GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

# Prediction of disease-related microRNAs by incorporating functional similarity and common association information

**K. Han[1]*, P. Xuan[2]*, J. Ding[2], Z.J. Zhao[1], L. Hui[2] and Y.L. Zhong[2]**

[1]School of Computer and Information Engineering,
Harbin University of Commerce, Harbin, China
[2]Department of Computer Science and Technology,
Key Laboratory of Database and Parallel Computing of Heilongjiang Province,
Heilongjiang University, Harbin, China

*These authors contributed equally to this study.
Corresponding authors: K. Han / P. Xuan
E-mail: hanke@hrbcu.edu.cn / xuanpinghdu@126.com

Genet. Mol. Res. 13 (1): 2009-2019 (2014)
Received January 2, 2013
Accepted August 9, 2013
Published March 24, 2014
DOI http://dx.doi.org/10.4238/2014.March.24.5

**ABSTRACT.** The identification of human disease-related microRNAs (miRNAs) is important for understanding the pathogenesis of diseases, but to do this experimentally is a costly and time-consuming process. Computational prediction of disease-related miRNA candidates is a valuable complement to experimental studies. It is essential to develop an effective prediction method to provide reliable candidates for subsequent biological experiments. In this study, we constructed a miRNA functional similarity network based on calculation of the functional similarity between each pair of miRNAs. Here, we present a new method (*DismiPred*) for predicting disease-related miRNA candidates based on the network. This method incorporates functional similarity and common association information to achieve an efficient prediction performance. *DismiPred* has been successfully shown to recover experimentally validated disease-related miRNAs for 12

common human diseases, with an F-measure ranging from 69.49 to 91.69%. Furthermore, a case study examining breast neoplasms showed that *DismiPred* could uncover novel disease-related miRNAs. *DismiPred* is useful for further experimental studies on the involvement of miRNAs in the pathogenesis of diseases.

**Key words:** Disease-related miRNAs; miRNA similarity network; Functional similarity; Common association

## INTRODUCTION

MicroRNAs (miRNAs) are a set of short (21-24 nt), non-coding RNAs that play important roles in gene regulation by targeting mRNAs for cleavage or translational repression (Bartel, 2004; Ambros, 2004). It has been shown that miRNAs usually participate in a set of important biological processes, including growth, hematopoiesis, organ formation, apoptosis, and cell proliferation. Furthermore, increasing evidence indicates that miRNAs play important roles in the development and progression of various human diseases (Iorio et al., 2005; Esquela-Kerscher and Slack, 2006; Latronico et al., 2007; Lynam-Lennon et al., 2009).

Experimental methods, such as qRT-PCR and microarray profiling, can successfully identify disease-related miRNAs (Barad et al., 2004; Lu et al., 2005; Gaur et al., 2007; Chen et al., 2009; Gutierrez et al., 2010). They can certainly highlight potential new miRNA-disease associations, which then need to be followed up by *in vitro* manipulation. However, these kinds of methods have high experimental costs and take a long time. Therefore, it is highly desirable to develop complementary computational methods that can quickly predict potential disease-related miRNA candidates for experimental studies.

Jiang et al. (2010), showed that functionally related miRNAs tend to be associated with phenotypically similar diseases. They constructed an miRNA network by establishing a functional relationship between two miRNAs based on their predicted target genes. They then integrated the miRNA network with a phenome network to infer potential miRNA-disease associations. The high number of false positives in miRNA target predictions (Bartel, 2009) considerably limits the efficacy of this method. It was also reported that if miRNAs are associated with a similar regulatory pattern in the same type of disease, their target genes might share common functional characteristics (Wang et al., 2010). Based on these results, Li et al., 2011, suggested prioritizing the miRNAs for a specific disease by estimating the functional consistency score (FCS) among their predicted target genes and the known target genes associated with the specific disease. This method was applied to 11 human diseases, including breast cancer and lung cancer. However, the limited number of known disease-related target genes limits the usefulness of this method for a few diseases. Overall, it is highly desirable to develop new methods that can efficiently predict miRNA candidates for more human diseases.

We developed a new prediction method called *DismiPred* that is based on the central hypothesis offered in several previous studies that miRNAs with similar functions are often involved in similar diseases, and vice versa (Goh et al., 2007; Lu et al., 2008; Bandyopadhyay et al., 2010; Wang et al., 2010). The functional similarity between two miRNAs can be successfully measured by the semantic similarity of their associated diseases (Wang et al., 2010). We constructed an miRNA functional similarity network based on functional similarity calcula-

tions. Subsequently, the disease-related miRNAs were predicted by considering the functional similarity and common association information. Our proposed method, *DismiPred*, provides a relevance score for each miRNA candidate. The higher the score, the closer the relationship is between an miRNA candidate and a specific disease. We applied *DismiPred* to 12 human diseases and ranked the miRNA candidates by their relevance scores. The analysis results indicate that *DismiPred* can identify potential disease-related miRNA candidates.

