Journal of Biomedical Informatics 80 (2018) 26–36

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Human disease MiRNA inference by combining target information based on heterogeneous manifolds

Pingjian Ding[a], Jiawei Luo[a,*], Cheng Liang[b], Qiu Xiao[a], Buwen Cao[a]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410083, China
[b] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

ARTICLE INFO

Keywords:
miRNA-disease associations
Heterogeneousnetwork
Manifold learning
Graph regularization
Target information

ABSTRACT

The emergence of network medicine has provided great insight into the identification of disease-related molecules, which could help with the development of personalized medicine. However, the state-of-the-art methods could neither simultaneously consider target information and the known miRNA-disease associations nor effectively explore novel gene-disease associations as a by-product during the process of inferring disease-related miRNAs. Computational methods incorporating multiple sources of information offer more opportunities to infer disease-related molecules, including miRNAs and genes in heterogeneous networks at a system level. In this study, we developed a novel algorithm, named inference of Disease-related MiRNAs based on Heterogeneous Manifold (DMHM), to accurately and efficiently identify miRNA-disease associations by integrating multi-omics data. Graph-based regularization was utilized to obtain a smooth function on the data manifold, which constitutes the main principle of DMHM. The novelty of this framework lies in the relatedness between diseases and miRNAs, which are measured via heterogeneous manifolds on heterogeneous networks integrating target information. To demonstrate the effectiveness of DMHM, we conducted comprehensive experiments based on HMDD datasets and compared DMHM with six state-of-the-art methods. Experimental results indicated that DMHM significantly outperformed the other six methods under fivefold cross validation and de novo prediction tests. Case studies have further confirmed the practical usefulness of DMHM.

1. Introduction

The expression of messenger RNAs is suppressed in a sequence-specific manner by microRNAs that consist of small endogenous non-coding RNAs [1,2]. A mounting number of studies have indicated that miRNAs are important components in the cell. Specifically, studies have proved the vital roles of miRNAs in various biological processes, such as cell growth [3], cell development [3], cell cycle regulation [4], cell apoptosis [5], stress responses [6] and tumor invasion [7]. Furthermore, the strong relationships between miRNAs and diseases have been verified by numerous biological observations and studies [8,9]. The accumulating knowledge of disease-related miRNAs could be of great help to achieve pathological classifications, individualized diagnoses and disease treatments [10]. However, it still remains challenging for biologists to discover the underlying miRNA-disease associations extensively. Consequently, there is an urgent need to develop powerful computational approaches that could effectively uncover miRNA-disease associations.

With the emergence of large-scale genomic, transcriptomic and proteomic data, network-based methods have recently become popular tools for predicting disease-related miRNAs [11,12]. Archana et al. [13] simultaneously considered motif, sequential and network features to infer immune specific miRNAs from other non-immune miRNAs. However, as we known, it is currently hard or even impossible to obtain negative immune system diseases-related miRNAs. Xu et al. [14] showed aberrant behavior in prostate cancer by analyzing miRNA-target associations for sequences in the constructed miRNA-target dysregulated network (MTDN). Sun et al. [15] determined miRNA-disease associations by using a two-stage filtration procedure (miRPD). Chen et al. [16] inferred different types of miRNA-disease relationships by using a Restricted Boltzmann machine (RBMMMDA). Chen et al. [17] proposed a semi-supervised method for predicting disease-related miRNAs based on a regularized least squares method (RLSMDA). RLSMDA could prioritize all miRNA-disease associations simultaneously. Xuan et al. [18] utilized weighted k nearest neighbors to prioritize miRNA-disease associations (HDMP). Chen et al. [19] predicted disease-related miRNAs using random walking with restart on the miRNA functional similarity network (RWRMDA). Chen et al. [20]

* Corresponding author.
E-mail address: luojiawei@hnu.edu.cn (J. Luo).

https://doi.org/10.1016/j.jbi.2018.02.013
Received 25 September 2017; Received in revised form 11 February 2018; Accepted 21 February 2018
Available online 24 February 2018
1532-0464/

uncovered potential miRNA-disease associations based on the within and between scores for many complex diseases (WBSMDA). Chen et al. [21] developed a hybrid graph-based recommendation approach to reveal disease-related miRNAs, which simultaneously considered the network structure and node attribution (HAMDA). Luo et al. [22] adopted a strategy of collective prediction method to predict potential miRNA-disease associations (CPTL). Chen et al. [23] integrated miRNA functional similarity [24], disease semantic similarity and Gaussian interaction profile kernel similarity into a heterogeneous graph to uncover disease-related miRNAs (HGIMDA). Li et al. [25] designed a matrix computational algorithm that could efficiently update the low-rank miRNA-disease matrix to identify the disease-miRNA associations (MCMDA). In particular, network medicine is conducive to completely understand the molecular complexities of diseases [26,27]. Meanwhile, many studies have established relationships between diseases and genes using network-based methods [28–30].

Many network-based methods for uncovering miRNA-disease associations are mainly based on the following two principles [11,12]: (1) miRNAs are related to complex diseases through the target genes being regulated by them [14,31]; (2) target diseases tend to be involved with miRNAs that are related to similar diseases of the target disease [32–34]. Recently, some related papers were published based on the second principle [35–37]. For most methods based on the second principle, miRNA functional similarity calculated by MISIM (miRNA similarity) [24] based on the known miRNA-disease associations was utilized to prioritize disease-related miRNAs. Although many researchers recognize these two principles, there is still a lack of systematic approaches to automatically integrate the target information into the prediction of disease-related miRNAs by taking these two principles together. In addition, inference of disease-related miRNAs and prediction of gene-disease associations have been considered to be two separate tasks in previous studies. However, it would be more reasonable to make direct inference of disease-miRNA associations by incorporating target information.

