Identification of Breast Cancer Biomarkers by Constructing Protein-Protein Interaction and miRNAs-mRNAs Networks

Azizeh Asadzadeh*, PhD, Bita Behboodian**, PhD, Parsa Bernoos*, BSC, Masoumeh Shojaei Barjuei*, MSC

*Department of Biology, Faculty of Basic Sciences, Nourdanesh Institute of Higher Education, Meymeh, Isfahan, Iran
**Department of Animal Science, Kashmar Branch, Islamic Azad University, Kashmar, Iran

Corresponding author
Azizeh Asadzadeh, PhD
Department of Biology, Faculty of Basic Sciences, Nourdanesh Institute of Higher Education, Meymeh, Isfahan, Iran
Email: az.asadzadeh@gmail.com
az.asadzadeh@nourdanesh.ac.ir

Abstract
Background: In cancer-related diseases, early detection and control of disease progression are very important for successful treatment. Breast cancer is a significant problem due to its high mortality rate in the female population worldwide. By the early diagnosis of breast cancer, the 5-year survival rate reaches 93 to 98%. In this study, to identify breast cancer biomarkers, we construct new protein-protein interaction (PPI) and miRNAs-mRNAs networks by analyzing upregulated and downregulated genes in breast cancer patients.

Method: In this in silico study, two gene expression profile datasets, with the accession numbers GSE42568 and GSE154255, were downloaded from the GEO database. GEO2R was used to obtain differentially expressed mRNA (DEMs) and miRNAs (DEMIs) based on |logFC|>2 and adjusted P-value <0.05. Gene Ontology and KEGG Pathway Enrichment Analysis were performed by EnrichR. STRING v9.1 and cytoHubba plugin in Cytoscape (v3.9.1) were used to investigate PPI network construction and identification of hub genes. Finally, key microRNAs (miRNAs) were predicted.

Results: After protein-protein interaction analysis, a total of 10 upregulated DEMs (DLGAP5, CCNB1, TTK, NUSAP1, RRM2, BUB1B, CDK1, CENPF, TOP2A, and ASPM) and 10 downregulated DEMs (PPARG, LIPE, CD36, FABP4, SCD, LPL, DGAT2, PNPLA2, ACSL1, and LEP) were screened as hub genes. Based on miRNAs-mRNAs networks, 4 key miRNAs including hsa-miR-182-5p, hsa-miR-96-5p, hsa-miR-335-3p, and hsa-miR-32-5p play a critical role in network regulation.

Conclusion: Our study presents PPI and miRNAs-mRNAs networks for identifying molecular biomarkers in breast cancer. The introduced biomarkers open a new approach to diagnostic and therapeutic indicators for clinical applications.

Keywords: Computer Simulation, MicroRNAs, Biomarkers, Protein Interaction Mapping, Breast Cancer
Introduction
Breast cancer occurs when the cells of the breast gland grow out of control.\(^1\) According to the available databases and previous research, breast cancer is the most common cancer and one of the main causes of cancer-related deaths. Its incidence is rapidly increasing in all regions of the world, especially in industrialized and developed countries.\(^2\)

Many risk factors are involved in the development of breast cancer. The probability of tumorigenesis in women with mutations in breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) genes is much higher than in normal people. Obesity is associated with hypercholesterolemia which increases fat cells in the breast tissue that ultimately leads to an increase in the concentration of estradiol. Changes in the female sex hormones such as estradiol and progesterone cause the growth of breast ducts and the synthesis of estrogen and progesterone receptors in the mammary glands. The effect of these sex hormones on breast cancer development has been proven. Other risk factors for breast cancer include: increasing age, family history of breast cancer, smoking, drinking alcohol, and radiotherapy.\(^3,4\) However, it is very difficult to know the pathways leading to this type of malignancy because multiple factors are involved in the formation and development of breast cancer.

A deep understanding of the molecular mechanism of breast cancer and identification of key genes is possible by constructing interactome networks. An interactome is a set of molecular and physical interactions between molecules in a particular cell. Different types of graphs are used to display interactome networks. In these graphs, various entities are represented by nodes, such as proteins or genes, and connections between the nodes are done by the edge.