## MATERIAL AND METHODS

### Human miRNA-disease association data

The human miRNA disease association data were downloaded from the human miRNA disease database (HMDD, updated on March 15, 2012), which records 4488 associations identified by experiment-based studies (Lu et al., 2008). The invalid association data, where the disease name or miRNA name is incorrect, were filtered out. The disease names are in accordance with the MeSH disease terms from the National Library of Medicine (http://www.nlm.nih.gov/). The miRNA names are consistent with the latest miRNA database, miRBase 19 (Griffiths-Jones et al., 2008). The association entries, including general disease names, were further removed from the data. For example, there are two association entries, hsa-let-7a, *Neoplasms*, and hsa-let-7a, *Breast neoplasms*. This indicates that hsa-let-7a miRNA is associated with *Neoplasms* and *Breast neoplasms*. *Neoplasm* is a general disease name and *Breast neoplasm* is a more specific disease name. Furthermore, *Breast neoplasm* is subordinate to *Neoplasm*. Simultaneously maintaining the two entries will decrease the accuracy in measuring miRNA functional similarity. Therefore, the hsa-let-7a entry in *Breast neoplasms* was retained, and that in *Neoplasms* was filtered out. After filtering, the 3932 miRNA-disease associations between 450 miRNAs and 262 diseases were retained as valid associations. We calculated the functional similarities for each pair of miRNAs based on these associations.

### Prediction of miRNA candidates associated with a specific disease

For a given disease *d*, there are experimentally validated disease *d*-related miRNAs in the 450 miRNAs covered by all the valid miRNA-disease associations. These are referred to as *known disease d-related miRNAs* (known miRNAs). Excluding the known miRNAs, there is no evidence to validate that the remaining miRNAs are associated with disease *d*. These are referred to as *unknown disease d-related miRNAs* (unknown miRNAs). Since the unknown miRNAs are probably associated with disease *d*, our goal to rank them according to the possibility of their being associated with disease *d*. To achieve this goal, we correlated an unknown miRNA, such as miRNA *u*, with a relevance score, *rscore*(*u*). The higher the *rscore*(*u*), the higher the possibility that miRNA *u* is associated with disease *d*. We then rank all the unknown miRNAs according to their relevance scores, and select the top ranked miRNAs as potential disease *d*-related candidates.

The process of predicting disease *d*-related miRNA candidates is illustrated in Figure 1. First, the functional similarity between each pair of miRNAs is calculated according to their associated diseases. Second, we estimate the relevance score of each unknown disease-*d* related miRNA by incorporating the functional similarity and common association informa-

tion. Third, all the unknown miRNAs are ranked by their functional relevance scores. The top ranked miRNAs are the potential disease *d*-related miRNA candidates.

