**LINC00520** promotes breast cancer development by low expression as a tumor suppressor and prognostic biomarker by regulating the **ESR2** expression level: integrated systems biology bioinformatics and experimental analyses

Ehsan Ezzati  
Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran

Safie Mosadeghi  
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran

Aliasghar Akbarinia  
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran

Soroor Horriat  
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran

Mohammad Rezaei  
Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran  
https://orcid.org/0000-0003-3888-5839

Mansoureh Azadeh (✉ mazadeh@phd.iaurasht.ac.ir)  
Department of Biology, Rasht Branch, Islamic Azad University, Rasht, Iran  
https://orcid.org/0000-0003-2031-4640

---

**Research Article**

**Keywords:** Breast Cancer, Bioinformatics, Microarray, RNA Interaction, Signaling pathway, Real-time PCR

**Posted Date:** August 10th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-194462/v1

**License:** ☐ ☛ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background

ESR2 can have a remarkable effect on the proliferation and tumorigenesis of breast cancer (BC) as a hub tumor suppressor by regulating the effect of estrogen. The abnormal changes in the expression of the co-expressed IncRNAs can have a regulatory effect on the ESR2 expression and cause BC status. In this research, the expression and correlation of ESR2 and a related IncRNA (LINC00520) were examined through different statistical methods.

Method

An integrated bioinformatics analysis was performed to obtain a reasonable prediction about the relation of potential biomarkers and design an experiment. Based on the Microarray data analysis, pathway enrichment, GO, and ceRNA interaction analyses, a collection of RNAs was selected for the real-time PCR experiment. Several statistical methods were performed to evaluate the difference in the expressions in the BC samples, correlation of RNAs, and the possibility of being prognostic and diagnostic biomarkers (ROC and clinicopathological analysis).

Result

It is demonstrated that ESR2 (logFC: -1.842, P. Value < 0.0001) and LINC00520 (logFC: -2.231, P. Value < 0.0001) are the two robust breast cancer potential tumor suppressor in BC population. Likewise, it is demonstrated that the expression of these two RNAs has a significantly positive correlation in tumor samples (r: 0.7, P. Value: 0.0008). ESR2 (AUC: 0.9775, P. Value < 0.0001) and LINC00520 (AUC: 0.8450, P. Value: 0.0002) are the two potential diagnostic biomarkers of BC. Furthermore, expression of LINC00520 has a significant correlation with the stage (stage II) and tumor size (less than 5 cm) in the BC patients.

Conclusion

It is highly possible that the abnormal changes in the expression of LINC00520 can significantly help BC development by affecting the expression of ESR2 and playing a tumor suppressor role in Isfahan BC population. Furthermore, LINC00520 could be considered as a significant prognostic (clinicopathological analysis) and diagnostic (ROC analysis) biomarkers of BC.

1. Introduction

Estrogen is a major regulator of mammary gland development and a crucial player in BC etiology [1], [2]. Estrogen works by attaching to its intracellular high-affinity receptor (ER), which comes in two types: ERα and Erβ. Both ER types are nuclear ligand-regulated transcription factors that serve as estrogen signaling
transducers. The $ESR1$ gene on chromosome 6 (6q25.1) encodes human ER, while the $ESR2$ gene on chromosome 14 encodes human ER (14q23.2). $ESR1$ and $ESR2$ are both expressed in normal breast tissues and BC, respectively [2], [3].

$ESR2$ is a member of the estrogen receptor family and the nuclear receptor transcription factor superfamily. The gene product is found in the nucleus, cytoplasm, and mitochondria and has an N-terminal DNA binding domain and a C-terminal ligand-binding domain (www.genecards.org). Previous studies on the BC patients and BC cell lines revealed that this gene has an anti-proliferation effect, and decreasing the expression of $ESR2$ helps BC tumorigenesis [4], [5]. Previous studies have also discussed the role of essential polymorphisms in the $ESR2$ expression level and BC development [6]–[8]. However, the accurate function of this gene in BC development is still unclear. So, it is reasonable to investigate the expression pattern of this gene in the different populations, with the different clinicopathological situations.

