J. Appl. Environ. Biol. Sci., 8(7)36-45, 2018

© 2018, TextRoad Publication

ISSN: 2090-4274
Journal of Applied Environmental
and Biological Sciences
www.textroad.com

In Silico Analysis to Predict Structure, Sequence Motif and Expression Level of Breast Cancer Biomarker Molecules: Protein and miRNA

Romana Siddique*, Jannatul Ferdous

Biotechnology Programme, Department of Mathematics and Natural Sciences,
BRAC University, 66 Mohakhali, Dhaka-1212, Bangladesh

Received: January 19, 2018
Accepted: April 19, 2018

ABSTRACT

Breast cancer is an important public health issue since it has become the reason of 69% cancer caused death in women throughout the whole world and 15% of cancer death in Bangladesh. According to recent findings, biomarker studies can help in the betterment of diagnosis, treatment and recurrence problem by constructing a biomarker panel. The project was designed to find the basic properties -structure, sequence motif and expression level of potential biomarker molecules of breast cancer. Only protein and miRNA were selected as biomarkers since they were easily collectable from body fluid. 11 proteins and 7 miRNAs were selected, as they were the one showing high specificity and sensitivity as biomarkers. The protein molecules were - ER, ER Beta, PR, TTR, Ki67, HSP60, Her2, CyclinD1, Cyclin E, P53 and CEA. The miRNAs were- miR10b, miR21, miR145, miR155, miR191, miR 382 and miR425. Bioinformatics approach was the fundamental base of this research to detect properties of these biomarkers. For structure SWISS MODEL Workspace (protein), mfold (miRNA), for sequence motif MEME, and to check expression level GEO Profiles were used. In the end of this study it was seen that CEA, TTR, ER, PR, ER Beta, Cyclin E and Ki67 were the proteins that could be a potential biomarker for breast cancer screening panel. CyclinD1, Her2, P53 along with miR155 are potential biomarkers for breast cancer staging and miR10b, miR21 can be potential biomarker for ER silencing treatment.

KEY WORDS: miRNA, cancer biomarker, P53, CEA, Cyclin E, Ki67

1. INTRODUCTION

Being a significant contributor to overall morbidity and mortality, breast cancer is by far the most frequent cancer among women in both developed and developing countries with an estimated 1.38 million new cancer cases (Kulasingam, 2008). And with 15% cancer death in women in Bangladesh (Sacha, 2014) and incidence rate 22.5 per 100000 in females. (Rai, 2012). It has become a hidden burden in Banglqdesh. Studies have proved that, this condition is the result of the absence of suitable diagnostic or screening test for an early detection of clinically relevant breast cancer. The current screening methods used to detect breast tumors either benign or malignant, include clinical breast examination (CBE), mammography and ultrasound. (Kulasingam, 2008). All processes have their own limitations and also a very high false negetive rates (Kulasingam, 2008).

One of the most promising ways to achieve methods with improved sensitivity and specificity is through the use of cancer biomarkers. (Kulasingam, 2008). A joint venture on chemical safety, led by WHO with the United Nations and the International Labor Organization has defined a biomarker as "any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease."(Mandal, 2013.) Among all the different molecules in a human body- Protein and miRNA are known to perform the best as biomarker for specific diseases. Protein molecules are known as the best of biomarker molecules since they can easily be traced, studied and evaluated (Gam, 2010).

In the present study 11 protein molecules and seven miRNAs were selected (Table 1 & 2) to predict their structure, sequence motifs and expression level in cancer state. Protein molecules were :Estrogen Receptor (ER), Estrogen Receptor beta (ERbeta), Progesterone (PR), Transthyretin (TTR), Cyclin E, Cyclin D1, Her2(Human Epidermal Growth Receptor),Carcinoembryonic Antigen (CEA), P53, Heat Shock Protein60 (HSP60). When it comes to miRNA, Patterns of miRNA expression plays a very important role in oncogenesis. Because of their distinct patterns of expression associated with cancer type, remarkable stability in blood and other body fluids, miRNAs are considered to be highly promising cancer biomarkers (Zhao, 2010). Among the miRNAs, seven were selected for this study – miRNA10B, miRNA21, miRNA145, miRNA155, miRNA191, miRNA382, miRNA425.

Corresponding author: Romana Siddique, Biotechnology Programme, Department of Mathematics and Natural Sciences, BRAC University, 66 Mohakhali, Dhaka-1212, Bangladesh. Email: romanasiddique@gmail.com
Telephone:+880-2-8824051-4 Ext.4060 Fax:+880-2-58810383

The aim of this study was to explore the structure,motifs and expression level of some selected cancer biomarkers so that these can be used as a possible therapeutics in cancer treatments and for early brest cancer screening.