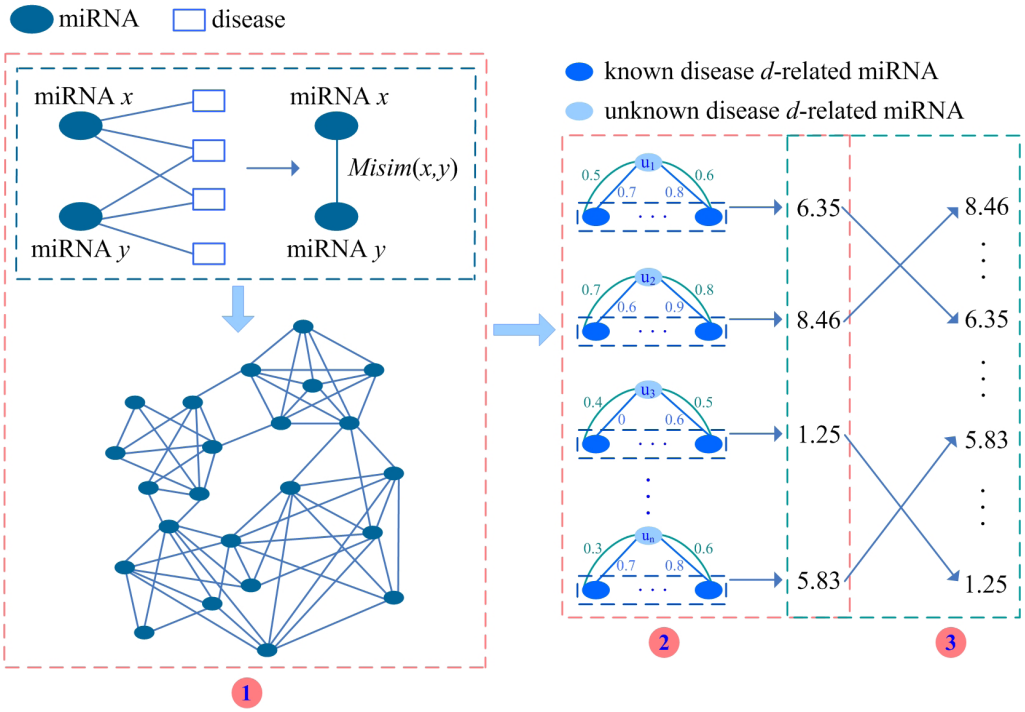



**Figure 1.** Process of predicting disease *d*-related miRNA candidates. Step 1: calculate the functional similarity of each pair of miRNAs and construct the miRNA functional similarity network; Step 2: estimate the relevance score of each unknown disease *d*-related miRNA, a blue figure represents the functional similarity of two miRNAs, and a green figure represents the common association coefficient of two miRNAs; Step 3: rank all the unknown miRNAs by their relevance scores and select the top ranked unknown miRNAs as the potential disease *d*-related miRNA candidates.

## Construction of an miRNA functional similarity network

It is well established that two miRNAs with higher functional similarity are often associated with a group of similar diseases, and vice versa (Goh et al., 2007; Lu et al., 2008; Bandyopadhyay et al., 2010; Wang et al., 2010). The functional similarity of two miRNAs has been successfully estimated by the semantic similarity of their associated diseases (Wang et al., 2010). However, the estimated functional similarity was obtained based on associations from an earlier version of the HMDD database released in September 2009. Since then, the HMDD database has been updated many times, and there are hundreds of newly reported miRNA-disease associations. To estimate the functional similarity based on the associations from the latest version of HMDD, we implemented the miRNA functional similarity calculation program. The functional similarity is calculated based again on valid associations in the latest version of HMDD (March 2012). As shown in step 1 of Figure 1, the diseases associated with two miRNAs, such as miRNA *x* and *y*, are extracted from the respective valid 3932 as-

sociations. The functional similarity between miRNA $x$ and $y$ is calculated using our implemented program, and denoted as $Misim(x,y)$. We calculate the functional similarity for each pair of miRNAs. If the functional similarity value of a pair of miRNAs is greater than 0, an edge is added to connect them. Thus, we construct the miRNA functional similarity network.

## Estimation of relevance score

The key to predicting disease $d$-related miRNAs is to estimate the relevance score of each unknown miRNA. For a given disease $d$, assume the known miRNA set is $V = \{v_1, v_2,\ldots, v_m\}$, and $v_i$ $(1 \le i \le m)$ represents a known disease $d$-related miRNA. The unknown miRNA set is $U = \{u_1, u_2,\ldots, u_n\}$, and $u_k$ $(1 \le k \le n)$ represents an miRNA that is probably associated with disease $d$.

The relevance score of an unknown miRNA is composed partly of functional similarity information and partly of common association information. For an unknown miRNA, such as $u \epsilon U$, and one of its neighboring known miRNAs, such as $v \epsilon V$, since $v$ is associated with disease $d$, $u$ is likely to associate with disease $d$. The higher the functional similarity between $u$ and $v$, the more probable is an association between $u$ and disease $d$. Therefore, we sum up the functional similarities between $u$ and each of the known disease $d$-related miRNAs. The sum is divided by the number of known disease $d$-related miRNAs, which becomes the first part of the functional relevance score of the unknown miRNA, $u$.

In addition, we consider the topological structure of the miRNA functional similarity network to explore the possibility that miRNAs $u$ and $v$ are associated with a group of similar diseases. As shown in Figure 2, we assume that miRNAs $s_1$, $s_2$, and $s_3$ are neighbors of miRNA $u$, and its neighbor set is denoted as $S_u = \{s_1, s_2, s_3\}$. Also, miRNA $s_2$, $s_3$, $s_4$, and $s_5$ comprise a neighbor set of miRNA $v$, $S_v = \{s_2, s_3, s_4, s_5\}$. The neighbors of miRNA $u$ and $v$ comprise the set $S_{u,v} = S_u \cup S_v = \{s_1, s_2, s_3, s_4, s_5\}$. Now assume $T_u$ represents the functional similarity vector of miRNA $u$ with respect to $S_{u,v}$. The $i$th element of $T_u$, $T_u(i)$, is assigned as follows.

