





Evaluation of the Expression Level of Some Coding and Non-coding Genes in Breast Cancer Samples Based on Bioinformatics and Laboratory Studies

Elmira Rostamnejad¹, Ronak Rashidi¹, Zohreh Akbari¹, Mazaher Maghsoudloo², Mehrdad Hashemi ^{3,*}, Amirnader Emami Razavi⁴, Maliheh Entezari ^{3,**}, Majid Sadeghizadeh⁵

¹ Department of Genetics, Faculty of Advanced Science and Technology, Tehran Medical Science, Islamic Azad University, Tehran, Iran

² Key Laboratory of Epigenetics and Oncology, the Research Center for Preclinical Medicine, Southwest Medical University, Luzhou, Sichuan, China

³ Farhikhtegan Medical Convergence Sciences Research Center, Farhikhtegan Hospital Tehran Medical Sciences, Islamic Azad University, Tehran, Iran

⁴ Iran National Tumor Bank, Cancer Institute, Tehran University of Medical Sciences, Tehran, Iran

⁵ Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

* Corresponding Author: Farhikhtegan Medical Convergence Sciences Research Center, Farhikhtegan Hospital Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.
Email: drmehashemi@gmail.com

** Corresponding Author: Farhikhtegan Medical Convergence Sciences Research Center, Farhikhtegan Hospital Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.
Email: entezarimal@yahoo.com

Received 2023 July 30; **Revised** 2023 November 20; **Accepted** 2024 February 20.

Abstract

Background: Breast cancer (BC) is the leading cause of cancer-associated mortality in women worldwide. However, the molecular mechanism underlying the process is still unclear. In this regard, bioinformatics studies play a decisive role in facilitating the path of biological investigations and can ultimately lead to the identification of better molecular candidates for further study.

Objectives: Due to the abnormal expression of many coding and non-coding genes in all types of cancers and their relationship with various mechanisms of carcinogenesis, this study aimed at evaluating the expression levels of certain coding and non-coding genes involved in BC based on bioinformatics findings and laboratory investigations.

Methods: Gene expression dataset, module extraction, functional enrichment analysis, protein-protein interaction network construction, and RT-qPCR were performed based on bioinformatics methods and laboratory investigations. Additionally, the promoter region mutations of these genes were investigated, using sequencing of extracted DNAs from formalin-fixed paraffin-embedded (FFPE) tumor tissues.

Results: A module was selected as a candidate for further investigation. Estrogen receptor 1 (ESR1) and forkhead box A1 (FOXA1) showed the highest degrees in the PPI network with 9 and 7 links, respectively. Furthermore, the expression levels of the FOXA1 gene, RNA component of mitochondrial RNA processing endoribonuclease (RMRP), and nuclear enriched abundant transcript 1 (NEAT1) were significantly upregulated in the tumor group compared to the control group (in order, $P = 0.044$, $P = 0.014$, and $P = 0.0004$). The tumors of patients with positive metastasis displayed significantly higher levels of NEAT1 and RMRP expression compared to those of negative metastasis samples ($P < 0.05$). Moreover, the expression level of RMRP dramatically decreased in HER2-positive patients compared to negative samples ($P = 0.011$). Finally, no mutations were observed in the promoter sequencing of positive metastasis samples compared to normal samples.

Conclusions: The upregulation levels of all three examined genes may correlate with BC progression. Therefore, they could potentially be used as biomarkers for detecting BC development.

Keywords: Breast Cancer, Biomarker, Bioinformatics Analysis

1. Background

One of the effective approaches in the diagnosis and management of patients with several different cancer types is the detection of tissue or blood biomarkers. Tumor biomarkers identify the biological differences between cancer and normal conditions (1-3). For this purpose, a wide range of genes has always been studied. Recently, extensive studies have been conducted on the potential of regulatory non-coding RNAs (ncRNAs) with a length of more than 200 nucleotides to predict cancer because non-coding regions of the genome play important regulatory roles in various cellular processes, such as transcription, post-transcriptional settings, gene silencing, and DNA demethylation. Indeed, small and large ncRNAs are involved in regulating gene expression through the interaction of their specific sequences with regulatory regions. These ncRNAs fall into different categories depending on the number of nucleotides in their body (4). On the other hand, the analysis of cancer mutations throughout the genome reveals a wide range of functional mutations within the genome, both in coding and non-coding regions that have extensive effects on the expression level of genes. These mutations can be prognostic and diagnostic markers for many cancers (5, 6). Additionally, breast cancer (BC) is the most common global malignancy in women nowadays. Every year, the GLOBOCAN statistics show that approximately 1.7 million women are diagnosed with this cancer worldwide, with a related mortality rate of nearly 1 063 in Iran (7). Therefore, bioinformatics methods can help researchers worldwide to initially evaluate genes, their characteristics, and functional mechanisms. It can be said that this method plays a decisive role in facilitating the path of biological studies and ultimately leads to the identification of better molecular candidates for further studies. In this respect, based on the co-expression network of genes using the WGCNA method, this study aimed to survey a group of coding genes, especially non-coding genes involved in BC. Eventually, for the validation of those results, further laboratory tests were conducted (8-13). The selected module, based on the bioinformatics analysis, contains 103 genes, 3 of which are long ncRNAs named metastasis associated lung adenocarcinoma transcript 1 (MALAT1), nuclear enriched

abundant transcript 1 (NEAT1), and X-inactive specific transcript (XIST), and 100 coding genes. One of the most critical coding genes in BC is estrogen receptor 1 (ESR1). Among those genes in the cyan module, NEAT1 as a long ncRNA, forkhead box A1 (FOXA1) as a coding gene, and RNA component of mitochondrial RNA processing endoribonuclease (RMRP) as a long ncRNA were selected for further laboratory tests due to their importance in BC based on previous studies.

2. Objectives

In other words, this study aimed at investigating the changes in the expression levels of those genes as well as their relationship with mutations in their promoter regions in a group of Iranian patients with BC so that the results can be used to confirm or reject previous hypotheses. The importance and role of the mentioned genes have been explained in detail in the discussion section.

3. Methods

3.1. Gene Expression Dataset and Pre-processing

The RNA-Seq read count data related to individuals with BC and normal samples were obtained from the UCSC Xena browser (<https://xenabrowser.net/>) (12). It consisted of 1217 samples from the Cancer Genome Atlas (TCGA). In the normalization step, genes with an average expression lower than one, including genes with zero expression, were removed. Then, the edgeR package was used on the data with the trimmed mean of the M-values (TMM) method to normalize the data (9). Furthermore, the normalized data were transformed to a Log₂ Scale after adding a value of one to avoid an undefined measure of log zero (10). Additionally, an outlier data detection analysis was performed.