One of the most important graphs that shows the interaction between proteins is protein-protein interactions (PPIs). This interactome is the basis of biological processes inside the cell; thus, any disruption in its function causes many different diseases, especially cancer. Moreover, constructing and analyzing PPIs is very helpful for designing target-based drugs.\(^5\)

The miRNA-mRNA network is another useful interactome for understanding how protein interaction networks are regulated by miRNAs in various cancers. MicroRNAs (miRNAs) are small molecules consisting of 22 nucleotides with an abundance of 1% of the animal genome.\(^6\) These oligonucleotides are non-coding RNA molecules and their main function is post-transcriptional regulation. They can regulate many essential biological processes such as differentiation, and apoptosis.\(^7\) Due to the relationship between the function of miRNA in turning on and off key genes in the tumorigenesis pathway, their important role in the formation and development of cancer cannot be ignored. Nowadays, these small nucleotides are used as biomarkers to detect cancers.\(^8\)

In recent years, the use of computer-based methods provides efficient information for researchers in various fields of biology such as analysis of molecular networks, vaccine and drug design. Cell signaling mechanisms and gene expression profiles, as the basis of malignancies, can be predicted by bioinformatics methods.

The purpose of the present study was to construct a new PPI network and miRNAs-mRNAs network for introducing diagnostic biomarkers in breast cancer.

Materials and Method
**Microarray data collection**
In this in silico analysis, the Gene Expression Omnibus (GEO) database was used to obtain suitable gene chips for checking differentially expressed mRNAs, and
miRNAs. Two gene expression profile datasets, with the accession numbers GSE42568 and GSE154255, were selected. The GSE42568 dataset contains 104 breast cancer and 17 normal breast samples and the microarray platform used for this dataset was the Affymetrix Human Genome U133 plus 2.0 Array. The GSE154255 dataset with Agilent 046064 Unrestricted Human miRNA V19.0 Microarray consists of 10 breast cancer tissues and 10 normal breast tissues.

**Differentially expressed mRNAs and miRNAs analysis**

GEO2R was used to obtain differentially expressed genes (DEGs) including mRNAs and miRNAs. To estimate the changes, we defined |logFC|>2 and adjusted P-value < 0.05.

**PPI network construction and hub genes detection**

To investigate interactions between proteins, STRING v9.1 software (available from https://string-db.org) was employed. STRING is one of the widely used biological databases with free access in molecular biology, which is used to search for relationships between genes and proteins. Two lists of names of 1040 downregulated and 578 upregulated differentially expressed mRNAs (DEMs) were inserted separately and two networks were formed based on a score of less than 0.4. After transferring the PPI network in TSV format to Cytoscape (v3.9.1) software, key genes for two downregulated and upregulated DEMs were obtained.

**Gene ontology and pathway analysis of upregulated and downregulated genes**

Gene ontology and KEGG pathway analysis are two very useful and important methods for extracting common gene categories in the functions. EnrichR in https://maayanlab.cloud/Enrichr/ was used for GO and KEGG pathway assessment.

**The miRNA-mRNA network construction**

To determine the target genes of the differentially expressed miRNAs (DEMIs), the latest version of the advanced miRNA target prediction algorithm called DIANA-microT-CDS 2023 was used. This algorithm identifies miRNA binding sites in two parts (3’-UTR and CDS). In the next step, the common genes of two datasets, target genes of DEMIs and DEMs were obtained. Finally, the miRNA-mRNA network was drawn using Cytoscape.

**Results**

**Differentially expressed mRNAs and miRNAs analysis**

In the GEO2R analysis, 1040 downregulated and 578 upregulated differentially expressed mRNAs (DEMs) in GSE42568 were detected. For the GSE154255 dataset, 13 upregulated and 21 downregulated small non-coding RNAs, as the differentially expressed miRNAs (DEMIs), were obtained.

**PPI network construction and hub genes detection**

Through the Cytoscape software, the PPI network was constructed. The network for upregulated DEMs includes 351 nodes as genes and 2606 edges as interactions (Figure 1A) and for downregulated DEMs includes 555 nodes and 3123 edges (Figure 2A). To obtain the ten key genes with increased and decreased expression, the cytoHubba plugin in the Cytoscape software was used. Initially the score of all nodes was calculated, and the interaction network of the top ten genes was constructed (Figure 1B and 2B). The 10 top hub genes with a high degree in the network for upregulated DEMs included: disks large-associated protein 5 (DLGAP5), Cyclin B1 (CCNB1), TTK protein kinase (TTK), Nucleolar and spindle-associated protein 1 (NUSAP1), Ribonucleotide reductase M2 (RRM2), BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B), Cyclin Dependent Kinase 1 (CDK1), Centromere
protein F (CENPF), DNA Topoisomerase II Alpha (TOP2A), and Abnormal spindle-like microcephaly-associated (ASPM). The 10 key genes associated with downregulated DEMs are as follows: Peroxisome proliferator activated receptor gamma (PPARG), Lipase E, Hormone Sensitive Type (LIPE), cluster of differentiation 36 (CD36), Fatty Acid Binding Protein 4 (FABP4), stearoyl-coenzyme A-desaturase (SCD), Lipoprotein lipase (LPL), Diacylglycerol O-acyltransferase 2 (DGAT2), Patatin-like phospholipase domain containing 2 (PNPLA2), Acyl-CoA Synthetase Long Chain Family Member 1 (ACSL1), and leptin gene (LEP).