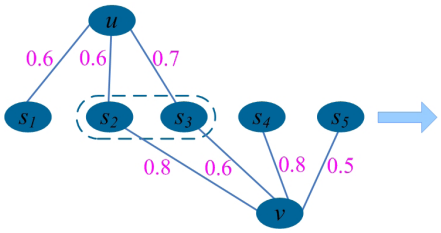
$$T_u(i) = \begin{cases} Misim(u, s_i) & if\ there\ is\ an\ edge\ between\ u\ and\ s_i \\ 0 & otherwise \end{cases} \quad \text{(Equation 1)}$$

Thus, we have vector $T_u = \{0.6, 0.6, 0.7, 0, 0\}$. In the same way, we obtain the functional similarity vector of miRNA $v$ with respect to $S_{u,v}$, $T_v = \{0, 0.8, 0.6, 0.8, 0.5\}$. It is clear that the more neighbors are commonly associated with miRNA $u$ and $v$, the greater the functional similarity between miRNA $u$ and the common neighbors, and between miRNA $v$ and the common neighbors. This means the greater the possibility that miRNAs $u$ and $v$ are associated with a group of similar diseases. Therefore, the cosine value between the two vectors, $T_u$ and $T_v$, is calculated as the common association coefficient, which is denoted as $Cosim(u, v)$. The value of $Cosim(u, v)$ in Figure 2 is 0.595. Obviously, the greater the $Cosim$ value, the higher the possibility that miRNA $u$ is associated with disease $d$. Thus, the average $Cosim$ value between miRNA $u$ and each of the known disease $d$-related miRNAs becomes the second part of its relevance score. The $Misim$ value is obtained directly, based on the diseases associated with miRNAs $u$ and $v$, and the $Cosim$ value is inferred indirectly from their neighbors. Since the former is more important than the latter, the latter is assigned lower weight, $w \epsilon (0,1)$.

In order to explore the possibility that each unknown miRNA, such as $u_k \epsilon U$, is associated with disease *d*, we sum up all the *Misim* values and the *Cosim* values between $u_k$ and each of the known disease *d*-related miRNAs. The sum is divided by the number of known disease *d*-related miRNAs to determine its functional relevance score. The relevance score of $u_k$ is denoted as $rscore(u_k)$, defined as follows,

$$rscore(u_k) = \frac{\sum_{i=1}^{m} Misim(u_k, v_i)}{m} + w \cdot \frac{\sum_{i=1}^{m} Cosim(u_k, v_i)}{m} \quad \text{(Equation 2)}$$

where $v_i$ represents a known disease *d*-related miRNA, and *w* is the weight of the *Cosim* value. To find a suitable *w* value, *w* values from 0.1 to 1 are tested by performing 5-fold cross-validation. We found that *DismiPred* has a better prediction performance when $w = 0.2$ than other values. Therefore, we set *w* as 0.2 in this study.




$$T_u = \{0.6, 0.6, 0.7, 0, 0\}$$
$$T_v = \{0, 0.8, 0.6, 0.8, 0.5\}$$
$$Cosim(u, v) = \cos(\vec{T}_u, \vec{T}_v) = \frac{\vec{T}_u \cdot \vec{T}_v}{\|\vec{T}_u\| \cdot \|\vec{T}_v\|}$$
$$= \frac{0.6 \times 0.8 + 0.7 \times 0.6}{\sqrt{0.6^2 + 0.6^2 + 0.7^2} \cdot \sqrt{0.8^2 + 0.6^2 + 0.8^2 + 0.5^2}}$$
$$= 0.595$$

**Figure 2.** Calculating the common association coefficient between an unknown miRNA and a known disease *d*-related miRNA.

The functional relevance score of each unknown miRNA is calculated. We ranked all the unknown miRNAs by their relevance scores. The top ranked miRNAs are the potential disease *d*-related miRNA candidates. The detailed disease-related miRNA prediction algorithm is shown in Figure 3.

## RESULTS AND DISCUSSION

### Prediction performance of *DismiPred*

To validate the prediction performance of *DismiPred*, we performed 5-fold cross-validation on 12 human diseases that are associated with at least 60 miRNAs. With 5-fold cross-validation, the 450 miRNAs are divided into 5 equal subsets, 4 of which are used as known information to predict disease *d*-related miRNA candidates, while the 5th subset is used for testing the prediction performance of *DismiPred*.

