# Underlying Effect of miRNA Patterns for Identification of Human Diseases

Ranjeet Kumar Rout
Applied Statistics Unit
Indian Statistical Institute,
Kolkata-700018, India

Pabitra Pal Choudhury
Applied Statistics Unit
Indian Statistical Institute,
Kolkata-700018, India

Beauty Barman
Sikkim Manipal Institute of
Technology, Sikkim, India

Saurabh Singh
Sikkim Manipal Institute of
Technology, Sikkim, India

Uttari Chakraborty
Sikkim Manipal Institute of
Technology, Sikkim, India

## ABSTRACT
MicroRNAs (miRNAs) originally an unappreciated class of small ribonucleic acids (RNAs) were diagnosed as moderate biological modifiers. Over the years rigorous research on microRNAs established them as a powerful regulator of various cellular processes like occurrence of different diseases. In this paper cellular automata approach is used by replacing the miRNA sequences with binary forms and using an evolutionary rule those sequences can be converted into distinct images. This is immensely helpful as the cellular automata images of the miRNA sequences help in deciphering various vital aspects which were previously unknown and thereby rendering it possible to find a pattern. Using this concept it becomes easier to distinguish amongst the different miRNA images and hence identify the influence of the segments of the miRNA sequences on the pattern of the corresponding images**.**

## Keywords
miRNA, cellular automata, State Space Images.

## INTRODUCTION
The role of miRNAs in normal cellular development and metabolism is of huge importance. Just as miRNA is involved in the normal functioning of eukaryotic cells, so has the dysregulation of miRNA been associated with diseases [4]. The loss or gain of miRNA function can be caused by a single point mutation in either the miRNA or its target. There are proofs which bring to light that miRNAs are crucial in cell growth, tissue differentiation, cell proliferation, embryonic development and cell proliferation. Also the mutation and the dysregulation of miRNAs lead to diseases like cancer, Fragile-X mental retardation syndrome, diabetes, chronic hepatitis, AIDS, obesity and the like. As explained in research papers [1] and [2] by using a specific rule on a specific type of micro-organism, its image can be analyzed for any distinguishing patterns. Moreover how exactly each of the iterations of rule gives rise to a specific space time evolution image is given in [3]. MicroRNAs (miRNAs) are a family of short non-protein-coding RNAs having diverse functions including regulation of cellular differentiation, proliferation and apoptosis [9]. It is interesting to speculate that miRNA expression signatures have been shown to be promising biomarkers for cervical cancer [10] prognosis [11]. A Cellular Automata (CA) based classifier to recognize the coding of a DNA succession has been proposed in [12][13].

In this paper a specific rule is used to understand how the miRNA sequences influence the patterns of their corresponding evolutionary space time images. Therebyexploring the potential roles miRNAs can play in a variety of diseases and helping to correlate the miRNAs responsible for human diseases. The rule in consideration is 172 which is one of the 3-variabe cellular automata rules [7] and as compared to other rules the images obtained from the rule 172 is very distinctive. When rule 2, rule 18, rule 29, rule 96, rule 98 and rule 172 were applied to the same disease corresponding to the same miRNA sequence, the obtained images are shown in Fig 1.

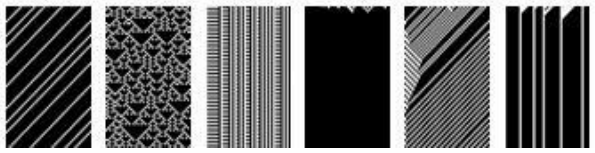

**Fig 1: Images of Rule 2, Rule 18, Rule 29, Rule 96, Rule 98 and Rule 172 respectively.**

As evident from the images above in Fig.1 the patterns obtained by applying the rules specified above are not clearly distinguishable. Thus, comparing the images of the different miRNA sequences becomes much easier when rule 172 is taken instead of the other rules as the images are easily distinguishable from one another and some decision may be taken on the patterns, if any. Moreover on comparison with rule 184 as discussed in the research paper [2], whose evolutionary images helped in contributing significantly to distinguish between *SARS corona virus* and *non-SARS corona virus*, however could not be used to decipher much when the same was applied to human disease-causing miRNA. For example, in case of the disease *Activation of caspases cascade*, the space time evolution image corresponding to the respective miRNA is as follows:

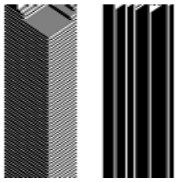

**Fig 2: Images of Rule 184 and Rule 172 respectively.**

Hence, it can be easily observed that by using rule 172 one can easily state some features about the miRNA sequence, using the patterns seen on the image. The same cannot however be said in case of rule 184 as the image does not give any clear indications of any observable pattern 'CAAAUUCGUAUCUAGGGGAAUA' is the miRNA sequence for which the above set of images in Fig. 2 are obtained. Thus, it can be explained for rule 172 that the existence of parallel vertical lines is due to the repeated alternating pattern of the characters C and G whereas A and U attribute to black spaces. This is better explained in section 2 of this paper. Now with respect to rule 184 no such deductions can be made regarding the same miRNA sequence.

