Journal of Engineering Science and Technology
Vol. 16, No. 2 (2021) 1480 - 1491

COMPARISON OF CORRELATION METHOD WITH MODIFIED STFT FOR THE DETECTION OF miRNA BINDING AND SEED BINDING REGIONS IN BREAST CANCER

BINTHIYA S. GABRIEL*, TESSAMMA THOMAS

Department of Electronics, Cochin University
of Science and Technology, Kochi, Kerala, India
*Corresponding Author: binthiya_gabriel@cusat.ac.in.mi

Abstract

MicroRNAs are efficient in controlling different functions including both genetic and metabolic, transcriptional control of gene expression by miRNA and the control of cholesterol synthesis and lipoprotein secretion in the liver, respectively. Digital Signal Processing methods, being flexible, have been widely applied to genomic analyses to give an easier way of understanding the details of the cells. A new approach for identifying the binding regions of miRNA using a modified Short-time Fourier transform (STFT) method was presented in this paper. Breast cancer specific mRNAs, RAC3 GTPase and BRCA1, are used for the analysis. The modified STFT recognizes the approximate binding regions of all the miRNAs considered, with respect to the ground truth. A perfect base pairing between miRNA target and the miRNA seed is necessary for studies related to silencing using miRNA expression. The identification of miRNA seed binding position is critical in that it is considered as an important determinant for the identification of targets. The modified STFT method is compared with the correlation method. The results obtained indicates that the correlation method is better than the modified STFT method as it gives all the exact miRNA binding as well as seed binding positions close to the ground truth.

Keywords: Binding region, Normalised correlation, Normalization, Seed region, STFT.

1. Introduction

Proximity of miRNAs to other genes within the genome and their location in introns has been found to influence the levels of gene expression. In the past few decades, cancer was considered to be the result of the changes in the expression of the gene that code for proteins. The association of miRNAs with the formation of tumours, have been recently confirmed [1].

The three-prime untranslated regions (3'UTR) is a section of the mRNA that follows the translation termination codon. It consists of sequences that are able to degrade or stabilize the transcript of mRNA. This region also contains sequences that signal, the need for addition to be made in product of translation. The role of the structural characteristics of 3'UTR in gene expression is crucial. The length of this untranslated region is significant; longer the length of the region, the lower the gene expression levels.

A specific miRNA may target several mRNAs and a specific mRNA may bind to several miRNAs, making target identification a major challenge. A co-operative repression effect is observed especially when the target regions of a miRNA cluster together. Identifying the targets of miRNA experimentally is a crucial step in obtaining prediction results that are reliable.

The methods available for recognizing the targets are mainly based on matching the patterns and they provide a redundant set of probable binding sites. Most of the algorithms already available for predicting the miRNA targets are based on the search for a complementarity of sequence between miRNA and the target. Cammaerts et al. [2] observed that these algorithms produced a large number of miRNA targets (gene) making it less possible to carry out experimental validations for all the targets.

cDNA microarray has been widely used for studying the gene expression. However, since one miRNA can repress the production of many proteins, proteomic data analysis would also be needed for the accurate detection of gene targets. This interaction between the miRNA and the target is importantly influenced by the seed region. The seed regions are regions present within the miRNA binding regions. Although there are several factors that pave a way to the binding between miRNA and mRNA, the impact of binding is determined by the seed sequence within the miRNA. The seed region consists of a continuous string of at least 6 to 8 nucleotides. miRNA recognizes its target by the degree of complementarity between the seed region, and few specific binding sites found along the sequence of mRNA.

The most significant prerequisite for proper repression, is a perfect seed pairing. The strength of base pairing contributes to the stability of the RNA-target duplex and is important in miRNA repression. This is confirmed by Govindarajan et al. [3].

The objective of this paper is to compare between the results of the binding region obtained from both the modified STFT method and the normalised correlation method based on the ground truth. The results obtained for both the methods are close to the ground truth. The result used for the comparison are the binding regions obtained from the peak intercepts of the spectral content values in the modified STFT method and binding regions obtained from the region of maximum correlation/coincidence from the point where the maximum strength in correlation is obtained for the correlation method.

2. Literature Studies

The recent developments in high throughput technologies relating to molecular genetics have brought a great impact to bioinformatics and computational biology. The application of these technologies includes the measurement of the gene and protein expression profiles in an integrated manner. Research on the huge amounts of data both genomic and proteomic, helps to understand the complex interactions between proteins and genes.

The significance of signal processing methods is attributed to their role in extracting, processing and interpreting the information present in both genomic and proteomic data. The importance of genomic signal processing is rising as it has been recognized that characterizing genomic and proteomic regulations require various disciplines related to signal processing. Systems biology demands a deeper understanding of systems theory. This entails the various methods of signal processing. Bioinformatics and Genomic Signal Processing both use computational methods to solve different biological problems [4].

