Research Article   Open Access

Construction of a Cerna Network in Breast Cancer and Analysis Correlation with Immune Infiltration

Yongfang Xie*, Yani Zhao, Huang Xiaorong, Yu Zhang, Zhang Langlang and Shu Kunxian***

Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, 400005, Chongqing, China

*Corresponding author: Yongfang Xie, Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, 400005, Chongqing, China, Tel: 0086-23-62460064, E-mail: xiey168f@163.com
Shu Kunxian, Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, 400005, Chongqing, China, E-mail: Shukx@cqupt.edu.cn

Abstract

Background: Breast cancer(BC) is a complex and aggressive malignant tumor with various pathogenesis and high mortality. More and more evidences showed that endogenous RNA (ceRNA) plays an important role in regulating the occurrence and development of BC. However, there is still a lack of BC ceRNA regulatory networks that can be used in clinical research. The purpose of this article is to conduct targeted gene research, identify differentially expressed RNAs(DERNAs) and perform functional enrichment analysis, and provide new ideas for the treatment of BC.

Method: From the Gene Expression Omnibus (GEO) databases, four databases were filtered. Among them, GSE122063, GSE36980 and GSE11855 were employed for bioinformatics analysis. GEO2R was employed to identify DERNAs with the predefined criterion (P.value < 0.05 and | log FC| ≥ 2). The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed for differentially expressed mRNAs (DEmRNAs) via FunRich and DAVID. Moreover, we constructed a protein-protein interaction (PPI) network by STRING database and ceRNA regulatory network. Finally, AutoDock4.0 software was used to study the interaction between the target protein and the drug through molecular docking.

Results: Compared with the control group, 791 DERNAs were isolated from GSE5764 , including 67 DElncRNAs, 9 DEmiR-NAs and 715 DEmRNAs. Functional enrichment analysis indicated that DEmRNAs is mainly related to biological functions such as cytokine activity and protein binding, and can affect cell growth and communication. The ceRNA regulatory network includes 15 dysregulated lncRNAs, 2 miRNAs and 77 mRNAs. Drug target analysis showed strong docking effect of Tamoxifen, Lapatinib, Eribulin and Afimoxifene with COMP and PIGR, the potential sites for the treatment of BC. We built a risk model based on ceRNA regulatory network.

Conclusion: In this paper, we are currently focusing on drug target analysis, identity DERNAs and construct ceRNA regulatory network, which is helpful for the pathogenesis exploration and early prevention and treatment of breast cancer.

Keywords: Breast Cancer; ceRNA Regulatory Network; Molecular Docking; Immune Infiltration

Introduction

Breast cancer (BC) is a common and aggressive malignant tumor that originates from breast tissue, especially in women. Every year, nearly 1.3 million people are diagnosed with BC, and 450,000 people died from BC in the world [1]. The incidence of BC is getting younger, while the morbidity and mortality rates remain high. Breast cancer is highly heterogeneous, with numerous subtypes, and clinical tests typically assess the expression of four proteins, including Ki67 (a protein associated with breast cancer proliferation), Estrogen Receptor (ER), Human epidermal growth factor receptor 2 (HER2), and Progesterone Receptor (PR). Depending on the expression, if cancer cells are negative for the receptor, this indicates that the receptor is not present in the tumor tissue, e.g. cells with PR receptor are PR-negative or PR-; conversely, if cancer cells are positive for the receptor, this indicates that this receptor is present in the tumor tissue, e.g. cells with PR receptor are PR-positive or PR+. According to these criteria, breast cancer was classified into four subtypes in clinical studies[2-4], as shown in Table 1. Similarly, breast cancer is highly metastatic, with 20-30% of breast cancer patients develop-

ing metastases after diagnosis and treatment, with sites including bone, lung, liver and brain[5,6]. Whereas metastasis leads to a decrease in overall patient survival, the 5-year overall survival rate is 22.8% for bone metastases and 16.8% for lung metastases, and brain metastases are the most risky factor among patients with metastases[7-9]. Currently, BC diagnosis and treatment methods include mammography, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound imaging, and other imaging methods[10-13].Trials have shown that mammography can reduce breast cancer mortality by 20%, and ultrasound can also detect cancer patients, but the positive predictive value of ultrasound is only 3-8%[14, 15]. MRI, although proven to have a high cancer detection rate in breast cancer imaging modalities, is neither applicable nor cost-effective in screening large numbers of women at high risk for breast cancer [16]. Therefore, defects in sensitivity and low specificity at the time of diagnosis can cause patients to miss the best time for treatment. In addition, the high metastasis and invasion of BC also put forward more stringent requirements for clinical treatment. Considering the complexity and the unclear pathogenesis of the clinical and pathological symptoms of BC, it is still necessary to conduct in-depth research so as to fully grasp the regulatory mechanism of BC.

Table 1: Four subtypes of breast cancer

subtype	Clinical Diagnostic Index				Relapse rate%
	ER	PR	HER2	Ki67	
Luminal A	+	+	-	-	30–70
Luminal B	+	+	-	+	10–20
HER2 overexpression	-	-	+	NA	5–15
Basal-like	-	-	-	NA	15–20

Table Note: + indicates that the receptor is present, - indicates that the receptor is not present, and NA (not applicable) indicates that it is not applicable

As a type of RNA that does not encode protein, NcRNA includes long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and circular RNAs (circRNAs) and so on [17]. More and more evidences have shown that non-coding RNA (ncRNA) has played an important role in many human tumors including breast cancer by involving cell development, proliferation, differentiation and apoptosis [18]. Firstly, proposed the competing endogenous RNA hypothesis, named ceRNA, suggesting that the presence of MRE can relieve the inhibition of miRNA to the target mRNA

by combining with miRNA. An increasing number of studies suggested that lncRNA can act as ceRNA or sponge to indirectly regulate mRNA expression by blocking miRNA [19]. In addition, ceRNA regulatory network analysis can be used for mining potential biomarkers, providing valuable information for exploring the mechanism of tumor occurrence and development. For example, LINC0092 and C2orf71 has been used as new prognostic indicators for estrogen receptor positive (ER+) and ER negative (ER-) breast cancer [20], RP11-434D9.1, LINC00052, BC016831

and IGKV are related to the occurrence of TNBC(triple negative breast cancer), which can be used as a potential biomarker for clinical diagnosis or treatment targets of TNBC [21]. Besides, lncRNA can be used as a new biomarker to determine the subtype of breast cancer [22]. These studies demonstrated that ceRNA network disorder is related to the occurrence of breast cancer, and the construction of ceRNA regulatory network is an effective way to study the occurrence of cancer. However, the current research on ceRNA regulatory networks related to breast cancer is scarce

In the current study, we aimed to conduct research on BC targeted genes. Through the GEO database, we identified differentially expressed RNA by bioinformatics methods and constructed lncRNA-miRNA-mRNA regulatory network. And through molecular docking technology to find the relevant target information that is helpful for early prevention and diagnosis of breast cancer provides theoretical support for new drug development, and also opens the way for precision treatment.

Materials and methods

The Cancer Genome Atlas Program

GEO(https://www.ncbi.nlm.nih.gov/geo/query/acc. cgi) is a public open gene expression profile database established and maintained by the National Center of Biotechnology Information (NCBI) in 2000, which can store high-throughput gene expression data sets [23]. Potential reasonable RNA-Seq databases were selected based on the following criterion:

(1) The research organism is homo sapiens;

(2) Specimens with RNA;

(3) Human BC and control Expression data from breast;

(4) Expression profiling by array;

(5) Supported by GEO2R analysis.

