The Role of Drug Resistance Causing IncRNAs in Breast Cancer: In-silico Analysis

Amirhossein Hajaliassgary Najafabadi
University of Isfahan

Mahdieh Khojasteh
University of Isfahan

Kamran Ghaedi (kamranghaedi@sci.ui.ac.ir)
University of Isfahan

Research Article

**Keywords:** IncRNA, Drug resistance, IncRNA connection network, Breast cancer, Genomics, Bioinformatics, In-silico analysis

**Posted Date:** December 29th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1164574/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](https://creativecommons.org/licenses/by/4.0/)
Abstract

**Background:** Breast cancer is the most common cancer in women globally. LncRNAs are non-coding RNAs that play an essential role in biological pathways. Many IncRNAs have been discovered to influence cancer medication resistance. As a result, identifying how IncRNAs may cause drug resistance is vital.

**Method:** Breast cancer TCGA RNA-seq data was applied in this study. We used the PharmacoGX package to explore IncRNAs with drug resistance or sensitivity effect through GDSC and CCLE data. Differential gene expression analysis (DGE) was used to find dysregulated IncRNAs (P<0.01). Survival analysis was performed to identify IncRNAs associated with patient survival, and a model based on them was developed. Multivariate cox regression analysis and ROC curve analysis were applied to assess the model. The TCGA-BRCA and two independent datasets (GSE21653 and GSE20685) were used to study the relevance of IncRNAs in biological pathways. IncRNA-miRNA-mRNA interaction network was investigated. The connections of IncRNAs with MRPs were analysed through the correlation test. Finally, IncRNA and MRP mRNAs attachment sites were analysed through the LncRRIsearch tool.

**Result:** According to our data, thirty-eight IncRNAs were associated with cell drug response in breast cancer cells. **IL12A-AS1**, **AC137723.1**, **LINC00667**, **SVID-AS1**, **CYTOR**, and **MIR4435-2HG** linked to patient survival (P<0.05). **AC137723.1** and **LINC00667** were identified as good prognostic genes, while the others were discovered to have poor prognostic effects. Moreover, the risk score model separated patients perfectly, in which about 45% of high-risk patients were dead; by contrast, around 95% of low-risk patients could survive. ROC curve results proved that CYTOR, MIR4435-2HG, and LINC00667 are potential biomarkers in breast cancer with AUC >0.8. Pathway analysis revealed that CYTOR and MIR4435-2HG are highly correlated with the Epithelial-Mesenchymal transition pathway, while AC137723.1 and LINC00667 were negatively correlated with the pathway. **AC128688.2**, **CYTOR**, **TDRKH-AS1** and LINC00667 can participate in IncRNA-miRNA-mRNA networks. Also, **MIR22HG** might influence drug resistance by attaching to MRP mRNAs.

**Conclusion:** Our findings revealed 38 IncRNAs involved in cancer cell treatment resistance and sensitivity. They can participate in patients’ prognosis, diagnosis and cellular pathways. Also, they may influence cell drug response through connections with CSPs, IncRNA-miRNA-mRNA networks and MRPs.

Introduction

Aside from nonmelanoma skin cancer, breast cancer is the most common cancer in women around the world. Over 250 000 new breast cancer cases were discovered in the United States in 2017, and 12% of all women in the country would be diagnosed with the disease at some time in their lives. [1] Despite the fact that most patients respond well to treatment at first, they eventually develop more aggressive tumour forms that are resistant to chemotherapy and radiation. As a result, metastasised recurrence with a poor prognosis is still a major worry in breast cancer treatment. Currently, breast cancer is treated with a mix of surgeries and adjuvant therapies such as chemotherapy, radiation, hormone therapy, targeted therapy, or a combination of these [2, 3].

