miRNA-mRNA Binding And Seed Binding Region Detection Of Breast Cancer Specific mRNA Using Correlation Method

[1] Binthiya Suny Gabriel, [2] Tessamma Thomas
[1] Research Scholar, [2] Emeritus Scientist
[1] Department of Electronics
[1] Cochin University of Science and Technology, Kochi, Kerala, India

Abstract : The microRNA targeting has recognized the canonical miRNA-mRNA interactions that feature complementarity to the seed sequence, as a critical process. A large emphasis has been placed on the importance of seed pairing for a sequence complementarity between mRNA and microRNA. The application of correlation method for identifying the breast cancer specific mRNA--miRNA binding regions and also the seed binding region of this mRNA, is presented in this paper. The analysis was done using the mRNA, RAC3 GTPase. The application of correlation method to identifying the seed region, provide results close to that obtained for the seed region, using seed generation and matching technique.

Index Terms **- miRNA, mRNA, seed region, correlation, circular shift, reverse compliment.**

I. INTRODUCTION

microRNAs (miRNA) are molecules that are double stranded and antisense in nature and are 19–23 nucleotides long. An antisense sequence consists of nucleotide sequences that are complementary to a coding (or sense) sequence, for example, that of a messenger RNA (mRNA**)** molecule. The binding of the mRNA 3'UTR to the miRNA seed region leads to the mRNAs being repressed at the post transcriptional level, leading to mRNA decay. This process is dominated by the six to eight-base miRNA seed region. Further, miRNAs not only target genes, but also target significant parts of the pathways.

Though there are several factors that may contribute to the binding of miRNA and mRNA, the completeness of miRNA-mRNA binding, and the resulting impact that the miRNA has on the mRNA expression, is thought to be determined by the seed sequence of miRNA. The seed sequence is highly conserved in nature and the slightest variation in the sequence can change the spectra of the target. A seed region comprises a contiguous string of at least 6 nucleotides beginning at position two of the miRNA sequence. The core seeds are 6-mer (bases 2–7), 7-mer ("7-mer-A1" with bases 1–7, "7-mer-m8" with bases 2–8), and 8-mer (bases 1–8) from the 3' end [1]. Figure 1. shows how the core seeds of the miRNA, hsa-miR-197, are found, as per the seed generation and matching method [1].

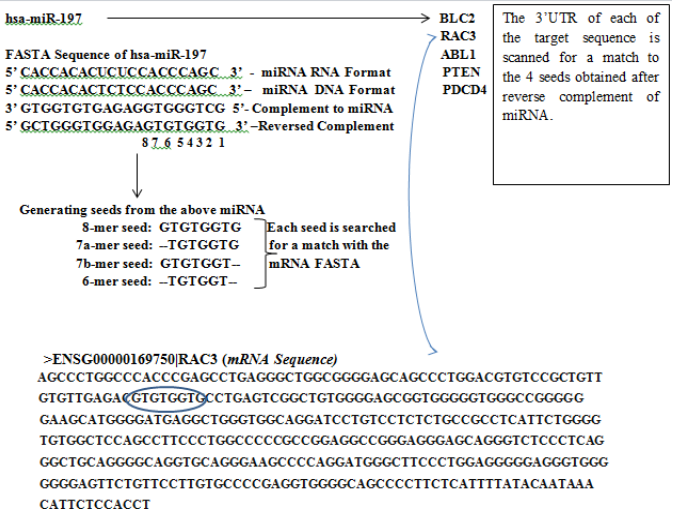

Figure 1 Seed Generation and Matching

In this study, a simple correlation method is developed to identify the seed binding region of the mRNA**.**

II. LITERATURE STUDIES

The seed region has been considered as the most important feature in miRNA target prediction. Researchers have been relying on computational tools to identify the target candidates for further validating the experiments. High-throughput novel technologies like Argonaute HITS-CLIP have done a detailed study of the miRNA-mRNA duplices. The interaction maps generated by these technologies help discriminate functional from the non-functional target sites.

Different seed paradigms have been applied by prediction algorithms to identify the target sites of miRNA. Therefore, a quantitative assessment of the target site prediction, has gained interest [2]. Recent years have seen tremendous progress in the identification of sequence features relevant to the prediction of targets such as the seed sequences. Despite the progress, the target prediction tools have only less optimum performance. The major obstacle seen in predicting the targets, computationally, is the lack of guidance from experimental observations [3].