Use the keywords: (("Breast Cancer") AND ("RNA" OR "Expression profiling by array") AND ("Homo sapiens")) to find potential datasets. Then, the databases of GSE5764, GSE7904, GSE8977 and GSE10810 were selected, and the samples were divided into BC group and control group for the next study, respectively, according to the predefined datasets inclusion criteria. Gene expression data from breast cancer are downloaded from TCGA (The Cancer Genome Atlas Program, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) [24] with clinical data and screened and organized for analysis.

Identification of Differentially Expressed RNA

In order to have a better view of RNA with significant expression changes in different samples, we identified differentially expressed RNAs(DERNAs) through GEO2R, a useful online analysis tool that comes with the GEO database [25]. Namely, enter a GSE number into the GEO accession box. After all samples (GSMs) were assigned to groups we wanted to compare (BC vs. Normal), we performed the test with default parameters and clicked "Save all results" to obtain all the DERNAs. The threshold for DERNAs is P.Value<0.05 and |log FC|≥2. The Venn diagram in FunRich_v1.3.1 [26] was used to show the common DERNAs identified by GEO2R in the four datasets.

Bioinformatics Analysis of Differentially Expressed RNA

In order to obtain a comprehensive understanding of mechanism and progression in BC patients, we performed Gene Ontology including biological process, molecular function, cellular component and Kyoto Encyclopedia of Genes and Genomes (KEEG) analysis on DERNAs using FunRich_v1.3.1 and DAVID to predict their possible functions and signaling pathways [27].

We adopted online tool STRING[28] (https://string-db.org/) to develop protein-protein interaction (PPI) network of differentially expressed mRNAs (DEmRNAs). Briefly,PPI network was showed by inputing all gene symbol into the search box and selecting correct organism (homo sapiens). Cytoscape_v3.6.1 is used to visualize the network [29].

Construction of ceRNA Regulatory Network

The combination of MiRNA with lncRNA was predicted by databases starbase [30], miRNA targeting genes intersection common mRNA were obtained through online databases mirDIP [31]. Organized and sort out the corresponding data of lncRNA-miRNA-mRNA. The established ceRNA network were visualized with Cytoscape_v3.6.1.

Statistical Methods

The genes in the ceRNA network were individually subjected to monofactor analysis using the R programming language, and the mRNAs were screened again using least absolute shrinkage and selection operator (Lasso) analysis, followed by cox regression of the screened variables to obtain a risk model.

The Kaplan-Meier method was used for survival analysis, and the predictive performance of the risk model was assessed by the time-dependent receiver operating characteristic (ROC) of the "SurvivalROC" R package. In addition, the Mann-Whitney test was used to assess differences between sample groups, and the Kruskal-Wallis test and Dunn's multiple test were used for multiple comparisons to calculate the median, with P<0.05 being considered a statistically significant difference.

Molecular Docking

After downloading the protein structure corresponding to the 12 common DEmRNAs from RCSB database [32] (http://www.rcsb.org/pdb), we consulted the relevant literatures to get related treatment drugs of BC (Tamoxifen, Lapatinib, Eribulin and Afimoxifene) as ligands for molecular docking (table 2).

Both ligands and protein receptors were prepared using AutoDock_v4.0 [33]. Before the molecular docking, we pre-treated protein receptors (PDB ID: 3FBY, 1XED, 1O7Y) and drug molecules (hydrogenation, charging, etc.) and calculated the grid energy. the GirdBox grid region covered the studied target proteins was used as the center (the active site is unknown). In the docking process, we employed the genetic algorithm for regional search, the initial population was set to 150, the maximum number of energy assessments was 2.500000, and other parameters were selected by default. After the docking, binding energy (BE) was used as the criterion to select the optimal conformation. The greater energy indicated the stronger interaction between molecules. We selected this conformation as the optimal result of docking. Finally, the results was imported into PyMOL_v2.3.3 to visualize the conformation of the complex and analyze the interaction between protein and drug [34].

Table 2: Related properties of small drug molecules

Drug	Formula	Molecular weight (g·mol^{-1})	Number of Rotatable Keys
Tamoxifen	C26H29NO2	371.524	8/32
Lapatinib	C29H26ClFN4O4S	581.069	11/32
Eribulin	C40H59NO11	729.908	24/32
Afimoxifene	C26H29NO2	387.523	9/32

Results

Identification of Differentially Expressed RNA

Through searching and inclusion criterion, we found four databases of GSE5764, GSE7904, GSE8977 and GSE10810G from GEO (Table 3). All RNA-Seq data are from GPL570 platform. After processed by GEO2R, we obtained DERNAs with the criterion (P. Value<0.05 and |log FC|≥2) (Figure 2). Meanwhile, 12 DEmRNAs (SFRP1, PLPP4, PIGR, MT1M, MMP11, INHBA, EMP1, CXCL10, COMP, COL11A1, COL10A1 and C2orf40) identified by GEO2R were used for subsequent drug target analysis (Figure. 2).

Table 3: Differentially expressed RNAs for breast cancer

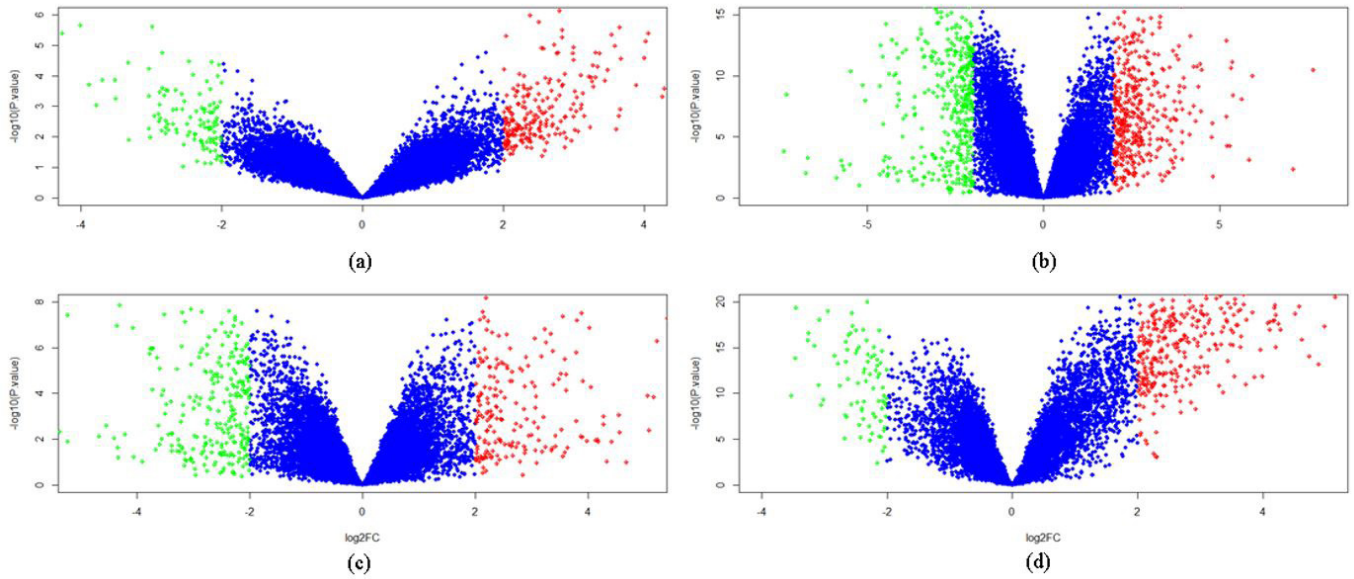| Database	Samples	DERNA	Expression	DEmiRNAs	DElncRNAs	DEmRNAs
GSE5764	30（10/20）	336	up	2	10	218
			down	0	3	103
GSE7904	62（43/19）	791	up	4	36	366
			down	5	31	349
GSE8977	22(7/15)	402	up	0	20	157
			down	2	11	212
GSE10810	58(31 /27)	391	up	4	15	287
			down	0	0	85