Several proteins found on the membranes of cancer cells in breast cancer can promote resistance to several anticancer medicines. Chemotherapeutic multidrug drug resistance (MDR) mechanisms have been revealed for HER2 proteolysis, Mucin-4 overexpression, and PTEN phosphatase loss. [4]. The ATP-binding cassette (ABC) transporter family is also one of the most well-known proteins associated with MDR. ABC transporters are a large protein family involved in receiving and exporting a wide variety of substrates. Evidence has accumulated that ABC proteins, such as **ABCB1**, **ABCC1**, and **ABCG2**, play a role in the MDR phenotype of breast cancer cells. [5, 6].

LncRNAs are a type of non-coding RNAs that have evolved over time and have a length of roughly 200 nucleotides. LncRNAs play a vital role in a range of biological pathways and cellular functions, including embryogenesis, cell proliferation, and tumorigenesis, by modifying gene expression at the transcriptional, post-transcriptional, and translational stages. These RNAs interact with mature RNAs or proteins in the cytosol and play a role in epigenetic alterations and chromatin remodelling, transcriptional control, and RNA processing in the nucleus. In the last ten years, numerous studies have shown that IncRNAs have played critical roles as oncogenes or tumour suppressors. [7–9]. Multiple IncRNAs, including HOTAIR, MALAT1, GASS, PC3, and H19, have been linked to tumorigenesis, migration, and prognosis in cancers [10–12]. Through the STAT3/INOS80 pathway, the IncRNA **PTCSC3** influences drug resistance in anaplastic thyroid carcinoma. Furthermore, via the **mir-143/PLK1/BUBR1** axis, IncRNA **CCAT1** is a biomarker for esophageal cancer proliferation and therapeutic resistance. In addition, the **mir-134-5p/MBTD1** axis promotes osteosarcoma cell apoptosis and treatment resistance via the IncRNA **TTN-AS1** [13–15]. Therefore, it is revealed that IncRNAs are able to participate in MDR, tumorigenesis, and cancer prognosis. In this study, our purpose was to determine which IncRNAs are essential in drug resistance and then investigate their impact on prognosis by providing a risk score model based on their expression levels. Also, we aimed to identify potential biomarkers and discover how they impact drug resistance in breast cancer.

Material And Method

**Data Processing and Preparation**

The main data in this study was TCGA RNA-seq data and its linked clinical data in BRCA cases, which were obtained using the TCGA-Biobricks R package. The data contained 1222 samples, in which 113 were normal, and 1109 were patients. The data were normalised using the edgeR package through the CPM-based (At least 70% of the samples had a minimum CPM of 10) TMM approach. Furthermore, the data was converted to a log2 scale and applied in the study.

**Drug-resistance and sensitivity analysis:**

CCEL and GDSC data were used to explore the role of each IncRNA expression in drug resistance. The GRAY dataset in the PharmacoGX package was used to explore the correlation between drugs and genes in breast cancer using R. The FDA-approved drugs were checked through U.S. Food & Drug Administration website (https://www.accessdata.fda.gov/scripts/cder/daf/). The IncRNAs with a significant association with the FDA-approved drugs (FDR<0.01) were chosen for the study. The study then focused on IncRNAs, with all other genes being excluded.
**Investigation of Differentially Expressed Genes**

Applying both clinical and expression data, all samples were divided into normal and patient groups, and a differential gene expression analysis (DGE) was executed on them using the limma package in R. All the genes except lncRNAs were excluded from DGE outcomes. Then differentially expressed lncRNAs with significant drug resistance or sensitivity effect were selected. The adj. P.val<0.01 was considered the threshold for selecting lncRNAs linked to drugs. These lncRNAs were known as our candidate genes in the study.

**Survival Analysis**

First, all normal samples were excluded from data, and then patients were divided into dead and alive groups. Patients with unknown tumours and tumour-free conditions were excluded from the dead group (896 samples remained; 60 dead and 836 alive). Afterwards, univariate cox proportional hazard regression was applied to discover lncRNAs that impact patients' prognosis. Only candidate lncRNAs with LogRank<0.05 were selected. The identified candidate lncRNAs were applied to calculate a risk score for each patient using the following formula:

\[
\text{Riskscore} = \sum_{j=1}^{n} W_j \times \exp(j)
\]

Where \(W_j\) is the univariate coefficient for lncRNA \(j\), \(\exp(j)\) is the expression value of lncRNA \(j\) in patient \(i\), and \(n\) is the number of testing lncRNAs.