For a given disease *d*, to assess whether *DismiPred* reflects the relationship between miRNAs and disease *d*, we determined whether the known disease *d*-related miRNAs are ranked higher the list. The precision (P), recall (R), and F-measure were calculated to dem-

**ALGORITHM:** Prediction of a given disease *d*-related miRNA candidates
**Input:** a specific disease *d*
known disease *d*-related miRNA set $V=\{v_1,v_2,\ldots,v_m\}$
unknown disease *d*-related miRNA set $U=\{u_1,u_2,\ldots,u_n\}$
**Output:** ranked disease *d*-related miRNA candidates and their functional relevance scores

```
For the ith miRNA x_i (x_i ∈ V ∪ U, 1≤i≤m+n)
  For the jth miRNA x_j (x_j ∈ V ∪ U, i<j≤m+n)
    calculate the functional similarity between miRNA x_i and x_j, Misim(x_i, x_j)
    If Misim(x_i, x_j)>0
      add an edge to connect miRNA x_i and x_j
    End If
  End For
End For
For the kth unknown disease d-related miRNA u_k (1≤k≤n, u_k ∈ U)
  Initialize the functional relevance score of u_k, rscore(u_k), as 0
  For the ith known disease d-related miRNA v_i (1≤i≤m)
    If the functional similarity between u_k and v_i is greater than 0
      rscore(u_k)= rscore(u_k)+Misim(u_k,v_i)
    End If
    calculate the common association coefficient between u_k and v_i based on
    the miRNA functional similarity network
    If the common association coefficient between u_k and v_i is greater than 0
      rscore(u_k)= rscore(u_k)+Cosim(u_k,v_i)
    End If
  End For
  rscore(u_k)= rscore(u_k)/m
End For
All the unknown miRNAs are ranked by their functional relevance scores
The top ranked miRNAs are the potential disease d-related miRNA candidates
```

**Figure 3.** Algorithm of predicting the miRNA candidates associated with disease *d*.

onstrate the prediction performance. Suppose that, in the test set, the number of miRNAs associated with disease *d* is $n_d$, and there are *p* correctly predicted miRNAs in the top *m* ranked miRNAs related to disease *d*. This means that there are *p* known disease *d*-related miRNAs in the top *m* ranked miRNAs. The precision, recall, and F-measure at rank *m* are defined as:

$$\text{Precision} = \frac{p \text{ correctly predicted miRNAs}}{m \text{ ranked miRNAs}} \qquad \text{(Equation 3)}$$

$$\text{Recall} = \frac{p \text{ correctly predicted miRNAs}}{n_d \text{ disease } d\text{-related miRNAs}} \quad \text{(Equation 4)}$$

$$\text{F-measure} = \frac{2}{1/\text{precision} + 1/\text{recall}} \quad \text{(Equation 5)}$$

The greater precision means that there are more correctly predicted miRNAs in the top ranked *m* miRNAs. The greater recall ensures a more complete set of known disease *d*-related miRNAs having higher rankings. To obtain PR curves, the precision and recall values are calculated for the increase of *m* with step 1. Note that the F-measure reflects both precision and recall by their harmonic mean. Thus, a greater F-measure indicates an overall more accurate prediction performance, with better precision and recall.

The valid 3932 miRNA-disease associations cover 262 diseases. The maximum number of miRNAs associated with a disease is 195, and the minimum number is 1. We tested 12 human diseases that have at least 60 associated miRNAs. Figure 4 shows the PR curves for 12 diseases. The corresponding maximum F-measures in their PR curves are listed in Table 1. The highest F-measure (91.69%) was obtained with *hepatocellular carcinoma*, and the lowest (69.49%) was with *stomach neoplasms*. The results indicate that *DismiPred* can successfully recover the known disease *d*-related miRNAs.

