Original Article

DOI: 10.5582/bst.2014.01031

Combination of microRNA expression profiling with genome-wide SNP genotyping to construct a coronary artery disease-related miRNA-miRNA synergistic network

Lin Hua[1,2,*,], Hong Xia[1,2,*], Ping Zhou[1,2], Dongguo Li[1,2], Lin Li[1,2]**

[1] *School of Biomedical Engineering, Capital Medical University, Beijing, China;*
[2] *Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China.*

Summary **In recent years, microRNAs (miRNAs) were found to play critical roles in many important biological processes. On the other hand, the rapid development of genome-wide association studies (GWAS) help identify potential genetic variants associated with the disease phenotypic variance. Therefore, we suggested a combined analysis of microRNA expression profiling with genome-wide Single Nucleotide Polymorphism (SNP) genotyping to identify potential disease-related biomarkers. Considering functional SNPs in miRNA genes or target sites might be important signals associated with human complex diseases, we constructed a miRNA-miRNA synergistic network related to coronary artery disease (CAD) by performing a genome-wide scan for SNPs in human miRNA 3' -untranslated regions (UTRs) target sites and computed potential SNP cooperation effects contributing to disease based on potential miRNA-SNP interactions reported recently. Furthermore, we identified some potential CAD-related miRNAs by analyzing the constructed miRNA-miRNA synergistic network. As a result, the predicted miRNA-miRNA network and miRNA clusters were validated by significantly high interaction effects of CAD-related miRNAs. Accurate classification performances were obtained for all of the identified miRNA clusters, and the sensitivity and specificity were all more than 90%. The network topological analysis confirmed some novel CAD-related miRNAs identified recently by experiments. Our method might help to understand miRNA function and CAD disease, as well as to explore the novel mechanisms involved.**

Keywords: GWAS, target, sites, interaction

1. Introduction

MicroRNAs (miRNAs) are endogenous small non-coding RNA molecules that can regulate gene expression at the posttranscriptional level by binding with 3'-untranslated regions (UTRs) of the target mRNAs through base pairing. About 80% of miRNA genes are located within introns of defined transcription units, and their expression is frequently correlated with expression profiles of their host genes (*1*). Currently, it is estimated that a miRNA may regulate hundreds of target genes and most of human protein coding genes are regulated by miRNAs. miRNA plays critical roles in many important biological processes, and it is therefore suggested that miRNAs are also involved in human complex diseases such as cancers (*2*), heart diseases (*3*), nervous system damage (*4*) and so on.

On the other hand, in recent years, genome-wide association studies (GWAS) have become an increasingly effective tool to identify genetic variation associated with the risk of complex diseases (*5,6*). However, currently identified genetic variants collectively can explain only a small proportion of disease phenotypic variance, and the noise causes many of the identified signals to be false positive loci. A noteworthy observation is that miRNAs are wide and key regulators of gene expression,

*These authors contributed equally to this works.
***Address correspondence to:*
Dr. Lin Hua, School of Biomedical Engineering, Capital Medical University, Beijing, 100069, China.
E-mail: hualin7750@139.com

miRNA-related Single Nucleotide Polymorphisms (SNPs) including SNPs in miRNA genes and target sites therefore may function as regulatory SNPs through modifying miRNA regulation to affect phenotypes and disease susceptibility (*7,8*). More and more evidence has shown that SNPs in target sites or miRNA genes are associated with diseases (*7,8*). For example, Chen and Rajewsky uncovered that SNP density in conserved miRNA sites was lower than in conserved control sites, suggesting that a large class of computationally predicted conserved miRNA target sites is under significant negative selection (*9*). Their results also implicate that SNPs located in miRNA binding sites are likely to affect the expression of the miRNA target and might contribute to the susceptibility of humans to complex diseases. Recent advances in genetic studies have systematically identified and analyzed human polymorphisms in miRNAs and/or miRNA target sites (*8,10*). However, most of them focus on SNPs in target sites and their effects or disease-related miRNAs, while few of them mentioned the understanding of the synergistic regulation of miRNAs and their potential targeted SNPs cooperative effects contributing to disease. With the rapid accumulation of disease-related miRNAs, it is increasingly needed to uncover their functional relationships contributing to diseases at a system biology level.

Different from those constructed miRNA-miRNA networks depending on co-occurrence in 3'-UTRs of the same miRNA target genes (*11*), or using the correlations between the gene silencing scores of individual miRNAs or the combinatorial effects of co-expressed miRNAs in the modulation of a given pathway (*12*), or using the functional information and protein interaction data of predicted miRNA targets (*13*), we attempted to construct a miRNA-miRNA synergistic network related to coronary artery disease (CAD) by performing a genome-wide scan for SNPs in human miRNA 3'-UTR target sites and computed SNP cooperative effects contributing to disease based on the potential miRNA-SNPs interactions reported recently (*7*). In this process, miRNA expression profiling data and genome-wide SNP genotype data were integrated. As a result, predicted a miRNA-miRNA network was validated by significantly high interaction effects of CAD-related miRNAs. Our method can help to understand miRNA function and CAD disease, as well as to explore novel miRNA biomarkers and infer novel mechanisms connecting miRNAs to functions.

2. Materials and Methods

2.1. *Data Source*

2.1.1. *CAD related genome-wide SNP genotype data*

For GWAS data, we selected the genome-wide Wellcome Trust Case Control Consortium (WTCCC) data (*http://www.wtccc.org.uk*), which examined about 2000 individuals for seven major diseases, using a shared set of about 3000 controls. In the current study, only CAD data was used. As a part of the WTCCC, 482, 428 genome-wide SNPs were genotyped. The Standard WTCCC thresholds were applied for SNP quality control (genotype call rate > 95%, Hardy Weinberg equilibrium p-value > 5×10^{-7}, Minor Allele Frequency (MAF) > 0.01). In this analysis, we took those SNPs with $p < 0.01$ in single association tests as potential risk SNPs in consideration of relaxing the cutoff which can avoid losing some true positive loci. In fact, we have found some significant CAD-related risk SNPs confirmed by experiment, such as rs4299376 ($p = 0.004$), rs2505083 ($p = 0.008$) and rs46522 ($p = 0.008$) (*14*), have become non-risk SNPs in GWAS according to the significant level of 0.0001 in previous studies (*15*).

2.1.2. *CAD-related miRNA expression profiling data*

For miRNA expression profiling data, we selected the platform evaluated by Illumina (Illumina Inc, San Diego, CA, USA) Human v2 miRNA expression beadchip (GEO association: GPL8179, *http://www.ncbi.nlm.nih.gov/geo/*). The dataset GSE28858 includes 1,146 miRNA expression values of 12 CAD patients and 12 controls. We downloaded miRNA expression values after log2 transformation.