Table 1: Protein markers associated with Breast cancer

Protein	Accession number	Taxonomic name
ER	AAI28574.1	Estrogen receptor 1
PR	BAC06585.1	Progesterone receptor
HER 2	AAA75493.1	human epidermal growth factor receptor 2
CEA	CAE75559.1	carcinoembryonic antigen
KI67	NP_002408.3	Ki-67 protein
CYCLIN D1	AAH23620.1	Cyclin D1
CYCLIN E	NP_001229.1	G1/S-specific cyclin-E1
ER BETA	AAV31779.1	estrogen receptor 2 (ER beta)
TTR	CAG33189.1	Transthyretin
P53	BAC16799.1	Tumor protein p53
HSP60	AAF66640.1	heat shock protein HSP60

Table 2: miRNA markers associated with Breast cancer

miRNA	GI number	Taxonomic name
MIR10B	262206216	Homo sapiens microRNA 10b
MIR21	262205659	Homo sapiens microRNA 21
MIR145	262205329	Homo sapiens microRNA 145
MIR155	269846817	Homo sapiens microRNA 155
MIR191	262205347	Homo sapiens microRNA 191
MIR382	262206264	Homo sapiens microRNA 382
MIR425	262205357	Homo sapiens microRNA 425

2. METHODS

In the current study , all the sequences information for the miRNA were retrieved from miRBase (www.mirbase.org) and sequence information for proteins were retrieved from NCBI databse (https://www.ncbi.nlm.nih.gov/) and Uniprot(www.uniprot.org/) .In silico analysis for structure, expression level and sequence motifs were carried out for selected proteins and miRNAs.

2.1 Structure prediction

Mfold web server version 3.5 (unafold.rna.albany.edu/) was used to predict secondary folded structure of miRNA (Zucker M., 2003).

To predict protein structures, first the FASTA sequences were retrieved from a database, Uniprot (www.uniprot.org/). The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. UniProt is a collaboration between the European Bioinformatics Institute (EMBL-EBI), the SIB Swiss Institute of Bioinformatics and the Protein Information Resource (PIR)(www.uniprot.org/). Then blast was done using the Basic Local Alignment Search Tool (BLAST) (https://www.ncbi.nlm.nih.gov/BLAST/) with these sequences to find their best suited templates. After that alignment was checked with this sequence and their template with Clustal omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). And finally this alignment result was given as an input in the homology modelling website, Swiss model workspace (https://swissmodel.expasy.org/workspace/) which is a web-based integrated service dedicated to protein structure homology modelling. It assists and guides the user in building protein homology models at different levels of complexity (Arnold K, 2006).

2.2 Finding Sequence Motif

MEME Suite (meme-suite.org/) was used to discover sequence motif in both the cases of protein and miRNA molecules. The MEME Suite is a software toolkit with a unified web server interface that enables users to perform four types of motif analysis: motif discovery, motif–motif database searching, motif-sequence database searching and assignment of function. It offers a significantly expanded set of programs for these tasks compared with the earlier web server (Bailey, T. L., 2006).

2.3 Expression Level Observation

To find out the expression level of the selected proteins and miRNA molecules in different conditions of breast cancer, GEO (https://www.ncbi.nlm.nih.gov/geo/) of NCBI was used. GEO represents to Gene Expression Omnibus. The Gene Expression Omnibus (GEO) is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomic data sets(Barrett T,2013 The GEO Profiles (www.ncbi.nlm.nih.gov/geoprofiles/) database stores gene expression profiles derived from curated GEO Datasets (https://www.ncbi.nlm.nih.gov/gds). Each Profile is presented as a chart that displays the expression level of one gene across all Samples within the Dataset.

3. RESULTS

3.1 Structure Prediction for miRNAs

Basically in this study, finding the secondary hairpin like structure of the miRNAs was the main focus. Knowing this kind of structure specifically is important because the folding in a certain way leads to the solution of different unanswered question in binding and functioning of that miRNA molecule. Secondary structure for seven miRNA moleculaes were predicted using Mfold (Fig.1)

