Breast Cancer Specific miRNA-mRNA Binding Region Analysis using Signal Processing Techniques

Binthiya Suny Gabriel[1] and Tessamma Thomas[2]
[1]Research Scholar, Department of Electronics, Cochin University of Science and Technology, Kochi, Kerala, India
binthiya_gabriel@hotmail.com
[2]Professor (Rtd.), Department of Electronics, Cochin University of Science and Technology, Kochi, Kerala, India
tessamma1@gmail.com

Abstract: A new approach for identifying the binding region between the miRNA and mRNA using the STFT method with Blackman window of length 24 is presented in this paper. The seed region within the miRNA is also identified using simple correlation methods. mRNA, RAC3 GTPase, is used for the analysis. Results of binding region and seed region identification based on the methods mentioned above, provide values closest to the data available at Microcosm website.

Keywords: miRNA, mRNA, binding region, seed region, STFT, correlation with circular shift.

Introduction

microRNAs (miRNA) are short double stranded 19–23 nucleotides long molecules antisense in nature, which are both produced and processed endogenously. An antisense sequence has a sequence of nucleotides complementary to a coding (or sense) sequence, which may be either that of the strand of a DNA double helix which undergoes transcription, or that of a messenger RNA molecule. The main function of miRNA involves regulating the gene expression by exhibiting perfect or nearly perfect base pairing with the messenger RNA (mRNA) it targets, thus inhibiting its expression at post-transcriptional level by either degrading the mRNA or repressing the translation process. Lin-4 was the first miRNA that was discovered in C elegans to regulate the expression of lin-14 protein coding gene.

Further research indicated that more than 30% of the protein-coding genes in humans are regulated by miRNAs [1]. MicroRNAs have been under investigation particularly because of its roles related to the progression and development of cancer. Fig. 1 indicates how the seed region match results in either deadenylation, degradation of mRNA or in the translation process being inhibited [2].

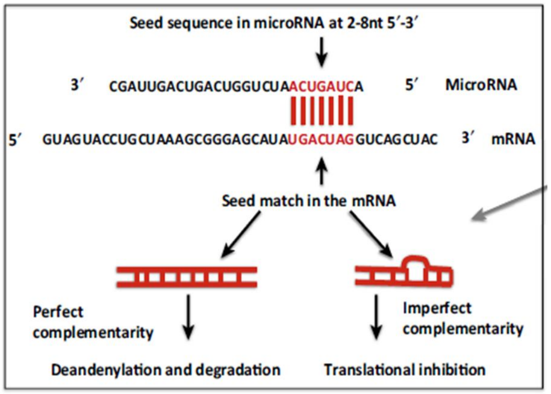

Seed sequence in microRNA at 2-8nt 5′-3′
3′ CGAUUGACUGACUGGUCUAACUGAUCA 5′ MicroRNA
5′ GUAGUACCUGCUAAAGCGGGAGCAUAUGACUAGGUCAGCUAC 3′ mRNA
Seed match in the mRNA
Perfect complementarity
Imperfect complementarity
Deandenylation and degradation
Translational inhibition

Fig. 1 Degradation of mRNA and Translational Inhibition

In the deadenylation process, miRNAs channel their targets to the cellular 5’ to 3’ mRNA decay pathway. In this pathway mRNAs are first deadenylated by the deadenylase complex followed by decapping. The deadenylation process involves removing an adenylate group from a protein. The loss of 5’ cap allows the major cytoplasmic exonuclease to degrade mRNAs. Translational inhibition is established when the decrease in protein product is greater than the observed decrease in mRNA.

Binding Region and Seed Region

Though there are several factors that contribute to the binding between miRNA and mRNA. The resulting impact is determined by the seed sequence in miRNA [3], [4].

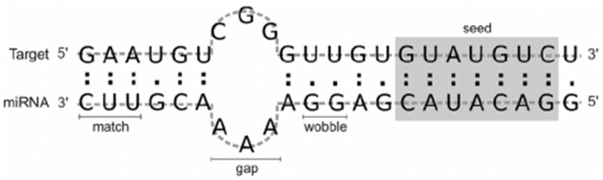


Fig. 2 Seed Region

The seed region, as in Fig. 2, consists of a contiguous string of at least 6 nucleotides. miRNA recognizes its target by the degree of complementarity between the special nucleotide motif (seed region, which is found embedded within the mature miRNA) and some specific binding sites found along the mRNA sequence. Longer seeds which are 7 or 8 nucleotides in length are considered to have greater efficacy as it is related to mRNA repression.

Moreover, since mRNA degradation is also dependent on precise binding, miRNAs that bind to target genes with longer seed region is considered to be more significantly associated with mRNA degradation.