# 1. METHOD

These studies have produced a large number of miRNA-disease associations and shown that the mechanisms of miRNAs implicated in diseases are very complex. Therefore, a large-scale analysis of the space evolution diagrams offers a platform to understand the mechanisms of miRNAs in disease. A miRNA sequence generally consists of nucleic acids or bases in succession. The four nucleic acids are Adenine (A), Cytosine (C), Guanine (G) and Uracil (U). In order to cope with the difficulty of finding any characteristic feature of any miRNA sequence which is very long, the images from the space time evolution of cellular automata is used. Firstly, the miRNA sequences are coded in a binary mode wherein a nucleotide sequence is coded as follows: A = 00; U =11; C =10; G =01; As, Adenine (A) and Uracil (U) are complimentary to each other hence their corresponding binary sequences are complimentary as well. Similar, is the case of Cytosine(C) and Guanine (G). Hence using the above scheme, a miRNA sequence is transformed into a series of digital signals. For example, the sequence ''UACAGUG'' is transformed to ''11001000011101''. Assuming the array 'mat' stores each miRNA sequence and 'a', 'b', 'c' indicates cells of 'mat', where 'a' is the left neighbour, 'b' is the current cell and 'c' is the right neighbor. The procedure is as follows:

1. For initial condition i=1, $a = mat(1, length(mat))$, $b = arr(j-1, i)$ and $c = arr(j-1, i+1)$.
2. For final condition i=length (mat), $a = arr(j-1, i-1)$, $b = arr(j-1, i)$ and $c = mat(1, 1)$.
3. For all other condition, $a = arr(j-1, i-1)$, $b = arr(j-1, i)$ and $c = arr(j-1, i+1)$.

The rule used for evolution of the images must be able to distinguish the images of the miRNA sequences. It was found that 172nd rule distinctly helped in distinguishing the images among all the 256 kinds of evolving rules. It was observed that when the time was around 300 the structure of the image was fixed.

When transforming the 2D array (matrix) into a binary image with visualization techniques, the basic bitmap format is chosen because its property is easily handled. In this way, if the matrix element was zero, the color of the counterpart pixel bit will be black; otherwise, white. Moreover sometimes the obtained image is too huge in size for some long sequences. As such the compression of the image is needed that is actually to highlight the characteristic of the image.

# 2. RESULT AND DISCUSSION

In this paper it has been found that among the 256 evolving rules some are better than the others in building evolutionary images for a given miRNA sequence. For example, Rule 172 is the most befitting for finding patterns in the miRNA sequencesresponsible for human diseases. The miRNA sequence and image produced is one to one mapping if the rule and time for the evolution are constant. Different rules have been applied to analyze the 43 human diseases, but only when Rule 172 is used, the images of the different miRNA sequences are most distinguishable. These images are mainly with vertical parallel lines pattern with triangles (which may be complete or incomplete) of varied magnitude attached to them. By analysing the different images corresponding to the miRNA sequences, few remarkable patterns for the class of miRNA affecting a disease were deciphered. It has been observed that if the image is black or is predominantly black then the miRNA sequence has higher frequency of occurrence of the character 'A' (i.e., 'AA', 'AAA', and 'AAAA'). Similarly, if the image is white or is predominantly white then the miRNA sequence has higher number of U's (i.e., 'UU', 'UUU', and 'UUUU') because A and U are complimentary to each other. In the miRNA sequence if the frequency of occurrence of the characters A and U, or A and G, or A and C are same, like 'AAAAAAAAAAAUUUUUUUUUU' then the triangles are observed in the image. Since A and U are complimentary to each other, similar results are obtained both the cases as observed. if no triangle is seen at all in the images then it indicates presence of combinations of the characters 'UA', or 'UG' or 'GC' taken together.

From the space time diagrams few patterns which are common to a specific disease were deciphered. For example: the disease Cardio-genesis the pattern "AAAG" was a recurring pattern in all the miRNA's responsible for it. Thereby, proving that the miRNA's affect that very disease. Similarly few other patterns recurring in some diseases are tabulated in Table 1 along with their corresponding images.

**Table 1.Diseases and its patterns**

| Diseases | Patterns | Fig. |
|---|---|---|
| 1. Cardio-genesis | "AAAG" | (a) |
| 2. Carbohydrate metabolism | "CUGG" | (b) |
| 3. Akt pathway | "CUU" | (c) |
| 4. Apoptosis | "UUG" , "AAA" | (d) |
| 5. Anti- cell proliferation | "UAGG" | (e) |
| 6. Activation of caspases cascade | "GUA", "GAU" | (f) |
| 7. Adipocyte differentiation | "GUG" | (g) |
| 8. Smooth muscle cell fate | "GAA" | (h) |
| 9. Haematopoiesis | "UGG" | (i) |

In Fig 3(a) "CUGG" is present in all the miRNA's affecting Carbohydrate metabolism and the observed image contains more number of stripes attributed to the recurrence of 'C's and 'GG'. In Fig 3(b) the pattern "CUU" is recurrent and the image contains bisections in the triangles mostly which are to the left. However in Fig 3(c) there are occurrences of stripes

with triangles and bisections due to "UUG" but the stripes are fewer in number which can be explained due to the recurrence of "AAA". For Fig 3 (e) there are presences of stripes and triangles, however due to the presence of "UA" and "AU" some blank spaces are seen.

