

Computational study of 'HUB' microRNA in human cardiac diseases

Remya Krishnan¹, Achuthsankar S. Nair¹, Pawan K. Dhar^{*1,2}

¹Department of Computational Biology and Bioinformatics, University of Kerala, Thiruvananthapuram, Kerala - 695 581; ²School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067; Dr. Pawan K. Dhar - E-mail: pawan.dhar@mail.jnu.ac.in; Phone: +911126738887; *Corresponding author

Received July 29, 2016; Revised January 1, 2017; Accepted January 5, 2017; Published January 24, 2017

Abstract

MicroRNAs (miRNAs) are small non-coding RNAs ~22 nucleotides long that do not encode for proteins but have been reported to influence gene expression in normal and abnormal health conditions. Though a large body of scientific literature on miRNAs exists, their network level profile linking molecules with their corresponding phenotypes, is less explored. Here, we studied a network of 191 human miRNAs reported to play a role in 30 human cardiac diseases. Our aim was to study miRNA network properties like hubness and preferred associations, using data mining, network graph theory and statistical analysis. A total of 16 miRNAs were found to have a disease node connectivity of >5 edges (i.e., they were linked to more than 5 diseases) and were considered hubs in the miRNA-cardiac disease network. Alternatively, when diseases were considered as hubs, >10 of miRNAs showed up on each 'disease hub node'. Of all the miRNAs associated with diseases, 19 miRNAs (19/24= 79.1% of upregulated events) were found to be upregulated in atherosclerosis. The data suggest micro RNAs as early stage biological markers in cardiac conditions with potential towards microRNA based therapeutics.

Keywords: disease network, dys-regulation, miRNAs, statistical analysis

doi: 10.6026/97320630013017

Background:

MicroRNAs are non-coding RNAs that are transcribed but do not encode proteins. miRNAs have been reported to play a pivotal role in regulating key biological processes e.g., post-transcriptional modifications and translation processes. Ever since the discovery of miRNAs in 1993, they have been reported to influence a variety of human physiological conditions such as post transcriptional regulation of skeletal development [1], circadian rhythms [2], cancers [3-4], and neuro-degenerative diseases like Alzheimer's [5], cardiac disorders [6], etc. MicroRNAs are robust and diverse in functions and are also predicted to control human ageing process [7]. A more recent study has revealed that cardiac microRNAs can indirectly regulate the function of other cardiac microRNAs ie, microRNA-mediated microRNA regulation. A number of experimental as well as computational studies have supported the role of microRNA regulators in heart diseases like Myocardial Infarction, Cardiomyopathies, Atherosclerosis, Stenosis, etc.

Cardiovascular diseases (CVDs) are one of the leading causes of human deaths [8]. The global status report of the World Health Organization [9] reported an estimated 17.3 million people having died of CVD as of 2008. This number has been projected to increase to 23.3 million people dying of heart diseases alone by the year 2030. To understand a complex heart condition such as hypertrophic cardiomyopathy or atherosclerosis in terms of causes and treatment it's important to study the molecular inventory and explore the network organizes itself to produce a certain phenotype, in normal and diseased states. Using computational approach, an attempt was made to find an overall interaction pattern of experimentally validated microRNAs in heart conditions, based on previous reports associating microRNAs with clinical phenotypes. We were interested to find if microRNAs "hubs" existed in disease networks and, if yes, could they be used as biomarkers before / during the onset of disease conditions? Also, if micro RNA associations existed, could one use them as potential therapeutic agents to treat

cardiac conditions?

In this study, a total of 28 different human heart diseases were identified from the Human MicroRNA Disease Database

(HMDD: <http://202.38.126.151/hmdd/mirna/md/>) and mapped to 190 experimentally validated microRNAs associations.