**Gene ontology and pathway analysis of upregulated and downregulated genes**

The results of gene ontology and KEGG pathway analysis are shown in table 1. The upregulated genes are involved in spindle assembly checkpoint signaling, mitotic spindle assembly checkpoint signaling, ECM-receptor interaction, and proteoglycans in cancer while the downregulated genes are involved in the regulation of fat cell differentiation, positive regulation of cell differentiation, PPAR signaling pathway, and regulation of lipolysis in adipocytes.

**The miRNA-mRNA network construction**

A total number of 350 common genes between target genes of DEMIs and DEMs were obtained. After calculating the score of each node by the cytoHubba plugin, considering the 4 top miRNAs, a miRNA-mRNA network was created (Figure 3).

**Discussion**

In the present study, we identified 1040 downregulated and 578 upregulated genes in the DEMs study. We also detected 13 upregulated and 21 downregulated miRNAs in the DEMs analysis. After filtering differentially expressed mRNAs with cytoHubba plugin in Cytoscape software, 10 top hub nodes with a high degree for upregulated genes were obtained including: DLGAP5, CCNB1, TTK, NUSAP1, RRM2, BUB1B, CDK1, CENPF, TOP2A, and ASPM. Our study revealed that 10 key nodes associated with downregulated genes as follows: PPARG, LIPE, CD36, FABP4, SCD, LPL, DGAT2, PNPLA2, ACSL1, and LEP.

Based on the research, most of the obtained hub DEMs are involved in cell division, cell proliferation, and cell cycle regulatory pathways. The presence and role of these genes in vital processes confirm their role in carcinogenesis pathways. Stabilization of spindle formation is one of the main functions of DLGAP5. Initial studies on this protein showed that it is a cell cycle regulator. In many studies, the relationship between DLGAP5 and different types of cancers has been mentioned. Xu et al. investigated biomarker genes in breast cancer and introduced DLGAP5 as a potential oncogene. Some industrial chemical compounds such as bisphenol A cause breast cancer by targeting DLGAP5 and activating the IL-6/JAK2/STAT3 signaling pathway.