**Evaluation of Risk Score Model**

Clinical information, including cancer receptor condition (HER2, ER, and PR), T-stage, and age, were added to each patient, and a multivariate cox proportional hazard regression analysis was performed on prepared data through survival and survminer packages. The survival plot was drawn to illustrate the significance of the computed model based on lncRNAs. Also, the ROC curve analysis was performed on our lncRNAs to investigate the potential biomarkers among them.

**Gene Set Enrichment Analysis**

Spearman correlation test was executed between each of 6 lncRNAs in the risk score model and all other genes. The genes with a correlation coefficient >0.5 were utilised to perform pathway analysis through the EnrichR database (https://maayanlab.cloud/Enrichr/). The pathways with the highest q-value were selected. Afterwards, the expression matrix of two independent GEO datasets (GSE21653 and GSE20685) were downloaded, and the same test was run on them. Then, the genes of the same significant pathway among all datasets were selected, and a correlation test was executed among them and our lncRNAs.

**LncRNA-miRNA-mRNA Interaction Network Analysis**

Primary, all deregulated miRNAs in breast cancer were downloaded through the miRCancer database (http://mir cancer.ecu.edu/). The data contained the type of deregulation of miRNAs (up or down-regulated) and the origin article. All miRNAs were divided into upregulated and downregulated groups as well as candidate lncRNAs. Afterwards, the lncRNA-miRNA interactions for each lncRNA group and the opposite miRNA group in breast, mammary gland, and breast cancer tissues were investigated through the LncBase tool in the Diana database. Detected miRNAs were checked with mirCancer data, and only those were selected that were similarly dysregulated with the mirCancer list. Then, the miRNAs targets were searched through the mirWalk tool (http://mirwalk.umni-heidelberg.de) or corresponding article. We filtered the miRNAs with 95% CI that were linked to the three different databases in the miRWalk, and the identified miRNAs were checked with our differential expression data. We selected lncRNA-miRNA-mRNA interactions in two expression forms: up-down-up and down-up-down. The threshold for differentially expressed miRNA selection was adj. P.Val<0.01. Then, the significant pathways (q-value<0.05) related to the target miRNAs were discovered through the EnrichR database.

**LncRNA-mRNA Interaction Analysis**

Pearson correlation test between MRPs and candidate lncRNAs was performed. Two lncRNA clusters that had the highest correlation with the MRPs were selected for interaction analysis. The interactions were investigated through LncRRsearch [16], while the interaction energy threshold was set at 12 kcal/mol.

**Software and Statistics**

The Cox regression test was used to assess the connections between gene expression and survival rates, and the LogRank test was utilised to determine the significant level. The significance threshold for other parameters in this investigation was FDR<0.05, whereas adj. P.val<0.01 was used to identify significant lncRNAs through DGE analysis. R (v 4.0.4) and Prism GraphPad 9 were used for all of the analyses and illustrations. In addition, Cytoscape illustrated the relationship between medicines and candidate CSPs (v 3.7). The heatmaps were drawn using the pheatmap library in R. All the correlation tests were performed by the Pearson method. miRNA selection through the miRWalk database was executed by a 95% confidence interval.

**Results**

**Identification of 38 lncRNAs related to drug resistance and sensitivity of patients**

In order to evaluate the impact of lncRNAs on breast cancer cell drug response, TCGA-BRCA RNA-seq data and its clinical data were downloaded from the TCGA portal. Mentioned data contained 113 normal samples and 1109 cancer patients. After normalisation, only 26768 genes were considered to be probably important for further analysis. Furthermore, drug resistance analysis revealed about 300 significant (FDR<0.01) and FDA-approved lncRNA-drug connections (Table S1). Then DGE analysis was performed to detect lncRNAs that were significantly (adj. P.val<0.01) over or down expressed in cancer patients. As a result, we found 38 dysregulated lncRNAs linked to cell drug response, and we named them candidate genes (Fig. 1, Table S2). The outcomes revealed that
Docetaxel and Cetuximab are resistant to a vast number of IncRNAs, while Gefitinib, Epirubicin, Trametinib, Tamoxifen, and Gemcitabine are sensitive to a large number of IncRNAs.