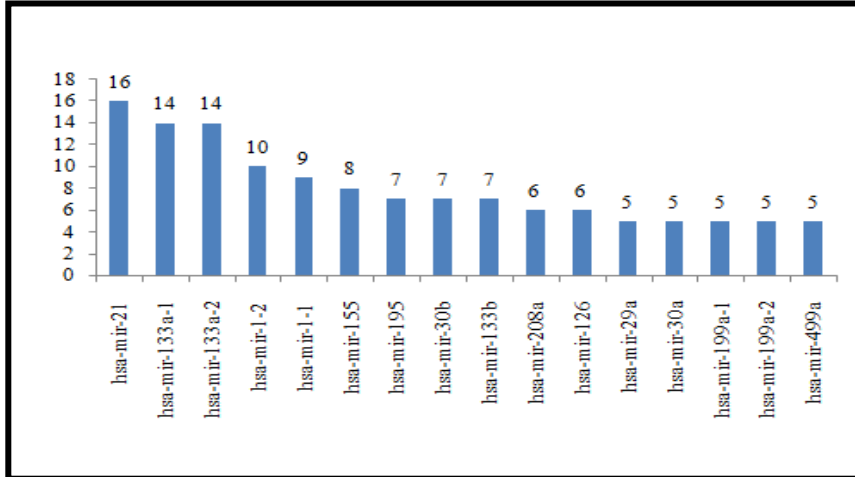


Figure 1: The frequency of microRNA- cardiac disease incidence

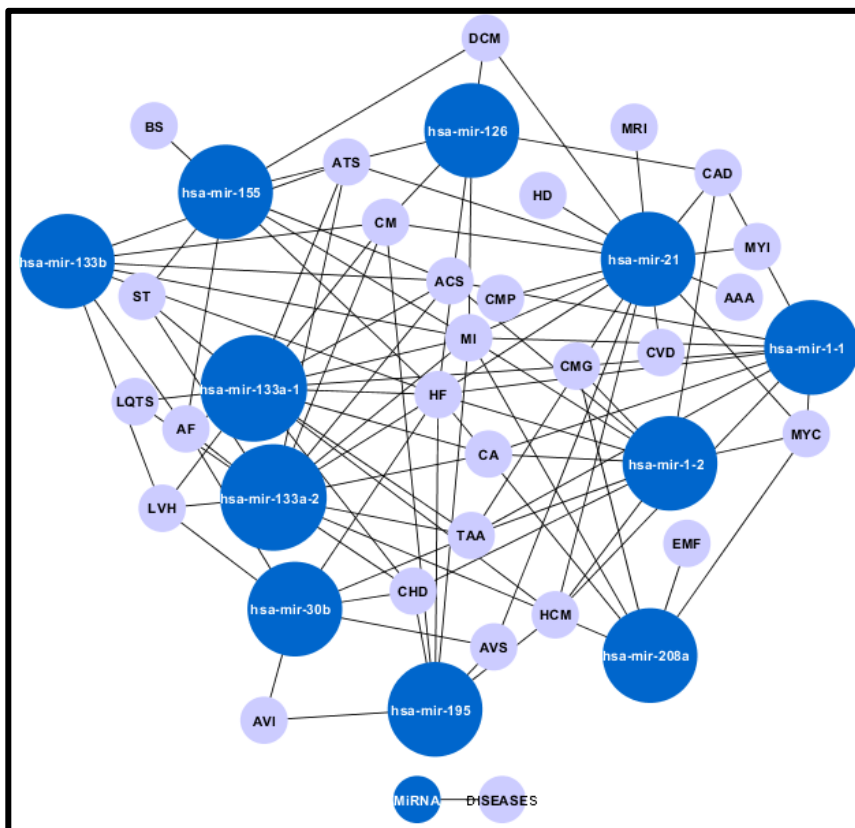


Figure 2: The microRNA-cardiac disease network. (nodes indicated in blue are hub micro RNAs and grey color nodes are diseases. All the blue nodes are hubs implicated in various cardiac diseases). The abbreviated names are read as follows: EMF- Endomyocardial Fibrosis; CMG- Cardiomegaly; CA- Cardiac arrhythmias; MYC- Myocardium; CAD- Coronary Artery Disease; ATS- Atherosclerosis; CM- Cardiac myocytes; LQTS- Long QT syndrome; ACS- Acute Coronary Syndrome; HCM- Hypertrophic cardiomyopathy; ST- Stroke; TAA- Thoracic Aortic Aneurysm; MI- Myocardial Infarction; CHD- Congenital Heart Defects; HF- Heart Failure; BS- Behcet Syndrome; AF- Atrial Fibrillation; LVH- Left Ventricular Hypertrophy; AVS- Aortic Valve Stenosis; MYI- Myocardial Ischemia; HD- Heart diseases; AAA- Abdominal Aortic Aneurysm; CMP- Cardiomyopathies; MRI- Myocardial Reperfusion Injury; CVD- Cardiovascular Disease