Researchers have identified several features for seed binding region that have found to help improve the efficacy of the binding site including AU rich composition of the nucleotide near the binding region [4].

Correlation method has found application in the identification of miRNA binding region [5]. In this method, the mRNA is analyzed, in order to locate the binding regions of various miRNAs, which bind to specific regions of mRNAs. Simple correlation methods are used for this. The 'microcosm' website provides data of different miRNAs and the regions where the miRNAs bind to an mRNA.

III. METHOD USED

III.1 Determination of binding region of mRNA:

First, the correlations between the selected mRNA and the various miRNAs that bind to the different regions of the mRNA were studied to find their binding regions.

The steps involved in the method are as follows:

1. Select the miRNA, and also the mRNA whose region of binding to the specific miRNA, is to be found.
2. Reverse compliment the miRNA [5]. An example is illustrated in Table 1 using the miRNA hsa-miR-197.

Table 1 Reverse Compliment of hsa-miR-197

5' CACCACACUCUCCACCCAGC 3'	miRNA RNA Format
5' CACCACACTCTCCACCCAGC 3'	miRNA DNA Format
3' GTGGTGTGAGAGGTGGGTCG 5'	Compliment of miRNA
5' GCTGGGTGGAGAGTGTGGTG 3'	Reversed Compliment of miRNA

3. Calculate normalised correlation between the mRNA and miRNA sequences, with lateral shifting by one base and find the location of maximum correlation to obtain the binding region of the mRNA to that specific miRNA.

III.2 Determination of seed-binding region of the mRNA:

The seed region of the miRNA are (as mentioned in section I), bases 1 to 7, 2 to 7, 1 to 8 and 2 to 8 from the 3' end of the miRNA [1].

The steps involved are as follows:

1. The normalised correlations were computed for the selected binding region of the mRNA, with circular shifting of the miRNA.
2. The maximum correlation indicates the region of the mRNA which binds to the seed region of the miRNA. The number of bases in the miRNA seed region which bind to the mRNA gives an indication of the complexity of the cancer cells.

IV. DATABASE

The details of the various miRNAs which bind to various positions of a specific mRNA are obtained from the microcosm website and the detailed sequences are obtained from the NCBI database. The breast cancer specific mRNA, RAC3 GTPase and its corresponding 12 miRNAs were used for analysis, in this study.

V. IMPLEMENTATION OF THE METHOD

Normalised correlation between the reverse complimented versions of one of the breast cancer specific miRNA sequence and the entire mRNA RAC3 GTPase sequence was done. This was repeated by shifting the miRNA by one base to obtain the region of maximum correlation/coincidence. Once this region was obtained, normalised correlation for that specific mRNA region was computed with circular shifting of the miRNA and the position of maximum correlation was noted. The seed binding region/sequence of mRNA corresponds to the region of the mRNA which got bonded to the known seed region in this position. The above steps were repeated for all the 12 miRNAs.

VI. RESULTS AND DISCUSSION

With respect to the 12 miRNAs, reported in the Microcosm website, to bind with the RAC3 GTPase mRNA, only 10 miRNAs showed their seed regions as 6-mer (bases 2-7), 7-mer ("7-mer-A1" with bases 1-7, "7-mer-m8" with bases 2-8), and 8-mer (bases 1-8) as reported by L.E. Mullany, 2016 and one of the miRNAs namely, hsa-miR-412, had its binding position from base four to nine. The 12th miRNA had its seed region shifted to the middle portion. Further studies have to be done to check if it is due to addition or deletion of bases. The details of the 11 best matching miRNAs with their sequences and reverse complimented seed sequence are shown in Table 2.

As per ref [5], mRNA binding regions were identified by the STFT method, and it was noted that the binding regions were the same as in Microcosm website. The seed regions were then determined using correlation method.