3.2. Co-expression Network Construction and Module Extraction

The normalized gene expression matrix was filtered, using mean expression analysis, and genes with a threshold greater than 11 were selected. The weighted gene co-expression network analysis (WGCNA) R package was, then, used to reconstruct the co-expression network (11). Pairwise gene correlations were computed,

using the biweight midcorrelation (bicor) measure. To create a scale-free network, the optimal power of β ($R^2 = 0.9$) was selected. The co-expression network was reconstructed as a signed network. The min module parameter in WGCNA was set to 50, and the dynamic tree-cut method was implemented to cluster the modules. The remaining parameters in WGCNA were set to their default values. The first principal component of the expression matrix of each module was considered the module eigengene (ME), representing the gene expression profiles in the modules. The pairwise correlation of ME was calculated, and modules with correlations greater than 0.85 were merged. Finally, the modules were labeled with different colors, and the grey module contained genes that could not be placed in other modules.

3.3. Functional Enrichment Analysis

To understand the biological mechanism of genes in the extracted modules, all of the modules were enriched, using the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis, performed with the gprofiler2 R package (12). Additionally, the DOSE package (13) was utilized for disease enrichment analysis in all modules. Finally, a module was chosen as the candidate module in BC.

3.4. Protein-protein Interaction Network Construction

The genes of the candidate module in BC were used to construct the protein-protein interaction network, using the online tool STRING v11 (<https://string-db.org>). Only the experimental interactions were considered, and genes without interactions were excluded. In this study, the minimum confidence score in STRING was set to 0.7, with the remaining parameters kept at their default values. The resulting gene network was visualized, using Cytoscape (version 3.9.1).

3.5. Sample Collecting

In this study, 30 formalin-fixed paraffin-embedded (FFPE) tumor tissues from patients with BC and 30 FFPE samples from adjacent normal tissues were selected for further laboratory evaluations. Informed consent was obtained from all participating subjects, including both patients and controls, and they were interviewed to gather demographic and clinical data.

The patients included in the study underwent surgical resection for BC at Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran. Only patients who had not received chemotherapy or radiotherapy were included. Clinical information, such as colonoscopy/pathology reports and follow-up data, was collected from medical records. All patients were followed up to confirm their clinical outcomes. The samples were classified based on average age, family history, tumor differentiation, progesterone/estrogen receptor (PR/ER), human epidermal growth factor receptor 2 (Her2) status, and metastasis.

The main focus of this study was to investigate the expression levels of FOXA1 and long non-coding RNAs (lncRNAs) NEAT1/RMRP in the tumor and control groups. After selecting the patients based on inclusion criteria, which included women with BC who had not undergone preoperative chemotherapy or radiotherapy, and confirming tumor histology through pathology reports, and excluding patients with prior cancers such as lymphoma or leukemia or those who received neoadjuvant chemotherapy or radiotherapy, FFPE tissue blocks were cut into 10 μm to 15 μm thick sections for mRNA extraction.

3.6. Extraction of DNA and RNA

DNA was extracted from paraffin-embedded tissues, using the Qiagen kit to determine the sequence of the gene promoter region. DNA and RNA extraction from FFPE tissues was performed, using the trizol technique, which is based on the phenol-chloroform method. To assess the quality of the extracted RNA and DNA, a 1.5% agarose gel was used. In eukaryotic samples, the observation of bands for 18s and 28s ribosomal RNA indicated the correct quality of RNA extraction. For quantification, concentration, and optical density (OD) measurement of the samples, a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Rockland, DE, USA) was used. The A260/A280 ratio was employed to assess the purity of the DNA and RNA, and values within the range of 1.8-2.0 were considered acceptable.

3.7. Reverse Transcription

The synthesis of cDNAs was carried out, using the PrimeScript RT Reagent kit (Takara, Shiga, Japan),

following the manufacturer's protocol. First, the RNA was incubated at 65 °C for 5 minutes, followed by rapid freezing for 3 minutes. For every 500 ng of RNA, 2 µL of buffer, 0.5 µL of random hexamers primer, 0.5 µL of RT enzyme, and 10 mL of RNase-free water were used.

The reaction was performed, using the Perkin-Elmer Gene Amp PCR System 2400. The reverse transcription reaction was conducted at 37 °C for 15 minutes, followed by a brief incubation at 85 °C for 5 seconds. The resulting cDNA was stored at -70 °C for future use.

3.8. Real-time PCR Analysis

Quantitative real-time reverse transcription-PCR was performed, using an ABI Step One RT-PCR thermal cycler (ABI Stepone, NY, USA) and a SYBR premix Ex Taq II kit (TaKaRa Biotechnology). The primers for FOXA1, lncRNA RMRP, lncRNA NEAT1, and the housekeeping gene GAPDH were designed, using primer Express software version 3.0.1 (Applied Biosystems, CA) and the IDT primer Quest Tool (Table 1). The PCR reaction was carried out in a final volume of 20 µL, which included 0.8 µL of each primer, 2 µL of cDNA template, 10 µL of SYBR premix Ex Taq II (Tli RNaseH Plus) (2×), 0.4 µL of ROX Reference Dye (50×), and 6 µL of sterile distilled water. The thermal profile for the PCR consisted of an initial denaturation step at 95 °C for 30 s, followed by 40 cycles of denaturation at 95 °C for 5 s and annealing/extension at 60 °C for 34 s. The relative expression levels of the target genes were determined by normalizing to GAPDH, using the $2^{-\Delta\Delta CT}$ method. Each measurement was performed in triplicate to ensure the accuracy and reliability of the results. Also, it should be noted that in this project, the primers were specifically designed as exon junction primers to target the intended regions. In addition, DNase was also used in the extraction phase to prevent DNA contamination.

3.9. DNA Sequencing

The promoter region sequencing method was employed to explore the relationship between mutations in the promoter region of these genes and changes in expression levels in the studied samples. DNA fragments of FOXA1, NEAT1, and RMRP measuring 245 bp, 96 bp, and 235 bp, respectively, were amplified, using the specific primers listed in Table 1. Two samples from the positive metastasis group, exhibiting the

highest expression changes, were selected for this analysis. To visualize the digested fragments, the PCR products were separated on a 2% agarose gel and stained with ethidium bromide. Subsequently, direct sequencing was performed, using fluorescent dideoxy sequencing on an ABI sequencing 3130 XL instrument, following the manufacturer's instructions. The single-strand PCR products were subjected to direct sequencing, using the BigDye™ Terminator v3.1 cycle sequencing kit and the ABI 3130XL genetic analyzer (Applied Biosystems). Finally, the resulting electropherograms were processed, using Lasergene software version 6 (DNASTAR) for analysis of the sequencing data.

3.10. Statistical Analysis

The mRNA expression levels of tumor tissues were reported as the mean ± standard deviation (SD). The differences in mRNA and lncRNA expression levels of FOXA1, RMRP, and NEAT1 genes among the established groups, including tumor differentiation grade, metastasis, and Her2 status were assessed, using One-way ANOVA and Mann-Whitney U analysis. All statistical analyses were conducted, using IBM SPSS Statistics software version 23 (IBM, SPSS, Chicago, IL, USA). Statistical significance was considered achieved when the P-value was less than 0.05 ($P < 0.05$).