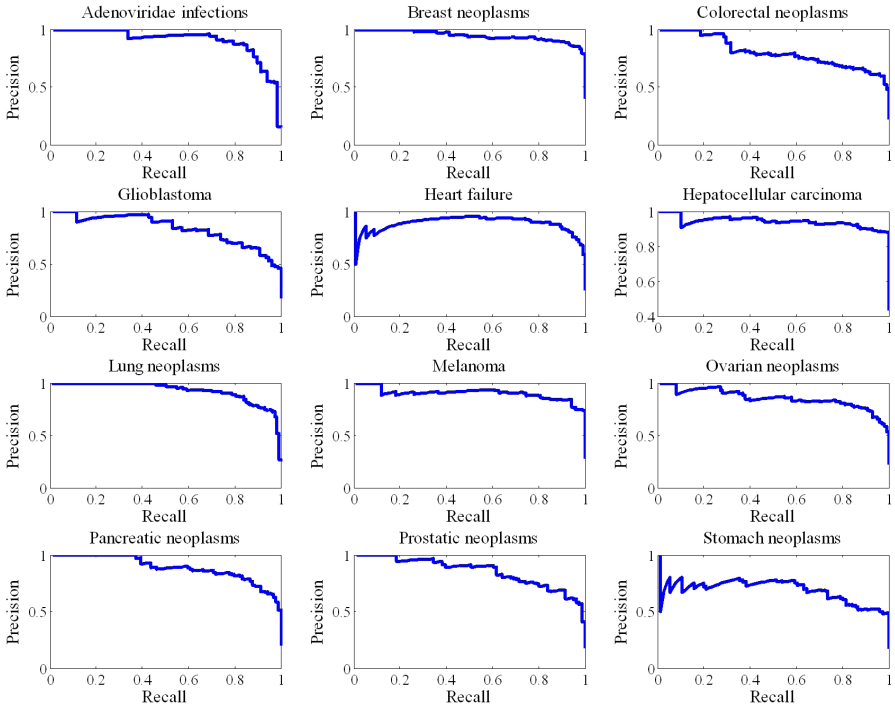



**Figure 4.** Precision versus recall for prediction of miRNA candidates associated with 12 human diseases.

**Table 1.** Prediction results of *DismiPred* for 12 human diseases.

| Disease name | No. of associated miRNAs | F-measure (%) |
|---|---|---|
| Adenoviridae infections | 68 | 84.24 |
| Breast neoplasms | 179 | 89.12 |
| Colorectal neoplasms | 98 | 74.49 |
| Glioblastoma | 77 | 74.69 |
| Heart failure | 112 | 86.76 |
| Hepatocellular carcinoma | 195 | 91.69 |
| Lung neoplasms | 113 | 83.91 |
| Melanoma | 126 | 87.39 |
| Ovarian neoplasms | 100 | 81.11 |
| Pancreatic neoplasms | 89 | 80.73 |
| Prostatic neoplasms | 81 | 77.17 |
| Stomach neoplasms | 76 | 69.49 |

## Case study: breast neoplasms

To further demonstrate the ability of *DismiPred* to uncover novel disease *d*-related miRNAs, we present a case study on breast neoplasms. Many researchers have shown that miRNAs play critical roles in breast neoplasms. Here, we provide a comprehensive analysis of the predicted breast neoplasm-related miRNA candidates.

To validate *DismiPred*'s ability to uncover novel, disease *d*-related miRNAs, *DismiPred* predicted the miRNAs based on the miRNA-disease association data in the earlier version of HMDD (January 2012). The newly reported breast neoplasm-related miRNAs in the latest HMDD (March 2012) were used to assess the predicted miRNA candidates. The top 15 miRNA candidates are shown in Table 2.