DSP in identification of binding region

The use of Digital Signal Processing has found applications in solving problems in genomics to provide information about several biomolecular sequences [5]. It has been observed that target identification is one of the major challenges in miRNA analysis, since a miRNA may control many different mRNAs. The currently available methods for target recognition are less reliable since the tools are mainly based on pattern matching systems. The application of cDNA microarray in identifying targets also has limitations as it has been recently shown that a miRNA regulates gene expression at the protein level only, and miRNA are not affected.

The seed or the critical recognition motif is supported by other functional elements functional miRNA. The available method for binding site identification does not allow to deeply investigate the effect of the seed context. Digital Signal Processing (DSP) methods have been applied for a long time in bio sequence analysis [6], [7].

Literature Studies

Genomic Background

The gene regions present in the DNA sequence forms the basis for all functions in an organism. The central dogma of molecular biology highlights the information transfer from DNA to mRNA and then to protein [8], [9]. Research has also considered the involvement of ncRNAs in numerous gene regulatory mechanisms, for research. It was observed that the RNA regulatory networks could also determine most of the complex characteristics and also played a significant role in disease [10]. Research had also been done in reviewing the emergence of the previously unsuspected world of regulatory RNA from a historical perspective [11].

The importance of RNAs has remained rather obscure. Except for the usual ones such as tRNAs (transfer RNAs) and rRNAs (ribosomal RNAs), functional noncoding RNAs (ncRNA) were considered to be rare. It is the discovery in 1998 by Fire and Mello, on how some small RNAs could switch off certain mRNAs (RNA-interference) that triggered the research on non-coding RNAs. Non coding RNA molecules are molecules that perform various functions without getting translated into proteins. Researchers found the importance of long non-coding RNAs in genomic regulation. Discussions had taken place in identifying lncRNAs as regulatory modules and these discussions had led to exploring the implications of these RNAs for disease pathogenesis [12].

MicroRNA (miRNA) is a 19-25 nucleotide long non coding RNA molecule that is typically encoded within introns, and has been discovered in metazoans, plants and viruses [13]. miRNAs oscillate between repression and stimulation in response to specific cellular conditions, sequences, and cofactors. In depth studies had also revealed that repression effect of miRNAs could be relieved when they are subject to different types of stress such as amino acid deprivation, oxidative stress and synaptic simulation [14].

The miRNA serves as a guide for the complex RISC to the target mRNA where it induces translational repression and accelerates mRNA deadenylation. Research work has been done in finding out the miRNA expression characteristics in specific situations like the DNA damage response [15]. The miRNAs are an important class of gene regulators that play a major role in several aspects of cellular functions including differentiation, cell cycle control and stemness. [16].

RNA interference (RNAi) is a biological process in which RNA molecules inhibit gene expression, typically by causing the destruction of specific mRNA molecules. miRNAs being smaller than protein coding genes, can regulate the translation of hundreds of genes through sequence-specific binding to mRNA, and depending on the degree of complementarity will result in the inhibition of translation and/or enhanced mRNA decay [17].

MicroRNAs can promote translation of specific mRNAs in quiescent (G0) mammalian cells. Translation upregulation by microRNAs has been observed as a result of two possible outcomes: activation by the direct action of microRNAs/microRNPs and relief of repression, where the action of a repressive microRNA or microRNP is abrogated. Several methods had been developed to identify the relationship between epigenetic alteration and dysregulation of miRNAs in cancer [18]. Several methods had been discovered in order to detect microRNAs from genome and next generation sequencing data with probability matrices and combined features. Several modules including mirExplorer-genome and mirExplorer-NGS had been designed to de novo predict pre-miRNAs from genome, and to discover miRNAs from NGS data, respectively.

The miRseqViewer helps visualise the sequence alignment, secondary structure and normalized read counts in synchronous multipanel windows. This application helped in the research to easily examine the relationships between the structure of precursor and the sequences and abundance of final products and thereby facilitated the studies on miRNA biogenesis and regulation [18]. The application of an online tool for miRNA genomics, miRBase, was used to facilitate studies of miRNA genomics wherein all the miRNAs are mapped to their genomic coordinates [19].

DSP methods applied to Genomics

Three base periodicity is quite pronounced in exons and is commonly used in Digital Signal Processing (DSP) based methods to locate the exonic regions [5].

DFT was used for spectrum analysis of biological data where initially the DNA sequence was mapped into a numeric sequence and spectrum of finite-length windowed DNA numerical sequences was computed. (c) Window function was translated by one or more bases and power spectrum was calculated along the length of the investigating DNA sequence [20].

Target identification is one of the major challenges in miRNA analysis, since a miRNA may control many different mRNAs. Recent findings had shown that one miRNA can repress the production of hundreds of proteins. This showed the need of proteomic data analysis for the accurate detection of miRNA gene targets. Digital Signal Processing (DSP) methods had been applied for in bio sequence analysis and helped the researcher to better study the role of the seed context [21].