4. Results

4.1. Gene Expression Dataset and Pre-processing

The read count data for BC samples was obtained from the Xena browser. This dataset consists of 1 104 primary solid tumor samples and 113 solid tissue normal samples. The demographic characteristics of the samples can be found in Table 2.

The downloaded dataset was normalized, using TMM. Then, the ensemble ids were converted into gene symbols. Finally, 28 638 genes were identified for further analysis. Also, a quality control analysis was performed, and no outlier sample was detected.

4.2. Co-expression Network Construction and Module Extraction

The genes with a mean expression lower than 11 were filtered out, resulting in a selection of 5 366 genes as the

Table 1. The Primer Sequences Used for RT-qPCR

Forkhead box A1	
Forward	5'-ACTGTGAAGATGGAAGGGCA-3'
Reverse	5'-CCGCTCGTAGTCATGGTGT-3'
Promoter region (forward)	5'-GAAGAGGAAGCCCAGAGC-3'
Promoter region (reverse)	5'-GTAGGTGCGAGCGTCTT-3'
Forward	5'-TGG AGG AGT CAG GAG GAA TAG -3'
Nuclear enriched abundant transcript 1	
Reverse	5'-GGC ATG GAC AAG TTG AAG ATT AG -3'
Promoter region (forward)	ATC GTC CCG TTG AGC AAT GA-3'-5'
Promoter region (reverse)	CAT CCC TCC CTG TCG CTA AC-3'-5'
Forward	CTACACACTGAGGACTCTGTTC-3'-5'
RNA component of mitochondrial RNA processing endoribonuclease	
Reverse	CAGCGGATACGCTTCT-3'-5'
Promoter region (forward)	ACTCTGCCCCGAGGTC-3'-5'
Promoter region (reverse)	-TAGTCTTAGAAAGTTATGCCCGAAA-3'-5'
Forward	5'-ACAACCTTGGTATCGTGAAGG-3'
Reverse	5'-GCCATCACGCCACAGTTTC-3'

Table 2. The Demographic Characteristics of Breast Cancer Samples

Sample Type	No. of Samples	Sex (Male/Female)	Age	Age Range	Vital Status (Alive/Dead)	Stage	Race
Primary tumor	1104	13/1091 (1.18% - 98/02%)	58.36	20 - 90	947/157 (85.78% - 14.22%)	I. 182 (16.49%); II. 628 (56.89%); III. 250 (22.64%); IV. 20 (1.81%); NA. 24 (2.17%)	Asian 61 (5.53%); Black 183 (16.58%); White 764 (69.19%); NA 96 (8.70%)
Solid tissue normal	113	1/112 (0.88% - 99.12%)	57.33	20 - 90	69/44 (61.06% - 38.94%)	-	Asian 1 (0.88%); Black 6 (5.31%); White 105 (92.93%); NA 1 (0.88%)

input data for the WGCNA. To ensure a scale-free network, the power β was set to 17. Subsequently, 15 modules were extracted, each containing a varying number of genes ranging from 57 to 1 870. The hierarchical clustering of the identified modules is depicted in [Figure 1](#).

4.3. Functional Enrichment Analysis

The identification of 15 modules in BC led to the selection of a specific module, the cyan module, as a potential candidate. To further evaluate the functional significance of the genes within this module, enrichment analysis was performed, using gprofiler2. This analysis aimed at assessing the enrichment of gene sets associated with various biological processes, pathways, and BC-related diseases within the genes of the cyan module.

The results of the gene enrichment analysis for the cyan module are presented in [Figure 2](#). The cyan module

consists of 103 genes, 100 of which are protein-coding genes, while 3 genes, namely MALAT1, NEAT1, and XIST, are ncRNAs. The cyan module exhibited significant enrichment in various biological processes and pathways relevant to BC.

For a comprehensive list of genes within the cyan module, please refer to appendix 1.

4.4. Protein-protein Interaction Network Construction and Identification of Modules

One hundred and three genes of the cyan module were inserted into the STRING database to extract the PPI network. The constructed network includes 20 and 25 nodes and edges, respectively. The average node degree in the network is 0.451, and the PPI enrichment P-value is 2.69e-07. The PPI network using Cytoscape was visualized, and the network is represented in [Figure 3](#). Estrogen receptor 1 and FOXA1 are the most degrees in the PPI network with 9 and 7 links, respectively. So, among these genes, FOXA1 and two ncRNAs, one of

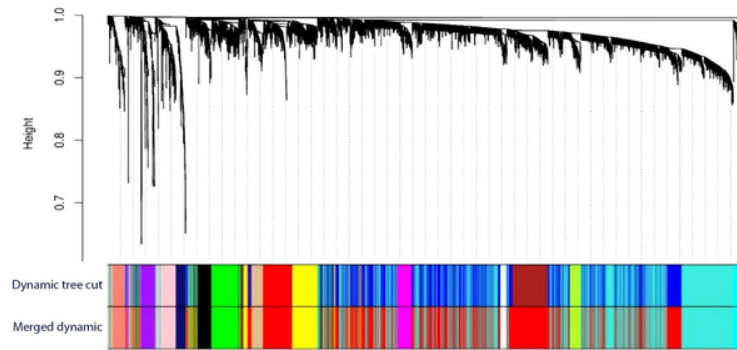


Figure 1. Cluster Dendrogram of extracted modules in breast cancer (BC). The row of dynamic tree cut shows the color of modules before the merged module analysis, and the merged dynamic row represents the color of modules after the merged module analysis.

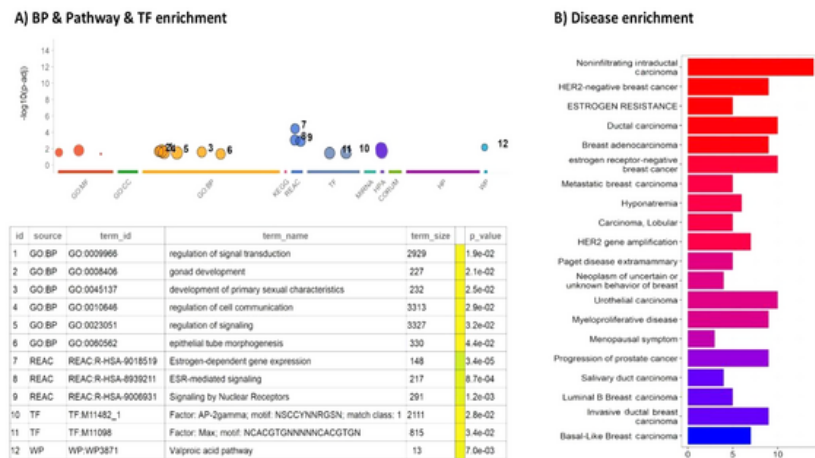


Figure 2. Enrichment analysis of genes in the cyan module. A, most prevalent enriched biological processes (BP), pathways (including KEGG and Reactome), transcription factors (TF), and Wikipedia pathways (WP) associated with the genes in the cyan module; B, list of disease enrichment in the cyan module, with the color of bars indicating the degree of significance of enrichment items.

which is in the cyan module, were validated and investigated in vitro step.