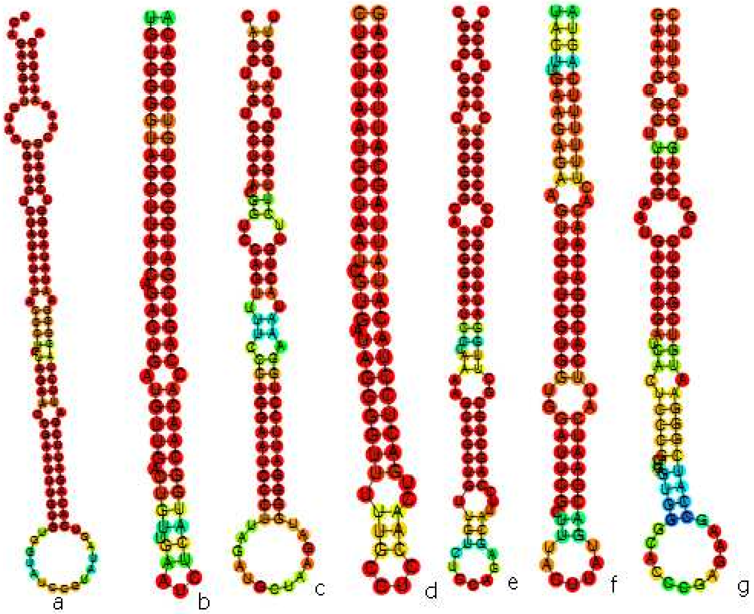

Fig 1: Hairpin structure of a) miR10B b) miR21 c) miR145 d) miR155 e) miR191 f) miR382 g) miR425

3.2 Motif Analysis for miRNA

For seven selected miRNAs, motif was observed (Fig.2) using MEME Suite (meme-suite.org/). Motif means a sequence that can have special importance biologically or functionally. Knowing motifs are important because they are recurrent and they indicate binding sites or functional sequence of that molecule.

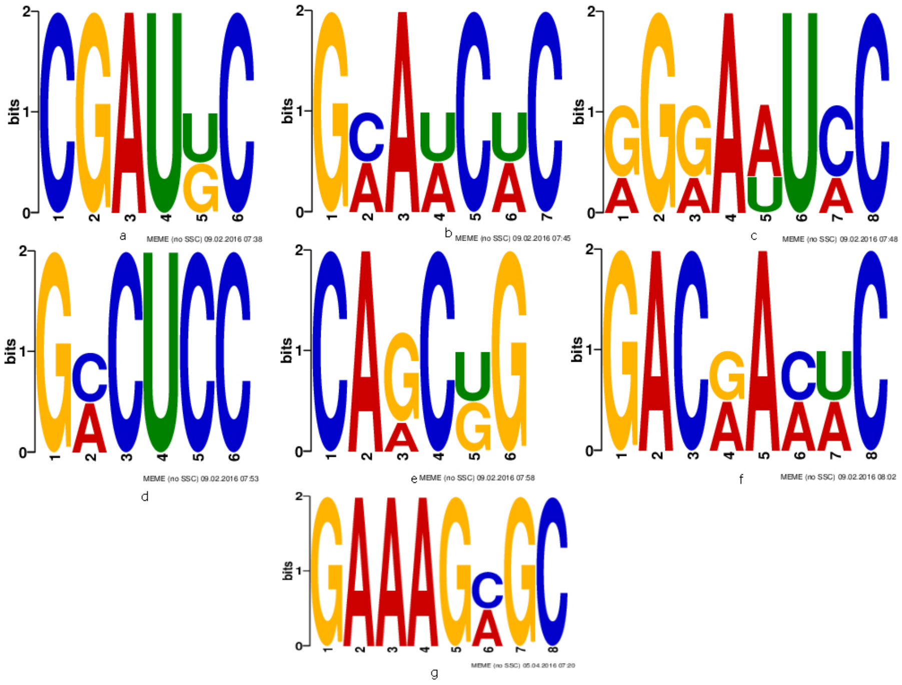

Fig 2: Important sequence motif of a) miR10B b) miR21 c) miR145 d) miR155 e) miR191 f) miR382 g) miR425

3.4 Analysis of Expression Level for miRNAs:

GEO tool was used to observe expression of seven miRNA molecules to observe expression pattern in different conditions of breast cancer. This is a tricky one to observe because this involves a multiple step to be expressed and also because this is an actual indicating property that makes the miRNA molecules a biomarker. Out of seven three miRNA molecule (miR10B, miR21, miR155) showed significant changes in their expression level. The expression level was measured in normal breast cancer patient and in patient whom were treated with ER mutation. It is seen that miR10B expression level is higher in the ER mutated cells and miR21 is low expressed in the ER mutated cells (Fig. 3).

This expression profile of miR155 showed expression in two different breast cancer cell lines along with a control. It was seen that in both cancerous cell lines (MDA-MB-436 and HCC 1954) miR155 is expressed differently than the normal cell line. It is expressed more in the MDA-MB-436 cell line and less in the HCC 1954 cell line (Fig. 3).