2.2. *Scan of 3'-UTR SNPs from the WTCCC CAD-related GWAS data*

In this analysis, we used an update SNPnexus (*http://www.snpnexus.org*) tool to perform the scan of 3'-UTR SNPs from the WTCCC CAD-related GWAS data. SNPnexus database provides annotation and genomic location on clones, contigs or chromosomes for both novel and public SNPs by incorporating a broader range of variations such as insertions/deletions block substitutions, and region-based analysis. In addition, we referred to the enhanced functional annotation involved in the update SNPnexus tool (*16*).

2.3. *Identification of differentially expressed miRNAs*

In this analysis, we applied Significance Analysis of Microarrays (SAM) method (*http://otl.stanford.edu*) to extract the statistical significance miRNAs between CAD and control groups. SAM can identify statistically significant miRNAs by carrying out miRNA specific t-tests and computing a statistic for each miRNA, which measures the strength of the relationship between miRNA expression and the phenotype (*17*). In this method, repeated permutations of the data are used to determine if the expression of any miRNA is significant

related to the phenotype. In this analysis, we used the criterion of $p < 0.05$ and False Discovery Rates (FDR) < 0.1 to determine miRNAs variously differentially expressed.

2.4. *Identification of differentially expressed miRNA and 3'-UTR SNP pairs*

In this analysis, consider that the gain of a functional miRNA target site will repress protein expression and affect physiological function and clinical phenotype, and thus we used miRNA and 3'-UTR SNP target gain predicted by the MiRNASNP tool (*http:// www.bioguo.org/miRNASNP*) (*7*) to extract differentially expressed miRNA and 3'-UTR SNP pairs. MiRNASNP tool combined results of two popular tools, TargetScan (*http://www.TargetScan.org/*) and miRanda *(http:// www.microrna.org*), which are regularly updated and considered to have relatively good performance. In this tool, the UTR sequences were divided into wild type and corresponding mutant type, and miRNA and 3'-UTR SNP target gain was defined as follows:

$$\text{Gain} = (\text{ST} \cap \text{SM}) - (\text{WT} \cup \text{WM}), \quad (1)$$

where ST are the target genes of SNP-miRNAs processed by TargetScan and SM are the target genes of SNP-miRNAs processed by miRanda; whereas WT are the target genes of wild-type miRNAs processed by TargetScan and WM are the target genes of wild-type miRNAs processed by miRanda (*7*). That means, if one miRNA-SNP pair was predicted in both ST and SM, but neither in WT nor WM, this pair is called miRNA and 3'-UTR SNP target gain. For each of miRNA-SNP pair, minimum hybridization energy of the miRNA–target interaction is calculated using RNAhybrid (*http:// bibiserv.techfak.uni-bielefeld.de/rnahybrid*) (*7,18*). In this analysis, we used the binding energy changes between wild-type miRNA/target and SNP-miRNA/target to evaluate the interaction strength of each miRNA-SNP pair. More energy change means sharper interaction.

2.5. *Construction of miRNA-miRNA synergistic network*

Here, for those 3'-UTR SNPs regulated by differentially expressed miRNAs, we used logistic regression methods based on PLINK software (*http://pngu.mgh.harvard.edu/~purcell/plink/*) to detect SNP-SNP cooperation effects contributing to coronary artery disease. We used 0.05 as the cutoff of significant interaction effect between two SNPs, and the multiple test corrections were not performed to avoid the loss. On one hand, we consider that Bonferroni correction assumes that the tests are independent, yet many interaction tests are highly correlated. Multiple test corrections might be overly conservative due to high correlations among the interaction tests, and adjusting for multiple testing may result in a decrease in power to detect weaker associations of susceptibility SNPs. On the other hand, when considering interaction effect in the logistic regression analysis, such as gene-gene interaction or gene-environment interaction, p-values can be relaxed (*19*).

For two given miRNA-SNP pairs: $miRNA_1$-SNP_1 and $miRNA_2$-SNP_2, if SNP_1 interacted with SNP_2 ($p_{12} < 0.05$), we defined an interaction score of $miRNA_1$ and $miRNA_2$:

$$S_{12} = E_1 \times E_2 \times (-\log(p_{12})) \quad (2)$$

where E_1 and E_2 are binding energy changes of $miRNA_1$-SNP_1 and $miRNA_2$-SNP_2, respectively. The greater S score means stronger interaction between two miRNAs. After assembling all miRNA pairs identified above, we generated the miRNA-miRNA synergistic network related to CAD disease. A node represents a miRNA, and two nodes are connected if the corresponding miRNA pair has an interaction action. Figure 1 shows the workflow to construct the miRNA-miRNA synergistic network.

2.6. *Extraction of miRNA clusters and Randomization tests*

To validate the constructed miRNA-miRNA network, we extracted miRNA clusters to evaluate their co-expression or combined strength. Note that highly connected nodes (*hubs*) characterized by those nodes with high degrees are often considered as important targets of networks (*20*), and sub-networks implicated by *hubs* have also been shown to be highly conserved in maintaining the housekeeping biological functions of cells (*15*). We therefore examined all of the miRNA clusters implicated by *hub* miRNAs to see their characteristics. Considering *hubs* might be different in terms of the network size, we did not directly define *hubs* as nodes with more than five interactions in the network like Taylor's study (*21*). Instead, we assumed the degree of nodes followed a Poisson distribution in a random network (*22*). To determine whether a miRNA is a *hub* miRNA, the following formula was used to compute its probability of degree $\geq t$ (*23*):

$$P(x \geq t) = 1 - P(x < t) = 1 - \sum^{t-1} \lambda^k \exp(-\lambda)/k! \ (\lambda = np_1, p_1 = m/C_n^2), \quad (3)$$

where n is the number of miRNAs and m is the number of interacting miRNA-pairs in the miRNA-miRNA network. In our analysis, a miRNA was considered as a *hub* when its P value was smaller than the probability of this rare event. This filtering criteria was supported by some previous studies (*15,24*).