Serpedin et al. [5] investigated the application of Genomic Signal Processing methods in detecting the coding regions. These methods include DNA to signal mapping and obtaining the spectral content of sections of the signal with STFT, using a sliding window having fixed length. STFT also finds application in the search for genomic repeats and also in finding the thermodynamic and bending properties of DNA by Fourier analysis. Several computer algorithms based on DSP methods including the existing STFT method have been applied for understanding the characteristics of RNA sequences. However, these algorithms provide less accurate and computationally complex solution with much background noise.

3. Method Used

This section discusses about the two methods analysed in this study: Modified STFT method and correlation method. These methods provide details about the position and sequence of the binding region. The correlation method also gives the details about the miRNA seed binding region. This section discusses about how these methods pave way to determining the binding region and the miRNA seed binding regions.

3.1. Determination of mRNA binding region using modified STFT method

Maggi et al. [6] considered applying the STFT to the indicator sequence (sequence obtained by replacing a nucleotide with its corresponding EIIP value) and selecting every sixth component to obtain the spectral content. In the modified STFT method, a normalizing technique is applied to the spectral content by dividing with the total spectral value. This is followed by scaling, to identify the peaks which give the binding regions. The regions where the binding occurs (binding regions), is selected by noting the value of only those peaks and the position of the peaks, which lie above the standard deviation of the total spectral content value. The length of the seed region is thought to consist of a string of at least 6 nucleotides. In order to find the seed region, the maximum value of the

normalised correlation using circular shift between miRNA (from ground truth) and the obtained binding region, is used.

RAC3 GTPase is a member of the GTPase superfamily that regulates an array that is diverse and includes cellular events. The RAC3 GTPase binds to a large number of proteins known as effectors and also plays a central role in the actin and microtubule cytoskeleton and the transcription of genes. Through these effects, RAC3 GTPase, which belongs to the Rho family, influence many normal cellular functions such as invasion and motility along with the progression and survival of the cell cycle. These breast cancer specific genes have roles that are far beyond regulating the actin cytoskeleton directly.

RAC3 GTPase acts as a molecular switch that help control the transduction of external signals in both nuclear and cytoplasmic effectors. In the past couple of years, the developments of genetic tools have helped in defining the specific roles of RAC3 GTPase in breast cancer.

The modified STFT method was also applied to BRCA1, another breast cancer specific mRNA, for the purpose of validation. Figure 1 shows the results obtained after applying the modified STFT method to BRCA1.

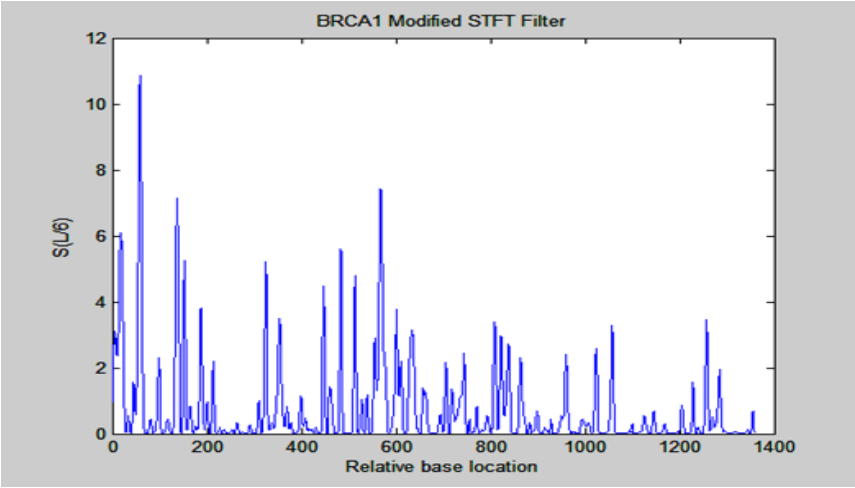

Fig. 1. Modified STFT plot data for BRCA1.

In the above graph, the x-axis denotes the location of the BRCA1 3'UTR bases. The length (number of bases) in the BRCA1 3'UTR is 1363 bases. The y-axis indicates the power spectral density values obtained by applying the modified STFT to the indicator sequence.

$$S[k]=|U[k]|^2 \quad (1)$$

where, $S[k]$ is the power spectral density/spectral content, and $U[k]$ is the array of values obtained by applying modified STFT method to indicator sequence.