Long non-coding RNAs have shown to have a variety of functions in gene transcription, post-transcription, translation, and epigenetic modification in a growing number of studies. IncRNA malfunction or abnormal expression is linked to a number of illnesses, including cancer [9]. According to a recent manual annotation of IncRNAs, there should be over 9640 IncRNAs, accounting for almost half of all protein-coding genes. Also, based on LNCipedia 2.0, there are already 32,183 annotated IncRNA in the human [10]. LncRNA can have some essential effects on gene expression, in direct or indirect ways, like ceRNA interactions [11]. The primary goal of this research is to demonstrate an integrated transcriptome interaction network related to $ESR2$ and evaluate the possible effect of hub nodes of this network on BC development, especially in Isfahan population.

2. Materials And Methods

2.1. Microarray analysis

The GSE42568 [12] dataset ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) was selected and processed to discover differentially expressed genes (DEGs) in BC samples compared to the control samples. This dataset was discovered using the GEO database. The DEG analysis was carried out using the limma [13] and GEOquery [14] packages. The ggplot2 and pheatmap packages were used to create microarray analysis plots and Figures. The NormalizeQuantiles command in R was used to normalize the raw data using the quantiles normalization approach. A total of 121 samples (17 control and 104 BC samples) were examined. Genes with logFC larger than the third quartile (0.20588) were identified as up-regulated, while genes with logFC less than the first quartile (-0.07747) were selected as low expressed genes. Based on the distribution of relative expression, data of all genes were investigated in this experiment. The level of significance was set at an adj.p.value of less than 0.001. Principal Component Analysis (PCA) was performed for the distinguishing of the control and tumor samples and deleting bad quality samples (prcomp command in R studio).

2.2. Bioinformatics analyses
Pathway enrichment analysis was performed by KEGG [15]–[17] and Reactome [18], [19]. miRNA-mRNA interaction analysis was conducted by miRWalk 3 [20], [21]. miRNA-IncRNA interaction analysis was made by NPInter [22]–[25]. More information about IncRNA was obtained from LncRNADisease v2.0 [26], [27].

The interaction between the IncRNA and gene was analyzed by IncRRIsearch [28]. Based on interaction energy, which is calculated using both accessibility energy and hybridization energy, IncRRIsearch makes predictions about local base-pairing interactions. In a nutshell, IncRRIsearch takes into account the impacts on stabilization energy resulting from hybridization between two RNA sequences as well as the energy for avoiding the creation of intramolecular double-stranded structure. For each RNA-RNA pair, IncRRIsearch produced numerous possibilities for local base-pairing interactions. Each local base-pairing interaction visualized by the VARNA algorithm [29]. The IncRRIsearch databases uses the linear algorithm of VARNA to illustrate the local base-pairing interaction of IncRNA with mRNA. Visualization of interaction network was drawn by Cytoscape 3.8.0. Gene ontology analysis was implemented by QuickGO online software (https://www.ebi.ac.uk/QuickGO/). The protein-protein interaction analysis was employed by STRING V11.00 [30].

2.3. The clinical features of tissue samples

All patients signed written consent forms, and all methods for the research in this study involving human samples were approved by the Al-Zahra Hospital Ethics Committee, Isfahan University of Medical Science. Samples of normal breast tissue and breast cancer from 20 individuals with breast cancer were compared in a case-control study. Specimen collection was performed by biopsy. None of the patients had ever received radiation or chemotherapy. Tissue samples were rinsed in distilled water and promptly frozen in liquid nitrogen for RNA Later solution (Invitrogen, USA) immersion for pathologist assessment. The clinicopathological characteristics of patients with breast and stomach cancer are listed in Table 1.
Table 1
Frequency of clinicopathological variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Status</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td>I</td>
<td>1 (5%)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>8 (40%)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>7 (35%)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>4 (20%)</td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 45</td>
<td>9 (45%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 45</td>
<td>11 (55%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Tumor size (TS)</td>
<td>&lt; 5cm</td>
<td>12 (60%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 5cm</td>
<td>5 (20%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>4 (20%)</td>
</tr>
<tr>
<td>Menopausal status</td>
<td>Yes</td>
<td>8 (40%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>12 (60%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Lymph node</td>
<td>Yes</td>
<td>12 (60%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>4 (20%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>4 (20%)</td>
</tr>
<tr>
<td>ER receptor</td>
<td>Positive</td>
<td>7 (35%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>5 (25%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>8 (40%)</td>
</tr>
<tr>
<td>PR receptor</td>
<td>Positive</td>
<td>6 (30%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>6 (30%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>8 (40%)</td>
</tr>
<tr>
<td>HER2/neu receptor</td>
<td>Positive</td>
<td>6 (30%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>6 (30%)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>8 (40%)</td>
</tr>
</tbody>
</table>