According to previous studies, CCNB1 is related to tumor cell proliferation and metastasis. The increased expression of CCNB1 protein in tumor breast tissues compared with normal breast tissues has been showed. Cyclin B1 was mentioned as an independent predictive factor for breast cancer. TTK, a protein kinase that plays an important role in mitosis, is an prognostic biomarker in breast, colorectal, and cervical cancers. NUSAP1 is a significant macromolecule in spindle stability by binding to microtubules. Qiu et al. reported that NUSAP1 participates in metastasis in breast cancer cells through AMPK/PPAR signaling pathway. Koyuncu et al. by analyzing gene expression changes in breast cancer cells indicated a positive correlation between BUB1B overexpression and more aggressive behavior. The protein
translated from the BUB1B gene is mitotic checkpoint serine/threonine kinase B which plays an important role in regulating the spindle assembly checkpoint.\textsuperscript{19} According to the reports, CENPF is a protein marker for the proliferation of tumor cells, and its expression is strongly changed in breast, prostate, Lung, and renal cancers.\textsuperscript{20-23} Centromere F protein is one of the important proteins in the cell cycle regulation.\textsuperscript{24} Given down-expressed genes, PPARG belongs to the superfamily of nuclear receptors. In a study by Li et al., a decrease in the expression level of PPARG was reported in breast cancer patients.\textsuperscript{25} Among downregulated DEMs, LIPE is expressed in adipose tissues and regulates the rate of lipolysis of diacylglycerol and monoacylglycerol. This gene and PNPLA2 have been introduced as a hub gene in breast ductal carcinoma in a study by Wu et al.\textsuperscript{26} Results of the GO and KEGG pathways demonstrated that most of the network genes are involved in signaling receptor interaction, cell differentiation, and regulation pathways. MicroRNAs are small molecules in cells and the bloodstream that mainly control gene expression by binding to messenger RNA. These oligonucleotides play regulatory roles in various developmental, cellular, and physiological processes. Comparing cancer samples with normal samples in vivo and in vitro experiments has shown that there are significant differences in the expression of microRNA.\textsuperscript{27} Given that changes in the expression of microRNAs are related to various diseases, they can be used as molecular biomarkers.\textsuperscript{28} In the present study, to identify biomarkers based on microRNAs, miRNAs-mRNAs network was constructed. According to the results obtained from the miRNA-mRNA network, among 13 upregulated and 21 downregulated small non-coding RNAs, four top miRNAs were identified as key regulators in the network. These miRNAs included hsa-miR-182-5p, hsa-miR-96-5p, hsa-miR-335-3p, and hsa-miR-32-5p. The expression of the ribonucleotide miR-182-5p is increased in prostate and colon cancers, and it is a biomarker of melanoma, endometrial, and lung cancers.\textsuperscript{29} Darbeheshti et al. investigated the key miRNA in breast cancer and showed that miR-182-5p expression in breast cancer patients is very high.\textsuperscript{30} Breast cancer cells increase the expression of miR-182-5p and transport it through extracellular vesicles, which ultimately leads to cancer spread.\textsuperscript{31} The function of microRNA-96-5p as a oncogene has been proven in ovarian cancer cells.\textsuperscript{32} Qin et al. reported that breast cancer migration mediated by activation of MEK/ERK signaling is associated with MiR-96-5p.\textsuperscript{33} According to Yin et al., this type of miRNA promotes cell proliferation and the progression of breast cancer by negatively regulating the tumor suppressor gene FOXO3.\textsuperscript{34} In recent years, miRNAs have been considered as key factors in the carcinogenesis process, tumor development, metastasis, and drug resistance, and their identification is very helpful in rapid diagnosis and progress of diseases.\textsuperscript{35-37} These non-coding RNA molecules affect various processes by directly binding to the 3'-UTR of target mRNA. Based on our miRNAs-mRNAs network, hsa-miR-182-5p, hsa-miR-96-5p, hsa-miR-335-3p, and hsa-miR-32-5p create a large number of overexpressed and underexpressed genes in breast cancer disease. After studying the genes affected by these four types of microRNAs, it was found that 125 genes from 578 upregulated differentially expressed mRNA, and 225 genes from1040 downregulated differentially expressed mRNA were changed. Among the nodes in network, there are cases where several MicroRNAs affect their expression. For example, hsa-miR-335-3p, and hsa-miR-32-5p increase the expression of SOX4 gene,
simultaneously. Zhang et al. observed that SOX4 interferes with the survival, migration and invasion of breast cancer cells in vitro and increases tumor growth and metastasis in vivo. Similar to our results, Liang et al. reported that abnormal expression of CACNA2D1 in breast cancer is closely related to the prognosis of breast cancer, and it may serve as target for the treatment of breast cancer. In our study, CACNA2D1 was changed by hsa-miR-182-5p, hsa-miR-96-5p, and hsa-miR-335-3p. Also, ITSN1 gene was downregulated by hsa-miR-182-5p and hsa-miR-335-3p. Xie et al. showed a decrease in the expression of ITSN1 in breast cancer tissues and identified this gene as a key biological target in breast cancer.

The present study has potential limitations that should be considered. The target genes of miRNA are predicted by the software database, and their relationship has not been confirmed experimentally; thus, the prospective clinical confirmation of the predicted genes in the miRNA-mRNA network will expand the results of this study.

Conclusion
We performed a combined analysis of messenger RNAs and micro RNAs expression profiles of breast cancer to identify diagnostic biomarkers. Our study demonstrated 10 upregulated DEMs, 10 downregulated DEMs, and four miRNAs as key biomarkers in breast cancer. Most of the obtained genes were involved in signaling pathways, receptor interaction, cell differentiation, and regulation pathways. The present study has the potential to enhance the diagnosis and prognosis of breast cancer, and introduce innovative techniques and technologies for its treatment.