BRCA1 is a tumour suppressor gene and the mutations of these genes are found to be a major cause of breast cancer. However, this gene is involved in the stability of the genome. BRCA1 gene is involved in cell cycle checkpoints. Investigations done on

various organisms for BRCA1, have paved way to providing knowledge of the role of BRCA1 in breast cancer and the evolution of BRCA1.

3.2. Correlation method for determining miRNA binding and seed binding regions

The seed types are of 4 types: 6-mer (bases 2-7), 7-mer (7-mer-A1 having bases 1-7, and 7-mer-m8 having bases 2-8), and 8-mer (bases 1-8). This is validated by Borrayo et al. [7]. The roles of miRNA sequences in controlling the expression and survival of target mRNA, has been applied by Mullany et al. [8] in both the recognition of target and the regulation. However, recent studies done by Laura et al. [9] have shown that the roles of miRNA sequences have gone beyond this seed region. The chosen miRNA was reverse complimented. The normalised correlation between the chosen mRNA and the above reverse complimented miRNA sequence were calculated, with lateral shifting by one base, to obtain the location of maximum correlation, which corresponds to the binding region.

Binthiya and Tessamma [10] observed that the seed regions are present in the first 8 bases of the miRNA and hence only a one-to-one match for this short length of bases with respect to the miRNA and binding region of the mRNA is required.

4. Implementation

As per the methods mentioned in section 3, the miRNA binding and seed binding regions are obtained for cancer specific mRNAs RAC3 GTPase and BRCA1. These are tabulated in Tables 1 and 2. The program codes for both correlation and modified STFT methods were designed using the built-in utility of MATLAB.

5. Database

The lists of the miRNAs that target an mRNA are obtained from the exiqon website [https://www.exiqon.com/miRSearch] and the miRNA sequences are obtained from the miRBase website [https://www.mirbase.org/]. The 3'UTR of the breast cancer specific mRNAs are obtained from the UCSC Genome browser website [https://www.genome.ucsc.edu/]. The breast cancer specific mRNAs RAC3 GTPase and BRCAI and the miRNAs that target both these mRNAs were used for analysis, in this study. The ground truth is taken from the CNR-ISMAC, Institute for the Study of Macromolecules website [11].

6. Results and Discussion

The number of miRNAs binding to RAC3 and BRCA1 were chosen as 37 and 35 respectively, for the study. The comparative details of the binding regions with respect to the modified STFT and correlation methods, for RAC3 and BRCA1 respectively, has been shown in Tables 1 and 2. Comparison between the two methods has also been done using standard deviation and range plots.

Considering the correlation method, seeds of all the 37 miRNAs binding to RAC3 (Table 1), have exactly the same binding position as that in the ground truth (columns 4 and 8); but 2 of the seeds have length less by 1 base (Sl. Nos. 21 and 29). Columns 9 and 10 give the seed sequence and seed type with respect to the correlation method.

Table 1. Comparative study on binding region details with respect to modified STFT and correlation methods for RAC3.