**A model consisting of 6 IncRNAs can predict patients’ survival**

We utilised univariate cox proportional hazard regression to evaluate our candidate genes in terms of prognosis. Results demonstrated that only six of our candidate genes were related significantly to patients’ prognosis (LogRank<0.05). Also, among identified genes, two were correlated to patients’ good prognostic, while three IncRNAs were linked to the poor survival rate of patients (Table 1). Afterwards, a risk score model was computed through the formula below:

\[
Riskscore = 0.230 \times \exp_{\text{ENSG00000244040}} - 0.450 \times \exp_{\text{ENSG00000264569}} - 0.254 \times \exp_{\text{ENSG00000263753}} + 0.257 \times \exp_{\text{ENSG00000224597}} + 0.375
\]

Then we calculated risk scores for each patient to divide them into two high and low-risk groups. Through Kaplan-Meier analysis, we realised that the first quartile is the best threshold to group patients(Fig. 2a). Our model demonstrated that only about 5% of low-risk patients died, while the survival rate for the high-risk group was around 57% (Fig. 2b). These results reveal that our model correctly distinguishes the patients into high or low-risk groups.

**CYTOR, LINC00667, and MIR4435-2HG are potential biomarkers for breast cancer**

We utilised multivariate cox regression analysis to indicate that the computed model is independent of other clinical parameters. Therefore, clinical factors including T-stage, Age, and Hormone receptor status were added to the data that included patients’ risk scores. Multivariate results showed that our risk score model, T-stage and age, are independently related to patients’ poor survival, while hormone receptors did not significantly link to patients’ survival (Table 2). Thus, these results reveal the accuracy of our prognostic index. In addition, a ROC curve analysis was performed to evaluate our selected IncRNAs in terms of their potential in patients’ diagnoses (Fig. 3). The outcomes revealed that the expressions of all our model IncRNAs were significantly different among normal and patients (P<0.0001). Moreover, LINC00667, MIR4435-2HG and CYTOR indicated significant areas under the curve (>0.8). Therefore, it was indicated that our selected IncRNAs have a substantial role in the patients’ diagnosis and prognosis.

**Table 1**  
Risk Score Model IncRNAs

<table>
<thead>
<tr>
<th>IncRNA</th>
<th>HR</th>
<th>95%CI</th>
<th>P.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL12A-AS1</td>
<td>1.26</td>
<td>1.10-1.43</td>
<td>0.0051</td>
</tr>
<tr>
<td>AC137723.1</td>
<td>0.64</td>
<td>0.43-0.94</td>
<td>0.0115</td>
</tr>
<tr>
<td>LINC00667</td>
<td>0.77</td>
<td>0.62-0.96</td>
<td>0.0278</td>
</tr>
<tr>
<td>SVIL-AS1</td>
<td>1.29</td>
<td>1.05-1.59</td>
<td>0.0195</td>
</tr>
<tr>
<td>CYTOR</td>
<td>1.45</td>
<td>1.12-1.88</td>
<td>0.0049</td>
</tr>
<tr>
<td>MIR4435-2HG</td>
<td>1.64</td>
<td>1.27-2.11</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Table 2**  
Survival Analysis Results