Table 1: Disease specific dysregulation pattern

S. No.	Disease	% involvement	Up/Down regulation	p value
1	Coronary artery disease	75 % (15/20)	UP	0.0001
2	Heart Failure	53 % (70/132)	UP	0.001
3	Hypertrophic cardiomyopathy	52.1 % (12/23)	UP	0.0009
4	Myocardial Infarction	52.7 % (19/36)	UP	0.026
5	Myocardial Ischemia	71.4 % (5/7)	UP	0.001
6	Myocardial reperfusion injury	100 % (5/5)	UP	0.0001
7	Myocarditis	100% (2/2)	UP	0.024
8	Stroke	88.8% (8/9)	UP	0.0001
9	Thoracic aneurysm	60% (9/15)	DOWN	0.0006

The p values in the bold are extremely significant

Table 2: Novel cluster associations of microRNAs

miRNA associations	Disease type	Dys-regulation pattern	Type
1-1, 1-2 => 133a-2	TAA, CMG, HCM	D, D, D	Validated [2], [4], [19], [20]
1-1,133a-2 =>1-2,133a-1	TAA, CMG, HCM	D, D, D, D	Validated [2], [4], [19], [20]
133 a-1, 133b => 133a-2	AF	D, D, D	Predicted
133a-1, 155 => 133a-2	LVH, HCM, DCM	D, D, D	Predicted
155 ==> 133a-1	DCM, HCM	D, D	Predicted
155 ==> 133a-1, 133a-2	LVH, HCM, DCM	D, D, D	Predicted
155 ==> 133a-2	DCM, HCM, LVH	D, D	Predicted

TAA = Thoracic Aortic Aneurysm, CMG = Cardiomegaly, AF = Atrial Fibrillation LVH = Left Ventricular Hypertrophy, HCM = Hypertrophic Cardiomegaly, DCM = Dilated Cardiomyopathy D = Downregulated

Discussion

MicroRNAs that occurred most frequently in human cardiac diseases obtained from HMDD are shown in the **Figure 1**. From the data, several hub microRNAs with more than 5 disease associations were found. Based on the derived data, **Figure 2** shows a bi-partite graph based representation of the microRNA-cardiac disease network generated using Cytoscape.

In order to perform the Fisher's two-tailed test, 191 microRNAs associations with 30 cardiac diseases were recorded as 388 disease incidences, as one microRNA was found to dysregulate more than one cardiac disease. The entire dataset comprised of 388 disease incidences, 224 microRNA upregulation events and 164 down-regulation events. MicroRNAs linked with cardiac diseases like atherosclerosis ($P = 0.0001$); coronary artery diseases ($P = 0.0001$); cardiomegaly ($P=0.0002$) and other statistically significant associations (**Table 1**) show distinct dysregulation pattern in a disease specific manner. Up to 19 miRNAs (19/24=79.1% of total upregulated events) were found to be up regulated in atherosclerosis. This emphasizes that the products of genes linked with a disordered associations, interact more with disordered molecules than with others. Hence during or before the onset of a cardiac disease significant number of microRNAs are attracted to interact with one another with distinct dysregulation patterns making the altered physiological state to act as hubs. Thus microRNAs can be used as biomarkers for identifying such states.

To understand the hidden patterns and trends in large-scale data,

Knowledge discovery or data mining methods can be employed. Data mining uses several methods to identify the implicit information, one of which is Association rule mining (ARM). ARM employs algorithms such as Apriori [15], FP Growth [16], Tree Projection [17], etc. We have used the FP Growth algorithm for association rule mining in order to identify novel patterns of associations of all the microRNAs implicated in human cardiac diseases. Most algorithms used in ARM are derivatives of the Apriori algorithm. Since the Apriori algorithm has disadvantages in candidate set generation; the FP Growth algorithm was developed to overcome this problem. The FP Growth algorithm uses a partition-based technique and identifies frequent patterns by pattern fragment growth. SPMF (Sequential Pattern Mining Framework) generated 50 association rules and 26 frequent item sets from 191 microRNAs with a pre-defined threshold support and confidence of 0.2 and 0.6 respectively. Only the biologically significant validated and novel associations are listed in **Table 2**.

MicroRNAs are master regulators of gene functions. They can manifest in robust ways, hence being able to predict the nature of their behavior in a particular disease phenotype is a boon. Computational biology algorithms, techniques and bioinformatics tools have enhanced the significance of predictive science. In this study, predictive behavior of microRNAs in cardiac diseases has been concluded. This is a milestone in adding to the therapeutic potentials of microRNAs in terms of gene-based therapy. Signature expression patterns marks a flag post for where the experimentalists ought to look for. This study

also proves the potential of miRNA genes to function as molecular markers, which can be considered as a pre-diagnostic method to predict the onset of a disease. However the assessment of the above can be done only by experimental validation.