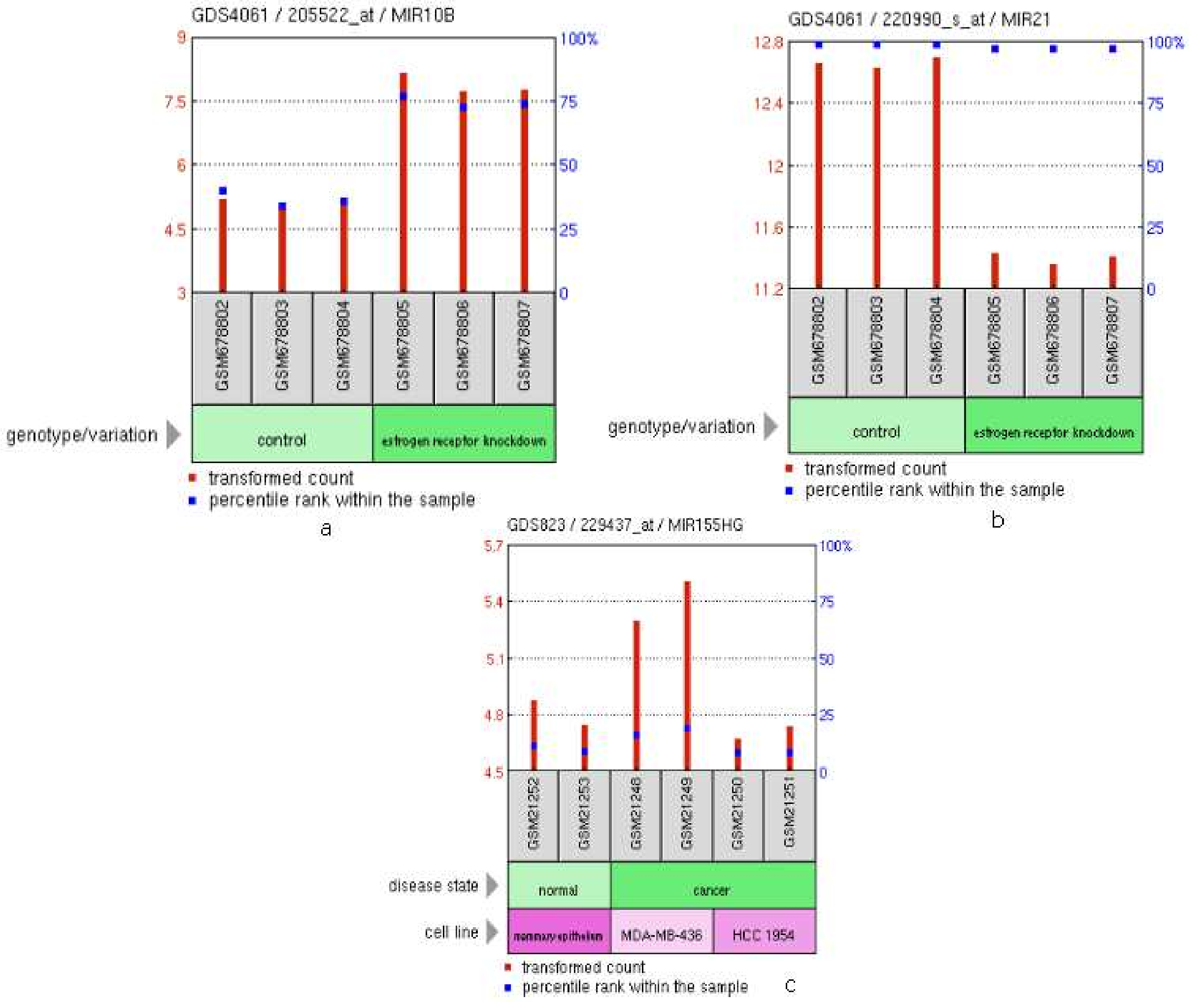

Fig 3: Expression level of a) miR10B b) miR21 c) miR155

3.5 Homology Modelling for Protein Biomarkers

Homology modelling was done using For the 11 protein biomarkers using Swiss model workspace (https://swissmodel.expasy.org/workspace/) (Fig.4). Predicting the homology model of these molecules can help in determining the structural motifs as well as site directed mutagenesis that might make them a candidate for bi-omarker panel of breast cancer.