<table>
<thead>
<tr>
<th>Clinical Parameter</th>
<th>Univariate CoxPH</th>
<th>P.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk score</td>
<td>1.92 (1.23-3.01)</td>
<td>0.0040 **</td>
</tr>
<tr>
<td>Age (60+ /60-)</td>
<td>1.04 (1.01-1.08)</td>
<td>0.0053 **</td>
</tr>
<tr>
<td>ER(Positive/Negative)</td>
<td>0.60 (0.20-1.76)</td>
<td>0.3560</td>
</tr>
<tr>
<td>HR(Positive/Negative)</td>
<td>0.65 (0.24-1.74)</td>
<td>0.3945</td>
</tr>
<tr>
<td>PR(Positive/Negative)</td>
<td>0.55 (0.19-1.62)</td>
<td>0.2774</td>
</tr>
<tr>
<td>T-stage (1,2,3,4)</td>
<td>3.76 (1.97-7.20)</td>
<td>&lt;0.0001 ****</td>
</tr>
</tbody>
</table>
The first cluster was positively correlated in which lncRNAs and MDR proteins. The results are shown in the illustrated heatmap (Figure 6A). We found two clusters highly correlated with known as lncRNAs might interact with MDR protein mRNAs, the main proteins in multidrug resistance in cancers. Firstly, we chose the To clarify the role of candidate lncRNAs in cell drug response differently, we analysed lncRNA-mRNA direct interactions. We believed that our candidate MIR22HG mRNAs, respectively.

downregulation of TDRKH-AS1 might upregulate Angiogenesis and TGF-beta Signaling pathways by protecting MIR22HG, CYTOR, MIR4435-2HG, and LINC00667 can inhibit hsa-miR-125b-5p, hsa-miR-1301-3p, hsa-miR-16-5p, hsa-miR-19a-3p, and hsa-miR-92a-3p from targeting different mRNAs

Another way to discover the role of IncRNAs in cellular pathways and cell drug response is to discover their interactions with miRNAs. We hypothesised that overexpressed IncRNAs are able to upregulate mRNAs indirectly by attaching to the miRNAs that target them. Therefore, upregulation of these IncRNAs causes upregulation to their corresponding mRNA and vice versa. The IncRNA-miRNA-mRNA interactions are illustrated in Figure 5. Our search in LncBase (Diana tool) for overexpressed candidate IncRNAs revealed a number of IncRNA-miRNA interactions in which the interactions of AC128688.2 and hsa-miR-125b-5p, CYTOR and hsa-miR-1301-3p, and TDRKH-AS1 and two miRNA hsa-miR-16-5p and hsa-miR-19a-3p passed mirCancer data double-check. On the other hand, the same method selected LINC00667 connection to hsa-miR-92a-3p in the downregulated IncRNA group. The mRNA connections were checked through miRWalk and differential expression data (adj. P.val<0.01) for each group. We revealed that the AC128688.2 connection might cause upregulation in the G2-M Checkpoint pathway by protecting SLC7A1 and SUV39H1 from degradation. While CYTOR connection was able to affect no pathways through miRNAs, TDRKH-AS1 might upregulate Angiogenesis and TGF-beta Signaling pathways by protecting VEGFA and TGF1 from degradation, respectively. However, downregulation of LINC00667 may indirectly decrease UV Response Dn and Protein Secretion pathways by decreasing the amount of NFIB and ATP7A mRNAs, respectively.

MIR22HG can attach to MDR protein mRNAs

To clarify the role of candidate IncRNAs in cell drug response differently, we analysed IncRNA-mRNA direct interactions. We believed that our candidate IncRNAs might interact with MDR protein mRNAs, the main proteins in multidrug resistance in cancers. Firstly, we chose the ABC family and ABCG2 (also known as BCRP, Breast Cancer Resistance Protein) as the MDR proteins in this research. Then, we executed a Pearson correlation test between candidate IncRNAs and MDR proteins. The results are shown in the illustrated heatmap (Figure 6A). We found two clusters highly correlated with ABCG2 and ABCC9. The first cluster was positively correlated in which MIR22HG was present. The second cluster, which included MIR4435-2HG and CYTOR, was negatively correlated with mentioned MDRs. We searched each cluster member in the LncRisearch database to investigate their significant interactions with ABCC9 and ABCG2 (12kcal/mol was set as interaction threshold). Finally, we discovered MIR22HG interaction with both of the MDRs. The interaction location and bases are shown in Figure 6B.