Table 2 Seed Binding Regions and their Sequences

Sl. No.	miRNA Name	miRNA Sequence (5' - 3')	miRNA Seeds : as 6-mer, 7-mer and 8-mer	Microcosm Website			Correlation Method		
				Binding Region and Sequence of mRNA	Seed Binding Region, Sequence	Maximum Correlation Value at the binding region	Binding Region and Sequence of mRNA	Seed Binding Region, Sequence	Maximum Correlation Value at the binding region
1	mmu-miR-673-3p	UCCGGGGCUGAGUUCUGUGCACC	2-7 : - CAGCCC	25-47 GGCTGGCGGGGAGCAGCCCTGGA	38-43, CAGCCC	0.9679	25-47 GGCTGGCGGGGAGCAGCCCTGGA	38-43, CAGCCC	0.9713
2	hsa-miR-197	CACCACACUCUCCACCCAGC	1-8 : - GTGTGGTG	58-80 GTTGTGTTGAGACGTGTGGTGTC	71-78, GTGTGGTG	0.9587	58-78 GTTGTGTTGAGACGTGTGGTG	71-78, GTGTGGTG	0.979
3	mmu-miR-693-5p	CAGCCACAUCCGAAAGUUUUC	1-7 : - TGTGGCT	168-188 CCTCATTCTGGGGTGTGGCTC	181-187, TGTGGCT	0.9794	168-188 CCTCATTCTGGGGTGTGGCTC	181-187, TGTGGCT	0.9775
4	mmu-miR-690	AAAGGCUAGGCUCACAACCAAA	2-7 : - AGCCTT	175-196 CTGGGGTGTGGCTCCAGCCTTC	190-195, AGCCTT	0.987	175-196 CTGGGGTGTGGCTCCAGCCTTC	190-195, AGCCTT	0.9714
5	hsa-miR-663	AGGCGGGGCGCCGCGGGACCGC	1-7 : - CCCCGCC	190-210 AGCCTTCCCTGGCCCCCGCCG	203-208, CCCCGC	0.9558	189-210 CAGCCTTCCCTGGCCCCCGCCG	203-209, CCCCGCC	0.972
6	hsa-miR-526b	CUCUUGAGGGAAGCACU-UUCUGU	2-7 : - TCCCTC	220-243 AGGGAGCAGGGTCTCCCTCAGGGC	233-238, TCCCTC	0.9338	220-243 AGGGAGCAGGGTCTCCCTCAGGGC	233-238, TCCCTC	0.9692
7	hsa-miR-324-3p	ACUGCCCCAGGUGCUGCUGG	1-7 : - GGGGCAG	237-255 TCAGGGCTGCAGGGGCAGG	248-254, GGGGCAG	0.9515	236-255 CTCAGGGCTGCAGGGGCAGG	249-254, GGGCAG	0.9775
8	hsa-miR-500*	AUGCACCUGGGCAAGGA_UUCUG	2-7 : - GGTGCA	238-260 CAGGGCTGCAGGGGCAGGTGCAG	254-259, GGTGCA	0.9628	238-260 CAGGGCTGCAGGGGCAGGTGCAG	254-259, GGTGCA	0.9629
9	hsa-miR-412	CUUCACCUGGUCCACUAGCCGU	4-9 : - CAGGTG	239-261 AGGGCTGCAGGGGCAGGTGCAGG	252-257, CAGGTG	0.9759	239-261 AGGGCTGCAGGGGCAGGTGCAGG	252-257, CAGGTG	0.9688
10	hsa-miR-486-5p	UCCUGUACUGAGCUGCCCC	1-7:- TGCAGGG	245-262 GCAGGGGCAGGTGCAGGG	256-262,TGCAGGG	0.9622	248-266 GGGGCAGGTGCAGGGAAGC	256-262,TGCAGGG	0.9708
11	mmu-miR-687	CUAUCCUGGAAUGCAGCAAUGA	1-7:- AGGATGG	255-276 GTGCAGGGAAGCCCCAGGATGG	270-278, AGGATGG	0.9813	255-276 GTGCAGGGAAGCCCCAGGATGG	270-278, AGGATGG	0.975

The region of the mRNA that binds with the miRNA, the seed binding region and its sequence with respect to both the Microcosm website and the new correlation method, are also given in the table. Though the mRNA binding region with respect to the data given in Microcosm website and the correlation method differs by one or two base positions, the seed region is the same with respect to all the 11 miRNAs.

It can be noted that as the seed length varies from 6 to 8 in different miRNAs, the strength of the correlation coefficient increases. The length of the miRNA seed region is an indication of the complexity of cancer cells.