Tbe antinotch IIR filter has found application in identifying the period 3 region of the DNA sequences instead of the DFT. The antinotch filters were implemented very easily and efficiently using the Gray and Markel lattice structure, having only two multipliers. One of the multipliers controlled the sharpness of the filter peak, by controlling the pole radius. while the other multiplier was used for setting the antinotch frequency at 2π/3. Though this method had a compromise between the sharpness of the notch filter and the base-domain resolution achieved, the method was promising [6].

The applications of digital filters had been seen not only in predicting the gene but it also helped eliminate the background 1/f spectrum noise exhibited by nearly all DNA sequences. Even though the IIR antinotch method had been found to work well there was room for improvement. Further research showed that with a slight increase in the number of multipliers, it was possible to design filters with much better stop band attenuation. Such filters were needed to suppress the background 1/f noise mostly present in the DNA sequences of many organisms, due to long-range correlation between the base pairs [22].

Digital signal processing and control has been widely used in many areas of science and engineering. It provides practical and powerful tools to model simulate, analyze, design, measure, and control complex and dynamic systems such as robots and aircrafts. Gene networks are also complex dynamic systems which can be studied via digital signal processing and control. Unlike conventional computational methods, this approach was used by the researchers for not only modelling but also controlling gene networks since the experimental environment is mostly digital [23].

A new method was introduced based on a modified Gabor-wavelet transform (MGWT) for the identification of protein coding regions. This novel transform was tuned to analyze periodic signal components and presented the advantage of being independent of the window length. The proposed work compared the performance of the MGWT with other methods [24].

Discrete Wavelets could be employed to decompose genomic DNA sequences followed by data-dependent thresholding algorithms to remove the background. Following this, entropic segmentation method was applied to find boundaries between segments well-characterized genes. Before the wavelet decomposition, the genomic DNA sequences were digitized into numerical sequences based on their contents. The results showed that the wavelet approach was feasible and better than the knowledge-based methods in these cases [25].

Components of sequences at different scales of interest were extracted whose features were aligned with those of the original data sequence. This helped the researhers to detect transmembrane segments of proteins without first having to apply a smoothing procedure that would usually involve the choice of threshold parameters. Moreover, scale by scale wavelet decompositions of variance and correlation helped in highlighting hidden structures of single sequences and similarities among different sequences [26].

Protein sequence comparison is one of the most important areas in bioinformatics research. The conventional BLAST based approach focused on the local pairwise amino acid match. However, two protein sequences with low sequential identity showed similarities in physiochemical properties and tertiary structure, which indicated a functional correlation between the two proteins. The conventional sequence-based comparison, is challenged in identifying this kind of similarity. In this research, wavelet analysis became a useful alternative as it was able to capture the multi-scale information that enables the comparison of protein sequences at different resolutions [27].

As mentioned earlier, when some of the miRNAs bind with mRNAs, deadenylation may occur, leading to cancer. Our purpose is to find out such binding regions specific to breast cancer. Short Time Fourier Transform (STFT) is used for finding these binding regions [7].

Methods

STFT Method to identify the binding region

In this method, STFT with Blackman window is applied to filter out the most probable binding regions of ncRNAs in relation to the recognition of miRNA binding site. STFT is a signal processing technique used to analyze non-stationary signals. This technique of Fourier Transform involves segmenting a signal into narrow time intervals and then finding the Fourier Transform of each of the segments. A sliding window used along with STFT gives the advantage of time localization by dividing the signal into short overlapping sections.

The STFT of any sequence w(n) is given by :

$$U[k] = \sum_{m=0}^{L-1} x[m + rR]w[m]e^{\frac{-j\pi km}{N}} \qquad (1)$$

where,

L: Window Length. N: DFT Length, R: Shift Interval, r, k: integers such that $-\infty < r < \infty$ and $0 \leq k \leq N-1$.

The spectral content at the kth instant is given as:

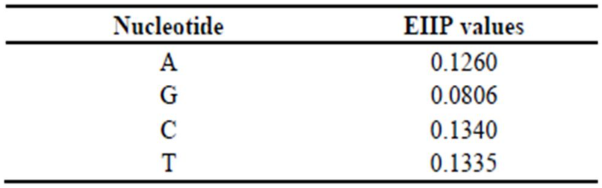
$$\mathrm{S}[k]=|U[k]|^2 \qquad (2)$$

Implementation of the STFT method to identify the miRNA binding region

Data used: Initially, the mRNA, RAC3 GTPase, which is a breast cancer specific mRNA, was used for the analysis. The detail of the sequence was taken from the NCBI database.