Discussion
LncRNAs have been linked to several stages of cancer development, according to the latest research. These IncRNAs interact with DNA, RNA, and cellular proteins, functioning as a key regulator in chromatin organisation, transcriptional control, and post-transcriptional regulation [17, 18]. Their role in several cancers, such as colorectal cancer (CRC) and breast cancer, has been revealed [19, 20]. For instance, risen expression of IncRNA DLEU1 in human CRC tissue causes chromatin remodelling through SMARCA1, which finally increases the amount of KPNAs and promotes cancer malignancy and poor survival of patients [21]. In addition, IncRNA SNHG20 is found in many cancers, including hepatocellular carcinoma, ovarian cancer, colorectal cancer, and bladder cancer, and plays a role in cancer genesis and progression through expression alterations in cancer cells [22]. Obviously, in most studies, the main focus was on IncRNAs interactions with other molecules and pathways; however, we tried to look at this issue from another perspective. Therefore, in this study, we aimed to identify poor prognostic IncRNAs linked to cell drug response with no focusing on different proteins and pathways to evaluate their role but introducing a model that can be applied to diagnose breast cancer through IncRNAs and effective drugs for treatment.

Our findings showed that seven drugs are significantly related to our candidate IncRNAs. Among them, a vast number of our candidate genes may cause resistance to Docetaxel. Breast cancer resistance to Docetaxel through IncRNA EPB41L4A-AS2 has already been shown [23]. Moreover, the role of IncRNAs in Docetaxel resistance has been indicated in prostate cancer. NEAT1 increases ACSL4 in PCa cells via inhibiting two miRNAs, which leads to Docetaxel resistance [24]. Our results are consistent with other studies and show that a number of IncRNAs cause resistance to this drug in breast cancer. Moreover, we have shown that Cetuximab is connected to three IncRNAs which reduce the sensitivity of cancer cells to this drug. Recent studies also suggest that different IncRNAs can contribute to Cetuximab resistance in colorectal cancer [25–27]. Among all drugs, Gemcitabine demonstrated the most connections to IncRNAs which cause sensitivity to this drug. Other studies also show that in different cancer types, resistance and sensitivity to Gemcitabine are under the impact of different IncRNAs [28–33].

There are other studies that have shown IncRNA signatures, which have the potential of being used as prognostic markers in breast cancer [34]. In addition, there are a vast number of studies that have introduced IncRNA models for gastric cancer, lung adenocarcinoma, Esophageal cancer, ovarian and cervical cancers [35–39]. Our model includes six IncRNAs, two good prognostic and four poor prognostics. The model could considerably separate high and low-risk patients based on their risk score and fatality rate. One of our poor prognostic IncRNAs is CYTOR (cytoskeleton regulator RNA), which has been suggested as biomarkers potential breast cancer [40]. We also confirm its capability for being a potential biomarker through our model. Moreover, its interaction with Sam68 and NCL can activate the NF-xb pathway in colorectal cancer and promotes cancer progression [41]. Other studies revealed that its interaction with β-catenin can cause metastasis in colon cancer [42]. Also, we suggest MIR4435-2HG as a prognostic and diagnostic marker for breast cancer both through ROC curve results and our risk score model. Other studies agree with the role of MIR4435-2HG in numerous cancers. It has been indicated that it can communicate with other IncRNAs to facilitate gastric cancer progression [43]. Furthermore, it might promote lung cancer by inducing the β-catenin signalling pathway [44]. IL12A-AST1 is one of the novel IncRNAs in our studies which there is little information about its expression and function in cancers. We realised its role in breast cancer prognosis as well as another study that suggests this IncRNA as a diagnostic marker in head and neck squamous cell carcinoma [45]. Additionally, SVIL-AST1 is known as a poor prognostic marker in our model with acceptable ROC curve results. It has been shown that the downregulation of this RNA might participate in abnormal DNA methylation in lung carcinoma [46]. In addition, this study is the first one that shows LINC00667 good prognosis impact on breast cancer. On the other hand, it has been shown that this IncRNA contains multiple miRNA binding sites that can promote gliomas malignancy through VM-associated proteins as well as its role in the lung, colorectal and esophageal cancers [47–50]. However, we have shown that our introduced model based on six IncRNAs, including LINC00667, is associated with poor prognosis independently. Therefore, we suggest that the whole model can ideally detect high and low-risk patients. The other good prognostic gene in our model is AC137723.1, a mysterious IncRNA with an unknown function. However, it is suggested that this IncRNA has an impact on prognosis and invasion in Clear cell renal cell carcinoma [51].