Figure1: The volcano plot of mRNA in GEO databases The order is GSE5764, GSE7904, GSE8977 and GSE10810. Green represents RNA down-regulation, red represents RNA up-regulation, and blue represent genes with no difference. The absolute value of log2FC indicate the level of these mRNAs
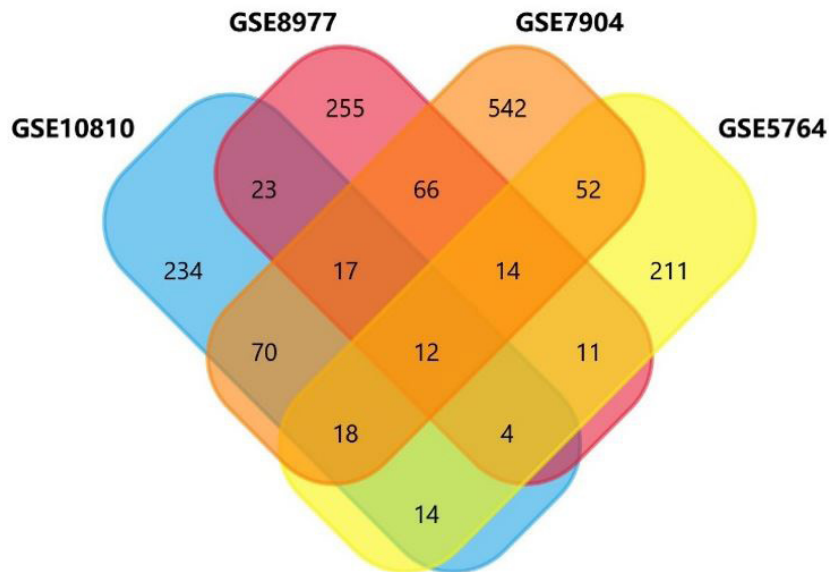

Figure 2: The 12 hub DEmRNAs across four datasets

Bioinformatics Analysis of Differentially Expressed Gene

The functional enrichment analysis of DEmRNAs was performed using DAVID and FunRich. DAVID enrichment factor with P.value<0.05 were thought to be significant. The top ten terms were visualized by R language. The results showed that DEmRNAs are related to cytokine activity and protein binding and other biological functions, and can affect cell growth and communication (Figure 3 and Figure4).

We also performed enrichment analysis on 12 common DEmRNAs with DAVID (Figure 5) and found that these DEmRNAs are related to the two pathways of protein digestion and absorption and ECM-receptor interaction, and are related to endoderm cell differentiation, negative regulation of cell growth, and biological functions such as the development of multicellular organisms, and can also predict drug response.
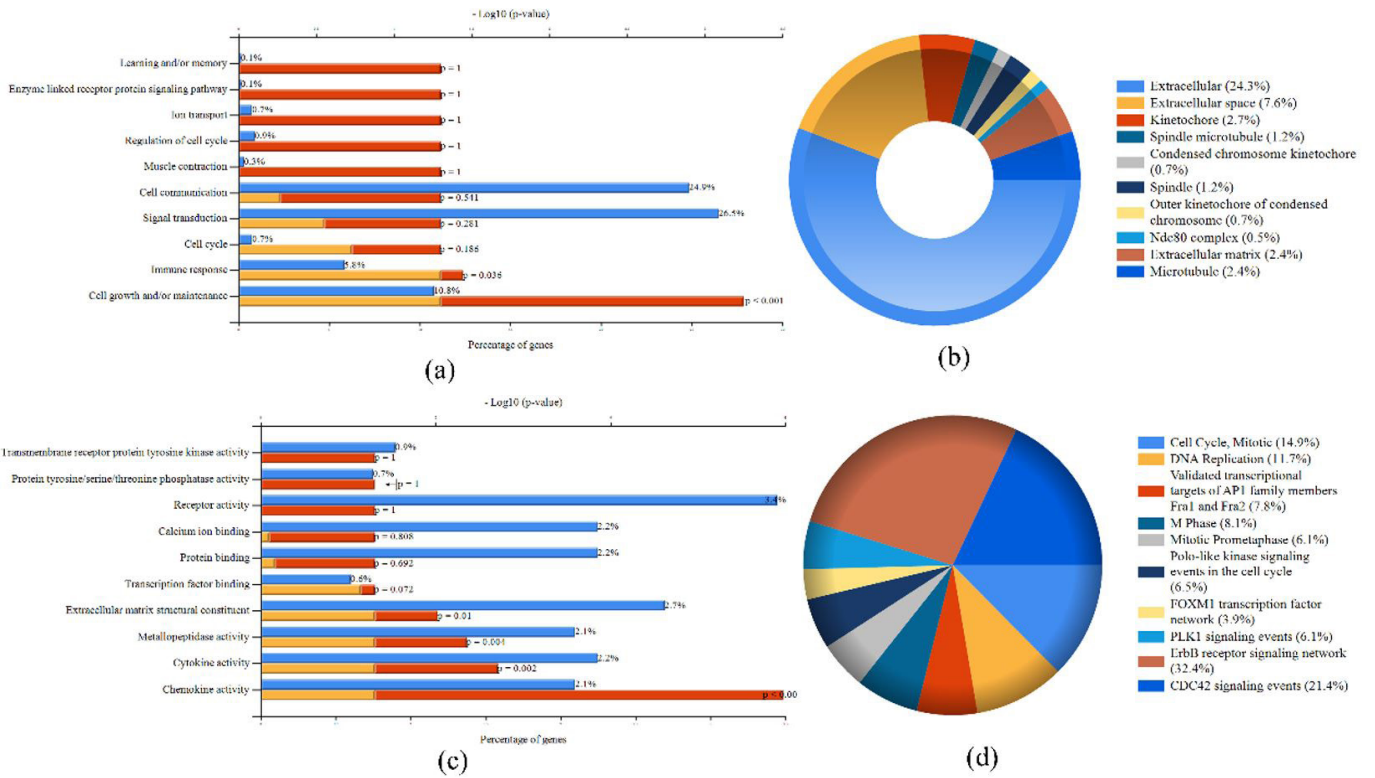
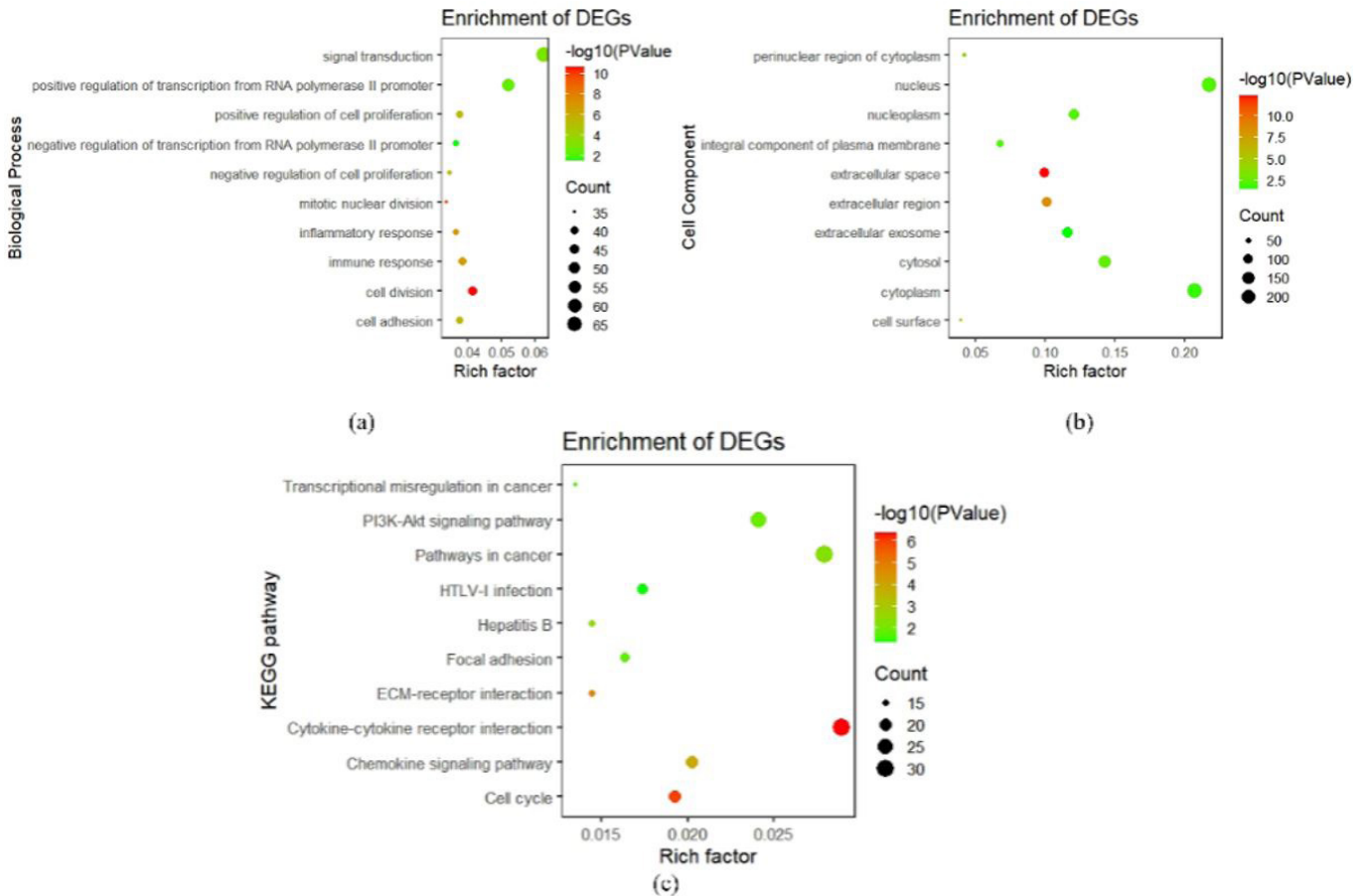**Figure 3:** Enrichment analysis of DEmRNA by FunRich. followed by MF, CC, BP and KEGG