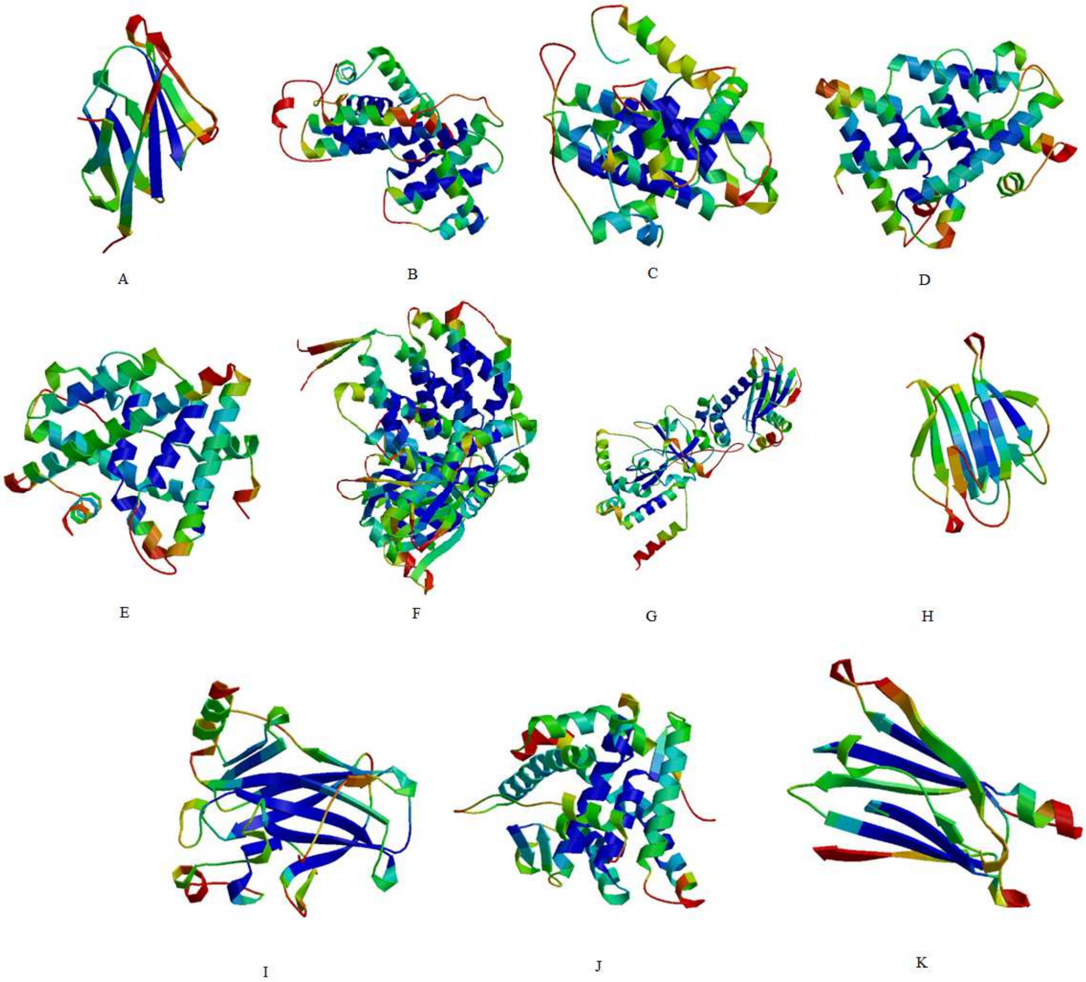

Fig 4: Homology model of a) CEA b) Cyclin D1 c) CyclinE d) ER e) ER Beta f) HSP60 g) HER2 h) Ki67 i) P53 j) PR k) TTR

3.6 MOTIF Analysis for protein biomarkers

Motifs for protein were also observed using MEME software (meme-suite.org/) was used and for each protein at least five motifs were observed. From these most significant motifs were selected (Fig.5). Knowing protein motifs are important because they give a clear information about the effects of sequence variation, protein interaction etc.

