Archaea and their potential role in human disease. *Infect. Immun.* 71, 591–596. doi: 10.1128/iai.71.2.591-596.2003
- Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13780–13785. doi: 10.1073/pnas.0706621104
- Furet, J. P., Kong, L. C., Tap, J., Poitou, C., Basdevant, A., Bouillot, J. L., et al. (2010). Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* 59, 3049–3057. doi: 10.2337/db10-0253
- Gevers, D., Kugathasan, S., Denson, L. A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005
- Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* 5, 82–91. doi: 10.1038/ismej.2010.92
- 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
- Hoppe, B., Grothoff, J. W., Hulton, S. A., Cochat, P., Niaudet, P., Kemper, M. J., et al. (2011). Efficacy and safety of *Oxalobacter formigenes* to reduce urinary oxalate in primary hyperoxaluria. *Nephrol. Dial. Transplant.* 26, 3609–3615. doi: 10.1093/ndt/gfr107
- Huang, Y. A., You, Z. H., Chen, X., Huang, Z. A., Zhang, S., and Yan, G. Y. (2017). Prediction of microbe-disease association from the integration of neighbor and graph with collaborative recommendation model. *J. Transl. 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. (2017). PBHMDA: path-based human microbe-disease association prediction. *Front. Microbiol.* 8:233. doi: 10.3389/fmicb.2017.00233
- Huang, Y. J., Nelson, C. E., Brodie, E. L., Desantis, T. Z., Baek, M. S., Liu, J., et al. (2011). Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J. Allergy Clin. Immunol.* 127, 372.e3–381.e3. doi: 10.1016/j.jaci.2010.10.048
- Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- Jung, J. W., Choi, J. C., Shin, J. W., Kim, J. Y., Park, I. W., Choi, B. W., et al. (2016). Lung microbiome analysis in steroid-naïve, ukrainian asthma patients by using whole sputum. *Tuberc. Respir. Dis.* 79, 165–178. doi: 10.4046/trd.2016.79.3.165
- 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
- Larsen, N., Vogensen, F. K., van den Berg, F. W., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., et al. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* 5:e9085. doi: 10.1371/journal.pone.0009085
- Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444, 1022–1023. doi: 10.1038/4441022a
- Li, H., Qi, T., Huang, Z. S., Ying, Y., Zhang, Y., Wang, B., et al. (2017). Relationship between gut microbiota and type 2 diabetic erectile dysfunction in Sprague-Dawley rats. *J. Huazhong Univ. Sci. Technol. Med. Sci.* 37, 523–530. doi: 10.1007/s11596-017-1767-z
- Li, N., Qiu, R., Yang, Z., Li, J., Chung, K. F., Zhong, N., et al. (2017). Sputum microbiota in severe asthma patients: relationship to eosinophilic inflammation. *Respir. Med.* 131, 192–198. doi: 10.1016/j.rmed.2017.08.016
- Long, J., Cai, Q., Steinwandel, M., Hargreaves, M. K., Bordenstein, S. R., Blot, W. J., et al. (2017). Association of oral microbiome with type 2 diabetes risk. *J. Periodontol. Res.* 52, 636–643. doi: 10.1111/jre.12432
- Ma, J., Prince, A. L., Bader, D., Hu, M., Ganu, R., Baquero, K., et al. (2014). High-fat maternal diet during pregnancy persistently alters the offspring microbiome in a primate model. *Nat. Commun.* 5:3889. doi: 10.1038/ncomms4889
- Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2017). An analysis of human microbe-disease associations. *Brief. Bioinform.* 18, 85–97. doi: 10.1093/bib/bbw005
- Mason, M. R., Preshaw, P. M., Nagaraja, H. N., Dabdoub, S. M., Rahman, A., and Kumar, P. S. (2015). The subgingival microbiome of clinically healthy current and never smokers. *ISME J.* 9, 268–272. doi: 10.1038/ismej.2014.114
- Miranda, T. S., Feres, M., Retamal-Valdes, B., Perez-Chaparro, P. J., Maciel, S. S., and Duarte, P. M. (2017). Influence of glycemic control on the levels of subgingival periodontal pathogens in patients with generalized chronic periodontitis and type 2 diabetes. *J. Appl. Oral Sci.* 25, 82–89. doi: 10.1590/1678-77572016-0302
- Moore, W. E., and Moore, L. H. (1995). Intestinal floras of populations that have a high risk of colon cancer. *Appl. Environ. Microbiol.* 61, 3202–3207.
- Neish, A. S. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology* 136, 65–80. doi: 10.1053/j.gastro.2008.10.080
- Peng, L. H., Yin, J., Zhou, L., Liu, M. X., and Zhao, Y. (2018). Human microbe-disease association prediction based on adaptive boosting. *Front. Microbiol.* 9:2440. doi: 10.3389/fmicb.2018.02440
- Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., et al. (2009). The NIH human microbiome project. *Genome Res.* 19, 2317–2323. doi: 10.1101/gr.096651.109
- Proberg, M., Bendtsen, F., Vind, I., Petersen, A. M., and Gluud, L. L. (2016). The association between the gut microbiota and the inflammatory bowel disease activity: a systematic review and meta-analysis. *Scand. J. Gastroenterol.* 51, 1407–1415. doi: 10.1080/00365521.2016.1216587
- Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513, 59–64. doi: 10.1038/nature13568
- Qu, J., Zhao, Y., and Yin, J. (2019). Identification and analysis of human microbe-disease associations by matrix decomposition and label propagation. *Front. Microbiol.* 10:291. doi: 10.3389/fmicb.2019.00291
- Ripsin, C. M., Kang, H., and Urban, R. J. (2009). Management of blood glucose in type 2 diabetes mellitus. *Am. Fam. Physician* 79, 29–36.
- Romero-Espinoza, J. A., Moreno-Valencia, Y., Coronel-Tellez, R. H., Castillejos-Lopez, M., Hernandez, A., Dominguez, A., et al. (2018). Virome and bacteriome characterization of children with pneumonia and asthma in Mexico City during winter seasons 2014 and 2015. *PLoS One* 13:e0192878. doi: 10.1371/journal.pone.0192878
- Rosner, J. L. (2014). Ten times more microbial cells than body cells in humans? *Microbe* 9:47. doi: 10.1128/microbe.9.47.2
- Schwabe, R. F., and Jobin, C. (2013). The microbiome and cancer. *Nat. Rev. Cancer* 13, 800–812. doi: 10.1038/nrc3610
- Sobhani, I., Tap, J., Roudot-Thoraval, F., Roperch, J. P., Letulle, S., Langella, P., et al. (2011). Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One* 6:e16393. doi: 10.1371/journal.pone.0016393
- Sommer, F., and Backhed, F. (2013). The gut microbiota—masters of host development and physiology. *Nat. Rev. Microbiol.* 11, 227–238. doi: 10.1038/nrmicro2974
- Stefanelli, T., Malesci, A., Repici, A., Vetrano, S., and Danese, S. (2008). New insights into inflammatory bowel disease pathophysiology: paving the way for novel therapeutic targets. *Curr. Drug Targets* 9, 413–418. doi: 10.2174/138945008784221170