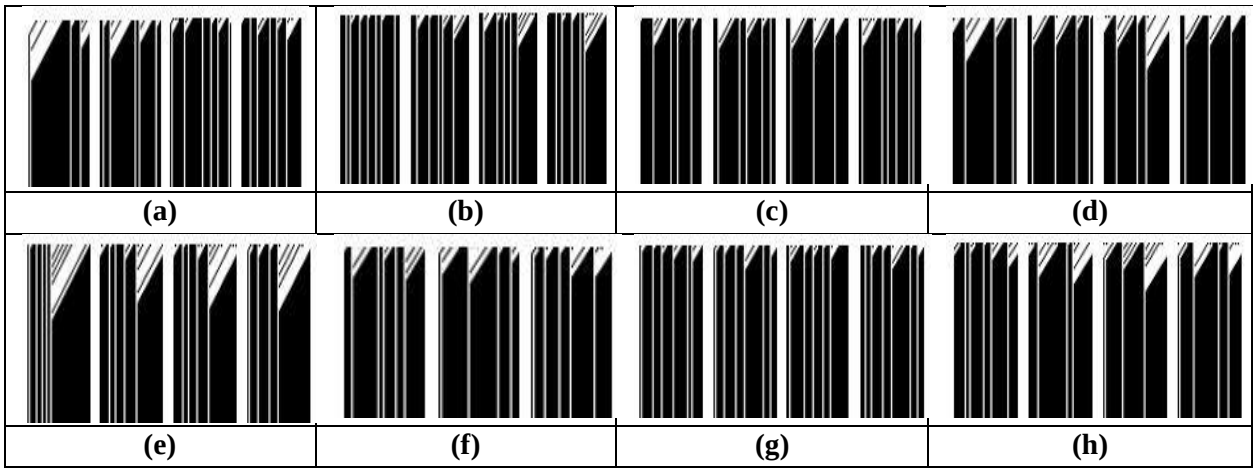


**Fig 3: Images of Images of different diseases as given in Table 1.**

## 3. Conclusion

It is demonstrated through this study that the cellular automata approach can be used by converting the complicated biological sequences, with the help of an evolutionary rule, into distinct images. These images help in finding a relationship between the various disease-causing miRNA and thereby finding a pattern among them. Hence it is possible to recognize the impact of the miRNA sequences on the pattern of the images.

## 4. REFERENCES

[1] X. Xiao et al. Bio-Informatics Research Center, "Using cellular automata to generate image representation for biological sequences",( 2005)Amino Acids. Vol. 28,( 2005) I-1, pp 29-35 .

[2] X. Xiao et al. Bioinformatics Research Center. "Using cellular automata images and pseudo amino acid composition to predict protein subcellular location" (2006) Amino Acids. , Vol. 30,( 2006) I-1, pp 49-54.

[3] Stephen Wolfram; Statistical Mechanics Cellular Automata; (1983)Reviews of Modern Physics, Vol. 55,.

[4] E. v. Rooij; *The Art of MicroRNA Research*; (2011)Circulation Research Journal of The American Heart Association, 108(2)( 2011):219-34. doi: 10.1161/CIRCRESAHA.110.227496Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[5] Sannella, M. J. 1994 Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[6] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, An Analysis of Human MicroRNA and Disease Associations. PLoS ONE 3(10)(2008) e3420. doi:10.1371/journal.pone.0003420

[7] Xi Chenet et. al, "Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases",(2008). Cell Research Cell Research (2008) **18**:997–1006. doi: 10.1038/cr.2008.282.

[8] S. WolForm, "A New Kind of Science", (2002)Wolfram Media, Inc.,

[9] D.P. Bartel MicroRNAs: genomics, biogenesis, mechanism, and function Cell, 116 (2004), pp. 281–297.

[10] Hu Xiaoxia, J.K. Schwarz, J.S. Lewis Jr. *et al.* **A microRNA expression signature for cervical cancer progression** Cancer Res., 70 (2010), pp. 1441–1448.

[11] G. Reshmi, P.M. Radhakrishna **Beyond HPV: Oncomirs as new players in cervical cancer** FEBS Lett., 582 (2008), pp. 4113–4116

[12] P. Maji and P. P. Chaudhuri, "Fuzzy Cellular Automata for Modeling Pattern Classifier," IEICE, (2004).

[13] P. Maji and P. P. Chaudhuri (2004), FMACA: A Fuzzy Cellular Automata Based Pattern Classifier, Proceedings of 9th International Conference on Database Systems, Korea, pp. 494–505, 2004.