4.5. Clinical and Pathological Characteristics of Patients

Demographic and clinical characteristics of BC tissue samples is shown in Table 3.

4.6. Gene Expression Analysis

Based on the results obtained by the real-time PCR method, the expression levels of FOXA1 gene, RMRP, and

NEAT1 lncRNAs were significantly upregulated in the tumor group compared to the control group (respectively, $P = 0.044$, $P = 0.014$, and $P = 0.0004$) (Figures 1, 2, and 3).

Details about the mean values of relative expressions of FOXA1, RMRP, and NEAT1 and their relationship with pathological factors including tumor differentiation and metastasis have been demonstrated in Figures 4 to 6.

The expression levels of the FOXA1 gene showed a significant increase in patients with grades I, II, and III

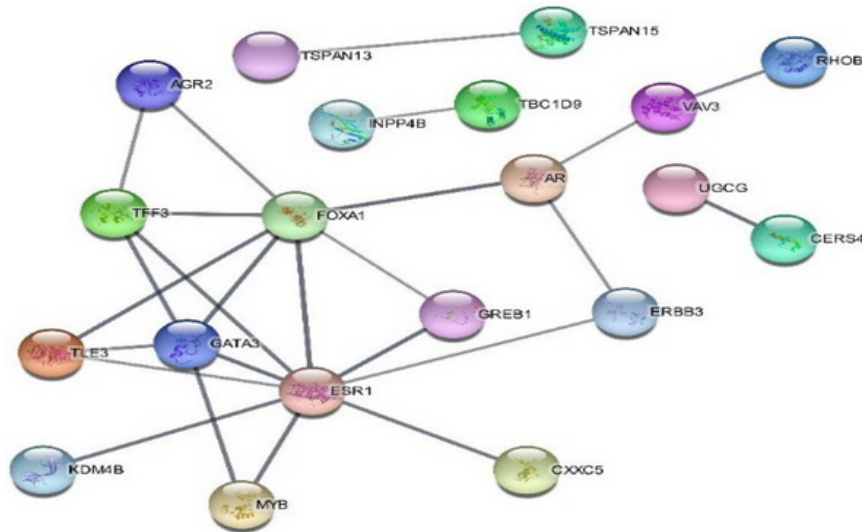


Figure 3. Network of genes extracted using STRING: Circles represent genes and edges indicate the relationships among them.

Table 3. Demographic and Clinical Characteristics of Breast Cancer Tissue Samples

Sample Type	No. of Samples	Sex (Male/Female)	Range Age	Average Age	Vital Status (Alive/Dead)	PR Status (+/-)	ER Status (+/-)	Human Epidermal Growth Factor Receptor 2 Status (+/-)	Metastasis Status (+/-)	Family History (+/-)	Tumor Differentiation Status (1,2,3,4)
Primary tumor/ solid tissue normal	30/30	0/30 (0%-100%)	25 - 78	55.3	27/3(90%-10%)	9/21(30% -70%)	12/18(40% -60%)	8/22(26.6% -73.3%)	8/22(26.6% -73.3%)	17/13(56.6% -43.3%)	7(23.3%) - 11(36.6%) - 10(33.3%) - 2(6.6%)

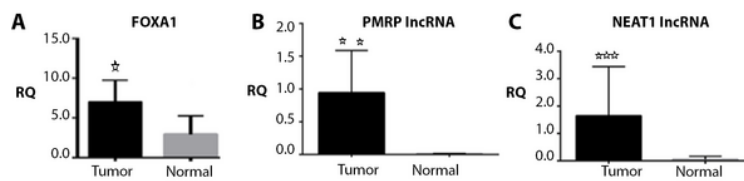


Figure 4. Relative quantification (RQ) established by RT-qPCR analysis. Gene expression levels of forkhead box A1 (FOXA1) (A); RNA component of mitochondrial RNA processing endoribonuclease (RMRP) (B); and nuclear enriched abundant transcript 1 (NEAT1) (C); genes differed significantly between patients with tumor samples and normal.

compared to normal samples, as well as in patients with positive metastases compared to those with negative metastases ($P < 0.05$, Figure 5). According to Figure 5, the gene expression levels of RMRP varied significantly between patients with and without metastasis. Tumors from patients with positive metastasis displayed higher

levels of RMRP expression compared to those from negative metastasis samples ($P < 0.05$, Figure 5). Additionally, the NEAT1 gene was significantly upregulated in patients with metastasis compared to negative samples. Tumors from patients with positive metastasis exhibited approximately a 6-fold higher level

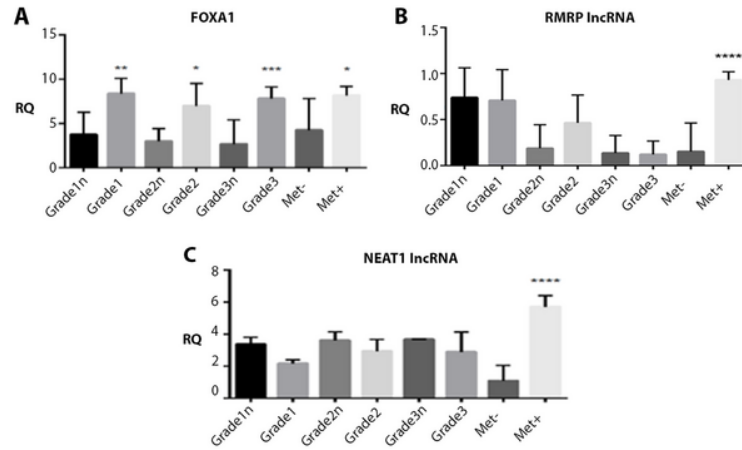


Figure 5. Forkhead box A1 (FOXA1) mRNA, RNA component of mitochondrial RNA processing endoribonuclease (RMRP), nuclear enriched abundant transcript 1 (NEAT1), relative quantification (RQ) established by RT-qPCR analysis according to tumor differentiation (histological grade) and metastasis status between tumor and normal samples.