**Table 2.** The top 15 breast neoplasm-related miRNA candidates in the ranked list.

| miRNA name | Description | References |
|---|---|---|
| **hsa-let-7i** | In human breast cancer, hsa-let-7i, hsa-mir-101, and hsa-mir-191 are significantly downregulated, as compared with normal breast tissue. | Iorio et al., 2005; Xin et al., 2009 |
| **hsa-mir-16** | Hsa-mir-16 and hsa-mir-15a are clustered in within 0.5 kb on chr 13. Also, they belong to a miRNA family (mir-15). Hsa-mir-15a is associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000070 |
| hsa-mir-101 | In human breast cancer, hsa-let-7i, hsa-mir-101, and hsa-mir-191 are significantly downregulated, as compared with normal breast tissue. | Iorio et al., 2005; Xin et al., 2009 |
| **hsa-let-7b** | Hsa-let-7b is a member of miRNA family (let-7). Let-7 regulates self renewal and tumorigenicity of breast cancer cells. | Yu et al., 2007 |
| hsa-mir-142 | In the genes-to-systems breast cancer (G2SBC) database, 4 of top 10 hsa-mir-142's predicted target genes are associated with breast cancer. | Mosca et al., 2010 http://www.itb.cnr.it/breastcancer/index.html |
| hsa-mir-99a | Hsa-mir-99a and hsa-let-7c are clustered in within 10 kb on chr 21. Hsa-mir-99a and hsa-mir-100 belong to a miRNA family (mir-99). Hsa-let-7c and hsa-mir-100 are associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/mirna_summary.pl?fam=MIPF0000025 |
| hsa-mir-106a | Hsa-mir-106a, hsa-mir-18b, hsa-mir-20b, hsa-mir-19b, and hsa-mir-92a are clustered in within 10 kb on chr X. The latter 4 miRNAs are associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000113 |
| hsa-mir-130a | In the G2SBC database, 2 of top 10 hsa-mir-130a's predicted target genes are associated with breast cancer. | Mosca et al., 2010 http://www.itb.cnr.it/breastcancer/index.html |
| hsa-mir-148a | Hsa-miR-148a expression is downregulated in human breast cancer. | Lehmann et al., 2008 |
| hsa-mir-191 | In human breast cancer, hsa-let-7i, hsa-mir-101, and hsa-mir-191 are significantly downregulated, as compared with normal breast tissue. | Iorio et al., 2005; Xin et al., 2009 |
| **hsa-mir-29c** | Hsa-mir-98 and hsa-mir-29c were upregulated greater than 2-fold in primary breast cancer compared with normal adjacent tumor tissues. | Yu et al., 2007 |
| hsa-mir-98 | Hsa-mir-98 and hsa-mir-29c were upregulated greater than 2-fold in primary breast cancer compared with normal adjacent tumor tissues. | Yu et al., 2007 |
| hsa-mir-192 | Hsa-mir-192 and hsa-mir-194 are clustered in within 10 kb on chr 11. Hsa-mir-192 and hsa-mir-215 belong to an miRNA family (mir-192). Hsa-mir-194 and hsa-mir-215 are associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/mirna_summary.pl?fam=MIPF0000063 |
| **hsa-mir-27b** | Hsa-mir-27b and hsa-mir-24 are clustered in within 10 kb on chr 9. Hsa-mir-27b and hsa-mir-27a belong to an miRNA family (mir-27). Hsa-mir-24 and hsa-mir-27a are associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/36 mirna_summary.pl?fam=MIPF00000 |
| **hsa-mir-193a** | Hsa-mir-193a and hsa-mir-193b belong to an miRNA family (mir-193). Hsa-mir-193b is associated with breast neoplasms. | http://www.mirbase.org/cgi-bin/mirna_summary.pl?fam=MIPF0000082 |

The miRNAs in bold are the new reported breast neoplasm-related miRNAs.

First, the earlier HMDD (January 2012) does not contain the following 6 miRNAs: hsa-let-7i, hsa-mir-101, hsa-mir-148a, hsa-mir-191, hsa-mir-29c, and hsa-mir-98. Biological experiments have shown that these miRNAs are significantly upregulated or downregulated in human breast neoplasms versus normal breast tissue (Iorio et al., 2005; Lehmann et al., 2008; Yan et al., 2008; Xin et al., 2009). *DismiPred* successfully found these miRNAs because of their higher rankings.

Second, some miRNAs are often found in genomic clusters. The clustered miRNAs are usually transcribed together and are likely associated with similar diseases (Baskerville and Bartel, 2005; Wang et al., 2010). Five miRNAs (hsa-mir-16, hsa-mir-99a, hsa-mir-106a, hsa-mir-192, and hsa-mir-27b) are clustered with other miRNAs that have been experimentally validated to be associated with breast neoplasms.

Third, homologous miRNAs are gathered into the same miRNA family by RFam (Gardner et al., 2009). The seed regions of miRNA sequences from the same family (normally 2-8 nucleotides from the 5'-end) are almost identical. Since the seed of an miRNA is commonly required to be perfectly complementary to the target mRNAs for cleavage or translational repression, miRNAs of the same family likely regulate a common set of mRNA targets. Therefore, it is more likely that they are associated with similar diseases (Yu et al., 2007; Wang et al., 2010). The 6 miRNAs (hsa-mir-16, hsa-let-7b, hsa-mir-99a, hsa-mir-192, hsa-mir-27b, and hsa-mir-193a) and other breast neoplasm-related miRNAs belong to the respective miRNA families. It indicates that these miRNA candidates are more probably associated with breast neoplasms.

Fourth, the genes-to-systems breast cancer (malignant breast neoplasms) database, G2SBC (Mosca et al., 2010), is usually used to assist in studying breast cancer. For the hsa-mir-142 and hsa-mir-130a miRNAs, at least 2 of the top 10 of their predicted target genes are real breast cancer-related genes. It shows that these 2 miRNAs are more likely to participate in breast cancer-related biological processes.