Then, we extracted miRNA clusters implicated by *hubs*, which include all miRNAs linked to the *hubs* directly. We computed the average S scores for each of

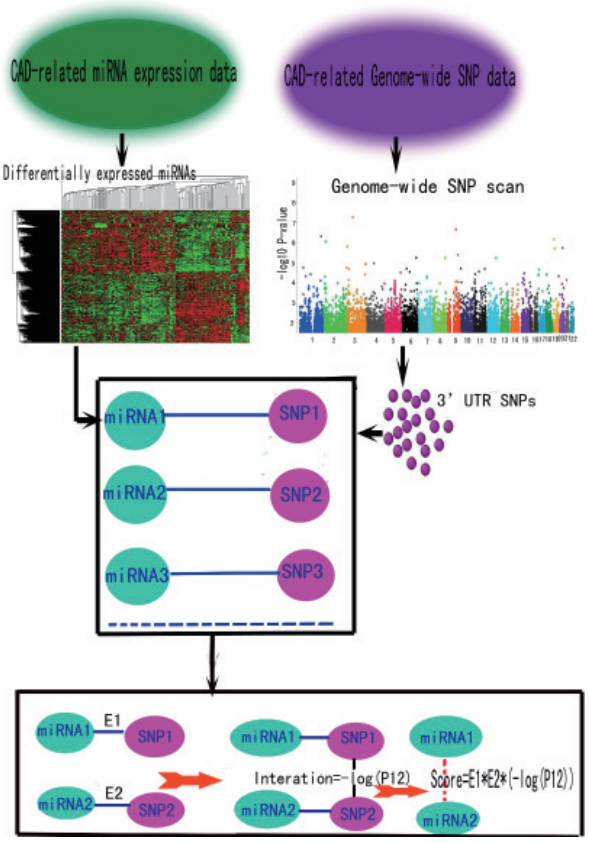

Figure 1. The workflow to construct the miRNA-miRNA network. The process involves three main steps. First, we identified differentially expressed miRNAs using CAD-related miRNA expression profiling data, and performed a scan of 3'-UTR SNPs from the WTCCC CAD-related genome-wide SNP genotype data. Second, we used the MiRNASNP tool to extract differentially expressed miRNA and 3'-UTR SNP pairs and their corresponding binding energy changes. Finally, we performed logistic regression to detect the significant interaction of 3'-UTR target SNPs of differentially expressed miRNAs, and then identified miRNA pairs based on the corresponding SNP-pairs. After assembling all miRNA pairs identified, we constructed the miRNA-miRNA synergistic network related to CAD disease.

the miRNA clusters. To evaluate the significance of co-expression or combined strength of miRNAs involved in these clusters, we performed the randomization tests. For each miRNA cluster, we randomly selected the same number of miRNA pairs from all SNP (miRNA)-SNP (miRNA) pairs, and calculated the corresponding average S score. This procedure was repeated 1,000 times, and the significant *P*-value is the fraction of the average S scores under random conditions, which is greater than the value in the real condition.

3. Results

3.1. *Identification of differentially expressed miRNAs*

According to the criterion of $p < 0.05$ and FDR < 0.1, we applied the SAM method to identify 398 significantly differentially expressed miRNAs. Where, 6 out of the 12 previously identified CAD-related miRNAs in miRNA disease association database-HMDD (Human MicroRNA Disease Database, *http://202.38.126.151/hmdd/tools/hmdd2.html*) (*25*) exhibited statistically significant differential expression between cases and controls, and they were miR-126 ($p = 0.0007$), miR-130a ($p = 0.0278$), miR-21 ($p = 0.0005$), miR-222 ($p = 0.0181$), miR-340 ($p = 6.11E-05$) and miR-624 ($p = 0.0012$). A growing amount of evidence has shown their association with CAD. For example, a recent report suggested that vascular miR-126 is consumed during transcoronary passage, and the differential regulation of circulating miRs during the transcoronary passage might provide important insights to exploit their role as cardiac biomarkers (*26*). We listed those significantly up-regulated (Fold_change > 1.5) or down-regulated (Fold_change < 1/1.5) differentially expressed miRNAs in Table S1 (*http://www.biosciencetrends.com/docindex.php?year=2014&kanno=5*).

3.2. *Identification of differentially expressed miRNA and 3'-UTR SNP pairs*

Using the SNPnexus tool, we performed a scan of 3'-UTR SNPs from the WTCCC CAD-related GWAS data, and 3,521 3'-UTR SNPs were identified. Among these 3'-UTR SNPs, only 43 (1.2%) were significant ($p < 0.01$). Then, we used miRNA and 3'-UTR SNP target gain predicted by the MiRNASNP tool to extract 126 differentially expressed miRNA and 3'-UTR SNP pairs. Among these miRNA and 3'-UTR SNP pairs, the average of binding energy changes caused by SNPs in 3'-UTR was 12.32 kcal/mol, where the miR-770-5p and rs9991 pair was up to 36.3 kcal/mol, suggesting a greater affect of the differentially expressed miRNA-target SNP binding.

3.3. *Construction of miRNA-miRNA synergistic network*

In this analysis, we used logistic regression methods based on PLINK software to detect SNP-SNP cooperative effects contributing to coronary artery disease. As a result, 7,875 SNP-pairs arose from 126 3'-UTR SNPs targeted by differentially expressed miRNAs. According to our criterion of significant interaction, 281 SNP-SNP pairs were selected. While, one pair of SNPs: rs175634 and rs1043515 showed the most significant interaction ($p = 0.000113$). There are a certain number of SNPs showing interaction with multiple other SNPs. For example, SNP rs10802805 interacts with another 10 SNPs, such as rs2112812 ($p = 0.0017$), rs9537 ($p = 0.0083$), rs6052935 ($p = 0.0236$) and so on.

For two given miRNA-SNP pairs: $miRNA_1$-SNP_1 and $miRNA_2$-SNP_2, and the corresponding interaction effect of SNP_1 and SNP_2, we computed the interaction score of $miRNA_1$ and $miRNA_2$ using the S measure as defined in the Methods section. For overlapped