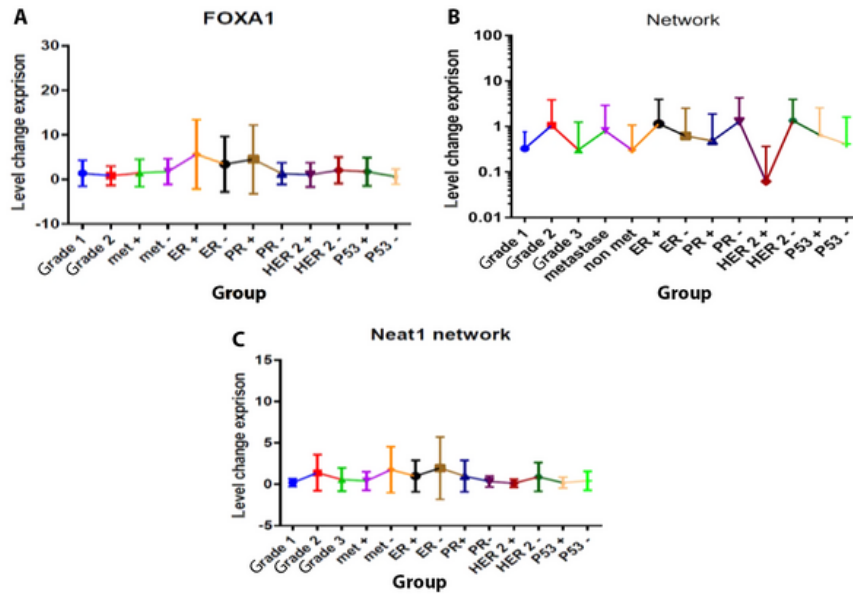


Figure 6. Forkhead box A1 (FOXA1) mRNA (A); RNA component of mitochondrial RNA processing endoribonuclease (RMRP) (B); nuclear enriched abundant transcript 1 (NEAT1) (C); relative quantification (RQ) established by RT-qPCR analysis compared to the expression level of other tumor markers in the total statistical population.

of NEAT1 expression compared to those from negative metastasis samples ($P < 0.05$, Figure 5). No other statistically significant correlations were found between tumor differentiation grade and the expression levels of RMRP and NEAT1.

4.7. Sequencing Analysis for Mutation Status of the Patients

To investigate the mutation status in the promoter region of FOXA1, RMRP, and NEAT1, and its relationship

with the expression levels of these genes in patients, direct sequencing was performed on genomic DNA extracted from FFPE samples in all cases, using the Deoxyfluorescent method. Our analysis revealed that none of the patients had mutations in the promoter region of these genes. As a result, no correlation was found between the expression levels of these genes and the presence of mutations in their promoter regions. However, to draw accurate conclusions, further studies with a larger sample size are required.

5. Discussion

As mentioned, this research focused on the *in vitro* evaluation and validation of three genes: Forkhead box A1 as a coding gene, NEAT1, and RMRP as long ncRNAs known to be involved in BC. These genes were selected based on previous research and their potential as diagnostic biomarkers, particularly for their relevance to BC. Additionally, it is worth noting that there are other important genes within the cyan module that are also associated with BC.

Approximately 75% of all BCs are hormone receptor-positive (HR+), specifically estrogen receptor-positive (ER+) (14). The ESR1 gene, located on chromosome 6, is a coding gene that encodes the estrogen receptor (ER) and functions as a transcription factor (15). Consequently, the rate of estrogen exposure can serve as a risk factor for ER+ BC. Higher levels of estrogen expression can lead to an increased number of ER receptors, which ultimately contributes to the excessive proliferation of breast epithelial cells. Therefore, based on the bioinformatics analysis conducted in this study, ESR1 is identified as a crucial gene for BC diagnosis (16).

Forkhead box A1 is a molecular candidate that can serve as a biomarker for assessing the degree of tumorigenesis in patients with BC. It is located on the long arm of chromosome 14 and is associated with the expression of ER α +, PR+, and GATA3 proteins, as well as endocrine signaling in BC (17, 18). This gene is recognized as an oncogenic factor in hormone ER cancers, and mutations in its promoter region result in increased expression by enhancing the binding of E2F, a transcription factor (19). Forkhead box A1 plays a role in enhancing luminal differentiation, gene methylation, and the overexpression of the ER gene in BC. Similar to our investigation, several studies have reported elevated levels of FOXA1 gene and protein expression in BC

compared to normal mucosa, suggesting its potential as a diagnostic biomarker (20-25). Previous studies have also demonstrated a correlation between upregulated FOXA1 gene expression and metastatic status as well as increased dysplasia grade in BC (26, 27), which aligns with the findings of our study. Based on these results, FOXA1 gene expression contributes to BC progression through intricate mechanisms and holds promise as a potential diagnostic biomarker for the disease.

XIST is the first identified lncRNA known for its primary role in chromosome X inactivation. Over time, researchers have conducted numerous studies on this lncRNA, uncovering its crucial roles in various types of human cancers. Based on these studies, XIST has been found to have oncogenic roles in many cancers, including lung, gastric, glioma, pancreatic, and others. However, interestingly, in BC, certain osteosarcomas, and a specific type of hepatic cancer, XIST acts as a tumor suppressor (28). In a series of studies, the tumor suppressor function of XIST has been elucidated, primarily attributed to its role as a sponge for miR-155, which targets CDX1, a gene associated with angiogenesis. Through this mechanism, XIST inhibits cell growth (29). Given these findings, XIST is recognized as another member of the Cyan module and represents a promising candidate for further validation through laboratory studies.

RNA component of mitochondrial RNA processing endoribonuclease is a lncRNA that encodes a component of mitochondrial RNA-processing endoribonuclease RNA. It can be isolated from mitochondrial RNA at a primary site of mitochondrial DNA replication and is also activated in the nucleolus during the final stage of 5.8S rRNA processing. The mutation of RMRP was first discovered in cartilage-hair hypoplasia (CHH), which led to the down-regulation of this lncRNA in CHH. RNA component of mitochondrial RNA processing endoribonuclease also interacts with the telomerase reverse transcriptase catalytic subunit to form a ribonucleoprotein complex that produces double-stranded RNAs and acts as a small intervention RNA (30, 31).

The lncRNA RMRP is a tumor promoter, and its upregulation has been observed in many cancers, such as breast, gastric, and bladder cancers (32-34). Previous studies have shown that the upregulation of the RMRP gene can increase cell growth and invasion by targeting

miR-1-3p in non-small-cell lung cancer and induce the overexpression of c-MYC by sponging miR-34a-5p to promote cell growth and inhibit apoptosis in multiple myeloma (35, 36). Qi et al. indicated that the RMRP gene increases the proliferation, migration, and invasion of triple-negative BC via the miR-766-5p/YAP1 axis and suggested the RMRP gene as a novel target for BC treatment (37). In line with previous studies, our present study showed the overexpression of the RMRP gene in BC (32). Additionally, a significant difference in the expression of the RMRP gene between HER2+ and HER2- tumors was found. These findings suggest that the overexpression of the RMRP gene is an important factor in breast carcinogenesis.