The character string of the mRNA sequence was mapped from a symbolic form onto a numerical form using the EIIP values, shown in Table I. An EIIP value is defined as the average energy of the electrons which are localized within a nucleotide.

Table 1. Electron Ion Interaction Pseudo Potentials of Nucleotides



Nucleotide	EIIP values
A	0.1260
G	0.0806
C	0.1340
T	0.1335

The STFT is applied to the indicator sequence (sequence obtained by replacing each nucleotide with its respective EIIP values) thus obtained using equation (1), to predict the binding sites of miRNA. The recognition between miRNA and the target site is based on motifs having a length of 6 nucleotides. The calculation of the spectral content is done using equation (2). S[L/6] is obtained by selecting every sixth component of the spectral component, after having applied the sliding window throughout the length of the sequence. A similar approach is adopted for identifying the exonic regions wherein every third component of the spectral content is selected. The peaks in the spectra correspond to the binding sites. A sampling value of '6' is selected based on the average dimension of seed region.

The length of the window affects the identification of the coding region. The optimal value for the length of the window is dependent on the number of exons in the miRNA sequence considered. Based on the maximal length of a miRNA, of 21-23 base pairs, a Blackman window having length L=24 is chosen.

Fig. 3 shows the spectral content obtained by applying STFT with Blackman sliding window to RAC3 protein sequence, which is breast cancer specific, extracted from the ensembledatabase.

[https://asia.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000169750;r=17:82031624-82034204].

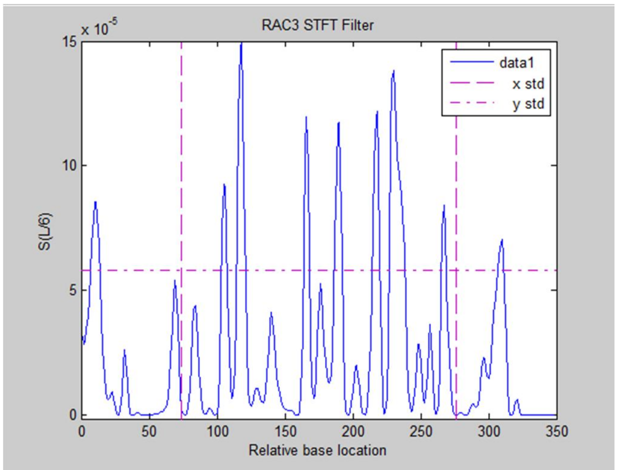
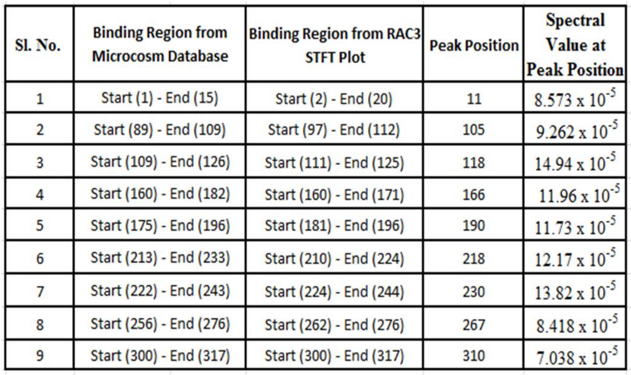


Fig. 3 Spectral content of RAC3 filtering

RAC proteins are members of the Rho GTPase family that act as molecular switches in regulating a number of biological processes including cell motility and cell cycle progression. RAC3 GTPase plays an active role in making its effector protein kinases hyperactive in human breast cancer-derived epithelial cell lines. The nine peaks above the dotted line, representing the standard deviation of the RAC3 mRNA sequence are found to be the region where the seed regions of various miRNAs can bind to the mRNA. The nine peak regions obtained from Fig. 3 are represented in Table 2.

Table 2. Peaks Obtained From the RAC3 Spectral Content



Sl. No.	Binding Region from Microcosm Database	Binding Region from RAC3 STFT Plot	Peak Position	Spectral Value at Peak Position
1	Start (1) - End (15)	Start (2) - End (20)	11	8.573×10^{-5}
2	Start (89) - End (109)	Start (97) - End (112)	105	9.262×10^{-5}
3	Start (109) - End (126)	Start (111) - End (125)	118	14.94×10^{-5}
4	Start (160) - End (182)	Start (160) - End (171)	166	11.96×10^{-5}
5	Start (175) - End (196)	Start (181) - End (196)	190	11.73×10^{-5}
6	Start (213) - End (233)	Start (210) - End (224)	218	12.17×10^{-5}
7	Start (222) - End (243)	Start (224) - End (244)	230	13.82×10^{-5}
8	Start (256) - End (276)	Start (262) - End (276)	267	8.418×10^{-5}
9	Start (300) - End (317)	Start (300) - End (317)	310	7.038×10^{-5}

From Table 2, it is clear that the 9 peak regions obtained from Fig. 3 is almost close to the binding regions of the RAC3 mRNA sequence as given in the ground truth data of the Microcosm database.