Through pathway analysis, we found out that CYTOR and MIR4435-2HG are highly correlated. Both of the genes may have a substantial impact on the cell transition to mesenchymal from the epithelial state. A number of cell surface proteins (CSPs) can be seen among the genes with a high correlation with both of these genes. In our recent study, we have revealed the important role of CSPs in drug resistance to colorectal cancer (CRC). In addition, among the CSPs, many collagen family members can be seen that are very important in cancer development. Their role in resistance to Mirin and other chemotherapy drugs have been shown. We also have indicated the impact of TIMP1, BGN, and THBS2 on the poor survival of CRC patients, which are among the genes with a high correlation to the mentioned IncRNAs [52].

It is a topic of much thought that IncRNAs can contribute to gene expression and pathway regulation through interactions with miRNAs. miRNAs can cause degradation to different mRNAs and contribute to gene regulation. Some research has been conducted to discover IncRNA-miRNA-mRNA networks in colorectal and oral cancers [19, 53]. We discovered that TDRKHAST1 relation to has-miR-16-5p and has-miR-19a-3p can cause overexpression to a vast number of genes. We revealed that TDRKHAST1 was upregulated in breast cancer patients, and also, both of its related miRNAs are downregulated based on miCancer data, which led to their targets overexpression. Moreover, we detected that this IncRNA might participate in Angiogenesis and TGF-beta Signaling pathways dysregulation through VEGFA and TGFβ mRNAs, respectively. Other research has indicated that VEGFA overexpression in ovarian cancer can lead to chemotherapy resistance and is mediated by autophagy [54]. Therefore, we believe that TDRKHAST1 can indirectly cause drug resistance and sensitivity by mediating VEGFA gene expression indirectly.

In order to investigate the relationship between candidate IncRNAs and MDR proteins to evaluate their role in drug resistance, we performed a Pearson correlation test. The results revealed two clusters correlated to ABC9 and ABCG2. CYTOR and MIR4435-2HG showed a negative correlation with the MDRs. We think this negative correlation is because of the downregulation of ABCG2 and ABCCC9 and the overexpression of the IncRNAs in breast cancer. On the other hand, MIR22HG expression is raised in several human cancers (colorectal cancer, gastric cancer, hepatocellular carcinoma, lung cancer, and thyroid carcinoma), while it is considerably lowered in others (esophageal adenocarcinoma and glioblastoma). MIR22HG has also been shown to behave as a competitive endogenous RNA (ceRNA), participate in signalling cascades, interact with proteins, and interact with miRNAs as a host gene in carcinogenesis and tumour growth [55]. Moreover, we found out that this gene might interact with ABCC9 and ABCG2 mRNAs. However, we cannot exactly explain this interaction role; the positive correlation between them suggests that MIR22HG might stabilise the mRNA.
Overall, the findings through our study and the others suggest the vital role of lncRNAs in human cancer. While more studies are needed to investigate the exact role of our model lncRNAs in human breast tissue and other cancers, it can not be denied that they can obviously predict patients' survival rates. Therefore, our results also agree with the other studies about the role of lncRNAs in cancers drug resistance, prognosis and diagnosis, and we suggest more studies to be performed to determine their impact on cancer cells, whether in prognosis or diagnosis.

Conclusion

This study aimed to determine lncRNAs related to the drug resistance feature of cancer cells and investigate them in terms of biomarker capability and prognosis effect. Our findings showed that there are 38 lncRNAs important in drug resistance and sensitivity of cancer cells. Also, we introduced a model consisting of IL12A-AS1, AC137723.1, LINC00667, SVIL-AS1, CYTOR, and MIR4435-2HG that can perfectly separate patients with low and high survival risk. In addition, CYTOR, MIR4435-2HG, and LINC00667 are introduced as potential biomarkers in breast cancer. We revealed that the lncRNA-mRNA-miRNA network and IncRNA-MDR protein correlation and relation might contribute to breast cancer drug resistance.