For the purpose of validation, the new correlation method was applied to another sequence, Programmed Cell Death 4 (PDCD4), an mRNA specific to breast cancer. PDCD4 is a protein that suppresses tumors which gets targeted for degradation during the progression of tumor. The 3'UTR nucleotide sequence of PDCD4 which consists of 1918 nucleotides was obtained from the UCSC Genome Browser.

The miRNAs hsa-miR-155, hsa-miR-183, hsa-miR-23a, hsa-miR-96 and hsa-miR-21 are known to target PDCD4. The binding regions of these miRNAs are not available as ground truth. Table 3 below summarizes the seed binding regions and the sequences for PDCD4.

Table 3 Seed binding regions and their sequences for PDCD4 mRNA

Sl. No.	miRNA Name and Length	Correlation Method			
		Binding Region of mRNA	Seed Binding Region of mRNA, Sequence	miRNA Seed Region and Sequence	Maximum Correlation Value at the binding region
1	hsa-miR-21 (72)	213-284	245-251, AGCTACC	6-12 : -AGCTACC	0.9817
2	hsa-miR-96 (78)	311-388	347-352, AAAATC	6-11 : - AAAATC	0.9771
3	hsa-miR-155 (65)	1001-1065	1030-1036, ATTAACA	2-8 : - ATTAACA	0.9807
4	hsa-miR-23a (73)	1424-1496	1457-1462, CCCAGC	6-11 : - CCCAGC	0.8546
5	hsa-miR-183 (110)	1712-1821	1763-1769, GTCACAC	8-14 : - GTCACAC	0.978

In the case of PDCD4 mRNA, only the miRNA, hsa-miR-155, has the core seed region at 2-8 (7-mer-m8). It may be noted that the mRNA and miRNA lengths in PDCD4 are nearly 4 to 5 times that of RAC3. In this case also, as the length of the seed varies, the correlation strength increases. As in the case of RAC3 GTPase mRNA, for PDCD4 also, investigations need to be done to check if the addition or deletion of bases or the lengths of mRNA or miRNA, has an effect on the shift of the core seed region.

V. CONCLUSION

The miRNA- mRNA binding and seed binding region and the sequences of various mRNAs were obtained using the new correlation method. It was observed that as the length of the seed binding region sequence increases the value of the correlation

strength also increases. The results obtained could lead to the application of correlation method in further analysing the miRNA-mRNA-TF network relating to breast cancer progression.

VI. APPLICATION POTENTIAL OF THE WORK

Researchers have found interest in investigating the role of miRNA in cancer. The progression of cancer is governed by not only the miRNA- mRNA interaction, but also by the antagonist action of the Transcription Factor (TF). The seed region detection will help to analyze the probable region of action of the TFs and its effect.

VII. FUTURE WORK

It has now been accepted that miRNAs and TFs together weave an inter-regulatory network that takes on the responsibility of a combined regulation of the gene expression. Analysis of the topological patterns of these interactions using combination of statistical and signal processing methods would be the next phase of this research.

VIII. ACKOWLEDGEMENT

The authors gratefully acknowledge the financial support from Kerala State Council for Science, Technology and Environment (KSCSTE), Trivandrum.

REFERENCES

[1] L. E. Mullany, J. S. Herrick, R. K. Wolff, M. L. Slattery, MicroRNA Seed Region Length Impact on Target Messenger RNA Expression and Survival in Colorectal Cancer, *PLoS ONE* 11(4): e0154177, April 28, 2016.

[2] Ellwanger DC, Buttner FA, Mewes HW, Stumpflen V. The sufficient minimal set of miRNA seed types. Bioinformatics. 2011; 27(10):1346–50.

[3] Wang X. Composition of seed sequence is a major determinant of microRNA targeting patterns. Bioinformatics 2014; 30(10):1377–83.

[4] Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell. 2007; 27(1):91–105.

[5] Binthiya Suny Gabriel, Tessamma Thomas, Breast Cancer Specific microRNA- mRNA Binding Region Analysis Using Signal Processing Techniques. Recent Application and Trends in Modern Engineering, Proceedings of the IDES Joint International Conferences, July 2018, Pages 120-138.