At the first and the ninth peak regions, there are no miRNAs that bind exactly to that region of the mRNA. Hence, only the remaining peak regions are considered for further analysis. As mentioned in section B, Table II gives the approximate binding region of the various miRNAs to the specific mRNA sequence. As the next step to find the exact binding region (or the 'seed' section), correlation between the binding regions and their corresponding miRNAs were considered.

Correlation Method for seed region identification

This method analyses the binding region of the mRNA obtained from the above STFT method to locate the seed regions of various miRNAs, which bind to specific regions of mRNAs. Simple correlation methods are used for this. The different miRNA sequences which bind to an mRNA are obtained from the ‘microcosm’ website.

http://www.ebi.ac.uk/enrightsrv/microcosm/htdocs/targets/v5/.

As direct cross correlation and normalised cross correlation yielded only low values, normalised cross correlation with circular shift was considered in this work.

1) Cross Correlation: Cross Correlation measures the Cross Correlation measures the similarity between similarity between one sequence and lagged/shifted copies of another sequence as a function of the lag. If the two sequences have unequal lengths, the cross correlation can be done by appending zeros at the end of the shorter sequence so that it has the same length as that of the longer sequence.

Cross Correlation between any two sequences **x**(n) and **y**(n) is given by:

$$R_{xy}\,(J) = \sum_{n=1}^{N} x(n)y(n-J) \qquad (3)$$

where, N is the length of the sequences.

2) Circular Shift: In circular shift, as the terms are being shifted past a point, the sequence gets looped around the other end.

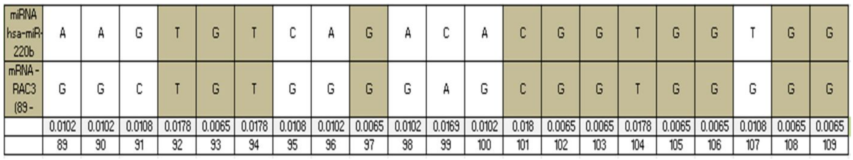
$$y(n) = x[\langle n-m \rangle_N] \qquad (4)$$

The above equation represents the circular shift of the sequence x[n] by m samples to the right.

3) Normalised Correlation with Circular Shifting: If x(n) and y(n) are 2 sequences, with y(n) being the circularly shifted one, then the normalised cross correlation is given by :

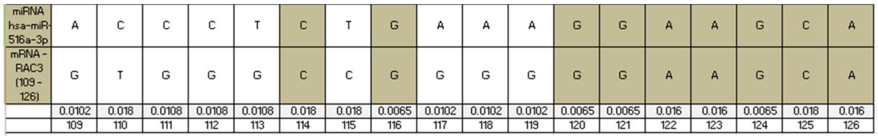
$$R'_{x,y}(J) = \frac{\sum_{n=0}^{N-1} x[n].y[n-J]}{(\sum_{n=0}^{N-1} x^2[n].\sum_{n=0}^{N-1} y^2[n-j])^{1/2}} \qquad (5)$$

The Figs. 4 & 5 shows the relationship between the miRNAs, hsa-miR-220b and hsa-miR-516a-3p with the different regions in the mRNA, RAC3. The highlighted regions are the binding regions.



miRNA hsa-miR-220b	A	A	G	T	G	T	C	A	G	A	C	A	C	G	G	T	G	G	T	G	G
mRNA - RAC3 (89 -	G	G	C	T	G	T	G	G	G	G	A	G	C	G	G	T	G	G	G	G	G
	0.0102	0.0102	0.0108	0.0178	0.0065	0.0178	0.0108	0.0102	0.0065	0.0102	0.0169	0.0102	0.018	0.0065	0.0065	0.0178	0.0065	0.0065	0.0108	0.0065	0.0065
	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109

Fig. 4 Correlation between hsa-miR-220b and mRNA RAC3(89-109)



miRNA hsa-miR-516a-3p	A	C	C	C	T	C	T	G	A	A	A	G	G	A	A	G	C	A
mRNA - RAC3 (109 - 126)	G	T	G	G	G	C	C	G	G	G	G	G	G	A	A	G	C	A
	0.0102	0.018	0.0108	0.0108	0.0108	0.018	0.018	0.0065	0.0102	0.0102	0.0102	0.0065	0.0065	0.016	0.016	0.0065	0.018	0.016
	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126

Fig. 5 Correlation between hsa-miR-516a-3p and mRNA RAC3(109-126)

Implementation of the Method

As part of this work, the normalised correlation between mRNA and miRNA was computed while circularly shifting miRNA throughout the mRNA sequence and computing the normalised correlation value each time it is shifted. In this method, the two inputs, miRNA and mRNA are taken to be of equal length i.e. the length of the binding region along mRNA is taken to be equal to the length of miRNA. The different steps to be carried out before finding the correlation between the miRNA and mRNA sequences are explained in the example below.