	Ground Truth			Modified STFT Plot Data		Correlation Method Results			
Sl. No. (1)	miRNA (2)	Length of miRNA (3)	Seed Binding Position in the UTR (4)	miRNA Binding Region in the UTR (5)	Peak in STFT plot (6)	miRNA Binding Position in UTR (7)	Seed Binding Position in the UTR (8)	Seed Sequence (9)	Seed Type (10)
1	hsa-miR-4687-5p	22	22-28	20-27	23	8-29	22-29	GAGGGCTG	1-8
2	hsa-miR-6829-5p	20	37-43	38-44	42	25-44 321-340	37-43 333-339	GCAGCCC GCAGCCC	2-8 2-8
3	hsa-miR-211-3p	21	78-85	77-91	84	65-85	78-85	GTCCCTGC	1-8
4	hsa-miR-4707-5p	23	113-119	111-125	118	97-119	113-119	GCCGGGG	2-8
5	hsa-miR-6840-5p	24	114-120	111-125	118	98-120	114-120	CCGGGGG	2-8
6	hsa-miR-6729-3p	21	128-135	125-134	129	115-135	128-135	GGGGATGA	1-8
7	hsa-miR-500b-3p	20	138-144	133-150	140	126-145	138-144	CTGGGTG	2-8
8	hsa-miR-676S-5p	25	166-172	158-171	166	148-172	166-172	CGCCTCA	2-8
9	hsa-miR-6820-5p	19	162-168	158-171	166	151-169	162-168	CTGCCGC	2-8
10	hsa-miR-4430	49	167-172	162-176	170	125-173	167-172	GCCTCA	2-7
11	hsa-miR-505-5p	22	183-190	182-196	190	169-190	183-190	TGGCTCCA	1-8
12	hsa-miR-3652	131	187-193	182-196	190	174-194	187-193	TCCAGCC	2-8
13	hsa-miR-6754-5p	22	79-85	77-91	84	64-85 64-85	79-85 79-85	TCCCTGG TCCCTGG	1-7 1-7
14	hsa-miR-1470	61	218-224, 285-289, 291-296	209-224	218	164-224 231-291 237-297	218-224 285-290 291-296	GGAGGGA GGAGGG GGAGGG	1-7 2-7 2-7
15	hsa-miR-4667-3p	21	218-224	209-224	218	204-224	218-224	GGAGGGA	1-7
16	hsa-miR-4725-5p	21	226-232	223-243	230	213-233	226-233	CAGGGTCT	1-8
17	hsa-miR-1184	23	242-248	243-253	248	227-249	242-249	GCTGCAGG	1-8
18	hsa-miR-6810-3p	23	245-251	243-253	248	230-252	245-251	GCAGGGG	2-8
19	hsa-miR-6801-3p	20	245-251	243-253	248	233-252	245-251	GCAGGGG	2-8
20	hsa-miR-210-5p	22	246-253	243-253	248	232-253	246-253	CAGGGGCA	1-8
21	hsa-miR-4749-3p	20	247-253	243-253	248	234-253	247-252	AGGGGC	2-7
22	hsa-miR-6804-3p	22	252-259	253-261	257	238-259	252-259	CAGGTGCA	1-8
23	hsa-miR-500a-3p	22	253-259	253-261	257	239-260	253-259	AGGTGCA	2-8
24	hsa-miR-3130-3p	21	254-260	253-261	257	241-261	254-260	GGTGCAG	2-8
25	hsa-miR-6752-3p	21	257-263	253-261	257	243-263	257-263	GCAGGGA	1-7
26	hsa-miR-4283	17	264-270	262-276	267	254-270	264-270	AGCCCCA	1-7
27	hsa-miR-185-3p	22	264-270	262-276	267	249-270	264-270	CAGCCCCT	1-8
28	hsa-miR-6852-5p	21	266-273	262-276	267	253-273	266-273	CCCCAGGA	1-8
29	hsa-miR-6742-3p	20	267-273	262-276	267	254-273	267-272	CCCAGG	2-7
30	hsa-miR-6783-3p	22	267-273	262-276	267	253-274	267-273	CCCAGGA	2-8
31	hsa-miR-1343-3p	22	267-273	262-276	267	253-274	267-273	CCCAGGA	2-8
32	hsa-miR-939-3p	21	267-273	262-276	267	253-273	267-273	CCCAGG	2-7
33	hsa-miR-671-5p	23	276-282	276-283	279	261-283	276-282	GGCTTCC	2-8
34	hsa-miR-6887-3p	21	285-291	284-291	288	272-292	285-291	GGAGGGG	2-8
35	hsa-miR-6865-3p	21	292-298	291-300	296	279-299	292-298	GAGGGTG	2-8
36	hsa-miR-6775-5p	25	319-326	316-324	321	302-326	319-326	TGCCCCGA	1-8
37	hsa-miR-6727-5p	23	320-326	316-324	321	305-327	320-327	GCCCCGAG	1-8

Table 2. Comparative study on Binding region details with respect to modified STFT and Correlation Methods for BRCA1.