Figure 4: Enrichment analysis of DEmRNA by DAVID, followed by BPCC, BP and KEGG. Among them, the abscissa is the enrichment factor, and the ordinate is the term of enrichment analysis. The size of the circle represents the number of genes, and the color of the circle is the visualization of the value of -log10(10)
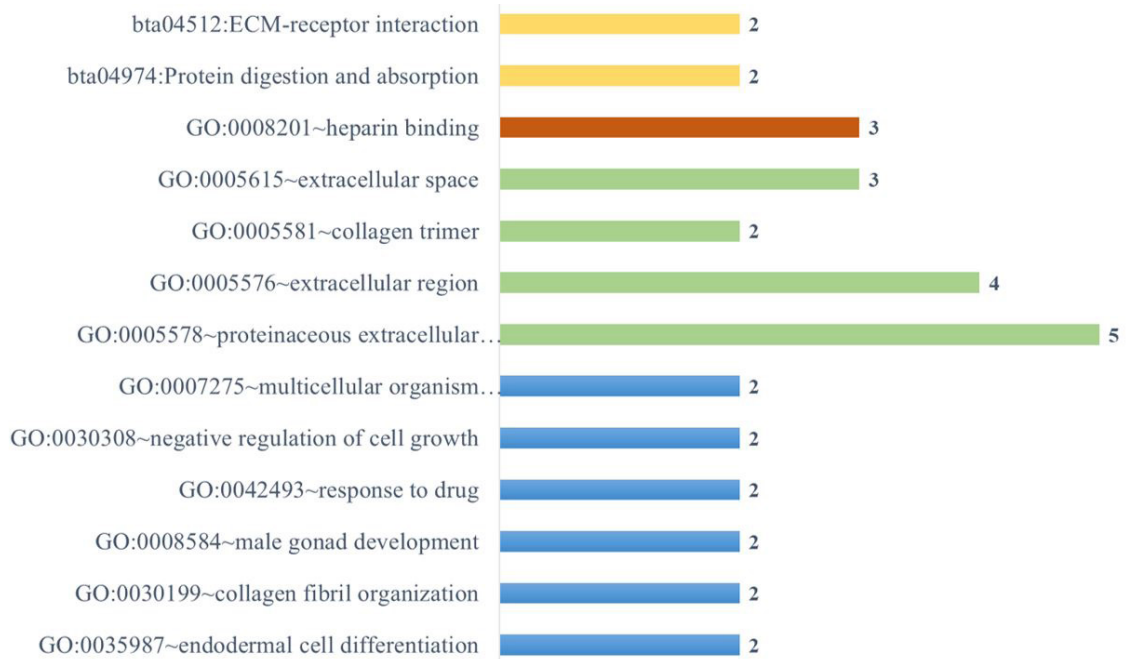
Figure 5: Enrichment analysis of common DEmRNAs by DAVID. Yellow represents KEGG, brown represents MF, green represents CC, and blue represents BP