2.4. Real-time PCR experiment
The RNA content of BC tissue samples and standard breast tissue equivalents from the same individuals were collected and extracted using Trizol reagent according to the manufacturer’s instructions (Promega Co., Madison, WI, USA).

Thermo Fisher Scientific, Waltham, MA, USA, provided the first-strand cDNA synthesis package, which was used to synthesize cDNA according to the manufacturer’s procedure. The cDNA products were processed at 20°C for the expression investigation of the \( ESR2 \) and \( LINC00520 \) genes and GAPDH as a control gene. For each gene, customized primers were designed using oligo software (version 7, Table 2).

### Table 2
The primer table of \( ESR2 \), \( LINC00520 \), and GAPDH genes.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( ESR2 ) (Forward)</td>
<td>AACACCTGGGACACCTTTCTC</td>
</tr>
<tr>
<td>2</td>
<td>( ESR2 ) (Revers)</td>
<td>CCTTCACACGACCAGACTCC</td>
</tr>
<tr>
<td>7</td>
<td>( LINC00520 ) (Forward)</td>
<td>TTAACAGTTTTTGGGTTCTGG</td>
</tr>
<tr>
<td>8</td>
<td>( LINC00520 ) (Revers)</td>
<td>GACTTGATGGGATTGTGGGG</td>
</tr>
<tr>
<td>9</td>
<td>( GAPDH ) (Forward)</td>
<td>TACAGGAACAGGGTGGGTG</td>
</tr>
<tr>
<td>10</td>
<td>( GAPDH ) (Revers)</td>
<td>ATTCAGTGTGGGTGGGGGAC</td>
</tr>
</tbody>
</table>

#### 2.5. Statistical analysis

GraphPad Prism (version 8) software was used to undertake statistical analysis of real-time PCR data using the \(-ddCt\) method. GraphPad was used to create the expression graphs. To ensure the normality of the expression, the Kolmogorov-Smirnov test was used. The expression data were subjected to a paired t-test to determine our findings’ significance and compare the mean of two control and tumor samples. The IBM SPSS Statistics was used to prepare the clinicopathological frequency table. Pearson correlation test was performed to evaluate the correlation of possible biomarkers.

### 3. Results

#### 3.1. Bioinformatics analyses

##### 3.1.1. Microarray analysis: Quality control

Microarray analysis was carried out on the GSE42568 microarray dataset. Principal component and correlation analysis were conducted to evaluate samples’ quality and the microarray data (Fig. 1–3). In the Fig. 1A, the raw microarray expression data was illustrated. After quantile normalization, the normalized expression data was illustrated in the Fig. 1B. In the x axis of Figure A, the microarray control and tumor samples are provided and in the y axis, the expression of all 54675 represented genes are provided. The noisy samples in the mentioned boxplots are the samples with different median, compared
to other samples. There were no bad quality samples in this dataset, based on the expression boxplots in the Fig. 1. In the Fig. 2, the PCA analysis was performed to find the bad quality samples, based on the first and second principal component of control and tumor samples. Based on the PCA analysis of raw and normalized data, three control samples and two tumor samples have not suitable quality for considering in the DEG analysis. The mentioned samples were deleted from the analysis. In the Fig. 3, the correlation between microarray samples were calculated and the heatmap was illustrated based on the correlation analysis. The control samples are provided in the first 17 columns (right) and the 104 tumor samples are represented in the 104 left columns. This analysis provided more evidence about the bad quality samples in this study and supported the result of PCA analysis. Finally, 14 control samples and 102 tumor samples were selected for the differential expression and statistical analyses (Fig. 4).