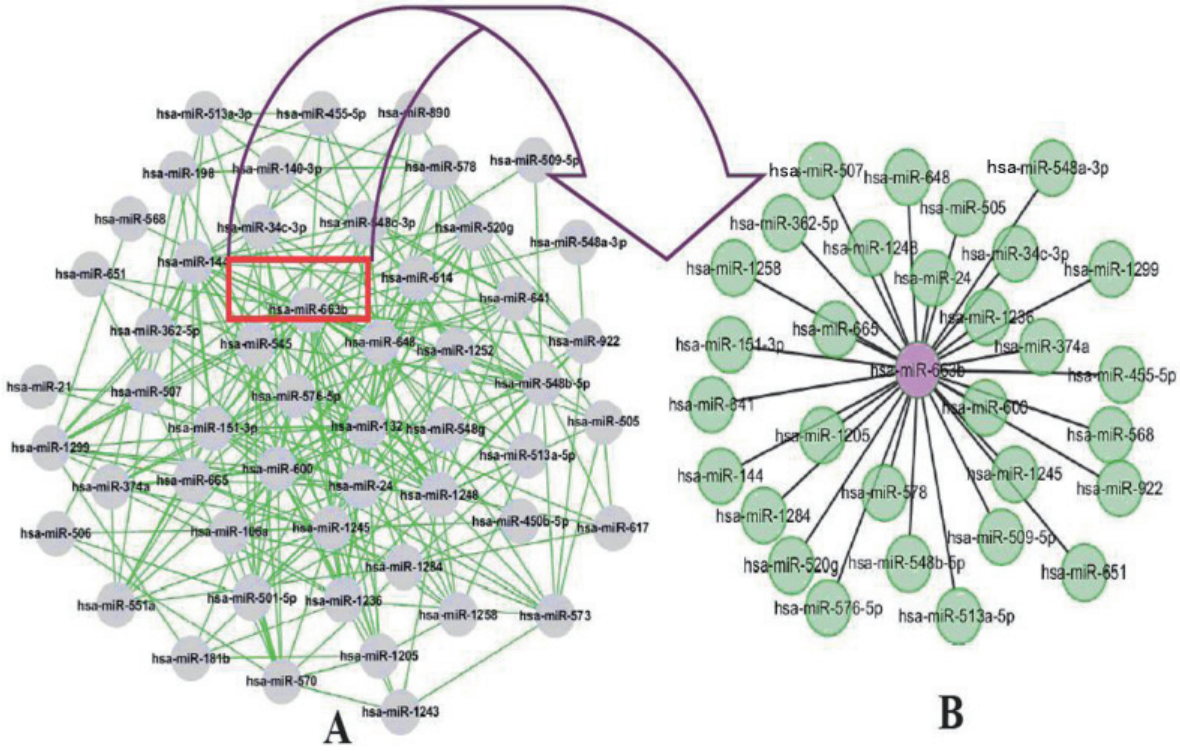

Figure 2. The constructed miRNA-miRNA synergistic network related to CAD. (A) CAD-related miRNA-miRNA network. Each node represents a miRNA, and two nodes are connected if the corresponding miRNA pair has an interaction. **(B)** A special example: miR-663b related miRNA cluster.

miRNA pairs, we took the maximum S scores as their interaction strength. After assembling all miRNA pairs, we can see that interacting miRNAs formed a miRNA-miRNA synergistic network related to CAD disease (See Figure 2A). For the whole network, the maximum interaction strength was seen from miR-665 and miR-132 (S = 1355.46), followed by miR-140-3p and miR-198 (S = 1319.96). The average S score was 188.68. To determine if this network is a small-world network, we constructed 1,000 random small-world networks using Cytoscape software (*http://www.cytoscape.org*) (*27*), and computed the average clustering coefficient and average diameter. We found that the network had a short diameter of 2.983, which is similar to that of a random small world network (2.987 ± 0.091, p = 0.144). In addition, we calculated the average clustering coefficient of the network, and obtained an average clustering coefficient of 0.322 which was much higher than for random networks (0.166 ± 0.014, **p** < 0.001) and approved the dense local neighborhoods of the network. This result suggests that immediate neighbors of a miRNA tend to be synergistic and contribute to disease (*28*).

With the topological characteristic analysis of this constructed miRNA-miRNA network, we can see that there were more poorly connected nodes than *hubs*. More than 60% of the nodes had degrees of ≤ 10. In other words, the degree shows a scale-free distribution, which means that most miRNAs only have a few connected miRNAs, but some miRNAs show an outstanding degree property. For example, miR-663b showed the highest degree (degree = 30, $p < 10^{-5}$), suggesting it can connect to a greater number of miRNAs than others. This is supported by newly published research in which they found miR-663b showed the highest sensitivity and specificity for discrimination of acute myocardial infarction cases from health controls of miRNA expression in peripheral total blood samples (*29*). In addition, miR-144 also showed a greater degree (degree = 18, $p = 0.0187$). A recent study has reported that ectopic expression of miR-144 augmented cardiomyocyte survival, which was further improved by over-expression of miR-144, compared to control cells in response to simulated ischemia (*30*). This result support our constructed networks not only shed light on the relationships between miRNAs, but also give insight into those *hub* miRNAs which are more evolutionarily conserved than non-*hub* miRNAs. It also can be explained that *hub* miRNAs are subject to selection pressure and constraints, due to their involvements in many biological processes and their multiple interacting miRNA partners.

3.4. *The validation of miRNA-miRNA network using miRNA expression value*

It is possible that miRNAs with similar functions tend to have similar expression profiles. To further

validate and evaluate the accuracy of our constructed miRNA-miRNA network and investigate the expression pattern of connecting miRNA-pairs, we analyzed their expression similarity using Pearson's correlation coefficients. As a result, we found that approximately 40% of miRNA pairs have positive or negative co-expression patterns ($p < 0.05$). In addition, we used logistic regression models based on miRNA expression values to analyze the cooperative effects of miRNA-pairs. Interestingly, more than half of miRNA-pairs (52.3%) are proved to be significant ($p < 0.05$). Specially, when investigating those miRNA-pairs including miR-663b; approximately 80% of them are significant (See Table 1 and Figure 2B). This result further supports that miR-663b is an important CAD association miRNA, which is more likely to interact with other miRNAs contributing to disease.

3.5. *Randomization test*

To evaluate the co-expression or combined strength of the miRNA groups involved in the network, we extracted miRNA clusters implicated by *hubs* from the constructed network. In our analysis, a *hub* miRNA with ≥ 17 partners in a random network ($p < 0.05$) was considered a rare event under the null hypothesis that *n* miRNAs were connected randomly. According to this criterion, when 6 *hub* miRNAs were extracted, they were miR-600 (degree = 17, $p = 0.034$), miR-548b-5p (degree = 17, $p = 0.034$), miR-1245 (degree = 19, $p = 0.0099$), miR-144 (degree = 18, $p = 0.0187$), miR-1248 (degree = 20, $p = 0.005$) and miR-663b (degree = 30, $p < 10\text{-}5$). For each miRNA cluster, we calculated the average S score to its 1000 random matched miRNA clusters. For miRNA clusters implicated by *hubs* miR-600, miR-548b-5p, miR-1245, miR-144, miR-1248, and miR-663b, each matched set consisted of 17, 17, 19, 18, 20, and 30 miRNAs randomly selected from 7,875 SNP (miRNA)-SNP(miRNA) pairs. According to the empirical distribution of S scores, we obtained a threshold value to determine the significance of each miRNA cluster. We found that excluding the miRNA cluster implicated by miR-548b-5p (S = 27.04, $p = 0.805$), other miRNA clusters implicated by *hubs* all showed outstanding S scores compared to random miRNA clusters: miR-600 (S = 138.48, $p < 0.001$), miR-1245 (S = 175.24, $p < 0.001$), miR-144 (S = 289.60, $p < 0.001$), miR-1248 (S = 206.52, $p < 0.001$) and miR-663b (S = 75.97, $p = 0.038$). In these identified miRNA clusters, some miRNAs have been proved to be associated with CAD. For example, in the miR-144 cluster, miR-545 has been found to show an up-regulation in premature CAD patients compared to controls in recent studies (*31*). In the miR-663b cluster,