Illustrating Example

Step 1: Selecting the inputs

miRNA: hsa-miR-516a-3p (of length 18)

mRNA: Corresponding binding region (109-126).

In the Fig.6, the shaded region is the binding region.

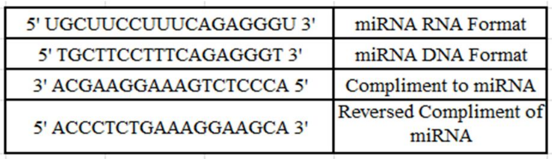
Step2: Reversed Compliment of the miRNA

The calculation of maximum correlation between microRNA and mRNA requires finding the reversed compliment of microRNA. The reverse compliment of hsa-miR-516a-3p is obtained as in Table 3.

3’ {miRNA)
UGGGAGACUUUCCUUCGU 5’
||:::|||||||
5’ (mRNA)
GTGGGCCGGGGGGAAGCA 3’

Fig. 6 Binding of a microRNA to its corresponding mRNA [Source: Microcosm website]

Table 3. Reversed Compliment Of Hsa-Mir-516a-3p



5' UGCUUCCUUUCAGAGGGU 3'	miRNA RNA Format
5' TGCTTCCTTTCAGAGGGT 3'	miRNA DNA Format
3' ACGAAGGAAAGTCTCCCA 5'	Compliment to miRNA
5' ACCCTCTGAAAGGAAGCA 3'	Reversed Compliment of miRNA

mRNA: RAC3 (109-126) : GTGGGCCGGGGGGAAGCA

miRNA: has-miR-516a-3p : ACCCTCTGAAAGGAAGCA [reverse complimented without any shift]

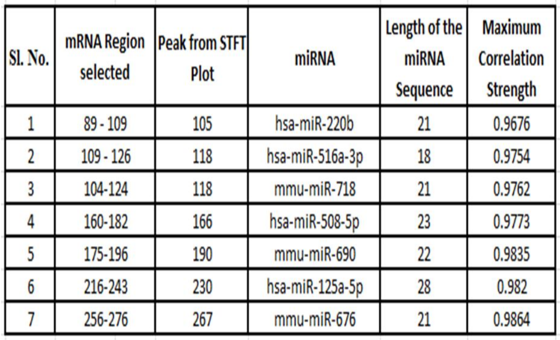
Step3: The normalised correlation was computed for the RAC3 region selected, with circularly shifting the miRNA.

Results and Discussions

The correlation strength is computed, using equation5, each time the miRNA is circularly shifted. For the example considered, since the miRNA is of length 18, the normalised correlation is calculated every time the miRNA sequence is shifted circularly to the right. The maximum strength from among the correlation values is noted. Table 4 below summarises the results obtained.

It was observed that the maximum correlation occurred when there was maximum coincidence between the reverse complimented miRNA and the mRNA sequences. It was also noted that, as per Microcosm database, the binding

Table 4. Correlation between the mRNA and miRNA sequences



Sl. No.	mRNA Region selected	Peak from STFT Plot	miRNA	Length of the miRNA Sequence	Maximum Correlation Strength
1	89 - 109	105	hsa-miR-220b	21	0.9676
2	109 - 126	118	hsa-miR-516a-3p	18	0.9754
3	104-124	118	mmu-miR-718	21	0.9762
4	160-182	166	hsa-miR-508-5p	23	0.9773
5	175-196	190	mmu-miR-690	22	0.9835
6	216-243	230	hsa-miR-125a-5p	28	0.982
7	256-276	267	mmu-miR-676	21	0.9864

region occurs in this region of maximum correlation. Moreover, the results of the binding region and seed region identifications based on the methods mentioned above; provide values closest to the data available at microcosm website.

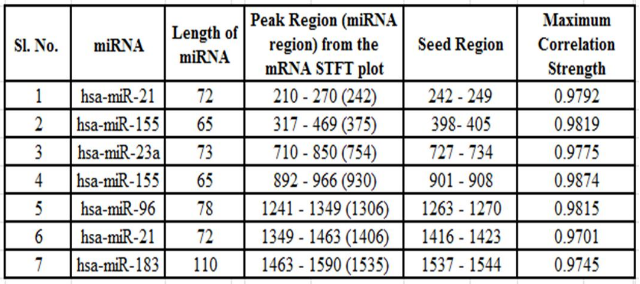
From this region of maximum correlation, the seed region can be identified using simple matching methods.