	Ground Truth			Modified STFT Plot Data		Correlation Method Results			
Sl. No. (1)	miRNA (2)	Length of miRNA (3)	Seed Binding Position in the UTR (4)	miRNA Binding Region in the UTR (5)	Peak in STFT plot (6)	miRNA Binding Position in UTR (7)	Seed Binding Position in the UTR (8)	Seed Sequence (9)	Seed Type (10)
1	hsa-miR-615-5p	22	30-37	29-39	34	16-37	30-36	GGACCCC	2-8
2	hsa-miR-99b-3p	22	43-49	39-50	45	28-49	43-48	GAGCTT	2-7
3	hsa-miR-4717-5p	22	54-60	50-71	59	40-61	54-60	GTGGCCT	2-8
4	hsa-miR-4419b	68	132-138	128-146	137	74-141	134-140	AGCCTGA	2-8
5	hsa-miR-6847-3p	22	208-215	206-221	213	194-215	208-214	CATGAGC	2-8
6	hsa-miR-125a-3p	22	234-241	232-242	237	220-241	234-240	TCACCTG	2-8
7	hsa-miR-7158-5p	24	272-279	270-283	277	256-279	272-278	ATTGAGC	2-8
8	hsa-miR-15b-3p	22	286-292	284-296	291	271-292	286-291	TGATTC	2-7
9	hsa-miR-6507-5p	21	307-313	305-316	310	294-314	307-313	TATTCTT	2-8
10	hsa-miR-18b-3p	22	335-341	333-340	337	321-342	335-341	TTAGGGC	2-8
11	hsa-miR-526b-5p	23	364-370	363-374	369	349-371	364-371	CTCAAGAG	1-8
12	hsa-miR-6823-5p	21	404-410	404-412	408	391-411	404-410	AACCCTG	2-8
13	hsa-miR-3667-3p	22	444-450	441-454	447	429-450	444-449	GGAAGG	2-7
14	hsa-miR-6814-5p	22	456-462	454-474	461	441-462	456-462	CTTGGGA	1-7
15	hsa-miR-345-5p	22	480-486	477-493	483	465-486	480-485	AGTCAG	2-7
16	hsa-miR-3145-5p	22	512-518	505-520	513	498-519	512-518	TTGGAGT	2-8
17	hsa-miR-6793-3p	20	525-531	520-533	529	512-531	525-531	TTGGGGA	1-7
18	hsa-miR-576-5p	22	550-557	547-561	555	536-557	551-556	TTAGAA	2-7
19	hsa-miR-512-3p	22	606-612	607-622	610	592-613	606-613	CAGCACTT	1-8
20	hsa-miR-29b-2-5p	22	655-661	650-672	658	641-662	655-662	GAAACCAG	1-8
21	hsa-miR-4743-3p	21	701-707	700-712	706	687-707	701-707	ACAGAAA	1-7
22	hsa-miR-6869-5p	22	744-750	727-750	744	730-751	744-750	CTACTCA	2-8
23	hsa-miR-4772-3p	22	793-799	786-802	793	779-800	793-799	GTTGCAG	2-8
24	hsa-miR-371a-3p	23	818-824	816-830	822	803-825	818-824	CGGCACT	2-8
25	hsa-miR-6879-3p	21	832-839	830-847	838	819-839	832-839	GGGTGACA	1-8
26	hsa-miR-1277-5p	24	911-917	909-923	915	895-918	911-918	ATATATTT	1-8
27	hsa-miR-95-5p	21	1018-1024	1014-1034	1023	1005-1025	1018-1024	TTTATTG	2-8
28	hsa-miR-7856-5p	21	1053-1059	1047-1066	1058	1039-1059	1053-1058	CTTAAA	2-7
29	hsa-miR-548c-3p	22	1141-1147	1137-1154	1146	1127-1148	1141-1147	GATTTTT	2-8
30	hsa-miR-545-3p	22	1163-1169	1162-1177	1168	1148-1169	1163-1169	TTGCTGA	1-7
31	hsa-miR-205-5p	22	1202-1208	1198-1216	1205	1187-1208	1201-1208	TGAAGGA	1-7
32	hsa-miR-629-5p	21	1226-1233	1222-1235	1229	1213-1233	1226-1233	TAAACCCA	1-8
33	hsa-miR-3690	23	1237-1243	1236-1249	1239	1221-1243	1237-1242	TCCAGG	2-7
34	hsa-miR-328-5p	23	1280-1286	1275-1293	1285	1265-1287	1280-1287	GCCCCCCC	1-8
35	hsa-miR-5003-5p	23	1337-1343	1329-1349	1342	1322-1344	1337-1343	TGTTGTG	2-8

Similarly, considering the correlation method, seeds of all the 35 miRNAs binding to BRCA1 (Table 2), have the same binding position as that in the ground truth; but 3 of the seeds have length less by 1 base (Sl. Nos. 5, 13 and 28).

In the case of modified STFT, considering the peak intercepts, all the seeds given in the ground truth are present in the intercept regions with respect to RAC3 and BRCA1. It may be noted that only 9 peaks (34.6%) for RAC3 and 24 peaks (33.3%) for BRCA1 are having values above standard deviation of the total spectral content.

But to find the exact miRNAs which bind to an mRNA at its peak positions and the exact binding region, a next level of processing is needed.

6.1. Standard deviation

The overall mean of the binding regions of all the miRNAs that bind to RAC3 (as mentioned in the ground truth data), was initially calculated.

$$\text{Overall mean, } O_{MG} = (a_{1G}+a_{2G}+a_{3G}+\ldots\ldots a_{NG})/N_G \quad (2)$$

where, a_{1G}, a_{2G}, a_{3G} etc., is the mean of individual binding regions mentioned in the ground truth and N_G is the total number of binding regions in the ground truth.

The mean of the individual binding regions, obtained using the correlation method, was calculated. The standard deviation of the binding regions obtained using correlation method, was calculated using the formula standard deviation of correlation method.

$$\sqrt{\sum((a_{1c}-O_{MG})^2+(a_{2c}-O_{MG})^2+\cdots(a_{Nc}-O_{MG})^2)/Nc} \quad (3)$$

where, a_{1c} is the mean of the first binding region obtained from the correlation method and N_c is the total number of binding regions obtained using the correlation method.