3.1.2. Microarray analysis: DEG analysis

Differential expression analysis was performed by limma package in R Studio. The DEG analysis revealed 5474 down-regulated genes and 4275 up-regulated genes in this dataset, based on the threshold. Among the down-regulated genes, ESR2 was selected as the significant low-expressed gene for the subsequent investigations (adj. P. Value: 7.2 * 10e-6, Fig. 5).

3.1.3. Pathway enrichment analysis

By KEGG and Reactome online databases, the pathway enrichment analysis of ESR2 was performed (Fig. 6). The relevant signaling pathways are indicated in Table III.

<table>
<thead>
<tr>
<th>KEGG</th>
<th>Reactome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endocrine resistance</td>
<td>ESR-mediated signaling</td>
</tr>
<tr>
<td>Chemical carcinogenesis - receptor activation</td>
<td>Generic Transcription Pathway</td>
</tr>
<tr>
<td>Breast cancer</td>
<td></td>
</tr>
<tr>
<td>Estrogen signaling pathway</td>
<td></td>
</tr>
<tr>
<td>Pathways in cancer</td>
<td></td>
</tr>
<tr>
<td>GnRH secretion</td>
<td></td>
</tr>
</tbody>
</table>

3.1.4. Gene ontology

Gene ontology analysis revealed that ESR2, located in the nucleus and mitochondrion, involves several essential biological processes and signaling pathways, like intracellular estrogen receptor signaling pathway, regulation of transcription, DNA binding, and zinc ion binding. A complete list of these processes is given in Table 4. These processes are also illustrated in Fig. 7.
<table>
<thead>
<tr>
<th>QUALIFIER</th>
<th>GO TERM</th>
<th>GO NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>enables</td>
<td>GO:0005496</td>
<td>steroid binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:00030284</td>
<td>estrogen receptor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0008270</td>
<td>zinc ion binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0004879</td>
<td>nuclear receptor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0003677</td>
<td>DNA binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0003700</td>
<td>DNA-binding transcription factor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0003707</td>
<td>steroid hormone receptor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0043565</td>
<td>sequence-specific DNA binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0003677</td>
<td>DNA binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0046872</td>
<td>metal ion binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0000978</td>
<td>RNA polymerase II cis-regulatory region sequence-specific DNA binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0004879</td>
<td>nuclear receptor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0030284</td>
<td>estrogen receptor activity</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0034056</td>
<td>estrogen response element binding</td>
</tr>
<tr>
<td>enables</td>
<td>GO:0019899</td>
<td>enzyme binding</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0030520</td>
<td>intracellular estrogen receptor signaling pathway</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0006355</td>
<td>regulation of transcription, DNA-templated</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0071392</td>
<td>cellular response to estradiol stimulus</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0045893</td>
<td>positive regulation of transcription, DNA-templated</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0051091</td>
<td>positive regulation of DNA-binding transcription factor activity</td>
</tr>
<tr>
<td>involved in</td>
<td>GO:0000122</td>
<td>negative regulation of transcription by RNA polymerase II</td>
</tr>
<tr>
<td>located in</td>
<td>GO:0005634</td>
<td>nucleus</td>
</tr>
<tr>
<td>located in</td>
<td>GO:0005634</td>
<td>nucleus</td>
</tr>
<tr>
<td>located in</td>
<td>GO:0005739</td>
<td>mitochondrion</td>
</tr>
<tr>
<td>located in</td>
<td>GO:0005634</td>
<td>nucleus</td>
</tr>
<tr>
<td>located in</td>
<td>GO:0005634</td>
<td>nucleus</td>
</tr>
</tbody>
</table>
3.1.5. RNA and protein interaction analyses