The DEmRNAs identified from the GSE122063 database was imported into the SRTING to construct PPI network. The isolated protein does not participate in the construction of the network, and hides it to obtain the protein interaction map of the DERNAs. From this PPI network, it was found that the numbers of node and edge are 683 and 6236, respectively. average node degree is 18.3, average local clustering coefficient is 0.44(Figure 6b). After importing 12 common DEmRNAs into the SRTING database, only 4 mRNAs were found to involve in the formation of protein network, namely COL10A1, COMP, MMP11 and COL11A1. From this PPI network, it was found that the number of nodes is 12, the number of edges is 4, the average node degree is 0.667, and the average local clustering coefficient is 0.278(Figure 6a).
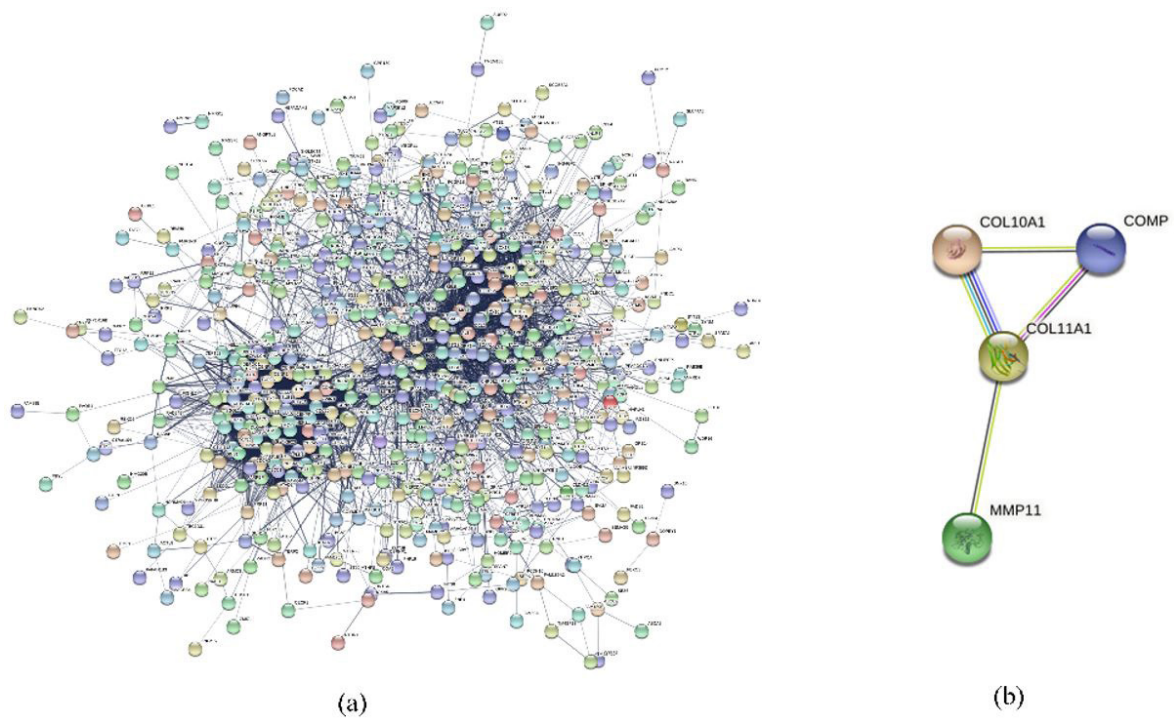

(a)

(b)

Figure 6: PPI network of DEmRNA

Construction of ceRNA regulatory network

Through the online databases mirDIP and ENCORI database, we got 15 lncRNA, 2 miRNAs, and 72 mRNA. Based on the relationship between miRNA and mRNA, lncRNA, we obtained ceRNA regulatory network and visualize it through Cytoscape (Figure 7).
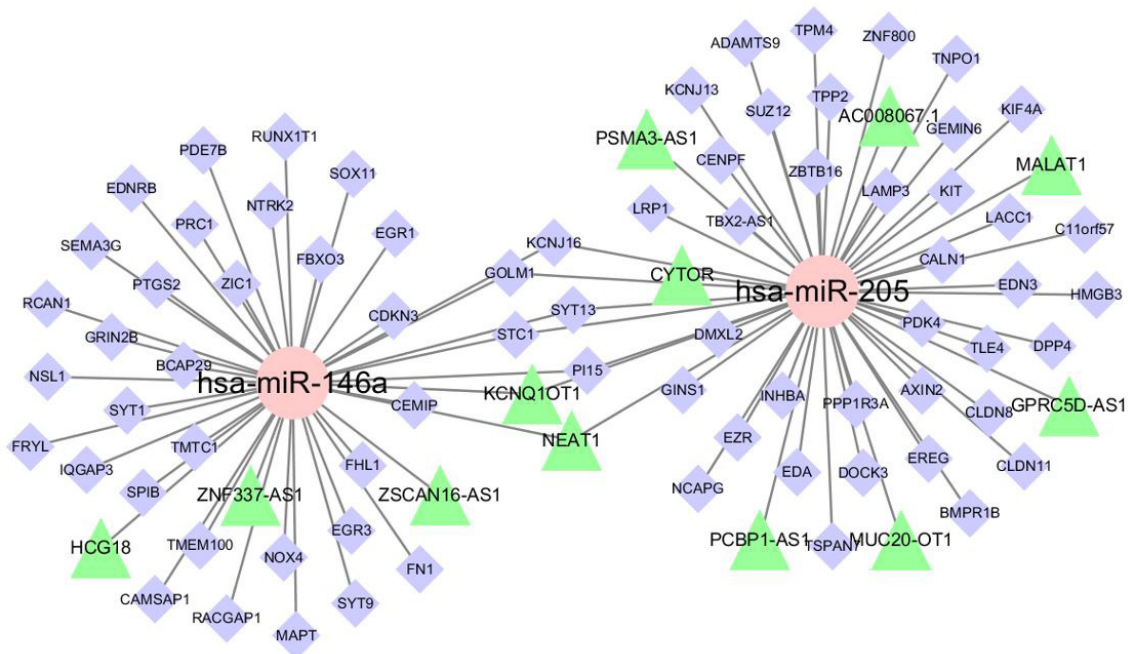

Figure 7: lncRNA–miRNA–mRNA regulatory network in BC. Among them, the red ellipse is miRNA, the green triangle is lncRNA, and the purple diamond represents the mRNA that binds to the target gene