The bar graphs, Figs. 2 and 3, highlight the overall mean and the standard deviation values respectively of both the methods for RAC3. The overall mean values obtained are 213.0625 for the correlation method and 214.459 for the MSTFT method. The standard deviation values obtained are 77 for the correlation method and 79 for the MSTFT method. The deviation of the correlation method is lesser showing its proximity to the data available in the ground truth.

Similarly, the standard deviation of the binding regions obtained using the Modified STFT method, was calculated using the formula standard deviation of MSTFT method.

$$\sqrt{\sum((a_{1s}-O_{MG})^2+(a_{2s}-O_{MG})^2+\cdots(a_{Ns}-O_{MG})^2)/Ns} \quad (4)$$

where, a_{1s} is the mean of the first binding region obtained from the MSTFT method and *Ns* is the total number of binding regions obtained using the MSTFT method.

The standard deviation values obtained for both the methods for RAC3 and BRCA1, has been shown below.

The bar graphs, Figs. 4 and 5, highlight the overall mean and the standard deviation values respectively of both the methods for RAC3. The overall mean values obtained are 634.24 for the correlation method and 641.26 for the MSTFT method. The standard deviation values obtained are 388.22 for the correlation method and 391.54 for the MSTFT method. In this case too, the deviation of the correlation method is lesser showing its proximity to the data available in the ground truth.

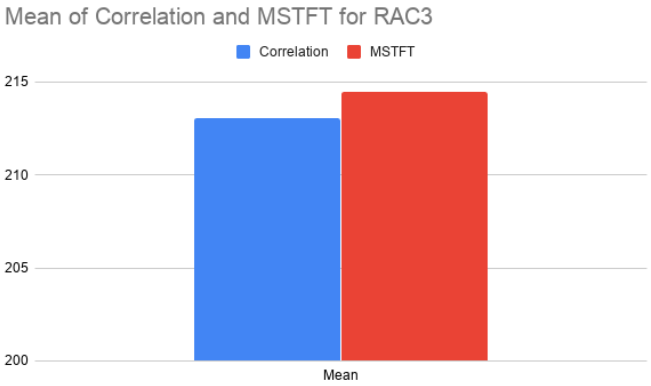

Fig. 2. Overall mean of correlation and modified STFT for RAC3.

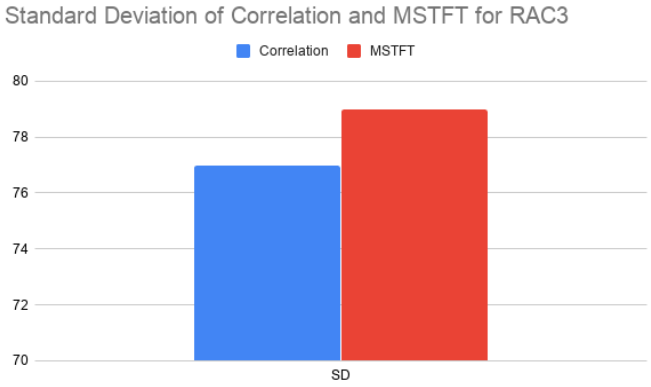

Fig. 3. Standard deviation of correlation and modified STFT for RAC3 (with respect to the ground truth).

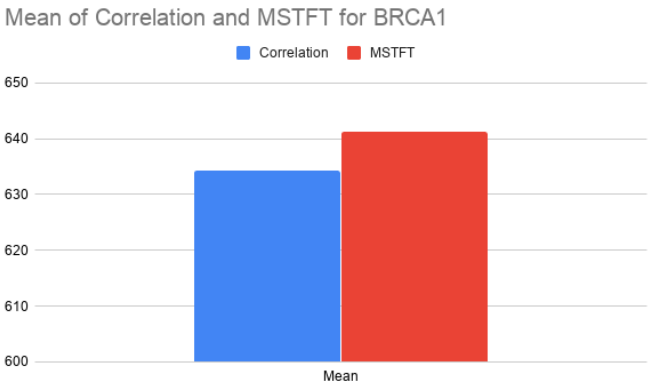

Fig. 4. Overall mean of correlation and modified STFT for BRCA1.