For validating the above methods, of identifying the miRNA binding region and the seed regions (in the miRNA), they were applied to a new sequence, Programmed Cell Death 4 (PDCD4), which is a breast cancer specific mRNA. PDCD4 is a tumor suppressor protein which is targeted for degradation during tumor progression. The PDCD4 3’UTR nucleotide sequence (1918 nucleotides), was obtained from the UCSC Genome Browser.

The miRNAs that target PDCD4 are hsa-miR-96, hsa-miR-21, hsa-miR-183, hsa-miR-23a, and hsa-miR-155. The miRNA sequences are available but their binding regions are not available as ground truth. Table V. below summarises the miRNA binding sites corresponding to the peaks and also the seed region.

As obtained for the RAC3 mRNA, in the case of PDCD4 mRNA also, the 7 peak regions correspond to the 5 miRNAs mentioned and from the correlation strength the seed regions were identified. It must be noted that hsa-miR-21, hsa-miR-155 binds to 2 regions each of the mRNA.

Table 5. Analysis of Mrna PDCD4 With Respect To Its miRNA Sequences



Sl. No.	miRNA	Length of miRNA	Peak Region (miRNA region) from the mRNA STFT plot	Seed Region	Maximum Correlation Strength
1	hsa-miR-21	72	210 - 270 (242)	242 - 249	0.9792
2	hsa-miR-155	65	317 - 469 (375)	398- 405	0.9819
3	hsa-miR-23a	73	710 - 850 (754)	727 - 734	0.9775
4	hsa-miR-155	65	892 - 966 (930)	901 - 908	0.9874
5	hsa-miR-96	78	1241 - 1349 (1306)	1263 - 1270	0.9815
6	hsa-miR-21	72	1349 - 1463 (1406)	1416 - 1423	0.9701
7	hsa-miR-183	110	1463 - 1590 (1535)	1537 - 1544	0.9745

Though 12 peaks were obtained from the STFT plot, only 7 peaks corresponding to the 5 available miRNAs were obtained, with miRNAs miR-21 and miR-155 corresponding to two peak regions each. From the correlation measure, the specific seed region was accurately determined. Further investigations have to be done to check whether there are other miRNAs which bind to the balance STFT peaks and also to spot the relevance of the peak strength.

Conclusion

The binding region of the miRNA to breast cancer specific mRNA, RAC3, was done using simple correlation method. The results obtained were close to the ground truth data. Validation was done using PDCD4 and correct miRNA binding and seed region could be determined.

Application Potential Of The Work

Cancer is a leading cause of death in a developing country like India. It is basically a disease of “genes gain bad”. Gene expression controls the way cell grow, divide and die. The identification of genomic regulatory elements is an important but unresolved problem in genome annotation. Investigation on the roles of miRNA in cancer represents a developing and promising research field in the war against cancer. As the emergence of new knowledge and technologies are continuously happening, novel and effective anti-cancer strategies are becoming increasingly possible. Clinical management of human cancers will greatly benefit from the development of miRNA-based diagnostic and therapeutic approaches, and the pharmaceutical industry is also welcoming a new challenging opportunity in this exciting process.

Future Work

Identification of the coding region is affected by the length of the window. If window size is adjusted, it is possible to get better results. Further analysis is being done on fixing the window size.

Acknowledgement

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

References

[1] Qureshi, N. Thakur, I. Monga, A. Thakur, M. Kumar, “VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets”, The Journal of Biological Database and Curation, Database, 2014, Pages1-10.

[2] J. Hayes, P. P. Peruzzi, S. Lawler, “MicroRNAs in cancer: biomarkers, functions and therapy”, CellPress, Trends in Molecular Medicine, Review Article, Volume 20, Issue 8, August 2014.

[3] M. R Mendoza, G. C da Fonseca, G. Loss-Morais, “RFMiRTarget: Predicting Human MicroRNA Target Genes with a Random Forest Classifier”, PLoS ONE, Volume 8, Issue 7, July 26, 2013.

[4] 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. doi:10.1371/journal.pone.0154177, April 28, 2016.

[5] D. Anastassiou, “Genomic Signal Processing”, IEEE Signal Processing Magazine, Volume 18, Issue 4, July 2001, Pages 8-20.

[6] P. P. Vaidyanathan, B. Jun Yoon, “The role of signal processing concepts in genomics and proteomics”, .J.Franklin Inst. 341, 2004, Pages 111-135.

[7] N. Maggi, P. Arrigo,”Optimize ncRNA targeting: A signal analysis based approach”, XIII Mediterranean Conference on Medical and Biological Engineering and Computing, IFMBE Proceedings, Volume 41, 2013, Pages 662-665.

[8] E. G. Moss, L. Tang, "Conservation of the heterochronic regulator Lin-28, is developmental expression and microRNA complementary sites", Science Direct, Developmental Biology 258 (2003), Pages 432-442.

