Methods 124 (2017) 69–77

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

A novel method of predicting microRNA-disease associations based on microRNA, disease, gene and environment factor networks

Wei Peng [a,*], Wei Lan [b], Jiancheng Zhong [c], Jianxin Wang [b], Yi Pan [d]

[a] Computer Center of Kunming University of Science and Technology, Kunming, Yunnan 650093, PR China
[b] School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, PR China
[c] College of Engineering and Design, Hunan Normal University, Changsha 410081, PR China
[d] Department of Computer Science, Georgia State University, Atlanta, GA 30302-4110, USA

ARTICLE INFO

Article history:
Received 30 March 2017
Received in revised form 2 May 2017
Accepted 28 May 2017
Available online 31 May 2017

Keywords:
MicroRNA
Disease
Environment factor
Gene
Biological network
Association
Random walk model

ABSTRACT

MicroRNAs have been reported to have close relationship with diseases due to their deregulation of the expression of target mRNAs. Detecting disease-related microRNAs is helpful for disease therapies. With the development of high throughput experimental techniques, a large number of microRNAs have been sequenced. However, it is still a big challenge to identify which microRNAs are related to diseases. Recently, researchers are interesting in combining multiple-biological information to identify the associations between microRNAs and diseases. In this work, we have proposed a novel method to predict the microRNA-disease associations based on four biological properties. They are microRNA, disease, gene and environment factor. Compared with previous methods, our method makes predictions not only by using the prior knowledge of associations among microRNAs, disease, environment factors and genes, but also by using the internal relationship among these biological properties. We constructed four biological networks based on the similarity of microRNAs, diseases, environment factors and genes, respectively. Then random walking was implemented on the four networks unequally. In the walking course, the associations can be inferred from the neighbors in the same networks. Meanwhile the association information can be transferred from one network to another. The results of experiment showed that our method achieved better prediction performance than other existing state-of-the-art methods.

1. Introduction

MicroRNAs (MiRNA) are an important class of single-stranded, non-coding RNA molecules about 21 nucleotides in length. They bind the 3′UTR of the target mRNAs and down regulate the mRNAs' expression by imperfectly base-paring with complementary sequence within mRNAs [1]. MicroRNAs can suppress the translation of target mRNAs, stimulate their degradation or induce their cleavage and prevent the protein product [2]. It is estimated that over half of protein coding-genes in human and other animals are targeted by microRNAs. MicroRNAs have been shown to be crucial in almost every biological process, including cell differentiation and growth, mobility and apoptosis [3]. Therefore, their deregulation appears to associate with various diseases, ranging from cancers to common diseases [4–6], i.e. cardiovascular diseases, schizophrenia. Studying alteration of expression profiles of disease-related microRNAs is helpful for disease therapies [7,8]. Recently, microRNAs are expected to become new generation of drugs [9]. With the development of high throughput experimental techniques, a large number of microRNAs have been sequenced. However, it is a still big challenge to experimentally identify which microRNAs are related to diseases due to high cost and time-consuming. Recently, some researchers have been focusing on designing computational methods to predict the associations between microRNAs and diseases.

Most of these computational methods predicted new microRNA-disease associations by referring to the known ones according to their similarities [10]. One of intuitive way is to construct microRNA similarity network where nodes are microRNAs and edges denote the functional similarities between microRNAs. The potential associations between microRNAs and diseases can be inferred according to the information of their functional similar microRNAs. This type of method is under the observation that the

Abbreviations: MFN, functional similarity network; DSN, disease semantic similarity network; GSN, gene-gene functional similarity network; ESN, environmental factor chemical structure similarity network.

* Corresponding author.
E-mail address: weipeng1980@gmail.com (W. Peng).

http://dx.doi.org/10.1016/j.ymeth.2017.05.024
1046-2023/

microRNAs exerting similar functions tend to associate with the common or similar diseases [11]. There were various methods to define the functional similarity of microRNAs. One of the ways that construct microRNA functional similarity network is based on disease similarity the microRNA related. After using this way to construct microRNA functional similarity network, Xuan et al. [12] selected the top k most similar neighbors on the network as the candidate microRNAs that were disease related. Considering previous methods only use local network information to infer disease related microRNAs, Chen et al. [13] have implemented a Random walking with Restart method (RWRMDA) on the whole microRNA functional similarity network to predict potential associations between microRNAs and diseases. Since microRNAs have the function of suppressing the expression of their target genes [14], microRNAs highly likely share similar functions if they target the same or functionally related genes [15]. Zeng et al. [16] have constructed a microRNA similarity network based on how significant that the two microRNAs share target genes. They also constructed other two types of microRNA similarity networks. The one was on the basis of the microRNA functional synergistic interactions collected from literatures and the other network was constructed by combing the former two networks. After that, they implemented Prince [17], PageRank with prior [18] and K-step Markov [18] on the three networks respectively to assign a score to each microRNA in the networks for interested diseases. Chen et al. [19] have constructed two microRNA similarity networks from two different perspectives. The one was constructed according to the functional similarity of microRNAs. The other was constructed according to the phenotype similarity of the microRNA associated diseases. After that, the Pearson correlation scores between the two types of similarities were used to infer the associations between microRNAs and diseases.