Acknowledgments
The authors would like to thank Dr. Afshin Fassihi for his helpful advice and send their gratitude to all members of the Isfahan Pharmaceutical Sciences Research Center for their support and encouragement.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors’ Contribution
A.A. conceived the presented idea, developed the theory and project administration, and performed computational analysis. A.A., B.B., P.B. and M.S.H. Collecting data, writing, and editing the manuscript. All authors have read and approved the final manuscript.

Conflict of Interest
None declared.

References


<table>
<thead>
<tr>
<th>Gene names</th>
<th>Gene Count</th>
<th>Type of DEGs</th>
<th>GO and KEGG pathway analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC20; CENPF; NUF2; ZNF207; BUB1B; TTK; TRIP13; ZWINT</td>
<td>8</td>
<td>upregulated genes</td>
<td>Spindle Assembly Checkpoint Signaling (GO:0071173)</td>
</tr>
<tr>
<td>CDC20; CENPF; NUF2; ZNF207; BUB1B; TTK; TRIP13; ZWINT</td>
<td>8</td>
<td>upregulated genes</td>
<td>Mitotic Spindle Assembly Checkpoint Signaling (GO:007094)</td>
</tr>
<tr>
<td>LAMA5; SDC4; TNC; FN1; HMMR; NPNT; COMP; COL1A1; COL1A2; SPP1; COL4A5; SDC1; AGRN</td>
<td>13</td>
<td>upregulated genes</td>
<td>ECM-receptor interaction</td>
</tr>
<tr>
<td>VAV3; SDC4; FN1; ITPR3; ANK3; MMP9; IGF1R; MAPK13; COL1A1; TIAM1; COL1A2; ERBB3; ERBB2; SDC1; FLNB; EZR</td>
<td>16</td>
<td>upregulated genes</td>
<td>Proteoglycans in cancer</td>
</tr>
<tr>
<td>LMO3; RARRES2; ZBTB16; ADIPOQ; ADIRF; TMEM64; FOXO1; MEDAG; ZFP36; SFRP1; BMP2; GPER1; LEP; PPARG; FERMT2</td>
<td>15</td>
<td>downregulated genes</td>
<td>Regulation Of Fat Cell Differentiation (GO:0045598)</td>
</tr>
<tr>
<td>TNXB; PKDCC; LAMA2; TWIST1; ADIRF; GDPD5; MEDAG; TMEM100; ZFP36; RRAS; CD36; APOB; SOX5; MMD; LMO3; RARRES2; ZBTB16; PLA2G2A; IGF1; TMEM64; TGFBR2; SFRP1; BMP2; ZEB1; RGCC; AGTR1; PPARG;CAMK1;FERMT2;DDR2</td>
<td>30</td>
<td>downregulated genes</td>
<td>Positive Regulation Of Cell Differentiation (GO:0045597)</td>
</tr>
<tr>
<td>ACSL1; ADIPOQ; LPL; NR1H3; ACSL4; SORBS1; FABP4; ACADL; FABP5; SCD; EHHADH; ME1; PLIN4; PLIN2; PPARG; ACADM; CD36; ANGPTL4; PLIN1; PCK1; PLTP</td>
<td>21</td>
<td>downregulated genes</td>
<td>PPAR signaling pathway</td>
</tr>
<tr>
<td>NPR1; PDE3B; PTGER3; NPY1R; IRS2; ADRB1; ABHD5; ADRB2; PTGS2; GNA11; FABP4; PLIN1; MGLL; PNPLA2</td>
<td>14</td>
<td>downregulated genes</td>
<td>Regulation of lipolysis in adipocytes</td>
</tr>
</tbody>
</table>

GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ECM: Extracellular matrix; PPAR: Peroxisome proliferator-activated receptors
Figure 1. This figure represents the networks of upregulated DEMs (A) and top 10 hub genes (B). Networks were created by Cytoscape software.

DEMs: Differentially expressed mRNAs
Figure 2. This figure represents the networks of down-regulated DEMs (A) and top 10 hub genes (B). Networks were created by Cytoscape software.

DEMs: differentially expressed mRNAs
Figure 3. This figure shows the miRNA-mRNA network construction, the green circles represent mRNAs, and the yellow rectangles represent miRNAs. 4 top miRNAs are has-miR-182-5p, has-miR-96-5p, has-miR-335-3p, and has-miR-32-5p. The miRNA-mRNA network is extracted using Cytoscape software.

miRNA: microRNA; mRNA: Messenger RNA; has: Homo sapiens