Building a risk model based on ceRNA regulatory network

Building a risk model based on mRNA

There were 1076 samples matching in TCGA database, 72 mRNA expression matrices in ceRNA network were extracted, and 18 genes were screened using lasso analysis: LRP1, CLDN11, AXIN2, NKAPD1, PPP1R3A, ZNF800, BMPR1B, EDN3, HMGB3, CALN1, GEMIN6, EGR3, ZIC1, SPIB, EDNRB, RCAN1, EGR1, BCAP29, for which we also set a 10-fold cross-validation to construct the optimal risk model as Risk Score=0.088*LRP1-0.042*CLDN11+0.053*AXIN2-0.084*NKAPD1-0.163*PPP1R3A-0.036*ZNF800+0.002*BMPR1B-0.021*EDN3+0.008*HMGB3-0.096*CALN1-0.096*GEMIN6-0.008*EGR3+0.028*ZIC1-0.049*SPIB-0.0189*EDNRB-0.022*RCAN1-0.039*EGR1-0.025*BCAP29. In addition, a multifactorial survival analysis based on model genes identified three genes, LRP1, ZIC1, and AXIN2, as risk genes with a risk ratio greater than 1, which are closely associated with breast cancer prognosis (Figure 8A). We used the R package max stat to calculate the optimal cutoff value of Risk Score, setting the minimum grouping sample size greater than 25% and the maximum sample size grouping less than 75%, and finally obtained the optimal cutoff value of -0.539204150358237, based on which the patients were divided into two groups, high and low, and the overall survival rate of patients in the low risk group was better than that in the high risk group (Figure 8B), and in addition the ROC curves at 1, 3 and 5 years showed the more accurate predictive ability of the model with the area under the curve of 0.60, 0.63 and 0.64, respectively (Figure 8C). We then analyzed the relationship between different risk scores and the follow-up time, events and changes in the expression of each gene in patients, and it was observed that the survival rate of patients decreased significantly with increasing risk scores, and as expected, all three genes, LRP1, ZIC1 and AXIN2, were risk factors, and the expression showed an up-regulation trend with increasing risk scores (Figure 8D). We also calculated the immune infiltrating cell score for each sample based on the expression profile of breast cancer tumor samples using the MP Counter method selected by the R package IOBR and found six types of immune cells in patients in the high-risk and low-risk groups: T cells, CD8+ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, myeloid dendritic cells, and two stromal cells, endothelial cells and fibroblasts, were significantly different in abundance (Figure 8E).
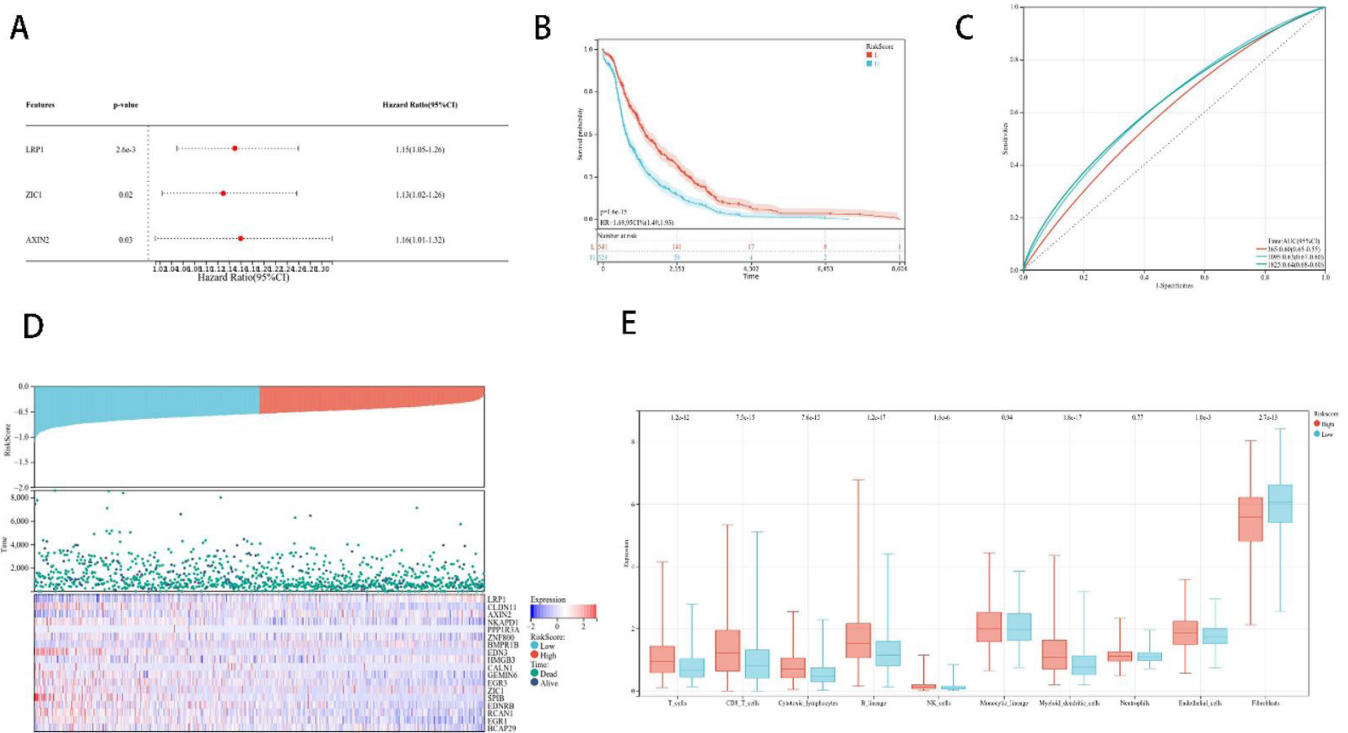
Figure 8: Construction of an mRNA risk model based on ce-RNA regulatory network in. (A) Forest plot of three risk genes LRP1, ZIC1, AXIN2, (B) Survival curves of breast cancer patients in two groups, (C) Time-dependent ROC curve of the risk model, (D) Distribution of survival status of samples with increasing risk scores and model gene expression, (E) Comparison of the abundance of 8 types of immune cells with 2 types of stromal cells under MPCounter algorithm in two groups of breast cancer patients

Independence of the Constructed Risk Model

Clinical characteristics of breast cancer were downloaded from TCGA and selected age, sex, histological type, TNM stage, stage, new event, cancer status, breast carcinoma estrogen receptor status, breast carcinoma progesterone receptor status, lymph node examined count, margin status, number of lymph nodes positive by he, and race along with the risk model for univariate survival analysis. In univariate analysis, age, cancer status, lymph node examined count, M, new event, race, and risk score were statistically significant, and then using them in multivariate survival analysis, risk Score, cancer status, new event were still statistically significant (Table 4), and the HR of risk score reached 6.4, which was an important influencing factor for patient prognosis.

When the risk score was analyzed together with cancer status and new event, it was found that there was no significant difference in risk score among patients with different status of cancer status and new event and the risk score could also classify patients with different status into high-risk and low-risk groups, and there was a significant prognostic difference between the two groups (Figure 9). These results suggest that the risk model we constructed has good independence in predicting breast cancer prognosis.

Table 4: Univariate and multivariate analyses of risk scores and characteristics

Characteristics	Univariate analysis			Multivariate analysis		
	Hazard. Ratio	CI95	P.value	Hazard. Ratio	CI95	P.value
Age	1.01	1.01-1.02	0.00010198	1	1-1.01	0.60009222
Cancer_Status	0.29	0.2-0.42	0	0.3	0.17-0.51	1.52E-05
estrogen_receptor	0.94	0.8-1.1	0.45603331			
he.positive.lymphnodes	0.99	0.97-1.01	0.22350717			
histological_type	1.06	0.98-1.15	0.14003869			
lymph_node	0.98	0.98-0.99	2.83E-05	0.99	0.98-1	0.05569107
M	0.25	0.1-0.59	0.00176032	0.38	0.09-1.55	0.17728732
margin_status	1.19	0.98-1.44	0.07976564			
N	1.04	0.97-1.13	0.27340616			
New_Event	0.37	0.28-0.47	0	0.41	0.28-0.6	3.41E-06
progesterone_receptor	0.96	0.84-1.11	0.62162759			
race	0.82	0.7-0.96	0.01335634	1.03	0.84-1.26	0.79408382
RiskScore	5.95	3.78-9.38	0	6.4	3.59-11.39	0
Sex	1.31	0.72-2.37	0.37739826			
Stage	1.04	0.95-1.15	0.36816163			
T	1	0.91-1.09	0.97460252			