The methods mentioned above were designed based on microRNA functional similarity networks which were constructed by using gene or disease or microRNA information. Some computational methods have been introduced based on target gene similarity network, where nodes denote genes and edges denote the functional similarities between genes. MicroRNAs suppressing the expression of their target genes and the dysfunctions of genes usually cause diseases. Some researchers detected microRNA-disease associations by comparing the similarity between microRNA target genes and disease related genes. Jiang et al. [15] have constructed a gene-gene function similarity network by integrating gene expression profiles, gene function information and protein domain interactions. After that they calculated a score for each interested microRNA-disease association according to the functional associations between the known disease genes and the microRNA target genes. The associations with high scores were regarded as candidate microRNA-disease associations. Shi et al. [20] have mapped microRNA targeted genes and disease related genes to protein–protein interaction (PPI) network respectively and constructed two gene similarity networks. They ranked the genes in the two networks separately by random walking with restart algorithm with different seeds. Then the p-value was adopted to measure the significant that the microRNA is associated with a disease.

Previous methods [12,13] employed the similarity of disease to weight the edges of microRNA similarity networks. Recently, some methods constructed microRNA similarity network and disease similarity network separately. Disease similarity network can be constructed according to the phenotype similarity or semantic similarity of diseases or the functional similarity of the genes the disease related. They thought similar microRNAs tend to associate with common diseases. On the other hand, similar diseases are highly likely caused by common microRNAs. On the basis of the two networks and some known microRNA-disease associations, some potential microRNA-disease associations were detected. Chen's group [21] have constructed two networks, microRNA functional similarity network and disease semantic similarity network. The potential microRNA-disease associations were detected by implementing a semi-supervised and global method, namely RLSMDA on the two networks simultaneously. The advantage of this method is that it can infer information from similar microRNAs or diseases and can work for diseases without known related microRNAs. The methods that construct microRNA similarity network and disease similarity network are various, i.e. microRNA functional similarity, microRNA sequence similarity, disease phenotype similarity, disease functional similarity, disease semantic similarity and so on. [16,10] have discussed the methods that calculated the similarity of microRNA or disease to predict the associations between microRNAs and diseases. Lan et al. [22] have developed a kernelized Bayesian matrix factorization (KBMFMDI) method that combined microRNA functional similarity, microRNA sequence similarity, disease functional similarity and disease semantic similarity to predict the microRNA-disease associations.

To further improve the prediction performance, more biological information should be introduced and more effective methods should be developed. Accumulated evidences show that microRNA expression can be influenced by environment factors (EF) [23–25], such as drug, alcohol, diet, stress, etc. Some environment factors may also cause diseases [16]. Ha et al. [26] have combined EF data to predict microRNA-disease associations. Their method was implemented on a microRNA functional similarity network which was constructed according to whether or not two microRNAs share common Environment Factors. However, their method ignored the internal relationship between environment factors and the intra-relationship between diseases and EFs. In fact, an EF-EF similarity network can be constructed according to the chemical structure similarity between EFs. The similar EFs not only usually regulate similar microRNAs but also cause similar diseases. Based on this observation, in our previous work, we have designed a method namely ThrRWMDE [27] to detect potential associations between microRNAs and diseases. ThrRWMDE constructed three different types of similarity networks according to microRNA similarity, disease function similarity and environmental factor similarity respectively, and implemented an unbalanced three random walking algorithm [28–30] on the three networks. ThrRWMDE was the first method for microRNA-disease association prediction based on three networks, which not only can flexibly infer information from different levels of neighbors in corresponding networks with respect to the topological and structural differences of the three networks, but also in the walking course, the association information within a network will be transferred among networks according to the associations between the nodes in different networks. ThrRWMDE constructed disease similarity network based on functional similarity of disease related genes, which made use of the gene-gene associations and gene-disease associations. However, it limits to ignore the relationship between genes and microRNAs.

As we all know, there are complex inter- and intra-relationships among microRNAs, diseases, genes and EFs. As Fig. 1 shown, 1) Functionally similar microRNAs tend to associate with common or similar diseases, environment factors and genes [19]. 2) The diseases with similar phenotypes are likely caused by common or similar microRNAs, genes and environment factors [24]. 3) Functionally similar genes are suppressed by common or similar microRNAs and cause common or similar diseases with high probability [11,31]. 4) Similar environment factors usually affect common or similar microRNAs and diseases [25].

In this work, we designed a novel method, namely UFourRW to comprehensively make used of the intra- and inter-relationship among microRNAs, diseases, genes and environment factors. Firstly