Table 1. miR-663b related miRNA cluster

$miRNA_1$ (SNP_1)	$miRNA_2$ (SNP_2)	Logistic *p*-values based on miRNA expression values	Score = $E_1 \times E_2 \times P_{12}$
miR-663b (rs6052935)	miR-507 (rs4312485)	0.034	0.708
miR-663b (rs6052935)	miR-548a-3p (rs6654)	0.016	99.10
miR-663b (rs6052935)	miR-1205 (rs165345)	0.103	10.21
miR-663b (rs6052935)	miR-1236 (rs10894557)	0.012	27.12
miR-663b (rs175634)	miR-1245 (rs1043515)	0.866	3.63
miR-663b (rs6052935)	miR-1248 (rs12218073)	0.040	188.91
miR-663b (rs175634)	miR-1258 (rs198413)	0.024	2.53
miR-663b (rs6052935)	miR-1284 (rs1052912)	0.301	1.13
miR-663b (rs6052935)	miR-1299 (rs10260499)	0.019	25.25
miR-663b (rs6052935)	miR-144 (rs11543230)	0.514	149.46
miR-663b (rs6052935)	miR-151-3p (rs12174237)	0.035	130.08
miR-663b (rs175634)	miR-24 (rs12214037)	0.047	1.37
miR-663b (rs175634)	miR-34c-3p (rs2810)	0.033	10.04
miR-663b (rs6052935)	miR-362-5p (rs10521099)	0.040	145.19
miR-663b (rs6052935)	miR-374a (rs3811610)	0.302	0.70
miR-663b (rs6052935)	miR-455-5p (rs10933164)	0.028	3.16
miR-663b (rs6052935)	miR-505 (rs10864675)	0.040	154.42
miR-663b (rs6052935)	miR-509-5p (rs1064395)	0.043	6.58
miR-663b (rs6052935)	miR-513a-5p (rs11153074)	0.026	231.30
miR-663b (rs175634)	miR-520g (rs4151045)	0.021	20.04
miR-663b (rs6052935)	miR-548b-5p (rs3737933)	0.039	48.11
miR-663b (rs6052935)	miR-568 (rs8035733)	0.012	15.66
miR-663b (rs6052935)	miR-576-5p (rs435382)	0.026	2.73
miR-663b (rs6052935)	miR-578 (rs10165660)	0.022	141.50
miR-663b (rs6052935)	miR-600 (rs1045291)	0.030	191.84
miR-663b (rs6052935)	miR-641 (rs6880907)	0.376	4.62
miR-663b (rs6052935)	miR-648 (rs8713)	0.033	182.80
miR-663b (rs6052935)	miR-651 (rs10516139)	0.026	10.28
miR-663b (rs6052935)	miR-665 (rs16854011)	0.919	279.78
miR-663b (rs6052935)	miR-922 (rs7677472)	0.047	201.19

Note: P_{12} indicates the significant interaction *p*-value of SNP_1 and SNP_2.

an increased level of miR-24 was found in idiopathic end-stage failing human hearts by northern blot analysis of the hypertrophy-regulated miRNAs (32). Moreover, when we computed the average correlation coefficients using miRNA expression values for these five miRNA clusters, the average correlation coefficients were all approximately 0.5 (See Table 2). This result suggests the identified CAD-related miRNA clusters have greater average S scores and greater relevance than other miRNA clusters expected by chance (See Figure 3).

3.5. *Validation of classification performances of five identified miRNA clusters using miRNA expression values*

To validate whether the identified miRNA clusters have better classification performance, we applied Random Forests (RF) method (*33*) to miRNA expression profiling data in which those miRNAs involved in miRNA clusters are taken as predictor variables to classify samples. RF is an ensemble classifier that consists of many decision trees and each tree depends on the values of a random vector sampled independently. In this analysis, we used 5-fold cross validation to assess the classification accuracy rate. All samples were divided into five sets and in each analysis one set is considered as testing data, whereas the others are training data. For Random Forests program, 5,000 trees were constructed. As a result, accurate classification performances were obtained for all miRNA clusters, and the sensitivity and specificity were

Table 2. The characteristics of CAD related miRNA clusters

miRNA clusters implicated by *hubs*	Degree of *hub*	S (Scutoff)	**p*-value	Average correlation coefficient
miR-600	17	138.48 (105.50)	< 0.001	0.512
miR-1245	19	175.24 (84.40)	< 0.001	0.546
miR-144	18	289.60 (106.80)	< 0.001	0.479
miR-1248	20	206.52 (81.80)	< 0.001	0.464
miR-663b	30	75.97 (73.50)	0.038	0.489

Note: * Significant *p*-value is the fraction of the average S scores in random conditions, which is greater than the value in the real condition.

