High AU content: a signature of upregulated miRNA in cardiac diseases

Richa Gupta, Neha Soni, Purbasa Patnaik, Ishita Sood, Rahul Singh, Kamal Rawal, Vibha Rani*

Department of Biotechnology, Jaypee Institute of Information Technology University, Noida, U.P-201307, India. Vibha Rani - E-mail: Vibha.rani@jiit.ac.in, Vibhasandeep@rediff.com; *Corresponding author

Abstract:
MicroRNAs have been implicated for the regulation of gene expression. These miRNA are a class of single stranded non coding RNAs, formed from endogenous transcripts and measure typically about 19-25 nucleotides in length. They are important regulators of the various biological and metabolic functions taking place in humans. Many miRNAs show tissue specific expression. Human heart is a complex organ which during various diseased and developed conditions shows differential expression of miRNA. Here, we overview the recent findings on miRNA in cardiac diseases and report the presence of high AU content in differentially expressed miRNA in developed and diseased condition of heart as compared to all the miRNA present in the human. A total of 905 human miRNA sequences taken from miRBase were computationally analyzed. Trend analysis was performed to study the influence of positional frequency of the nucleotides. This study will help us in understanding the significance of AU rich elements in miRNA during the development of cardiac diseases.

Key Words: Cardiac, miRNAs, Regulation, miRBase, AU content, Bioinformatics

Background:
Gene expression in mammalian system is influenced by various factors. One such factor is a class of small non coding RNA sequences- micro RNA (miRNA), small interfering RNA (siRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and many more. Some of these non-coding RNAs are involved in silencing the gene expression by the process of RNA interference. RNA interference is an effective mechanism for sequence specific regulation of gene expression [1].

MicroRNAs are small, non-coding regulatory RNA of around 22bp length that are the key regulators of gene expression. They bind to specific sites in the 3'UTR region of the transcribed genes and inhibit protein expression either by mRNA degradation or by inhibition of translational processes, by cleavage or by repressing translation by complementary binding [2]. They play significant roles in various important biological processes, including diseases, developmental processes, signal transduction, cell maintenance and differentiation [3, 4]. MiRNAs are evolutionarily ancient component of genetic regulation and are well conserved in eukaryotic organisms. Till date, more than 900 human miRNAs have been reported, and estimated to regulate more than one third of cellular messenger RNAs. Various experiments have been done which relate dysregulated expression of miRNAs to a variety of diseases, such as neurodegenerative diseases, cancer, cardiovascular diseases and viral infections [5, 6].

Many miRNAs are enriched in a tissue-specific manner [7], for example miR-1, miR-16, miR-27b, miR-30d, miR-126, miR-133, miR-143, and the let-7 family are abundantly expressed in adult cardiac tissue. The heart may also contain many other non-cardiomyocyte cell types, such as endothelial cells, smooth muscle cells, fibroblasts, and immune cells, which have distinct miRNA expression profiles. During cardiac diseases and developemental changes, miRNA pattern profiles show differential pattern. MiRNAs otherwise silenced in adult cardiac cells, tend to get re-expressed in several cardiac diseases [8, 9]. Chronic exposure to stress signals eventually leads to impaired function which finally results in cardiac failure. The involvement of miRNAs in this pathological process has been recently found [8, 10].

MicroRNAs control gene expression by inhibiting the translation and destabilizing the mRNAs. Emerging evidence suggests a direct link between miRNA and diseases. The mammalian cell sequences with a high AU content have been shown to cause mRNA instability [13].

Methodology:
The microRNA dataset comprises experimentally validated sequences derived from published data. The information was classified with respect to the association with their expression in heart diseases as reported in the literature. The sequence data was primarily derived from miRBase (http://www.mirbase.org/) for humans.

miRBase is a centralized online repository for nomenclature, sequence data, annotation and target prediction of all published miRNA. It provides a user-friendly interface for miRNA. It provides a range of data to facilitate studies of miRNA genomics. All miRNAs are mapped to their genomic coordinates. Clusters of miRNA sequences in the genome are highlighted, and can be defined and retrieved with any inter-miRNA distance. Sequences from the database as mature sequences in unaligned FASTA format were retrieved [11].

Sequence composition namely A%, U%, G%, C%, A+U% & G+C% was estimated manually and computationally using BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html), inhouse perl scripts as well as DNA SCANNER software (http://nldsps.jnu.ac.in/bioit/ccbb/dnascanner_help.html) [12].