Nuclear enriched abundant transcript 1 is a lncRNAs located in the nucleus and associated with Paraspeckles and nuclear domains. It plays a role in the maintenance of nuclear mRNA. This gene is a transcript of the multiple endocrine neoplasia locus, located on chromosome 11, and is expressed structurally in many non-neural tissues and cells (33). Nuclear enriched abundant transcript 1 is a key regulator of gene expression through the conservation of nuclear mRNA and transcription factors. It also serves as a regulatory biomarker in breast morphogenesis (34). The pivotal role of the BRCA1/lncRNA NEAT1 signaling pathway has also been identified in tumorigenesis and carcinogenesis (35). Long non-coding RNAs NEAT1 can enhance cell proliferation and growth in several cancers, including colorectal, pancreatic, and BC (36-39). It has been reported that lncRNA NEAT1 can promote epithelial-mesenchymal transition (EMT) in BC by regulating cell proliferation, migration, and invasion (39). Yan et al. found that the upregulation of the NEAT1 gene was correlated with advanced-stage and lymph node metastasis. Additionally, they considered the NEAT1 gene as a biomarker for the detection of BC (34, 40, 41). In the present investigation, a significant overexpression of the NEAT1 gene in breast tumors compared to normal tissues and in metastatic tumors was observed. In conclusion, the findings from our study suggest that the NEAT1 gene may play a crucial role in BC progression. Additionally, our results indicate that the NEAT1 gene has the potential to serve as a diagnostic biomarker for BC. Further investigations are necessary to elucidate the underlying mechanisms, by

which NEAT1 influences BC development and to evaluate its potential clinical utility as a diagnostic marker.

Metastasis associated lung adenocarcinoma transcript 1, also known as nuclear-enriched abundant transcript 2 (NEAT2), is a highly abundant lncRNAs found on human chromosome 11, primarily located in the nucleus. It has been observed that MALAT1 interacts with the spliceosome protein complex, influencing splicing processes (42-44). However, previous investigations have produced divergent results regarding the role of MALAT1 in cancer. Some studies have shown that MALAT1 is upregulated in various cancers, promoting tumor metastasis (45). On the other hand, another group of studies has indicated that MALAT1 acts as a tumor suppressor in BC, inhibiting cell migration and metastasis. Considering these contradictory findings, further research is needed to evaluate the potential of MALAT1 as a diagnostic biomarker for BC (46, 47).

Breast cancer at advanced stages, particularly stage IV, is often accompanied by metastasis, where cancer spreads to distant organs. In contrast, at earlier stages (I and II), the cancer primarily affects the breast tissue and neighboring areas. Accurate staging of the disease is crucial for effective disease management. Identifying genes that act as prognostic markers for each stage can greatly assist in patient care (48). In this study, it was observed that the expression levels of all three genes were elevated in the metastatic group compared to patients at other stages, suggesting their potential as promising metastasis markers. However, due to the limited sample size in this experiment, further investigations are needed to determine their prognostic value. Conducting larger-scale studies is essential to thoroughly investigate their potential as prognostic markers for BC.

5.1. Conclusions

Previous studies have established the significance of certain lncRNAs in BC progression. In the present investigation, the overexpression of the FOXA1 gene and lncRNAs RMRP and NEAT1 was observed in BC tumors with high-grade dysplasia and metastasis, as demonstrated in laboratory experiments. These findings suggest that the FOXA1, RMRP, and NEAT1 genes may serve as potential biomarkers for detecting BC progression. However, further research is required to

validate this concept and determine the clinical utility of these genes as biomarkers for BC detection and monitoring.

Supplementary Material

Supplementary material(s) is available [here](#) [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

Authors' Contribution: Study concept and design: E. Rostamnejad, R. Rashidi, Z. Akbari, M. Sadeghizadeh; acquisition of data: E. Rostamnejad, R. Rashidi, Z. Akbari, A. Emami Razavi; analysis and interpretation of data: M. Maghsoudloo, M. Entezari; drafting of the manuscript: E. Rostamnejad, Z. Akbari, R. Rashidi; critical revision of the manuscript for important intellectual content: M. Hashemi, M. Sadeghizadeh; statistical analysis: M. Maghsoudloo; administrative, technical, and material support: A. Emami Razavi; study supervision: M. Entezari, M. Hashemi.

Conflict of Interests Statement: The authors declare that there are no conflicts of interest regarding the content of this article.

Data Availability: It has not been declared by the authors.

Ethical Approval: This study was conducted with the approval of the Ethics Committee ([IR.IAU.PS.REC.1397.137](#)) at the cancer institute, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran.

Funding/Support: The authors have provided the project's required budget.

Informed Consent: Informed consent was obtained from all participating subjects, including both patients and controls, and they were interviewed to gather demographic and clinical data.

References

- Nagpal M, Singh S, Singh P, Chauhan P, Zaidi MA. Tumor markers: A diagnostic tool. *Nat J Maxillofac Surg*. 2016;7(1):17-20. [PubMed ID: 28163473]. [PubMed Central ID: PMC5242068]. <https://doi.org/10.4103/0975-5950.196135>.
- Peyravian N, Nobili S, Pezeshkian Z, Olfatifar M, Moradi A, Baghaei K, et al. Increased Expression of VANGL1 is Predictive of Lymph Node Metastasis in Colorectal Cancer: Results from a 20-Gene Expression Signature. *J Pers Med*. 2021;11(2). [PubMed ID: 33672900]. [PubMed Central ID: PMC7918343]. <https://doi.org/10.3390/jpm11020126>.
- Peyravian N, Larki P, Gharib E, Nazemalhosseini-Mojarad E, Anaraki F, Young C, et al. The Application of Gene Expression Profiling in Predictions of Occult Lymph Node Metastasis in Colorectal Cancer Patients. *Biomedicines*. 2018;6(1). [PubMed ID: 29498671]. [PubMed Central ID: PMC5874684]. <https://doi.org/10.3390/biomedicines6010027>.
- Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol*. 2020;3(1):56. [PubMed ID: 32024996]. [PubMed Central ID: PMC7002399]. <https://doi.org/10.1038/s42003-019-0741-7>.
- Jiang MC, Ni JJ, Cui WY, Wang BY, Zhuo W. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J Cancer Res*. 2019;9(7):1354-66. [PubMed ID: 31392074]. [PubMed Central ID: PMC6682721].
- Pezeshkian Z, Nobili S, Peyravian N, Shojaee B, Nazari H, Soleimani H, et al. Insights into the Role of Matrix Metalloproteinases in Precancerous Conditions and in Colorectal Cancer. *Cancers (Basel)*. 2021;13(24). [PubMed ID: 34944846]. [PubMed Central ID: PMC8699154]. <https://doi.org/10.3390/cancers13246226>.
- Farhood B, Geraily G, Alizadeh A. Incidence and Mortality of Various Cancers in Iran and Compare to Other Countries: A Review Article. *Iran J Public Health*. 2018;47(3):309-16. [PubMed ID: 29845017]. [PubMed Central ID: PMC5971166].
- Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-8. [PubMed ID: 32444850]. [PubMed Central ID: PMC7386072]. <https://doi.org/10.1038/s41587-020-0546-8>.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11(3):R25. [PubMed ID: 20196867]. [PubMed Central ID: PMC2864565]. <https://doi.org/10.1186/gb-2010-11-3-r25>.
- Haghjoo N, Moeini A, Masoudi-Nejad A. Introducing a panel for early detection of lung adenocarcinoma by using data integration of genomics, epigenomics, transcriptomics and proteomics. *Exp Mol Pathol*. 2020;112:104360. [PubMed ID: 31843580]. <https://doi.org/10.1016/j.yexmp.2019.10.4360>.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. [PubMed ID: 19114008]. [PubMed Central ID: PMC2631488]. <https://doi.org/10.1186/1471-2105-9-559>.
- Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res*. 2020;9. [PubMed ID: 33564394]. [PubMed Central ID: PMC7859841]. <https://doi.org/10.12688/f1000research.24956.2>.
- Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31(4):608-9. [PubMed ID: 25677125]. <https://doi.org/10.1093/bioinformatics/btu684>.

14. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;**61**(2):69-90. [PubMed ID: 21296855]. <https://doi.org/10.3322/caac.20107>.
15. Amend K, Hicks D, Ambrosone CB. Breast cancer in African-American women: differences in tumor biology from European-American women. *Cancer Res*. 2006;**66**(17):8327-30. [PubMed ID: 16951137]. <https://doi.org/10.1158/0008-5472.CAN-06-1927>.
16. DeVita VT, Lawrence TS, Rosenberg SA. *Cancer: principles and practice of oncology- advances in oncology*. Philadelphia, PA: Lippincott Williams and Wilkins; 2010.
17. Seachrist DD, Anstine LJ, Keri RA. FOXA1: A Pioneer of Nuclear Receptor Action in Breast Cancer. *Cancers (Basel)*. 2021;**13**(20). [PubMed ID: 34680352]. [PubMed Central ID: PMC8537079]. <https://doi.org/10.3390/cancers13205205>.
18. Tanaka K, Tokunaga E, Yamashita N, Sagara Y, Ohi Y, Taguchi K, et al. The relationship between the expression of FOXA1 and GATA3 and the efficacy of neoadjuvant endocrine therapy. *Breast Cancer*. 2017;**24**(3):384-92. [PubMed ID: 27473079]. <https://doi.org/10.1007/s12282-016-0714-3>.
19. Arruabarrena-Aristorena A, Maag JLV, Kittane S, Cai Y, Karthaus WR, Ladewig E, et al. FOXA1 Mutations Reveal Distinct Chromatin Profiles and Influence Therapeutic Response in Breast Cancer. *Cancer Cell*. 2020;**38**(4):534-550 e9. [PubMed ID: 32888433]. [PubMed Central ID: PMC8311901]. <https://doi.org/10.1016/j.ccell.2020.08.003>.
20. Al-Bedairy I, Shamsa M, Salim SA, Mahdi M, Dawood K, Al Faisal AH. FOXA1 expression in Iraqi women with ER+ breast cancer. *Baghdad J Biochem Appl Biological Sci*. 2021;**2**(2):106-19. <https://doi.org/10.47419/bjbabs.v2i02.43>.
21. He K, Zeng H, Xu X, Li A, Cai Q, Long X. Clinicopathological significance of forkhead box protein A1 in breast cancer: A meta-analysis. *Exp Ther Med*. 2016;**11**(6):2525-30. [PubMed ID: 27284343]. [PubMed Central ID: PMC4887757]. <https://doi.org/10.3892/etm.2016.3229>.
22. Hisamatsu Y, Tokunaga E, Yamashita N, Akiyoshi S, Okada S, Nakashima Y, et al. Impact of GATA-3 and FOXA1 expression in patients with hormone receptor-positive/HER2-negative breast cancer. *Breast Cancer*. 2015;**22**(5):520-8. [PubMed ID: 24415069]. <https://doi.org/10.1007/s12282-013-0515-x>.
23. Wang X, Feng Z, Li J, Han Y, Su L, Wang F, et al. Functional Variant rs4442975 Modulating FOXA1 Binding Affinity Can Influence Bone Marrow Suppression during Neoadjuvant Chemotherapy for Luminal A Type Breast Cancer. *Biomed Res Int*. 2019;**2019**:7073498. [PubMed ID: 30881995]. [PubMed Central ID: PMC6381587]. <https://doi.org/10.1155/2019/7073498>.
24. Dai X, Cheng H, Chen X, Li T, Zhang J, Jin G, et al. FOXA1 is Prognostic of Triple Negative Breast Cancers by Transcriptionally Suppressing SOD2 and IL6. *Int J Biol Sci*. 2019;**15**(5):1030-41. [PubMed ID: 31182923]. [PubMed Central ID: PMC6535797]. <https://doi.org/10.7150/ijbs.31009>.
25. Cheng TD, Yao S, Omilian AR, Khoury T, Buas MF, Payne-Ondracek R, et al. FOXA1 Protein Expression in ER(+) and ER(-) Breast Cancer in Relation to Parity and Breastfeeding in Black and White Women. *Cancer Epidemiol Biomarkers Prev*. 2020;**29**(2):379-85. [PubMed ID: 31871111]. [PubMed Central ID: PMC7007352]. <https://doi.org/10.1158/1055-9965.EPI-19-0787>.
26. Fu X, Pereira R, De Angelis C, Veeraraghavan J, Nanda S, Qin L, et al. FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proc Natl Acad Sci U S A*. 2019;**116**(52):26823-34. [PubMed ID: 31826955]. [PubMed Central ID: PMC6936436]. <https://doi.org/10.1073/pnas.1911584116>.
27. Tachi K, Shiraiishi A, Bando H, Yamashita T, Tsuboi I, Kato T, et al. FOXA1 expression affects the proliferation activity of luminal breast cancer stem cell populations. *Cancer Sci*. 2016;**107**(3):281-9. [PubMed ID: 26708273]. [PubMed Central ID: PMC4814260]. <https://doi.org/10.1111/cas.12870>.
28. Madhi H, Kim MH. Beyond X-Chromosome Inactivation: The Oncogenic Facet of XIST in Human Cancers. *Biomed Sci Lett*. 2019;**25**(2):113-22. <https://doi.org/10.15616/bsl.2019.25.2.113>.
29. Zheng R, Lin S, Guan L, Yuan H, Liu K, Liu C, et al. Long non-coding RNA XIST inhibited breast cancer cell growth, migration, and invasion via miR-155/CDX1 axis. *Biochem Biophys Res Commun*. 2018;**498**(4):1002-8. [PubMed ID: 29550489]. <https://doi.org/10.1016/j.bbrc.2018.03.104>.
30. Yang M, Ke H, Zhou W. LncRNA RMRP Promotes Cell Proliferation and Invasion Through miR-613/NFAT5 Axis in Non-Small Cell Lung Cancer. *Onco Targets Ther*. 2020;**13**:8941-50. [PubMed ID: 32982286]. [PubMed Central ID: PMC7494237]. <https://doi.org/10.2147/OTT.S255126>.
31. Chen Y, Hao Q, Wang S, Cao M, Huang Y, Weng X, et al. Inactivation of the tumor suppressor p53 by long noncoding RNA RMRP. *Proc Natl Acad Sci U S A*. 2021;**118**(29). [PubMed ID: 34266953]. [PubMed Central ID: PMC8307289]. <https://doi.org/10.1073/pnas.2026831118>.
32. Hussen BM, Azimi T, Hidayat HJ, Taheri M, Ghafouri-Fard S. Long Non-coding RNA RMRP in the Pathogenesis of Human Disorders. *Front Cell Dev Biol*. 2021;**9**:676588. [PubMed ID: 33996836]. [PubMed Central ID: PMC8120005]. <https://doi.org/10.3389/fcell.2021.676588>.
33. Shao Y, Ye M, Li Q, Sun W, Ye G, Zhang X, et al. LncRNA-RMRP promotes carcinogenesis by acting as a miR-206 sponge and is used as a novel biomarker for gastric cancer. *Oncotarget*. 2016;**7**(25):37812-24. [PubMed ID: 27192121]. [PubMed Central ID: PMC5122351]. <https://doi.org/10.18632/oncotarget.9336>.
34. Cao HL, Liu ZJ, Huang PL, Yue YL, Xi JN. LncRNA-RMRP promotes proliferation, migration and invasion of bladder cancer via miR-206. *Eur Rev Med Pharmacol Sci*. 2019;**23**(3):1012-21. [PubMed ID: 30779067]. https://doi.org/10.26355/eurrev_201902_16988.
35. Wang Y, Luo X, Liu Y, Han G, Sun D. Long noncoding RNA RMRP promotes proliferation and invasion via targeting miR-1-3p in non-small-cell lung cancer. *J Cell Biochem*. 2019;**120**(9):15170-81. [PubMed ID: 31050363]. <https://doi.org/10.1002/jcb.28779>.
36. Xiao X, Gu Y, Wang G, Chen S. c-Myc, RMRP, and miR-34a-5p form a positive-feedback loop to regulate cell proliferation and apoptosis in multiple myeloma. *Int J Biol Macromol*. 2019;**122**:526-37. [PubMed ID: 30389523]. <https://doi.org/10.1016/j.ijbiomac.2018.10.207>.
37. Qi L, Sun B, Yang B, Lu S. Long Noncoding-RNA Component of Mitochondrial RNA Processing Endoribonuclease Promotes Carcinogenesis in Triple-Negative Breast Cancer Cells via the Competing Endogenous RNA Mechanism. *J Breast Cancer*. 2021;**24**(5):428-42. [PubMed ID: 34652079]. [PubMed Central ID: PMC8561136]. <https://doi.org/10.4048/jbc.2021.24.e42>.
38. Feng Y, Gao L, Cui G, Cao Y. LncRNA NEAT1 facilitates pancreatic cancer growth and metastasis through stabilizing ELF3 mRNA. *Am J Cancer Res*. 2020;**10**(1):237-48. [PubMed ID: 32064164]. [PubMed Central ID: PMC7017733].
39. Li S, Hao J, Hong Y, Mai J, Huang W. Long Non-Coding RNA NEAT1 Promotes the Proliferation, Migration, and Metastasis of Human Breast-Cancer Cells by Inhibiting miR-146b-5p Expression. *Cancer*