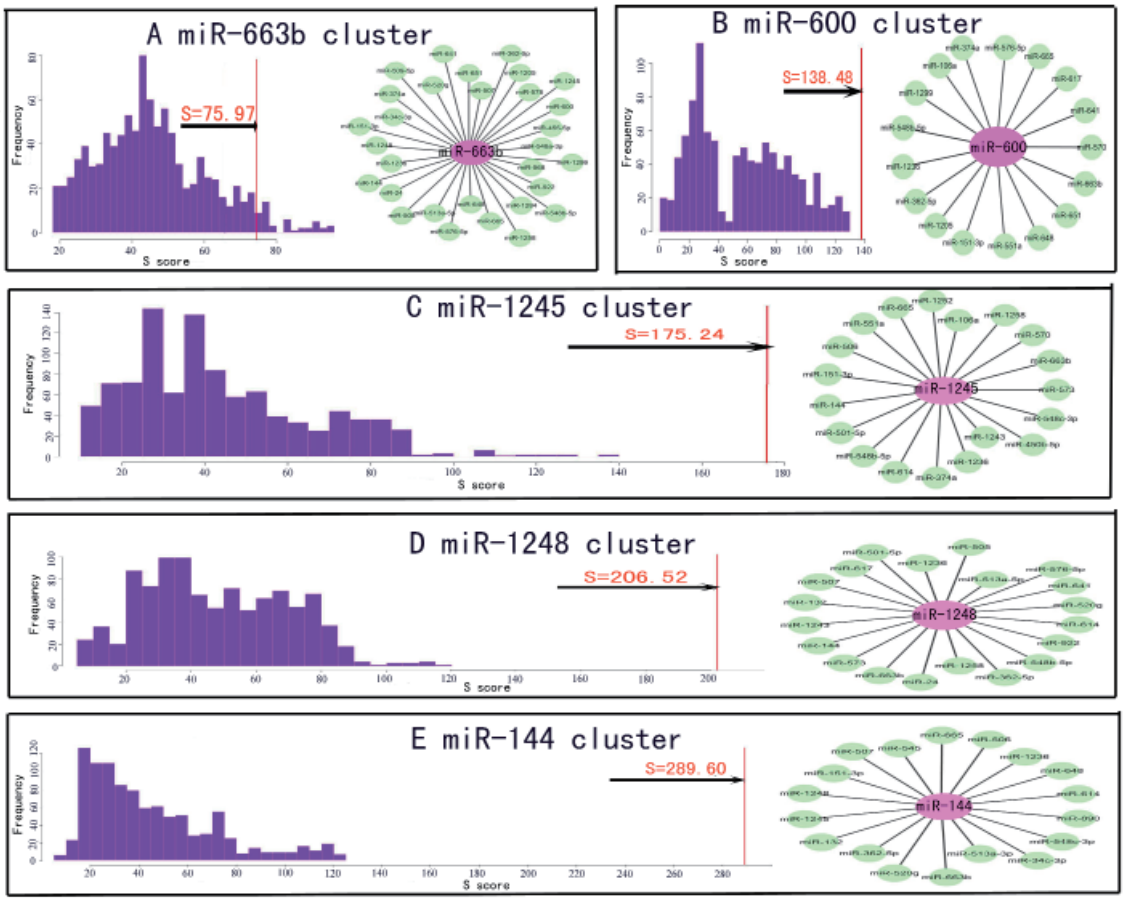

Figure 3. The distribution of average S scores for five miRNA clusters implicated by *hubs*. The random background distribution of average S scores for miRNA clusters implicated by miR-663b **(A)**, miR-600 **(B)**, miR-1245 **(C)**, miR-1248 **(D)** and miR-144 **(E)** are shown in purple colors. The red lines indicate the real average S scores of the corresponding miRNA clusters.

Table 3. Sensitivity, specificity, positive predictive value, negative predictive value, and area under ROC curve (AUC) for miRNA clusters used to detect CAD with Random Forest classifier

miRNA clusters implicated by *hubs*	The Prediction Property (95% Confidence Interval)					miRNA with the maximal MDG	MDG
	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Area under ROC curve (AUC) (%)		
miR-600	92.3(64.0-99.8)	100.0(71.5-100.0)	100.0(73.5-100.0)	91.7(61.5-99.8)	95.8(86.4-105.3)	miR-1236	2.858
miR-1245	92.3(64.0-99.8)	90.9(58.7-99.8)	92.3(64.0-99.8)	90.9(58.7-99.8)	91.6(78.5-104.8)	miR-1236	2.314
miR-144	100.0(73.5-100.0)	91.7(61.5-99.8)	92.3(64.0-99.8)	100.0(71.5-100.0)	96.2(87.4-104.9)	miR-1236	2.864
miR-1248	100.0(71.5-100.0)	92.3(64.0-99.8)	91.7(61.5-99.8)	100.0(73.5-100.0)	95.8(86.4-105.3)	miR-1236	2.237
miR-663b	100.0(73.5-100.0)	91.7(61.5-99.8)	92.3(64.0-99.8)	100.0(71.5-100.0)	96.2(87.4-104.9)	miR-1236	2.062

all more than 90%. The sensitivity, specificity, positive predictive value, negative predictive value, and area under the ROC curve (AUC) for five miRNA clusters used to detect CAD with Random Forest classifier are shown in Table 3.

Note that it is often interesting to know which miRNAs are important in classification in Random Forests program. There are two measures of importance in RF program, the mean decrease in accuracy and the mean decrease Gini index (MDG) (*33*). In this analysis, we used MDG to measure the risk of miRNAs in each miRNA cluster. Greater MDG will indicate that the degree of impurity arising from a category could be reduced furthest by this miRNA, and thus suggests an important miRNA. Interestingly, miR-1236 ranked first for all miRNA clusters (See Table 3). Recent evidence showed that miR-1236 may function as a negative regulator of vascular endothelial growth factor receptor (VEGFR)-3 (*34*).

3.7. *Comparison with MISIM tool*

We also attempted to validate our results to see whether miRNAs involved in miRNA clusters have potential functional similarity. Currently, the functional similarities of miRNAs are often predicted by indirectly inferring similarities of their corresponding target genes. These prediction methods include Gene Ontology (GO) or pathway enrichment analysis for miRNA-targeted gene sets (*35*). However, these methods are not applicable for miRNA genes because the function of most miRNAs remains unknown and no such function annotation database is available. Therefore, in this analysis, we used a recent developed tool, MISIM (*http://210.73.221.6/*), which is used to measure miRNA functional similarity based on human miRNA-disease association data and structures of the corresponding disease relationships (*13*). We used 0.7 as the MISIM threshold to determine whether two miRNAs have a link. In other words, those miRNA pairs with a MISIM coefficient greater than or equal to 0.7 were selected. The results of MISIM showed that miR-106a, miR-144, miR-520g, miR-24, miR-132, miR-505, miR-507 and miR-600 had a similar function (See Figure 4). Indeed, recently published data have shown some of these miRNAs are

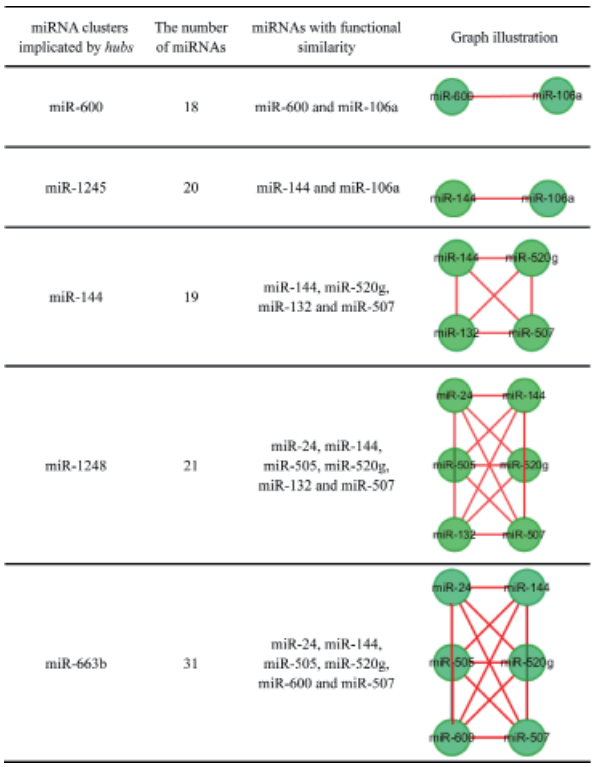

Figure 4. The evaluation of miRNA clusters based on functional similarity using MISIM tool. In graphical illustration, each green circle indicates a miRNA and each red line connects two miRNAs with similar function.

associated with CAD-related diseases, such as miR-132, miR-144, miR-106a and miR-24. For example, Katare *et al*. investigated the therapeutic activity and mechanistic targets of saphenous vein-derived pericyte progenitor cells (SVPs) in a mouse myocardial infarction model, and found SVP transplantation produced long-term improvement of cardiac function through a novel paracrine mechanism involving secretion of miR-132 and inhibition of its target genes (*36*). In addition, some miRNAs such as miR-24, were found to be up-regulated in human heart failure patients and transfection of neonatal cardiomyocytes with these miRNAs resulted in significant cardiomyocyte hypertrophy (*37*). These results suggest miRNA clusters identified by our method might have a potential functional aggregated trend related to CAD disease.

