OPEN ACCESS Freely available online

Gene Technology

Research Article

Quantification of miRNAs and Their Networks in the light of Integral Value Transformations

Sk. Sarif Hassana*

Applied Statistics Unit, Biological Sciences Division and Human Genetics Unit, Indian Statistical Institute, Calcutta, India

ABSTRACT

MicroRNAs (miRNAs), which are on an average only 21-25 nucleotides long, are key post-transcriptional regulators of gene expression in prokaryotes, metazoans and plants. A proper quantitative understanding of miRNA is required to comprehend their structures, functions, evolutions etc. In this paper, the nucleotide strings of miRNAs of three organisms namely Homo sapiens (hsa), Pan troglodytes (ptr) and Macacamulatta (mml) have been quantified and classified based on some characterizing features. A network has been built up among the miRNAs for these three organisms through a class of discrete transformations namely Integral Value Transformations (IVTs). Through this study, we have been able to nullify or justify one given nucleotide string as a miRNA. This study will help us to recognize a given nucleotide string as a probable miRNA, without the requirement of any conventional biological experiment. This method can be amalgamated with the existing analysis pipelines, for small RNA sequencing data (designed for finding novel miRNA). This method would provide more confidence and would make the current analysis pipeline more efficient in predicting the probable candidates of miRNA for biological validation and filter out the improbable candidates.

Keywords: miRNAs; Integral Value Transformations; Fractal dimension; Hurst Exponent and Mean Order

DESCRIPTION

Mature microRNAs (miRNAs) are a class of naturally occurring, small non-coding RNA molecules, of length~21-25 nucleotides [1]. Mature miRNAs are partially or completely complementary to one or more messenger RNA (mRNA) molecules. MiRNA binds to target, mostly at 3' UTR region in the metazoan although two exceptions have been reported recently [2]. In plants, targets can be located in the 3' UTR but mostly in the coding region. Perfect complementarity of MiRNA's seed region with the target region is an essential issue for successful target binding, though a recent study has shown that alternative is also possible [3].

The primary function of miRNAs is to down-regulate gene expression in a variety of manners, including translational repression, mRNA cleavage, and de-acetylation. They were first discovered in 1993 by Lee and colleagues, and the term microRNA was coined in [1]. Even in the prokaryotes miRNA like functional analogue have been reported [4]. Thousands of miRNAs genes found in intragenic regions or in anti-sense orientation to genes have since been identified in various organisms through random cloning and sequencing or computational prediction. Around 40% of miRNA genes lie in the introns of protein and non-protein coding genes or even in exons [5-7].

It is to be noted that several miRNAs have been reported to have links with certain types of cancer. Genetic signatures of miRNAs are also required for individualized cancer treatment strategies as reported by C. Hatzis et al [8,9]. A proper quantitative understanding of miRNAs is required to comprehend their structures, functions, evolutions etc. which will also help us to recognize a given string as a miRNA. There are many studies on DNA sequence quantification and representation in exon, intron level have been done.

In this paper nine mathematical features viz. Fractal dimensions (FD) of 4 binary indicator matrices, FD for DNA walk, variance & complexity measures, Hurst exponent and mean ordering of

Correspondence to: Sk. Sarif Hassana, Applied Statistics Unit, Biological Sciences Division and Human Genetics Unit, Indian Statistical Institute, Calcutta, India, E-mail: sarimif@gmail.com

Received: March 01, 2021; **Accepted:** March 15, 2021; **Published:** March 22, 2021

Citation: Sarif Hassana Sk (2021) Quantification of miRNAs and Their Networks in the light of Integral Value Transformations. Gene Technol. 10: 162.

miRNA strings were adumbrated using the miRNA databases of three organisms viz. hsa (Homo sapience) (Pan troglodytes) and mml (Macaca mulatta). Besides, a primary classification based on poly-string mean order is made for the miRNAs of the said three organisms. A network has been established among the miRNAs through Integral Value Transformations (IVTs) and validated via those nine mathematical features. Finally, seven examples of strings are given, six of which are probable and one is non-probable as per quantitative studies. In this section, some basics of Fractal and Integral Value Transformations (IVTs) have been discussed to warm up [10].

Our artificial world can be described easily through Euclidean geometric shapes, but there are many things in nature such as shape of cloud, geometry of lightening etc. could not be described through Euclidean geometry. Many mathematicians descended the challenge for a fair enough description of natural objects but after a long period in 1975, B. Mandelbrot took the challenge and gave the birth of a new geometry to describe nature which is known to us as 'Fractal Geometry' in short 'Fractal'. The precise definition of "Fractal" according to Benoit Mandelbrot is as "a set for which the Hausdroff Besicovitch dimension strictly exceeds the topological dimension". To gain a quantitative insight of Fractal, some fractal parameters namely Fractal dimension, Hurst exponent, Secularity, Lacunarity, Hurst exponent etc. are also introduced in the literature. A brief discussion follows about one of the well-known methods of calculating fractal dimension namely 'Box-Counting method'.

Box-Counting Method: This method computes the number of cells required to entirely cover an object, with grids of cells of varying size. Practically, this is performed by superimposing regular grids over an object and by counting the number of occupied cells. The logarithm of N(r), the number of occupied cells, versus the logarithm of 1/r, where r is the size of one cell, gives a line whose gradient corresponds to the box dimension [11].

METHODS AND RESULTS

Variance analysis

From the results as shown in table 4, it has been observed that the variance of hsa-miR string 46 is minimum and the variance of hsa-miR string 210 is maximum among all the miRNAs of the three species i.e. there are certain Human miRNAs which are adjacent to the mean and certain miRNAs which are distant from the mean. The intervals of variances of ptr and mm miRNAs are contained in the interval of variances of hsa-miRNA string.