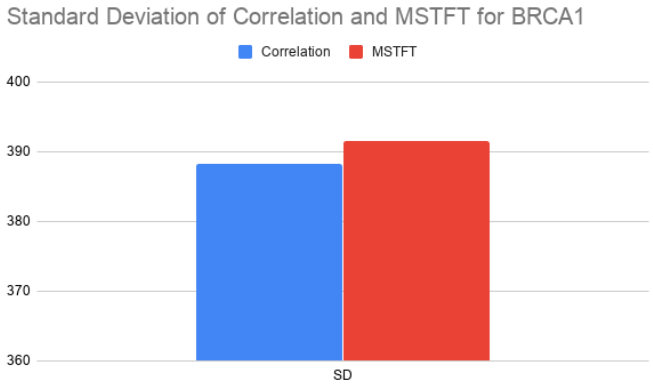

Fig. 5. Standard deviation of correlation and modified STFT for BRCA1 (with respect to the ground truth).

6.2. Range plots

The binding region range plots for MSTFT and correlation of RAC3 are shown in Figs. 6 and 7, respectively. The binding region ranges for the correlation method are higher giving it an advantage of being able to incorporate higher number of seed regions within a binding region. In table 1, Sl. No. 28, the binding region 253-273 has several seed binding regions 253-259, 254-260, 257-263, 264-270, 266-273, 267-272, 267-273. A greater number of seed binding regions within a binding region, makes it a prominent area to study the extent of repression caused by the respective miRNAs and thereby the extent of progression towards cancer.

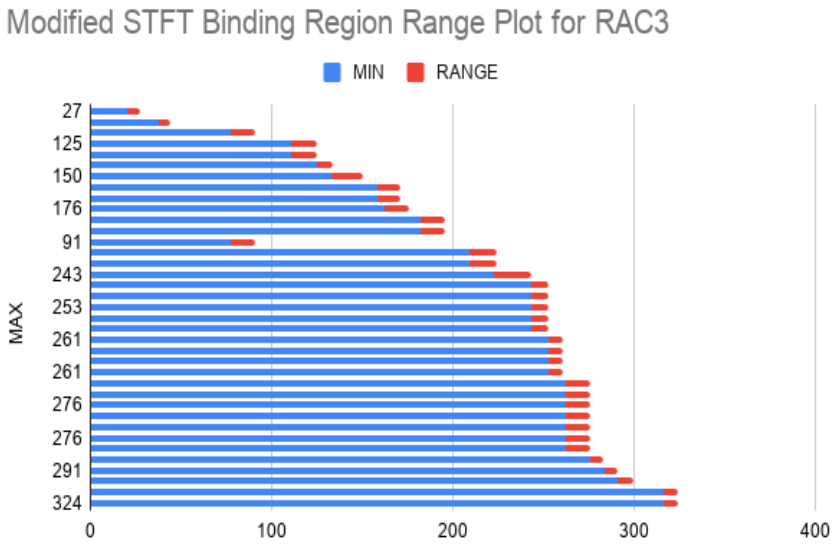

Fig. 6. Range plot for binding regions of modified STFT for RAC3.

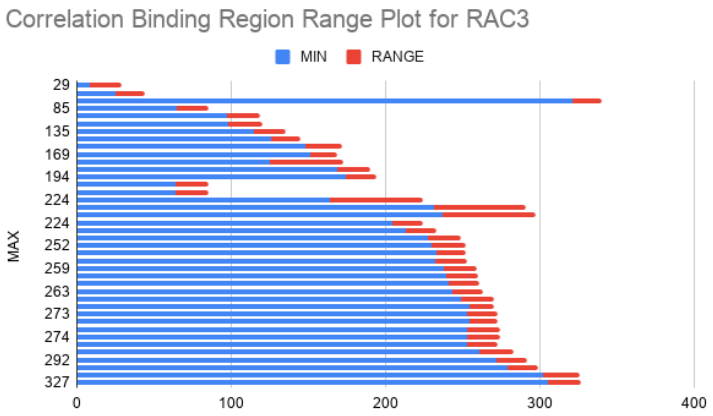

Fig. 7. Range plot for binding regions of correlation for RAC3.

6.3. The position of the seed binding region

In correlation method, the binding region obtained is exact. In order to find the seed region, only a one-to-one match for this short length of bases with respect to the miRNA and binding region of the mRNA is required.

6.4. Computational load

In modified STFT method, the computation required is more; FFT calculation is applied to the indicator sequence using MATLAB and is essential. In this method, the binding region obtained is approximate. The seed regions are obtained using normalised correlation. In order to find the seed region, the maximum value of the normalised correlation using circular shift between miRNA and the obtained binding region, is done.

However, in correlation method, the binding regions are obtained at the point of maximum correlation and the seed regions are obtained with just a one-to-one match between the miRNA and the mRNA binding region.