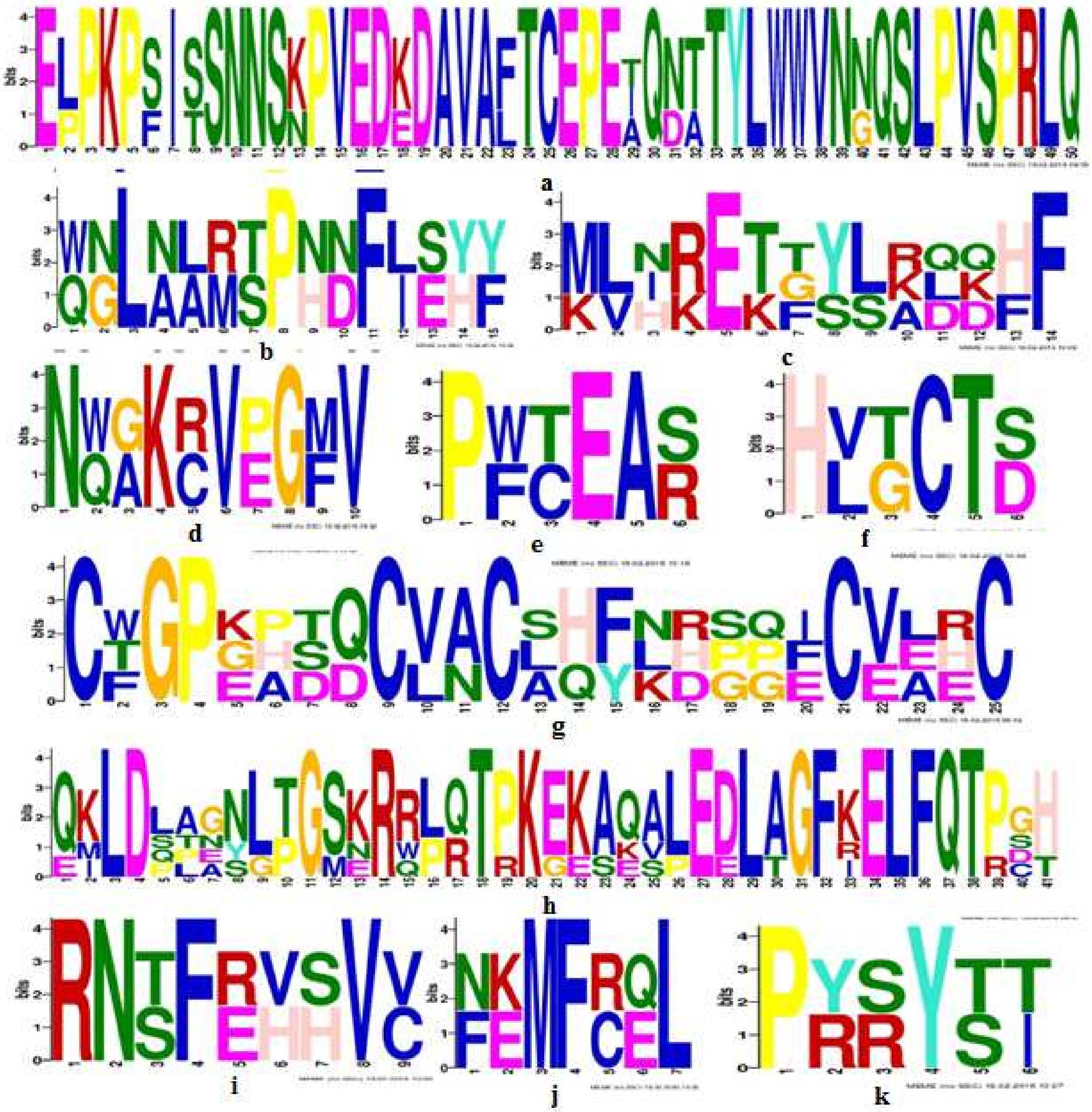

Fig 5: Important Sequence Motif of a)CEA b)Cyclin D1 c) Cyclin E d) ER e) ER Beta f) HSP60 g) HER2 h) Ki67 i) P53 j) PR k) TTR

3.7 Analysis of Expression Level for Proteins

The expression of CEA, CyclinE, ER, ER Beta, HSP60, Ki67 and TTR were almost similar. All of them were over-expressed in both the breast cancer cell lines (MDB-MB-436 and HCC 1954) (Fig.6). On the other side expression level of Cyclin D1, Her2 and P53 were almost similar (Fig.6). They were more expressed in the HCC 1954 cell line and less expressed in the MDB- MB- 436 cell line. Lastly PR was differently expressed than all as it was under expressed in both of the cell lines than the normal one.