In this paper, we propose a novel DMHM algorithm to accurately and efficiently infer disease miRNAs based on a heterogeneous manifold [38–40], which integrates miRNA-disease association inferences and disease-related gene predictions into one unified framework. Compared with previous methods based on a homogeneous miRNA network (Fig. 1(A)), homogeneous gene network (Fig. 1(B)), and a bi-layer heterogeneous network (Fig. 1(C)), the heterogeneous miRNA-gene-disease network in this method involves three different types of nodes (diseases, miRNAs and genes), as well as five types of relationships (Fig. 1(D)). Prior knowledge from existing databases [41–44] was used to construct disease-miRNA relationships, disease-gene relationships, miRNA-gene relationships and gene-gene relationships, while the disease semantic similarity was obtained to construct disease-disease relationships [45]. DMHM takes the first principle into account by utilizing miRNA-target interactions and disease-gene associations. Furthermore, the second principle is considered in the proposed algorithm by combining known miRNA-disease associations with disease similarity (Fig. 2(A)). The inference of disease-related miRNAs or genes could be formulated as a binary classification problem; that is, to decide whether a link exists on this heterogeneous network. A semi-supervised prediction algorithm based on a heterogeneous manifold that propagates information across the heterogeneous network is proposed to uncover the missing edges. The guilt-by-association [46] that has been validated to be effective in many previous studies [32,47,48] was also adopted in our method. As a unique characteristic, DMHM automatically incorporates target information into disease-associated miRNA prediction. Although the main goal of DMHM is to discover disease miRNAs, it is worth noting that new disease-gene relationships could be predicted simultaneously in heterogeneous biological networks as well.

As reported in previous papers [22,33], the performance of the second principle-based methods is always better than that of the first principle-based approach. To validate the performance of DMHM, we compared its performance with that of a homogeneous miRNA model (such as: MIDP [33]) (Fig. 1(B)) and bi-layers heterogeneous model (such as: RLSMDA (Regularized Least Squares for MiRNA-Disease Association) [32], MIDPE [33], WBSMDA (Within and Between Score for MiRNA-Disease association) [34], CPTL (Collective Prediction based on Transductive learning) [22] and HDMP (Human Disease-related MiRNA Prediction) [18]) (Fig. 1(C)) consisting of only diseases and miRNAs, which are all based on miRNA functional similarity. Furthermore, these six approaches are all related to the second principle. In the experiment, fivefold cross validation and de novo prediction were implemented to verify the effectiveness of DMHM on large-scale omics data from existing databases. Compared with the state-of-the-art methods, DMHM successfully achieved the best prediction results with the highest area under the receiver operating characteristics (AUC) and precision. By taking two diseases as case studies, we showed several top-ranked miRNA-disease associations that were validated by dbDEMC 2.0 (database of Differentially Expressed MicroRNA in Cancer) [49] and miR2Disease [50].

2. Materials

2.1. Network model for predicting miRNA-disease associations

Most computational methods that have been proposed for prioritizing or predicting disease-related miRNAs are based on network models, including homogeneous network models (Fig. 1(A) and (B)) and bi-layer heterogeneous network models (Fig. 1(C)). However, the goal of the proposed method was to rank the unlabeled nodes and find potential miRNA-disease associations in a complex heterogeneous miRNA-gene-disease network (Fig. 1(D)).

2.1.1. Homogeneous network model for disease miRNA inference

Traditionally, disease-related miRNA inferences based on a homogeneous network model (Fig. 1(A) and (B)), i.e., a model consisting of only one type of node, as well as links, mainly rely on a gene interaction network or miRNA similarity network. Homogeneous network-based methods can be divided mainly into two categories: (1) the miRNA network on which the prediction process is performed (such as MIDP [33] and RWRMDA [19]); and (2) prediction methods using a gene network based on the assumption that miRNAs are related to complex diseases by the target genes regulated by them (such as miR_DG [31] and Shi's method [51]).

2.1.2. Bi-layer heterogeneous network model for disease miRNA prediction

In contrast, a bi-layer heterogeneous network (Fig. 1(C)) for predicting disease-associated miRNAs usually consists of two types of nodes (disease and miRNA) and three types of links (disease-disease links, miRNA-miRNA links and disease-miRNA links). Given a heterogeneous network, $G_{DM} = \{V = \{D,M\}, E = \{R_{DD}, R_{MM}, R_{DM}\}\}$, $D = \{d_1, d_2, \ldots, d_{nD}\}$ and $M = \{m_1, m_2, \ldots, m_{nM}\}$ are the node sets consisting of diseases and miRNAs, respectively, while R_{DD}, R_{MM} and R_{DM} denote links between disease pairs, miRNA pairs and disease-miRNA pairs. These links are constructed by incorporating miRNA-miRNA similarities, disease-disease similarities and known disease-miRNA associations. The inference problem of disease-related miRNAs could then be formulated as uncovering the missing edges in the bi-layer heterogeneous network. Bi-layer heterogeneous networks involving miRNAs and diseases have recently become a popular tool for predicting miRNA-disease associations (such as CPTL [22], MIDPE [33], KRLSM [36], BRWH [37], RLSMDA [32], PBMDA [35] and NetCBI [52]).

2.1.3. Heterogeneous miRNA-gene-disease network model for identifying miRNA-disease associations

However, methods based on bi-layer heterogeneous networks can only consider miRNAs with at least one related disease as a candidate.