4. Discussion

Some recent studies have implicated that SNPs located at miRNA binding sites are likely to affect the expression of the miRNA target and might contribute to the susceptibility of humans to common diseases (*7,38*). In this paper, using the expression profiles of miRNAs and genome-wide SNP genotype data, we constructed a miRNA-miRNA synergistic network related to CAD by performing a genome-wide scan for SNPs in human miRNA 3'-UTR target sites and computed potential SNP cooperative effects contributing to disease based on the potential miRNA-SNPs interactions. Our method has effectively identified 5 miRNA clusters related to CAD, and a certain proportion of miRNAs identified have been reported to have association with coronary artery disease. For example, a recent study showed that loss of the miR-144 cluster limits ischemic preconditioning cardioprotection by up-regulating Rac-1-mediated oxidative stress signaling (*39*). Furthermore, some novel associations between miRNAs and CAD discovered recently have been predicted by our method, such as miR-663b.

It is worthy to note that the origin of our constructed miRNA-miRNA network is based on the genome-wide SNP-SNP interaction. In practice, when testing millions of SNP-pairs for interaction, novel statistic methods are required for getting high power, and a large sample size is required for estimating logistic regression parameters to avoid sparseness problems for modeling interaction effects. Fortunately, in our study, only 126 3'-UTR SNPs of approximately 5, 000 samples came into the interaction analysis, and therefore the "fast-epistasis" involved in PLINK could also perform well. Certainly, we have to point out that the absence of multiple testing corrections and the replication of interaction analysis might cause some false positive results. Therefore, further validation using other individual studies will be performed in the future.

Note that only 43 3'-UTR SNPs (1.2%) were significant ($p < 0.01$), which indicates that most SNPs involved in significant interactions have no marginal association. This result is consistent with those reported by Wu *et al*. (*40*), in which they found that 75% of interacting SNPs with *p*-values (for testing marginal association) larger than 0.2 and 44% of interacting SNPs with *p*-values (for testing marginal association) larger than 0.5. This strongly suggested that most significant SNP interactions have arisen from those SNPs with weak marginal association. However, an interesting observation was noted that the greatest binding energy changes was presented in miR-770-5p an rs9991 pair (36.3 kcal/mol), while miR-770-5p ($p = 0.0267$) and rs9991 ($p = 0.0043$) were all significantly associated with CAD. Although few matched data in this study can't prove the relevance between the binding energy changes and the significance of miRNAs or SNPs, this phenomenon suggests that those significant miRNAs or their corresponding targets should also be focused on when considering their interaction effects contributing to complex disease. Moreover, among 126 3'-UTR SNPs used in this analysis, 89 SNPs (70.6%) have relatively high MAF (MAF $\geq$ 0.10). Excluding 3 SNPs located in chromosome 18, 20, and 21 respectively, the average MAFs for the other 19 chromosomes were all greater than 0.14. Specially, SNP rs9991 which is significantly associated with CAD also has a relatively high MAF (MAF = 0.3115). This result agreed with a recent study in which Gong *et al*. demonstrated that those SNPs with a high population MAF, or high population frequency difference or undergoing positive selection pressure might be important candidates for population phenotype research and complex trait studies (*7*). Therefore, those SNP-associated miRNA target sites may be worthwhile to focus on in further experiments.

Furthermore, it should be pointed out when we investigated the relevance between S scores and the expression similarity for each connecting miRNA-pair involved in the constructed miRNA-miRNA network, we found there was no correlation between them ($p = 0.382$). This phenomenon also exists for identified miRNA clusters, such as miR-663b ($p = 0.920$) and miR-600 ($p = 0.963$). On one hand, this might be caused by the greater range of S scores (0.04-1355.46) due to the increased binding energy changes of miRNA-SNP pairs, or on the other hand, integration data might cause some potential errors due to different data types which contain different kinds of information. Moreover, we can see that some functionally similar miRNA pairs neither have high expression similarity nor belong to the same family or cluster.

Currently, although advanced laboratory instruments and computing systems developed to decipher the structure and function of genes, proteins, miRNAs and other substances in the human body, the limitations and the complexity of biomedical problems causes natural obstacles in understanding the etiology of complex diseases. Therefore, bioinformatics and computational data face a great challenge in discovering new biomarkers which can be used to detect, prevent and cure complex diseases by performing genomic data analysis. Fortunately, various high-throughput approaches such as genome sequencing technology, microarray technology and proteomic characterization of proteins and complexes have allowed us to gather vast amounts of data to construct cellular networks. Our future work will be enhanced by integrating more data types, such as miRNA expression profiling data, gene expression profiling data, protein-protein interaction data, disease data, genotype data and so on. Specially, we know that some novel findings obtained by bioinformatics analysis may not be the facts and need to be validated using experiment technologies. Therefore,

experiment validation is important and will further help confirm the identified biomarkers. In the past, we have used qRT-PCR, Western blots and luciferase assays to validate some breast cancer subtype-related miRNAs identified by silico analysis, and the report of this work is in press. In the future, we will use experimental technologies to further confirm these potential findings. We expect our study can provide more clues for other researchers to perform experiments and predictions for CAD-related biomarkers.

In conclusion, we have presented a framework to construct a miRNA-miRNA network using miRNA expression profiling data and genome-wide SNP genotype data. The network topological analysis confirmed some novel CAD-related miRNAs identified recently by experiment. Our method might help to undertand miRNA function and CAD disease, as well as explore the novel mechanisms connecting miRNAs and function.

Acknowledgements

This work is supported by Beijing Natural Science Foundation (Grant Nos. 7142015), National Science Foundation of China (Grant Nos. 31100905) and the Science Technology Development Project of Beijing Municipal Commission of Education (SQKM201210025008). This study is also funded by the excellent talent cultivation project of Beijing and the young backbone teacher's cultivation project of Beijing Municipal Commission of Education, and supported by the foundation-clinical cooperation project of Capital Medical University (11JL30, 11JL33 and 12JL75).