Last but not least, when our study was almost completed, HMDD was updated again on March 15, 2012, when 6 of 15 new miRNA candidates, including hsa-let-7i, hsa-mir-16, hsa-let-7b, hsa-mir-29c, hsa-mir-27b, and hsa-mir-193a, were supported by newly reported validated miRNAs. All the above results demonstrate that *DismiPred* is powerful in predicting potential disease-related miRNA candidates.

## CONCLUSIONS

In this study, we calculated the functional similarity for each pair of miRNAs based on the valid 3932 miRNA-disease associations. We also constructed an miRNA functional similarity network. The functional similarity and common association information were combined to efficiently predict disease-related miRNA candidates. Our proposed prediction method, *DismiPred*, has proven to be successful in identifying known experimentally validated disease-related miRNAs, and in predicting potential candidates for 12 human diseases. It indicates that *DismiPred* is a powerful new tool for experimental research on the association between miRNA and human disease.

## ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of China (Grant #61302139), the Natural Science Foundation of Heilongjiang Province (Grants #F201225, #F201206, #F201245), the Science and Research Foundation of Heilongjiang Education Department (Grants #12531154,

#12511401, #12531499), the Young Innovative Talents Research Foundation of Harbin Science and Technology Bureau (Grants #2012RFQXG082, #2012RFQXS094), the Starting Research Foundation for Doctors of Harbin University of Commerce (Grant #13DL005) and the Support Program for Young Academic Key Teacher of Higher Education of Heilongjiang Province.

## REFERENCES

Ambros V (2004). The functions of animal microRNAs. *Nature* 431: 350-355.

Bandyopadhyay S, Mitra R, Maulik U and Zhang MQ (2010). Development of the human cancer microRNA network. *Silence* 1: 6.

Barad O, Meiri E, Avniel A, Aharonov R, et al. (2004). MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.* 14: 2486-2494.

Bartel DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.

Bartel DP (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215-233.

Baskerville S and Bartel DP (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11: 241-247.

Chen Y, Gelfond JA, McManus LM and Shireman PK (2009). Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics* 10: 407.

Esquela-Kerscher A and Slack FJ (2006). Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* 6: 259-269.

Gardner PP, Daub J, Tate JG, Nawrocki EP, et al. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37: D136-D140.

Gaur A, Jewell DA, Liang Y, Ridzon D, et al. (2007). Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res.* 67: 2456-2468.

Goh KI, Cusick ME, Valle D, Childs B, et al. (2007). The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* 104: 8685-8690.

Griffiths-Jones S, Saini HK, van Dongen S and Enright AJ (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36: D154-D158.

Gutierrez NC, Sarasquete ME, Misiewicz-Krzeminska I, Delgado M, et al. (2010). Deregulation of microRNA expression in the different genetic subtypes of multiple myeloma and correlation with gene expression profiling. *Leukemia* 24: 629-637.

Iorio MV, Ferracin M, Liu CG, Veronese A, et al. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65: 7065-7070.

Jiang Q, Hao Y, Wang G, Juan L, et al. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* 4 (Suppl 1): S2.

Latronico MV, Catalucci D and Condorelli G (2007). Emerging role of microRNAs in cardiovascular biology. *Circ. Res.* 101: 1225-1236.

Lehmann U, Hasemeier B, Christgen M, Muller M, et al. (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J. Pathol.* 214: 17-24.

Li X, Wang Q, Zheng Y, Lv S, et al. (2011). Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res.* 39: e153.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435: 834-838.

Lu M, Zhang Q, Deng M, Miao J, et al. (2008). An analysis of human microRNA and disease associations. *PLoS One* 3: e3420.

Lynam-Lennon N, Maher SG and Reynolds JV (2009). The roles of microRNA in cancer and apoptosis. *Biol. Rev. Camb. Philos. Soc.* 84: 55-71.

Mosca E, Alfieri R, Merelli I, Viti F, et al. (2010). A multilevel data integration resource for breast cancer study. *BMC Syst. Biol.* 4: 76.

Wang D, Wang J, Lu M, Song F, et al. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26: 1644-1650.

Xin F, Li M, Balch C, Thomson M, et al. (2009). Computational analysis of microRNA profiles and their target genes suggests significant involvement in breast cancer antiestrogen resistance. *Bioinformatics* 25: 430-434.

Yan LX, Huang XF, Shao Q, Huang MY, et al. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* 14: 2348-2360.

Yu F, Yao H, Zhu P, Zhang X, et al. (2007). Let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 131: 1109-1123.