- Tadic, M., and Cuspidi, C. (2015). Type 2 diabetes mellitus and atrial fibrillation: from mechanisms to clinical practice. *Arch. Cardiovasc. Dis.* 108, 269–276. doi: 10.1016/j.acvd.2015.01.009
- Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi: 10.1038/nature07540
- van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500
- Ventura, M., O'Flaherty, S., Claesson, M. J., Turrioni, F., Klaenhammer, T. R., van Sinderen, D., et al. (2009). Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* 7, 61–71. doi: 10.1038/nrmicro2047
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241
- Wang, E., Zaman, N., McGee, S., Milanese, J. S., Masoudi-Nejad, A., and O'Connor-McCourt, M. (2015). Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.* 30, 4–12. doi: 10.1016/j.semcancer.2014.04.002
- 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
- Yamanishi, Y. (2013). Chemogenomic approaches to infer drug-target interaction networks. *Methods Mol. Biol.* 939, 97–113. doi: 10.1007/978-1-62703-107-3\_9
- Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., et al. (2009). Human gut microbiota in obesity and after gastric bypass. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2365–2370. doi: 10.1073/pnas.0812600106
- 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
- Zhou, M., Rong, R., Munro, D., Zhu, C., Gao, X., Zhang, Q., et al. (2013). Investigation of the effect of type 2 diabetes mellitus on subgingival plaque microbiota by high-throughput 16S rDNA pyrosequencing. *PLoS One* 8:e61516. doi: 10.1371/journal.pone.0061516
- Zhou, Y., Chen, H., He, H., Du, Y., Hu, J., Li, Y., et al. (2016). Increased *Enterococcus faecalis* infection is associated with clinically active Crohn disease. *Medicine* 95:e5019. doi: 10.1097/md.0000000000005019
- Zou, S., Zhang, J., and Zhang, Z. (2018). Novel human microbe-disease associations inference based on network consistency projection. *Sci. Rep.* 8:8034. doi: 10.1038/s41598-018-26448-8

**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.

Copyright © 2019 Niu, Qu, Wang and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.