[9] M. J. Turner, F. J. Slack, "Transcriptional control of microRNA expression in C. elegans: Promoting better understanding", RNA Biology, Volume 6, Issue 1, January/February/March 2009, Pages 49-53.

[10] J. Chang, E. Nicolas, D. Marks, C. Sander, A. Lerro, M. Annick Buendia, C. Xu, W. S. Mason, T. Moloshok, R. Bort, K. S. Zaret, J. M. Taylor, "miR-122, a Mammalian Liver-Specific microRNA, is Processed from hcr mRNA and May Downregulate the High Affinity Cationic Amino Acid Transporter CAT-1", RNA Biology 1:2, July/August 2004, Pages 106-113.

[11] J. Takamizawa, H. Konishi, K. Yanagisawa, S. Tomida, H. Osada, H. Endoh, T. Harano, Y. Yatabe, M. Nagino, Y. Nimura, T. Mitsudomi, T. Takahashi, "Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival", Cancer Research 64, June 1, 2004, Pages 3753-3756.

[12] T. R. Cech, J. A. Steitz,, "The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones", Cell 157, March 27 2014, Pages 77-94.

[13] K. Appasani, "RNA Interference: From Biology to Drugs and Therapeutics", Gene Expression Systems, Second International Conference on RNAi, RNA Biology, Volume 1 , Issue 2, May 2-4, 2004, Pages 118-121.

[14] X. Guo, Y. Wu & R. Hartley, "MicroRNA-125a represses cell growth by targeting HuR in breast cancer", RNA Biology 6:5, November/December 2009, Pages 575-583.

[15] X. Fan, L. Kurgan, "Comprehensive overview and assessment of computational prediction of microRNA targets in animals", Bioinformatics Advance Access, December 2014, Pages 1-15.

[16] L. He, G. J Hannon, "MicroRNAs: Small RNAs with a big role in gene regulation", Genetics, Nature, Reviews, Volume 5, July 2004, Pages 522-531.

[17] S. Saini, L. Dewan. "Application of discrete wavelet transform for analysis of genomic sequences of Mycobacterium tuberculosis", SpringerPlus, Volume 5, Issue 64, 2016.

[18] H. Suzuki, R. Maruyama, E. Yamamoto, M. Kai, "DNA methylation and microRNA dysregulation in cancer", Science Direct, Molecular Biology 6 (2012), Pages 567-578.

[19] S. Zafari, C. Backes, P. Leidinger, E. Meese, A. Keller, "Regulatory MicroRNA Networks: Complex Patterns of Target Pathways for Disease-related and Housekeeping MicroRNAs", Elsevier, Genomics Proteomics Bioinformatics 13 (2015), Pages 159-168.

[20] D. K. Shakya, R. Saxena, S. N. Sharma, “A DSP-Based Approach for gene prediction in eukaryotic genes”, International Journal of Electrical Engineering and Informatics, Volume 3, Number 4, 2011, Pages 480-487.

[21] N. Wong, X. Wang, ‘miRDB: an online resource for microRNA target prediction and functional annotations’, Nucleic Acids Research, November 5 2014.

[22] O. C. Kulkarni, R. Vigneshwar, V. K. Jayaraman, B. D. Kulkarni, “Identification of coding and non-coding sequences using local Holder exponent formalism”, Bioinformatics, Sequence Analysis, Volume 1, Issue 20, 2005, Pages 3818-3823.

[23] Y. Cai, X. Yu, S. Hu, J. Yu, “A Brief Review on the Mechanisms of miRNA Regulation”, Genomics Proteomics Bioinformatics, Volume 7, Issue 4, December2009.

[24] I. V. Bajic, ‘Digital Signal Processing Techniques in the analysis of DNA/RNA and protein sequences’, Report, Department of Electronic Engineering, 1998.

[25] J´s P. Mena-Chalco, H. Carrer, Y. Zana, R. M. Cesar Jr., “Identification of Protein Coding Regions Using the Modified Gabor-Wavelet Transform”, IEEE/ACM transactions on Computational Biology and Bioinformatics, Volume 5, Issue 2, April-June 2008, Pages 198-207.

[26] M. I. Almeida, R. M. Reis, G A. Calin, “MicroRNA history: Discovery, recent applications, and next frontiers”, Elsevier, Mutation Research 717 (2011), Pages 1-8.

[27] Tao Meng, Ahmed T. Soliman, Mei-Ling Shyu, Y. Yang, Shu-Ching Chen, S.S. Iyengar, J. S. Yordy, P. Iyengar, “Wavelet Analysis in Current Cancer Genome Research: A Survey”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Volume 10, Issue 6, November/December 2013, Pages 1442-1459.