Abbreviations

TCGA; The Cancer Genome Atlas, MDR; Multidrug Drug Resistance, CRC; Colorectal Cancer, CSP; Cell Surface Protein, GEO; Gene Expression Omnibus, DGE; Differential Gene Expression, GDSC; Genomics of Drug Sensitivity in Cancer, CCLE; The Cancer Cell Line Encyclopedia, ROC; Receiver Operating Characteristic, FDA; Food and Drug Administration, ER; Estrogen Receptor, PR; Progesterone Receptor, MRP; Multidrug Resistance Protein

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in The Cancer Genome Atlas (https://portal.gdc.cancer.gov/projects/TCGA-BRCA), and Gene Expression Omnibus (GSE21653 and GSE20685).

Competing interest

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors’ contributions

AHN and MK analysed the data in use in this study and contributed to drawing the figures. AHN wrote the major programming codes for the analysis. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References


**Figures**

**Figure 1**

*Connections between drugs and dysregulated lncRNAs.* All lncRNA-drug connections are depicted in which the numbers written on edges are Log fold changes. The red edges are resistance connections, and the green ones are sensitivity connections. The thickness of the edges shows the Estimate level. All drugs are FDA-approved, and FDR<0.01 is selected as the threshold level for connections.
Figure 2

**Six drug connected lncRNAs identified as poor prognostic genes.** (A) The risk score figures calculated based on the prognostic model can perfectly divide patients into low and high-risk groups, using the first quartile as the cut-off value. (B) The Kaplan-Meier graph shows that the fatality rate is about 5% in the low-risk group, while more than 40% in the high-risk group. The heatmap illustrates that our good prognostic lncRNAs lower when risk score increases, while poor prognostic lncRNAs expression grows with risk score increase.

Figure 3

**LncRNA biomarker discovery in breast cancer.** ROC curve analysis revealed that CYTOR and MIR4435-2HG are excellent potential biomarkers for breast cancer diagnosis, while LINC00667 is an outstanding one. (AUC; Area Under the Curve, $P<0.0001$).

Figure 4

**The role and function of identified lncRNAs in biological pathways.** (A) The correlation test on the TCGA-BRCA dataset between model lncRNAs and remaining genes identified that CYTOR and MIR4435-2HG are highly correlated with the Epithelial-Mesenchymal transition pathway. Other genes revealed no significant correlation to any pathway. The heatmap shows the highest correlated genes in the pathway with CYTOR and MIR4435-2HG. It is evident that good prognostic lncRNAs (AC137723.1 and LINC00667) are negatively correlated with the pathway genes and are located in the same cluster. (B-C) The same results are obtained from other independent datasets (GSE20685 at middle and GSE2163 at right). The UV Response Dn and Myogenesis were both detected as significant pathways. The heatmaps also show the same results; however, one of the lncRNAs (AC137723.1) was not found in the independent datasets.

Figure 5

**Drug resistance lncRNA-miRNA-mRNA interaction network.** The interaction network is depicted for candidate (drug resistance causing) lncRNAs. The green colour means RNA downregulation, while the red colour shows RNA overexpression. The adj.$P$-val<0.01 was used as the significant level of RNA differential expression.

Figure 6

**The connection of lncRNAs to MRPs.** (A) The heatmap depicts the Pearson correlation results in which two clusters with the highest correlation to ABCC9 and ABCG2 are highlighted. (B) MIR22HG (ENST00000574306, 506-529) attachment site to ABCG2 (ENST00000237612, 293-317, 5'UTR) is shown on top of the figure (-14.20 kcal/mol). MIR22HG (ENST00000574306, 1467-1503) attachment site to ABCC9 (ENST00000261200, 4304-4343) is shown on the bottom of the figure (-13.16 kcal/mol).

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.txt
- TableS2.txt
- TS3.docx
- TS4.docx