Table Note: Age: patient's age; Cancer_Status: TUMOR FREE, WITH TUMOR; estrogen_receptor: breast cancer estrogen receptor status, divided into positive and negative; he.positive.lymphnodes: positive lymph node count; histological_ type: divided into Infiltrating Lobular Carcinoma and others; lymph_node: lymph node count; M, N, T: TNM stage; margin_status: margin status: divided into Negative, Positive, Close; New_Event: whether cancer occurs again. progesterone_receptor: breast cancer progesterone receptor status, divided into Negative and Positive; race: race divided into WHITE, others; RiskScore: model risk score; Sex: gender; Stage: breast cancer staging.
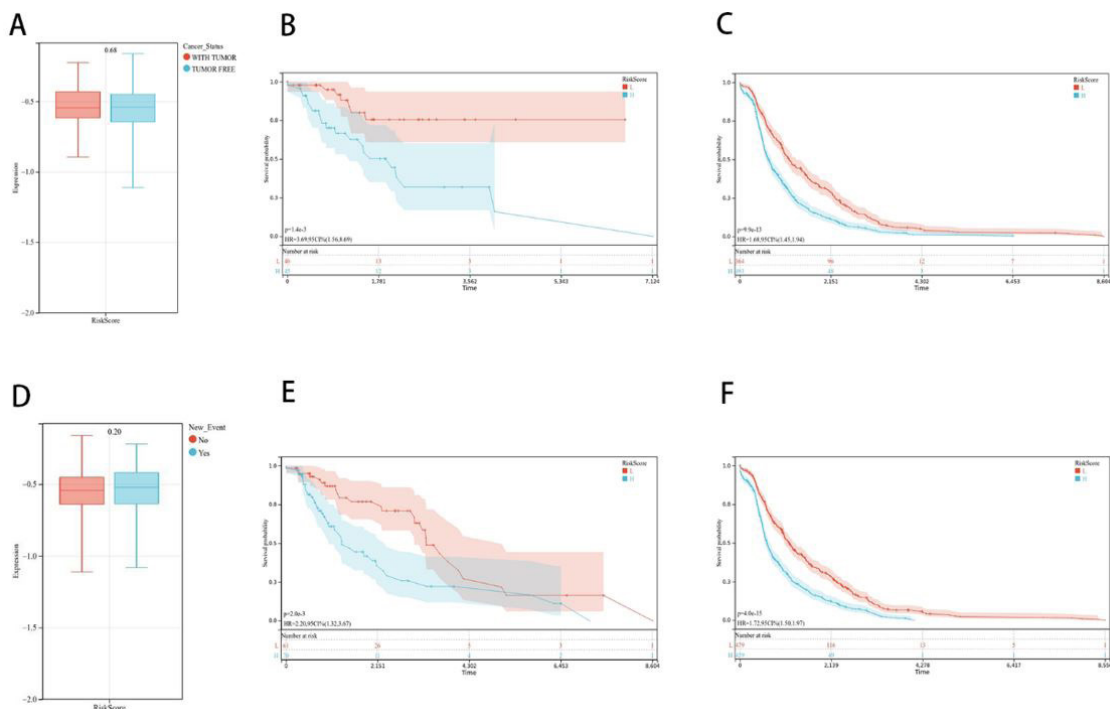

Figure 9: Association of risk score and clinical characteristics. (A) No significant difference was identified in patients with different Cancer status, (B) Survival curve of patients with tumor, (C) Survival curve of patient's tumor free, (D) No significant difference was identified in patients with different New Event, (E) Survival curves for patients with new event, (F) Survival curves for patients with no new events

Construction and Calibration of an Integrated Monogram

Two TBX2-AS1 and MUC20-OT1 lncRNAs that were significant for patient prognosis, with HR values of 1.21 and 1.19, respectively. Finally, to more accurately predict the prognosis of breast cancer patients, a risk model was constructed with clinical characteristics including cancer status, new event, TBX2-AS1, and MUC20-OT1. A nomogram was constructed, as shown in Figure 10. It was assigned specific scores based on the contribution of risk scores and pathological features to the prognosis of breast cancer. The overall C-index of the model was: 0.625,95% CI (0.603-0.646), P value=1.415e-29, and the observed overall survival was better matched with the actual survival at 3 and 5 years (Figure 10B), in addition the model was able to better classify patients into two groups of high and low risk, and the area under the curve reached 0.60, 0.67, and 0.69, suggesting that the risk model constructed based on ceRNA network can reliably and accurately predict the prognosis of breast cancer patients.
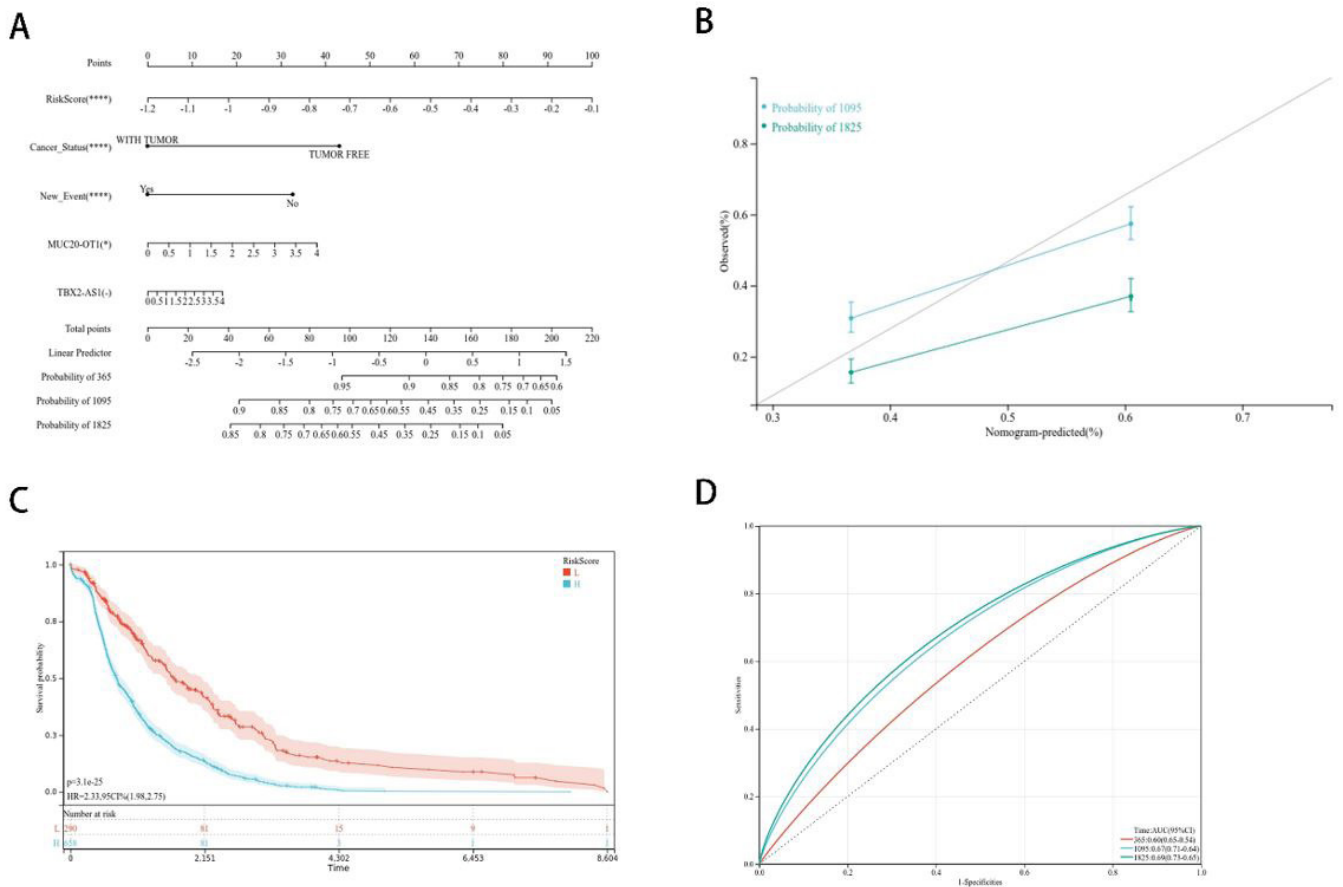