Hurst exponent of miRNA strings

Hurst exponent is referred to as the "index of dependence," and is the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction. It is a measure of long range correlation of one-dimensional time series.

Mean ordering of miRNA strings

A miRNA is a string constituting of different permutations of the base pairs A, C, U & G where repetition of a base pair is allowed. The classification of miRNAs is made based on ploy-string mean of miRNA strings. In poly-string mean, we have taken the count of homogeneous poly-string of different nucleotides of different lengths for all miRNAs. We can classify the miRNA sequences based on the ordering of poly-string mean of A, C, U, & G in the string. Given a string X, we calculate the mean of poly-strings consisting only of A, C, U & G separately.

Analysis of poly-string mean

There are 256 possible poly-strings of length four using four nucleotides A, U, C and G. It is worth noting that all the hsa-miRs, ptr-miR and mml-miRs have been classified into only 73, 71 and 61 classes respectively according to the order of poly-string mean. Mean ordering of the three species enlightens that the mean of all the four nucleotides are never equal. The classification of miRNAs based on mean ordering is given as in Dataset S5.So far hsa-miR, ptr-miR and mml-miR's quantifications are done. In the subsequent section, a mathematical network among the miRNAs has been deciphered in the light of IVTs.

Use of IVT for evolution of miRNAs

In this section, a network has been established through 4-adic IVTs. We have used nine IVTs from the set T 4, 1# to evolve each of the mml-miRs (without loss of generality). So corresponding to each of those miRNAs, there are nine distinct strings which have been blasted in the miRNA database (miRBase) to have significant similarities. Also all

The features for the blasted strings have been enumerated in Dataset S4. For example, let us consider mmu-mir-874 and the blast result for it is given in Table 6. Through these nine IVTs, it is possible to have a network among all the miRNAs over all the species with closest features as explained earlier.

Through this ample study of quantifications of miRNAs, is it possible to justify or nullify a given string as probable miRNA?

DISCUSSION

We have developed the analysis protocol, using nine mathematical features in characterizing miRNA strings, which is a novel approach of quantifying the strings pattern. Here we are reporting the pattern of all reported mature miRNA strings from Mir Base of three different organisms, viz. Homo sapiens, Pan Troglodytes and Macaca mulatta. On the basis of these observed features, we have established a network among different miRNAs of a given organism through implementation of IVTs and determined the bounded thresholds for all of the above mentioned mathematical features for miRNAs of the studied three organisms. On the basis of this extracted information this prescribed pipeline can now convincingly characterize a given string and define it whether as a probable mature miRNA or a non-probable mature miRNA string, at least for the 3 studied organisms, without any prior conventional

biological experiment(s). This adds to one prominent dimension to this study. Since efficient and successful prediction of miRNA is still one of the major issues in bioinformatics, and our developed protocol would add efficiency to already existing mature miRNA prediction tools, we believe our work is appropriate in present time frame.

Owing to increasing popularity of massively parallel sequencing techniques for RNA, a rapid increment in number of validated miRNA is occurring. Sequencing protocol for small RNA has been standardized both in SOLiDTM platform (Invitrogen) and Hi-Seq platform (illumina). First step of post sequencing analysis is to search for sequence match with the databases like miRBase for already validated miRNA sequences. The left out small RNA fragments are then filtered through an analysis pipeline to find out small number of most probable miRNA candidates for novel miRNA biological validation experiment(s). More and more efficient pipeline is required for more efficient prediction of probable miRNA candidate(s) to reduce the load of costly, cumbersome and may be redundant biological experiments in a greater way. Amalgamation of existing analysis pipeline with the proposed protocol will make the prediction of probable candidate(s) and exclusion of non-probable candidate(s), far more efficient. Above all, the proposed method can draw inference just reading around 26 nt long character strings only. So, this method would be applicable for very short read lengths, which are as small as 26 nt in length.

In near future, we would be extending our study for rest of organisms having quite a large number of already reported miRNA in Mir Base.

REFERENCE

1. Lee RC, Feinbaum R L, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-854.
2. Wang WX, Wilfred BR, Xie K, Jennings MH, Hu YH, Stromberg AJ, et al. Individual microRNAs (miRNAs) display distinct mRNA targeting "rules". RNA Biol. 2010;7(3): 373-380.
3. Peter M E. Targeting of mRNAs by multiple miRNAs: the next step. Oncogene. 2010;29(15): 2161-2164.
4. Tjaden B, Goodwin S, Opdyke JA, Guillier M , Fu DX, Gottesmanet S, et al. Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res. 2006;34 (9): 2791-2802.
5. Ruvkun, G. Molecular Biology: Glimpses of a Tiny RNA World. Science. 2001; 294(5543): 797-799.
6. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in C elegans. Science. 2001;294(5543): 858-862.
7. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res. 2004;14(10A): 1902-1910.
8. He L, Thomson JM, Hemann MT, Goodson S, Powers S, Cordon-Cardo C, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043): 828-833.
9. Hatzis C, Valero V, Booser DJ, Esserman L, Lluch A, Holmes F, et. al. A Genomic Predictor of Response and Survival Following Taxane-Anthracycline Chemotherapy for Invasive Breast Cancer J. Am. Med. Assoc. 2011;305(18):1873-1881.
10. Xiao Y, Chen R, Shen R, Sun J, Xu J. Fractal dimension of exon and intron sequences. J Theor Biol. 1995;175(1):23-26.
11. Shi-Yuan W, Feng-Chun T, Xiao L, Jia W. A Novel Representation Approach to DNA Sequence and Its Application. Signal Processing Letters IEEE. 2009;16(4):275-278.