mRNA-miRNA interaction analysis by miRWalk revealed the interacted miRNAs with ESR2. Based on the binding probability (score = 1) of miRNAs, 50 hub miRNAs were selected for the following investigations. Among the top 50 miRNAs indicated in Fig. 8, hsa-miR-106a-3p was selected for the subsequent investigations, based on the pairing energy. This miRNA has less binding energy, comparing to other miRNAs in this analysis. NPInter database revealed that LINCO0520 as a hub IncRNA has an interaction with hsa-miR-106a-3p. Based on this online software, mir-CLIP capture of a miRNA targetome uncovers IncRNA-miRNA interactions. So, this long non-coding RNA was selected as a competitive endogenous RNA for ESR2 that can indirectly affect the expression of this gene. So, the LINCO0520 - hsa-miR-106a-3p - ESR2 axis was selected as the hub ceRNA network, and the LINCO0520 and ESR2 were chosen for the experimental analyses. Besides, the direct interaction of LINCO0520 and ESR2 was revealed by the IncRRIsearch online software (Fig. 9). The general view (Fig. 9) and the linear view (Fig. 10) of local base-pairing interactions is provided in the manuscript. Based on IncRNA Disease, IncRNA LINCO0520 has an essential role in breast cancer.

According to the IncRRIsearch, there are 4 predicted local base-pairing interaction between LINCO0520 and ESR2 (sum of local base-pairing energy: -68.79 kcal/mol). All four local interactions are indicated in Fig. 10.

According to the STRING online software, ESR2 protein has biological interactions with other proteins, including NCOR1, NCOA1-3, MED1, DDX54, and SRC (Fig. 11).

3.2. qRT-PCR experiment

Real-time PCR analysis was performed to evaluate the expression pattern of ESR2 and IncRNA LINCO0520 in Isfahan BC samples, compared to control samples, in the different clinicopathological situations. In addition, the possibility of being a prognostic and diagnostic biomarker for these two RNAs is evaluated in this research. Based on the real-time PCR data analysis, ESR2 has a significantly low expression in Isfahan BC samples compared to control samples (logFC: -1.842, P. Value < 0.0001). Moreover, the IncRNA LINCO0520 has a significant low-expression in the BC samples (logFC: -2.231, P. Value < 0.0001, Figs. 12–13). Additionally, the expression level of ESR2 and IncRNA LINCO0520 had a significant positive correlation in the tumor samples (r: 0.7, P. Value: 0.0008, Fig. 14).

ROC and clinicopathological analysis were followed to understand the possibility of being a prognostic and diagnostic biomarker for ESR2 and LINCO0520. Based on ROC analysis, ESR2 and LINCO0520 can be the two remarkable prognostic biomarkers in Isfahan BC patients (Fig. 15).

Clinicopathological analysis was applied on the expression data of two ESR2 and LINCO0520 RNAs. This analysis was carried out on the samples with different Stages, Age, Lymph Node Metastasis, Menopausal, ER receptor, PR receptor, and Her2/neu receptor statuses. This analysis revealed that
LINC00520 had a significant correlation with the lower tumor size and stage II of BC (Fig. 16). So, this IncRNA can be a potential diagnostic biomarker in Isfahan BC population.

4. Discussion

To obtain valid and novel information and design a reasonable experiment, we performed an integrated bioinformatics analysis to find an essential gene for BC development and investigate some other biological factors that can affect the gene expression and the BC development process. Based on primary bioinformatics analysis, we demonstrated that ESR2 has a significant down-regulation in BC samples and has an essential role as a tumor suppressor for BC. Then, we found a novel IncRNA (LINC00520) that could, directly and indirectly, interact with ESR2 mRNA. So, we define a hypothesis that these two coding and non-coding RNAs could play an essential role in BC development by interacting with each other. Real-time PCR experiment revealed that these two RNAs have significantly low expression in Isfahan BC samples and can be two prognostic biomarkers for Isfahan BC. Furthermore, these two RNAs have a significant positive correlation in tumor samples. The correlation of LINC00520 with stage II of BC and the smaller tumor size has also been demonstrated.