References

1. Stone N, Pangilinan F, Molloy AM, Shane B, Scott JM, Ueland PM, Mills JL, Kirke PN, Sethupathy P, Brody LC. Bioinformatic and genetic association analysis of microRNA target sites in one-carbon metabolism genes. PLoS ONE. 2011; 6:e21851.
2. Martins M, Rosa A, Guedes LC, *et al*. Convergence of miRNA expression profiling, a-synuclein interacton and GWAS in Parkinson's disease. PLoS ONE. 2011; 6:e25443.
3. Ikeda S, Kong SW, Lu J, Bisping E, Zhang H, Allen PD, Golub TR, Pieske B, Pu WT. Altered microRNA expression in human heart disease. Physiol Genomics. 2007; 31:367-373.
4. Hutchison ER, Okun E, Mattson MP. The Therapeutic Potential of microRNAs in nervous system damage, degeneration and repair. Neuromolecular Med. 2009; 11:153-161.
5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362-9367.
6. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 2009; 5:e1000477.
7. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat. 2012; 33:254-263.
8. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: The implications for cancer research. Nat Rev Cancer. 2010; 10:389-402.
9. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet. 2006; 38:1452-1456.
10. Bhartiya D, Laddha SV, Mukhopadhyay A, Scaria V. miRvar: A comprehensive database for genomic variations in microRNAs. Hum Mutat. 2011; 32:E2226-E2245.
11. Shalgi R, Lieber D, Oren M, Pilpel Y. Global and Local Architecture of the mammalian microRNA–transcription factor regulatory network. PLoS Comput Bio. 2007; 3:e131.
12. Boross G, Orosz K, Farkas IJ. Human microRNAs co-silence in well-separated groups and have different predicted essentialities. Bioinformatcis. 2009; 25:1063-1069.
13. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatcis. 2010; 26:1644-1650.
14. Roberts R, Stewart AF. Genetics of coronary artery disease in the 21st century. Clin Cardiol. 2012; 35:536-540.
15. Huang W, Wang P, Liu Z, Zhang L. Identifying disease associations via genome-wide association studies. BMC Bioinformatcis. 2009; 10 (Suppl 1):S68.
16. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: A web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012; 40:W65-W70.
17. Taylor J, Tibshirani R, Efron B. The 'miss rate' for the analysis of gene expression data. Biostatistics. 2005; 6:111-117.
18. Krüger J, Rehmsmeier M. RNAhybrid: MicroRNA target prediction easy, fast and flexible. Nucleic Acids Res. 2006; 34:W451-W454.
19. Hand BD, Kostek MC, Ferrell RE, Delmonico MJ, Douglass LW, Roth SM, Hagberg JM, Hurley BF. Influence of promoter region variants of insulin-like growth factor pathway genes on the strength-training response of muscle phenotypes in older adults. J Appl Physiol. 2007; 103:1678-1687.
20. Barabási AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101-113.
21. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol. 2009; 27:199-204.
22. Blondel VD, Guillaume J-L, Hendrickx JM, Jungers RM. Distance distribution in random graphs and application to network exploration. Phys Rev E Stat Nonlin Soft Matter Phys. 2007; 76:101-108.
23. Jiang W, Li X, Rao S, Wang L, Du L, Li C, Wu C, Wang H, Wang Y, Yang B. Constructing disease-specific gene networks using pair-wise relevance metric: Application to colon cancer identifies interleukin 8, desmin and enolase

1 as the central elements. BMC Syst Biol. 2008; 2:1752-1766.
24. Shibolet O, Giallourakis C, Rosenberg I, Mueller T, Xavier RJ, Podolsky DK. AKAP13, a RhoA GTPase-specific guanine exchange factor, is a novel regulator of TLR2 signaling. J Biol Chem. 2007; 282:35308-35317.
25. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An Analysis of human microRNA and disease associations. PLoS One. 2008; 3:e3420.
26. De Rosa S, Fichtlscherer S, Lehmann R, Assmus B, Dimmeler S, Zeiher AM. Transcoronary concentration gradients of circulating microRNAs. Circulation. 2011; 124:1936-1944.
27. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular Interaction Networks. Genome Res. 2003; 13:2498-2504.
28. Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, Xu LD, Wang YY, Du L, Zhang YP, Jiang W, Li CQ, Xiao Y, Li X. MiRNA–miRNA synergistic network: Construction via co-regulating functional modules and disease miRNA topological features. Nucleic Acids Res. 2011; 39:825-836.
29. Meder B, Keller A, Vogel B, Haas J, Sedaghat-Hamedani F, Kayvanpour E, Just S, Borries A, Rudloff J, Leidinger P, Meese E, Katus HA, Rottbauer W. MicroRNA signatures in total peripheral blood as novel biomarkers for acute myocardial infarction. Basic Res Cardiol. 2011; 106:13-23.
30. Zhanga X, Wanga X, Zhua H, Zhu C, Wang Y, Pu WT, Jegga AG, Fan GC. Synergistic effects of the GATA-4-mediated miR-144/451 cluster in protection against simulated ischemia/reperfusion-induced cardiomyocyte death. J Mol Cell Cardiol. 2010; 49:841-850.
31. Bronze-da-Rocha E. MicroRNAs expression profiles in cardiovascular diseases. Biomed Res Int. 2014; 2014:985408.
32. van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, Richardson JA, Olson EN. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci U S A. 2006; 103:18255-18260.
33. Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, Floyd E, Zhao H. Pathway analysis using random forests classification and regression. Bioinformatcis. 2006; 22:2028-2036.
34. Jones D, Li Y, He Y, Xu Z, Chen H, Min W. Mirtron microRNA-1236 inhibits VEGFR-3 signaling during inflammatory lymphangiogenesis. Arterioscler Thromb Vasc Biol. 2012; 32:633-642.
35. Du Z, Li L, Chen CF, Yu PS, Wang JZ. G-SESAME: Web tools for GO-term-based gene similarity analysis and knowledge discovery. Nucleic Acids Res. 2009; 37:W345-W349.
36. Katare R, Riu F, Mitchell K, Gubernator M, Campagnolo P, Cui Y, Fortunato O, Avolio E, Cesselli D, Beltrami AP, Angelini G, Emanueli C, Madeddu P. Transplantation of human pericyte progenitor cells improves the repair of infarcted heart through activation of an angiogenic program involving micro-RNA-132. Circ Res. 2011; 109:894-906.
37. Catalucci D, Gallo P, Condorelli G. MicroRNAs in cardiovascular biology and heart disease. Circ Cardiovasc Genet. 2009; 2:402-408.
38. Yu Z, Li Z, Jolicoeur N, Zhang L, Fortin Y, Wang E, Wu M, Shen SH. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. Nucleic Acids Res. 2007; 35:4535-4541.
39. Wang X, Zhu H, Zhang X, Liu Y, Chen J, Medvedovic M, Li H, Weiss MJ, Ren X, Fan GC. Loss of the miR-144/451 cluster impairs ischaemic preconditioning-mediated cardioprotection by targeting Rac-1. Cardiovasc Res. 2012; 94:379-390.
40. Wu X, Dong H, Luo L, Zhu Y, Peng G, Reveille JD, Xiong M. A Novel statistic for genome-wide interaction analysis. PLoS Genet. 2010; 6:e1001131.

(Received February 26, 2014; Revised October 25, 2014; Accepted December 8, 2014)