Disparity in sequence length was eliminated through alignment and positional frequency for the four nucleotides was estimated using script MSA.pl for both all (control) and the upregulated (test) sequences. The relative frequency (RF) was computed by dividing the positional frequency

(PF) by the total number of sequences and further estimation was done by Trend Analysis [13].

RF was calculated for each position in aligned sequences and was plotted in MS-Excel software (**Figure 1**). The positional variation was estimated using Trend Analysis for both all (control) and upregulated (test) dataset. The function **COUNTIF** was used to count the number of nucleotide at each position and hence the positional frequencies were known. At each position of the sequence the relative frequency of each nucleotide was calculated by dividing the positional frequencies by the total number of sequences for both the groups which are the control and test miRNAs.

Relative frequency = Positional frequency/ Total No. of sequences

Using the relative frequencies calculated for each of the nucleotide at all positions the graphs were plotted to compare the variations in the pattern of position wise nucleotide frequency distribution between the upregulated test miRNA and the control miRNA. Then TREND ANALYSIS for each of the nucleotides was performed and added to these line graphs to compare the trend of nucleotide frequency distribution for each nucleotide base in test and control miRNA dataset.

Results and Discussion:
The upregulated miRNA sequences show a significant increase in the AU content. The average AU and GC content of control miRNAs were calculated to be 50% each by our analysis. On the other hand the AU and GC content of test miRNA sequences were found to be 56% and 44%. We observed that there is a significant increase in the AU content in the upregulated miRNA sequences (**Figure 1**). Analysis of the AU and GC content in the downregulated miRNA sequences were also done but difference was not significant (data not shown).

Trend Analysis: To visualize the trend of frequency distribution for each nucleotide base the trend analysis graphs were plotted for control and test sequences. For nucleotide "A" the trend analysis showed a concave downward curve for both control and test sequences with the trend line of test curve lying above the control (**Figure 2a**) indicating higher frequencies of this nucleotide in upregulated sequences. The trend for nucleotide "U" showed a concave downward curve for the control sequences whereas the test sequences showed almost a linear curve (**Figure 2b**). Nucleotide "G" showed concave downward curves for both test and control sequences (**Figure 2c**). For nucleotide "C" the trend of test

sequences was similar to that for G, as it showed a concave downward curve for both test and control sequences (**Figure 2d**). The comparative trends of normal and test sequences are shown in **Figure 2e and 2f** respectively. These trends noticeably indicate that since the frequency of nucleotide "A" and "U" both is higher in the test sequences, hence the average AU content becomes high altogether for these sequences.

The AUGC frequency distribution at each of the nucleotide positions was analyzed for control and compared with that for the test by plotting line graphs from the data calculated from these sequences (**See Supplementary material**) When these graphs were compared for the control and test sequences, the test graph were observed to have a considerable variation at each position and for each nucleotide, thus indicating role of sequence composition in the upregulation of miRNA in cardiac cells. Further control of test sequences was subjected to analysis using DNA SCANNER such as A-rule, thermodynamic parameters etc. to generate dinucleotide distributions of miRNA sequences.

Conclusion:
The AU richness in miRNA genes involved in cardiac diseases may be of importance but requires further studies. AU richness and other parameters may play a role in their prediction. This is a preliminary study where we hypothesise role of sequence based parameters in identification of upregulated miRNA involved in cardiac diseases.

AU-rich elements (AREs) can act as an important paradigm for post-transcriptional regulation as a control of cytoplasmic mRNA in the 3' untranslated region of transcripts encoding oncoproteins, cytokines and transcription factors. Many RNA-binding proteins have been shown to bind to AREs in vitro. Therefore higher AU content may reflect better binding of certain proteins and thus, a stress condition of heart [14].