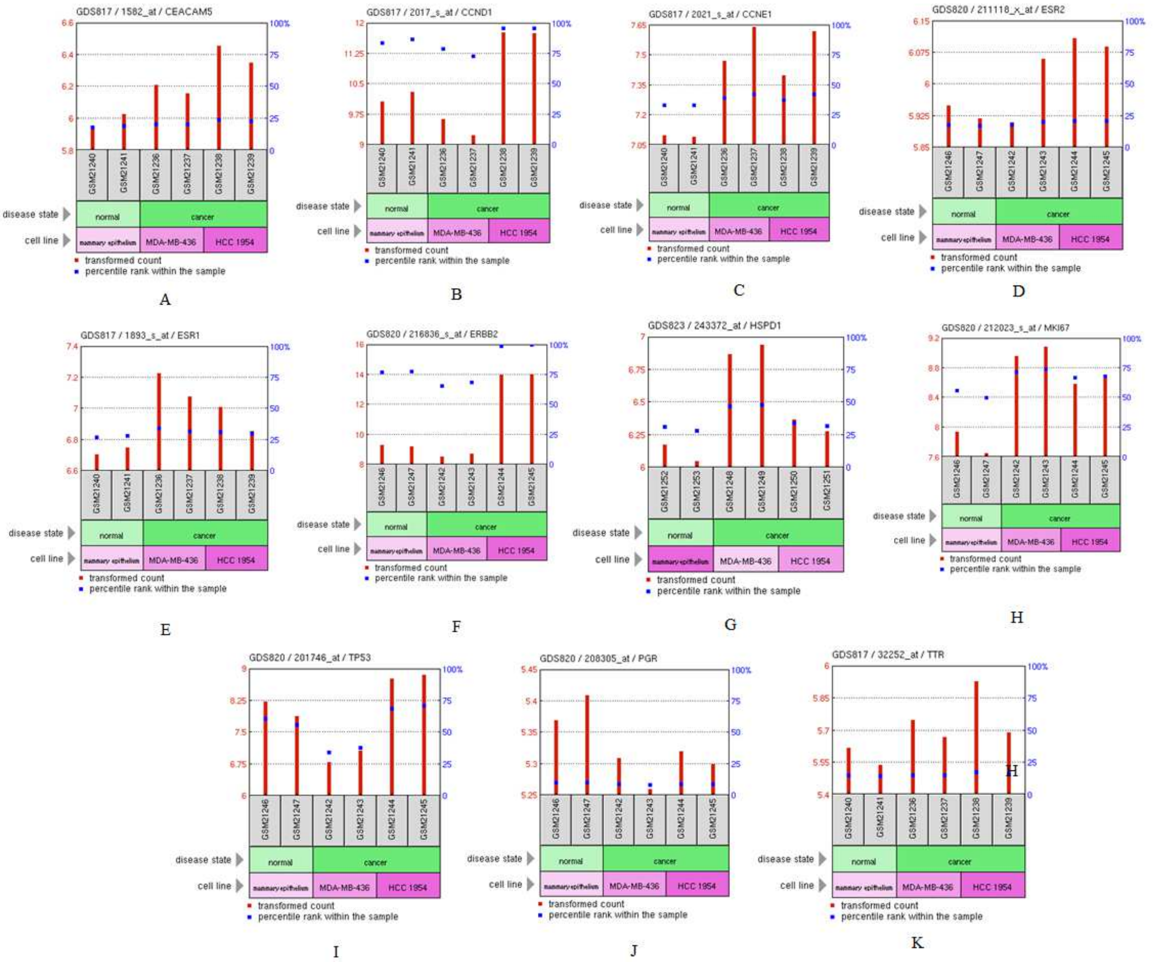

Fig 6: Expression level of a) CEA b) Cyclin D1 c) Cyclin E d) ER e) ER Beta f) HSP60 g) HER2 h) Ki67 i) P53 j) PR k) TTR

4. DISCUSSION

In the structure section secondary structure of miRNA and tertiary structure of proteins are shown. All the homology models are validated with QMEAN and the QMEAN score says all of them are more than 0.77% in a scale of 0 to 1.Sequence motif is the sequence that might have biological or functional importance. (Timothy, 1994).For each molecule maximum five motifs were commanded to be found and the standard length is from 6 to 50. Most of the motifs are present in two sites, but only a few are present in more than two sites. In the expression level observation results were taken from GEO Profile. From the huge store of different comparisons suitable results were taken. All these different cell lines are presented in the bottom light pink bars. And the diseased state are shown in the second bottom green bars. Above these two line of bars, ash colored bars show the name of the samples. In the long red lines that represents the transformed count of the expression level. This transformed count is from the actual experiment results as they were performed in affymetrix systems. And the blue squares presents their percentile rank among all the samples.

If these results are combined together a better biomarker panel could be decided with CEA, Cyclin E, ER, ER Beta, HSP60, KI67, TTR and PR. On the other hand Cyclin D1, Her2, P53 along with miR155can make a biomarker panel for breast cancer staging. mi10B and miR21 can play biomarker role in the ER silencing treatment systems.