As we mentioned in Fig. 6 and Table III, ESR2 involves several signaling pathways involved in BC. Based on the previous studies, Endocrine therapy is a beneficial method for blocking the ER pathway and helps the BC treatment [31]. Loss of ER expression, production of shortened isoforms of ERα and ERβ, and post-translational changes of ERα are all mechanisms of endocrine resistance in estrogen receptor (ER)-positive breast tumors [32]. Nevertheless, the main impress of ERα mutations in the endocrine resistance is still unclear [33]. About the estrogen signaling pathways, the pharmacological effects of co-expression of estrogen receptor and HER2 receptor are unknown, although they include relative resistance to hormone treatment and chemotherapy, as well as a better long-term outcome that was conferred by ER and a poorer long-term prognosis that was conferred by HER2 expression [34]. Nasun Hah et al. demonstrated that estrogen-dependent transcriptional regulation is influenced by all three RNA polymerases [35]. The GnRH secretion signaling pathway demonstrates that two forms of GnRH hormone are expressed in the breast tissue of humans, and especially in the BC, these two forms are overexpressed. [36]. M. Dowsett et al. showed that GnRH could inhibit aromatization for the premenopausal treatment of BC patients [37]. So, it is crucial to understand the expression pattern of ESR2 and other related and co-expressed coding and non-coding RNAs in the different BC populations.

As we indicated in Fig. 8, ESR2 and IncRNA LINC00520 have an endogenous competition to binding to the hsa-miR-106a-3p, as a ceRNA axis. The difference in the expression of each of these two RNAs can affect the expression of the other by changing the miRNA binding affinity [38]. About the mentioned micro-RNA, previous studies revealed that by sponging miR-106a-3p to control APC production and the Wnt/β-catenin pathway, LINC01133 as ceRNA suppresses gastric cancer development [39]. So, it is probable that this micro-RNA has an essential role in BC development.
About lncRNA *LINC00520*, previous studies revealed that it can have some essential role in the colorectal cancer [40], papillary thyroid carcinoma [41], head and neck squamous cell carcinoma [42] and lung cancer [43] by involving the different relevant signaling pathways, inducing different ceRNA axis, and serving as a ceRNA of different miRNAs. Especially about the BC, a previous study revealed that a genetic variant of *LINC00520* in the rs8012083 region could be used as a biomarker to the triple-negative BC [44]. Whitney S. Henry et al. showed that SRC, STAT3, And PI3K have an essential role in BC development by inducing IncRNA *LINC00520* [45].

In this experiment, we demonstrated the functional role of *LINC00520* and *ESR2* and the positive correlation of this two RNA in Isfahan BC samples for the first time. These results are consistent with the bioinformatics analyses (Microarray analysis, Pathway enrichment, Interaction analysis, and Gene Ontology). It is highly recommended that the same experiment be performed on a larger sample size for the different clinicopathological situations of BC samples, especially ER receptor-positive and negative samples. Due to our limitations in accessing the experimental equipment and lack of structural data in the online database, we could not evaluate the interaction of mRNA ESR2 and IncRNA LINC00520. We suggest that more accurate experiments validate the potential interaction of ESR2 and IncRNA LINC00520. Besides, it is highly recommended that the expression and biomarker analyses be performed on the *MED1, NCOR1, NCOA1-3, DDX54, NOS3, SRC, NR0B2* genes that we obtain from protein-protein interaction analysis and their related miRNAs and lncRNAs in Isfahan BC samples.

### 5. Conclusion

*LINC00520* is a novel prognostic biomarker for breast cancer patients as a significant low-expressed IncRNA. Furthermore, *LINC00520* regulates the expression level of ESR2, a potential tumor suppressor and low-expressed breast cancer biomarker. Based on the experimental approach (qRT-PCR) in this study, the expression level of ESR2 is significantly correlated to the expression level of LINC00520. Also, based on clinicopathological analysis, the expression level of LINC00520 has significant down-regulation in the patients with a tumor size of less than 5 cm and the patients in stage II of BC.

### Declarations

1. **Ethics approval and consent to participate:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Ethics Committee of Isfahan University of Medical Sciences (approval number: 3838988).

2. **Consent for publication:** Not applicable.

3. **Availability of data and materials:** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

4. **Competing interests:** The authors declare that they have no competing interests.

5. **Funding:** Not applicable.
6. **Authors' contributions:** EE, SM, AA, and SH performed the experiments, data analysis, and visualization. MA is the corresponding author and project manager and have revised the manuscript and validated the experiment. MR is the supervisor of this project and had a significant role in the conceptualization and revision of the manuscript. Ehsan Ezzati, Safiye Mosadeghi, Aliasghar Akbarinia, and Soroor Horriat are equally contributed in this project as first authors.