Sequence composition as well as sequence based patterns (thermodynamic features) plays a role in several biological functions. Sequence composition as well as sequence based patterns such as protein induced deformability, propeller twist etc. play a role in identification of genomic features such as Transcription start sites (TSS), insertion sites retrotransposons etc. Therefore it would be important to understand genomic neighbourhood of these upregulated miRNA. Further experiments are needed to understand the role of high AU content in upregulated miRNA.
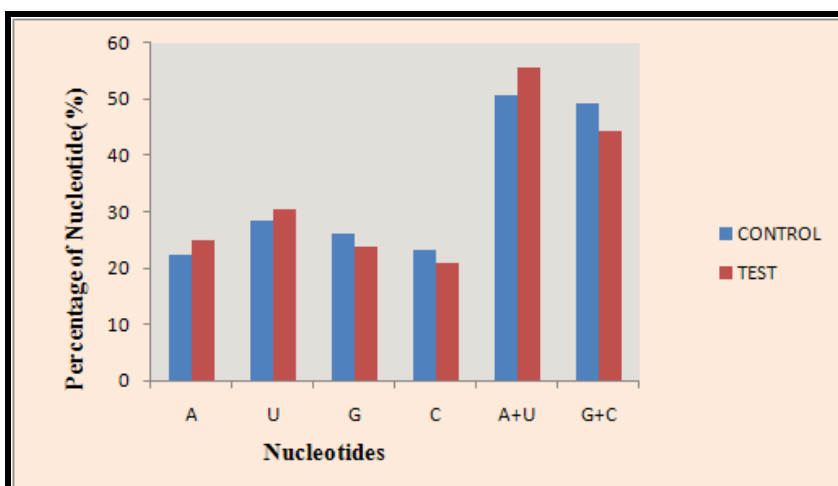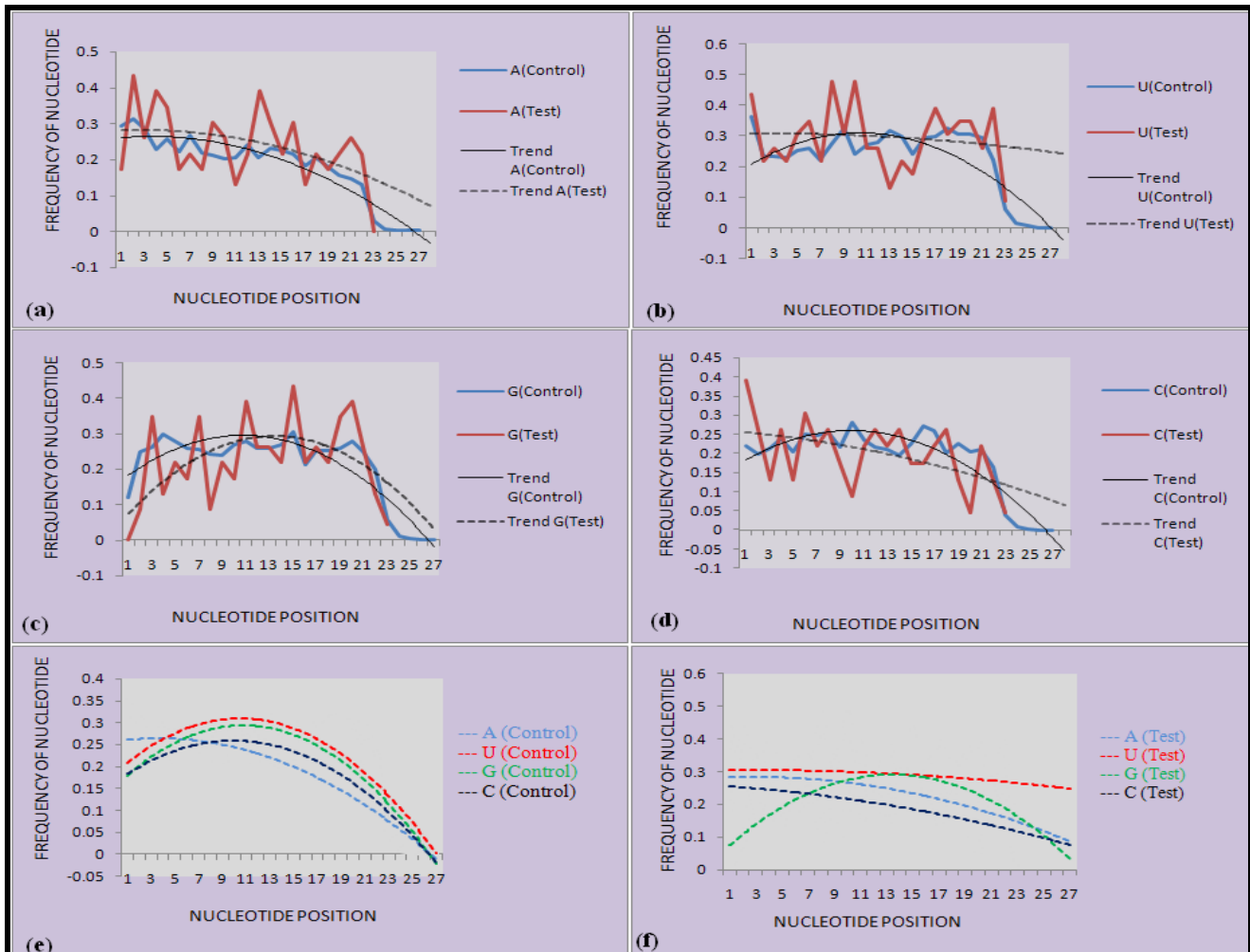