Currently individual experiments are being carried out to probe microRNA expression analysis with implications in several diseases. Even though data integration and updating schema is still a challenge however the bigger challenge lies in identifying and predicting the patterns such as miRNA dysregulation states, chromosome location based co-expression patterns, network modules etc.

Conclusion

Of 191 experimentally validated microRNAs, about 16 microRNAs were identified to act as "hubs" in the microRNA-cardiac disease network. The visualization was done by treating the network as a bi-partite graph based on graph theory analysis. The graph revealed "hubs" in two contexts: 1) disease 2) microRNAs. With reference to Table 1, it is evident that diseases tend to act as "hubs" that attract microRNAs with specific dysregulation patterns. This is statistical proof that if for example: In Stroke 88.8% (8/9) microRNAs are upregulated leading to a collective phenotypic condition leading to stroke, the statistical significance of which was confirmed by Fisher Two tail test similarly up to 19 miRNAs (19/24=79.1% of total upregulated events) were found to be up regulated in atherosclerosis. There may be several other players involved however the presence of "hub microRNAs" like miR 133 a-1; miR 133 a-2 and miR 155 in the stroke and atherosclerosis condition in an upregulated state enables the disease phenotype itself to act as a hub attracting other microRNAs that are mostly upregulated thus proving that the products of genes linked with a disordered associations, interact more with disordered molecules than with others [18]. Our study therefore emphasizes that judicious use of computational analysis techniques for large scale biological data analysis would reveal novel biological insights towards therapeutics which may not be apparent by individual experimental analysis alone.

References

- [1] Gámez B *et al.* J Mol Endocrinol. 2014 **52(3)**: 179 [PMID: 24523514]
- [2] Liu K & Wang R, J Theor Biol. 2012 **304**: 103 [PMID: 22554948]
- [3] Ikeda S *et al.* Mol Cell Biol. 2009 **29(8)**: 2193 [PMID: 19188439]
- [4] Iorio MV *et al.* Breast 2011 **20(3)**: S63 [PMID: 22015296]
- [5] Finnerty JR *et al.* J Mol Biol. 2010 **402(3)**: 491 [PMID: 20678503]
- [6] Jones JA *et al.* Circ Cardio Vasc Genet. 2011. **4(6)**: 605 [PMID: 22010139]
- [7] Kato M & Slack FJ Biol Cell 2008 **100(2)**: 71 [PMID: 18199046]
- [8] Mathers CD & Loncar D, PLoS Med, 2006. 3(11): e442. [PMID: 18923704]
- [9] www.who.int - Global status report on non-communicable diseases, Geneva, World Health Organization, 2011
- [10] Yoneki E *et al.* Social Nets 2008 7 doi: 10.1145/1435497.1435499
- [11] Shannon P *et al.* Genome Research 2003 **13(11)**: 2498 [PMID: 14597658]
- [12] Lui M *et al.* PLoS ONE 2008 **3**: e3420. [PMID: 18923704]
- [13] Asratian AS *et al.* Bipartite Graphs and their Applications, Cambridge Tracts in Mathematics 1998 131, Cambridge University Press, ISBN: 9780521593458.
- [14] Fournier-Viger P *et al.* 2013 URL: <http://www.philippe-fournier-viger.com/spmf/>
- [15] Agrawal R & Srikant R, Proc. of 20th Int'l conf. on VLDB, 1994 page 487 URL: <http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>
- [16] Han Jiawei *et al.* Data Mining and Knowledge Discovery, 2004. **8(1)**: 53 Kluwer Academic Publishers.
- [17] Agarwal R *et al.* 2001 **61**: 350 <http://charuaggarwal.net/jpdc.pdf>
- [18] Goh K *et al.* Proc Natl Acad Sci USA, 2007 **104(21)**: 8685 [PMID: 17502601]
- [19] Basson M Nat Med. 2007 **13(5)**: 541 [PMID: 17479098]
- [20] Wystub K *et al.* PLOS Genetics 2013 **9(9)**: e1003793. [PMID: 24068960]

Edited by P Kanguane

Citation: Krishnan *et al.* Bioinformatics 13(1): 17-20 (2017)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License