7. **Acknowledgment:** Not applicable.

**References**


**Figures**

**Figure 1**

Boxplot from the raw and normalized microarray samples, performing by *boxplot* command in R. A) Boxplot of the raw expression data showing the slight difference between the samples' median gene expression level. B) The boxplot of normalized expression data shows the same median for the expression level of genes in all samples. The same median of gene expression in every sample is acceptable evidence for the high quality of expression data.
Figure 2

PCA analysis from raw and normalized data for indicating the bad quality of samples in control and tumor group. A) PCA analysis of raw expression data. B) PCA analysis of normalized expression data revealed that there are five bad-quality samples in this microarray dataset.
Figure 3

The Pearson correlation analysis heatmap of normalized expression data of GSE42568.
**Figure 4**

The PCA analysis, after deleting 5 bad quality samples indicating the separation of control samples from tumor samples.

**Figure 5**

Volcano plot of the 54675 genes in the GSE42568. The down-regulation of \textit{ESR2} is indicated by a black point in the plot.
Figure 6
The signaling pathways in which ESR2 is involved, based on KEGG and Reactome analyses.

Figure 7
The different biological processes and cellular components that ESR2 protein located in, enables and involved in.

Figure 8
ESR2-related competitive endogenous RNA (ceRNA) network indicates the competitive endogenous long non-coding RNA (IncRNAs) and related microRNAs (miRNA). Based on miRNA-mRNA interaction analysis, hsa-miR-106a-3p has a strong RNA interaction with ESR2 mRNA and could regulate its expression level in the cell. Furthermore, based on NPInter online database, LINC00520 has the most significant interaction with hsa-miR-106a-3p. So, ESR2 and LINC00520 are two ceRNA regulated by hsa-miR-106a-3p). Dysregulation of LINC00520 indirectly affects the expression of ESR2 by increasing the binding affinity to miR-106a-3p. The yellow color indicates miR106a-3p as the miRNA with the strongest interaction with ESR2. The red color indicates other miRNAs interacted with ESR2. The blue color indicates IncRNA. The green color indicates mRNA (ESR2), the ceRNA network’s central node.

Figure 9
The local interactions of IncRNA LINC00520 and mRNA ESR2. The query RNA and the target RNA are shown in this picture as blue and red lines, respectively, while the expected interactions between the two RNAs are shown as gray or black lines. Color consistency indicates the intensity of interactions.
Figure 10

The local interactions of *LINC00520* and *ESR2*, based on the linear method of VARNA algorithm. The backbone is drawn in a straight line in the "linear algorithm" form, and arcs are used to link pairs of bases. (Base-pairing energy: A [-18.94 kcal/mol], B [-16.88 kcal/mol], C [-16.61 kcal/mol], D [-16.36 kcal/mol].)

Figure 11

Protein-protein interaction analysis of ESR2 indicates the hub proteins correlated to this gene, functionally or structurally.
Figure 12

Relative expression analysis of the *ESR2* and IncRNA *LINC00520* revealed that ESR2 and LINC00520 have a significant low-expression in the breast cancer samples, compared to control.
Figure 13

Relative expression analysis of ESR2 (A) and LINC00520 (B) revealed that the expression level of these two RNAs has a significant low-expression in most of BC cases.

\[ r: 0.6874 \]
\[ P. \text{ Value: 0.0008} \]

Figure 14

Pearson correlation analysis of ESR2 and LINC00520 revealed a significant positive correlation between the expression of these two RNAs.
Figure 15

ROC analysis of the ESR2 and LINC00520 expression level revealed that these two coding and non-coding RNAs can be the two diagnostic biomarkers of breast cancer.

![ROC curve: ROC of ESR2](image)

AUC: 0.9775
Std. Error: 0.01777
95% confidence interval: 0.9427 to 1.000
P value: < 0.0001

![ROC of lncRNA LINC00520](image)

AUC: 0.8450
Std. Error: 0.06450
95% confidence interval: 0.7186 to 0.9714
P value: 0.0002

Figure 16

Clinicopathological analysis of *LINC00520* expression level. IncRNA LINC0520 has a significantly low expression in the samples with the tumor size under 5cm and in the samples in stage II of BC compared to control samples.