Figure 1: A comparative analysis of the mononucleotide A, U, G & C and dinucleotides G+C & A+U, calculated for the normal and upregulated miRNA

Figure 2: Positional frequency distribution (a), (b), (c), (d) show the comparison between test (upregulated) and control (normal) miRNAs of nucleotide A, U, G and C respectively. The plots also show the trend lines for each of the nucleotide's positional frequency in case of both test and control datasets. (e) and (f) show the trends of all four nucleotides in normal and upregulated miRNAs respectively. The x-axis is the nucleotide position and y-axis is the frequency of occurrence of nucleotide.

References:

[1]   K Shrey *et al. Biochem Biophys Res Commun.*(2009) **386**: 273 [PMID: 19523438]

[2]   DP Bartel *Cell* (2004) **116**: 281 [PMID: 14744438]

[3]   B Zhang *et al. Comput Biol Chem* (2006) **30**: 395 [PMID:17123865]

[4]   B Zhang *et al. Gene* (2007) **397**:26 [PMID: 17574351]

[5]   Z Liu *et al. Int J Mol Sci* (2008) **9**: 978 [PMID: 19325841]

[6]   GA Calin *et al. Proc Natl Acad Sci U S A* (2004) **101**: 2999 [PMID: 14973191]

[7]   P Landgraf *et al. Cell* (2007) **129**: 1401 [PMID: 17604727]

[8]   T Thum *et al. Circulation* (2007) **116**: 258 [PMID: 17606841]

[9]   J Bauersachs *et al. Eur J Clin Invest* (2007) **37**: 829 [PMID: 17973779]

[10] D Catalucci *et al. Front Biosci* (2008) **13**: 312 [PMID: 17981549]

[11] SJ Griffiths *et al. Nucleic Acids Research* (2008) **36**: D154 [PMID: 17991681]

[12] K Rawal *et al. Nucleic Acids Research* (2006) **34**: 5752 [PMID: 17040894]

[13] AK Mishra *et al. Bioinformation* (2009) **4**: 151 [PMID: 20198191]

[14] SS Peng *et al. EMBO J* (1998) **17**: 3461 [PMID: 9628881]

Supplementary materials:
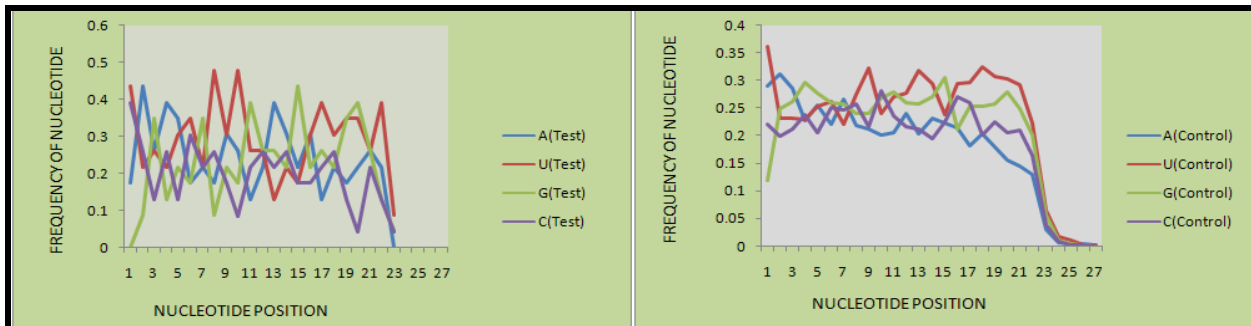
Nucleotide frequency distribution of miRNAs: (a) Nucleotide position wise AUGC distribution of all miRNAs. (b) The nucleotide position wise AUGC distribution for Upregulated miRNAs.