- Manag Res.* 2020;**12**:6091-101. [PubMed ID: 32801860]. [PubMed Central ID: PMC7382757]. <https://doi.org/10.2147/CMAR.S252295>.
40. Yan L, Zhang Z, Yin X, Li Y. lncRNA NEAT1 Facilitates Cell Proliferation, Invasion and Migration by Regulating CBX7 and RTCB in Breast Cancer. *Onco Targets Ther.* 2020;**13**:2449-58. [PubMed ID: 32273717]. [PubMed Central ID: PMC7102915]. <https://doi.org/10.2147/ott.s240769>.
 41. Li W, Zhang Z, Liu X, Cheng X, Zhang Y, Han X, et al. The FOXN3-NEAT1-SIN3A repressor complex promotes progression of hormonally responsive breast cancer. *J Clin Invest.* 2017;**127**(9):3421-40. [PubMed ID: 28805661]. [PubMed Central ID: PMC5669564]. <https://doi.org/10.1172/JCI94233>.
 42. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene.* 2003;**22**(39):8031-41. [PubMed ID: 12970751]. <https://doi.org/10.1038/sj.onc.1206928>.
 43. Yoshimoto R, Mayeda A, Yoshida M, Nakagawa S. MALAT1 long non-coding RNA in cancer. *Biochim Biophys Acta.* 2016;**1859**(1):192-9. [PubMed ID: 26434412]. <https://doi.org/10.1016/j.bbagr.2015.09.012>.
 44. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell.* 2008;**135**(5):919-32. [PubMed ID: 19041754]. [PubMed Central ID: PMC2722846]. <https://doi.org/10.1016/j.cell.2008.10.012>.
 45. Amodio N, Raimondi L, Juli G, Stamato MA, Caracciolo D, Tagliaferri P, et al. MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J Hematol Oncol.* 2018;**11**(1):63. [PubMed ID: 29739426]. [PubMed Central ID: PMC5941496]. <https://doi.org/10.1186/s13045-018-0606-4>.
 46. Kim J, Piao HL, Kim BJ, Yao F, Han Z, Wang Y, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat Genet.* 2018;**50**(12):1705-15. [PubMed ID: 30349115]. [PubMed Central ID: PMC6265076]. <https://doi.org/10.1038/s41588-018-0252-3>.
 47. Kwok ZH, Roche V, Chew XH, Fadieieva A, Tay Y. A non-canonical tumor suppressive role for the long non-coding RNA MALAT1 in colon and breast cancers. *Int J Cancer.* 2018;**143**(3):668-78. [PubMed ID: 29574704]. <https://doi.org/10.1002/ijc.31386>.
 48. Bafford AC, Burstein HJ, Barkley CR, Smith BL, Lipsitz S, Iglehart JD, et al. Breast surgery in stage IV breast cancer: impact of staging and patient selection on overall survival. *Breast Cancer Res Treat.* 2009;**115**(1):7-12. [PubMed ID: 18581232]. <https://doi.org/10.1007/s10549-008-0101-7>.