7. Conclusions

Out of the methods that were analysed namely, modified STFT method, and correlation method, for detecting the miRNA binding and seed binding regions for the breast cancer specific mRNAs, RAC3 GTPase and BRCA1, correlation method provided results exactly the same as that in ground truth. The modified STFT method, gives peaks, corresponding to all the approximate binding regions, with the seed positions within the peak intercepts. However, to find the specific miRNA which binds to a peak and the exact seed position, further processing is required.

Future Scope

The two methods considered in this study, have provided results that could help in further analysing the nature of the miRNA and mRNA interaction. Identifying the seed region is also crucial in evaluating this interaction. The correlation method has

advantage over the modified STFT method in finding not only the binding regions but also the seed binding regions of miRNA.

Identifying the seed region is crucial in evaluating the extent of breast cancer progression [12]. The next phase of this research would include finding the relation between seed length and the complexity of breast cancer cells.

Conflict of Interest

The authors declare that they have no competing interests.

Acknowledgement

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

Nomenclatures

a_{1c}	Mean of the first binding region obtained from the correlation method
a_{1G}	Mean of the first binding region in the ground truth
a_{1s}	Mean of the first binding region obtained from the MSTFT method
a_{2G}	Mean of the second binding region in the ground truth
a_{3G}	Mean of the third binding region in the ground truth
N_c	Total number of binding regions obtained using the correlation method
N_G	Total number of binding regions in the ground truth
Ns	Total number of binding regions obtained using the MSTFT method
O_{MG}	Overall mean (Ground truth)
S	Power spectral density / Spectral content
U	Array of values obtained by applying MSTFT to indicator sequence

Abbreviations

DSP	Digital signal processing
FFT	Fast Fourier transform
GTP	Guanosine triphosphate
STFT	Short-time Fourier transform
MSTFT	Modified short-time Fourier transform
mRNA	Messenger Ribonucleic acid
miRNA	Micro ribonucleic acid
RAC3	Ras-related C3 botulinum toxin substrate 3
BRCA1	Breast cancer type 1 susceptibility
3'UTR	Three prime untranslated regions

References

1. Catalanotto, C.; Cogoni, C.; and Zardo, G. (2016). MicroRNA in control of gene expression: an overview of nuclear functions. *International Journal of Molecular Science*, 17(10), 1-17.

2. Cammaerts, S.; Strazisar, M.; Rijk, P.D.; and Favero, J.D. (2015). Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. *Frontiers in Genetics*, 6(186), 1-12.
3. Govindarajan, R.; Duraiyan, J.; Kaliyappan, K.; and Palanisamy, M. (2012). Microarray and its applications. *Journal of Pharmacy and BioAllied Sciences*, 4(2), 310-312.
4. Bayat, A. (2002). Science, medicine, and the future: bioinformatics. *US National Library of Medicine: National Institute of Health*, 324(7344), 1018-1022.
5. Serpedin, E.; Garcia-Frias, J.; Huang, Y.; and Braga-Neto, U. (2009). Applications of signal processing techniques to bioinformatics, genomics, and proteomics. *EURASIP Journal on Bioinformatics and Systems Biology*, Article number: 250306, 1-2.
6. Maggi, N.; Arrigo, P.; and Ruggiero, C. (2013). Optimize ncRNA targeting: a signal analysis based approach. *IFMBE Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing*. Cyprus, 662-665.
7. Borrayo, E.; Mendizabal-Ruiz, G.; Vélez-Pérez, H.; Romo-Vazquez, R.; Mendizabal, A.P.; and Morales, J.A. (2014). Genomic signal processing methods for computation of alignment-free distances from DNA sequences. *PLOS ONE,* . 9(11): e110954.
8. Mullany, L.E.; Herrick, J.S.; Wolff, R.K.; and Slattery, M.L. (2016). MicroRNA seed region length impact on target messenger RNA expression and survival in colorectal cancer. *PLOS ONE*, 11(4): e0154177.
9. Laura, B.; Amy, E.; Pasquinelli, (2018). miRNA targeting: Growing beyond the seed. *Trends in Genetics*, 35(3), 215-222.
10. Binthiya, S.G.; and Tessamma, T. (2018). Breast cancer specific microRNA-mRNA binding region analysis using signal processing techniques. *Proceedings of the IDES Joint International Conference on Recent Applications and Trendsin Modern Engineering (RATME)*. Kochi, India, 120-136.
11. Salvatore, L. (2005). Retrieved January 10, 2019, from http://www.bi.ismac.cnr.it/index.html.
12. Huang, J.C.; Babak, T.: Corson, T.W.; Chua, G.; Khan, S.; Gallie, B.L.; Hughes, T.R.; Blencowe, B.J.; Frey, B.J.; and Morris, Q.D. (2007). Using expression profiling data to identify human microRNA targets. *Nature Methods*, 4, 1045-1049.