5. CONCLUSION

Breast cancer is a global curse. This is the most commonly encountered cancer in our country as well as in the whole world (Y. Baskin, 2010). With an objective to add a little help in the findings of better treatment and diagnosis system this study was designed to know more about breast cancer biomarker molecules like protein and miRNA. It is hoped that individually the information of these biomarker molecule can help in finding a new therapeutic agent, site directed mutagenesis, virtual screening. Also together they can make a panel of biomarkers with better specificity and sensitivity that is needed the most at this moment. (Li, 2002).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgement
The authors would like to thank Department of Mathematics and Natural Sciences, BRAC University.

REFERENCES

1. David P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function ,Cell, Vol. 116, 281–297
2. Jacques F., Hai-Rim S., Freddie B., David F., Colin M. and Donald M. (2008) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008,International Journal of Cancer
3. Story H., Love R., Salim R., Roberto A., Krieger L., and Ginsburg G. (2012) Improving Outcomes from Breast Cancer in a Low-Income Country: Lessons from Bangladesh ,International Journal of Breast Cancer Volume 2012, Article ID 423562, 9 pages doi:10.1155/2012/4 23562
4. Long D., Lee R., Williams P., Chan C., Ambros V.& Ding Y. (2007) Potent effect of target structure on microRNA function; doi:10.1038/nsmb1226
5. Liu J., Huang W., Yang H. &Luo Y. (2015) Expression and function of miR-155 in breast cancer, Biotechnology & Biotechnological Equipment, 29:5, 840-843, DOI: 10.1080/13102818.2015.1043946
6. Chung L., Moore K., Phillips L., Boyle F., Marsh D.and Baxter R. (2014) Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer, Breast cancer research
7. Li J., Zhang Z., Rosenzweig J., Wang Y, and Chan D., (2002), Proteomics and Bioinformatics Approaches for Identification of Serum Biomarkers to Detect Breast Cancer, , Clinical Chemistry 48:8 1296–1304
8. Alexander H., Stegner A., Mann C., Bois G., Alexander S., and Sauter E.(2004) ,Proteomic Analysis to Identify Breast Cancer Biomarkers in Nipple aspartate fluid Clinical cancer research ,Vol. 10, 7500–7510
9. Voduc K., Cheang M., Tyldesley S., Gelmon K., Nielsen T., Kennecke H., Oncol J. (2010) Breast Cancer Subtypes and the Risk of Local and Regional Relapse, 28:1684-1691.
10. Li J., Orlandi R., White C., Rosenzweig J., Zhao J., Seregni E., Morelli D.,Yu Y., Meng X., Zhang Z., Eric N., Fung T., and Chan D. (2005) Independent Validation of Candidate Breast Cancer Serum Biomarkers Identified by Mass Spectrometry, , Clinical Chemistry 51:12, 2229–2235.
11. Baskın Y. and Yiitbaı T. (2010) Clinical Proteomics of Breast Cancer , Current Genomics, 11, 528-536
12. Bhatt A., Mathur R., Farooque A., Verma A. & Dwarakanath B.(2010) Cancer biomarkers - Current perspectives
13. Kim B., Lee J., Park P., Shin Y., Lee W., Lee K., Ye S., Hyun H., Kang K., Yeo D., Kim Y., Ohn S., Noh D. and Kim C. (2009), The multiplex bead array approach to identifying serum biomarkers associated with breast cancer, Breast Cancer Research
14. Rai S. (2012) Structural Characterization of Potential Cancer Biomarker Proteins by
15. Kulasingam V. (2008) Identification And Validation Of Candidate Breast Cancer Biomarkers: A Mass Spectrometric Approach
16. Evangelia , Fourkala O. (2009) Risk factors and novel biomarkers in breast cancer

17. Rothschild S. (2014) microRNA therapies in cancer, , molecular and cellular therapies

18. Weige M.and Dowsett M. (2010), Current and emerging biomarkers in breast cancer: prognosis and prediction, Endocrine-Related Cancer

19. Misek D.and Kim E.(2011) Protein Biomarkers for the Early Detection of Breast Cancer, International Journal of Proteomics Volume 2011, Article ID 343582, 9 pages doi:10.1155/2011/343582

20. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridiza. tion prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.

21. Konstantin Arnold, Lorenza Bordoli, Jürgen Kopp, Torsten Schwede; The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics*, Volume 22, Issue 2, 15 January 2006, Pages 195–201, https://doi.org/10.1093/bioinformatics/bti770

22. Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, *34*(Web Server issue), W369–W373. http://doi.org/10.1093/nar/gkl198

23. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202–W208. http://doi.org/10.1093/nar/gkp335

24. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013Jan 41Database issueD991–5. [PMC free article] [PubMed][Reference list]