Figure 10: Construction and calibration of nomogram. (A) nomogram integrating risk score and clinical features, (B) calibration of the nomogram at 3 and 5 years, (C) Nomogram Survival Curve, (D) Nomogram ROC Analysis

Molecular Docking

According to the common mRNA obtained in the GEO database, three corresponding target proteins (COMP, PIGR and CXCL10) downloaded from the RCSB database were used as receptors for molecular docking (Table 5). In molecular docking, the ligands are Tamoxifen, Lapatinib, Eribulin and Afimoxifene. The docking results showed the poor docking effect of CXCL10 and Lapatinib and the strong docking effects of other target proteins with drug ligands, this result indciated that the three proteins can be used as potential sites for breast cancer therapy (Table 6). We imported the best conformation into PyMOL, and analyze the three-dimensional effect of the docking results (Figure 11).

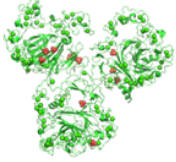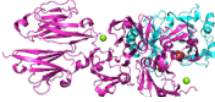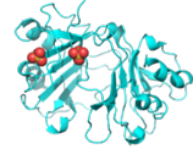**Table 5:** Related information of target protein

Protein	Structure	Expression	PDB	Method	Resolution
COMP,cartilage oligomeric matrix protein		down	3FBY	x-ray	3.15 Å
PIGR,polymeric immunoglobulin receptor		up	1XED	x-ray	1.9 Å
CXCL10,C-X-C motif chemokine ligand 10		down	1O7Y	x-ray	3.00 Å

Table 6: Interaction between target protein and drug

| Receptor	Ligand	Number of conformations	Binding energy kcal·mol⁻¹	Interaction		
				Ligand Atom	Amino Acid Residue	Bond Length / Å
COMP	Tamoxifen	25/50	-2.02	UNL1:HN	ASP248:OD2	2.9
	Lapatinib	5/50	-1.91	UNL1:HN	GLN364:OE1	2.1
	Eribulin	1/50	-5.64	UNL1:HN2	LEU713:O	2.2
	Afimoxifene	2/50	-3.31	UNL1:HO	ASP507:OD2	2.0
PIGR	Tamoxifen	6/50	-3.75	UNL1:O	ASN16:HD21	2.2
	Lapatinib	3/50	-4.21	UNL1:HN1	GLU14:OE2	2.1
	Eribulin	1/50	-4.49	UNL1:O	GLN81:HE22	2.1
	Afimoxifene	2/50	-3.72	UNL1:HO	SER19:O	6.8
CXCL10	Tamoxifen	4/50	-2.88	UNL1:N	GLU57:OE2	2.9
	Lapatinib	4	/	/	/	/
	Eribulin	2/50	-4.04	UNL1:O	LYS46:HZ2	2.2
	Afimoxifene	1/50	-4.85	UNL1:HO	GLN34:O	2.2
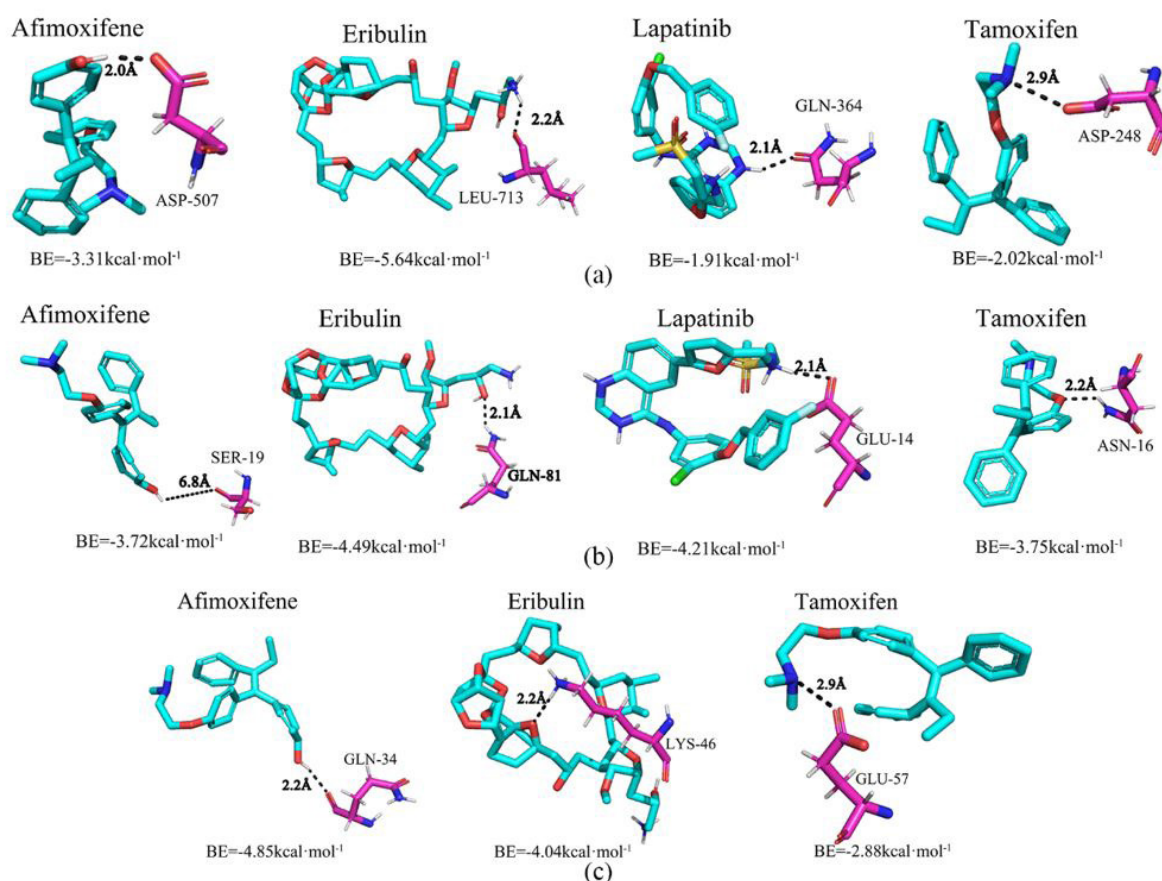
Figure 11: Schematic diagram of the docking of target protein and drug molecule, followed by COMP, PIGR and CXCL10

Discussion

BC is a common malignant tumor with extremely high mortality. Due to the lack of specific diagnostic and prognostic biomarkers, the survival rate of BC patients is very low. In order to study the clinical results and explore the regulatory mechanism of the initiation and progression of BC, the RNA expression data in the GEO database was used to construct a ceRNA regulatory network by conducting gene enrichment and protein interaction analysis using bioinformatics methods. Meanwhile, through molecular docking, the interaction between target protein and related therapeutic drugs was studied to find potential drug targets for the treatment of BC.

In this study, we conducted an in-depth analysis of 3 target proteins (COMP, PIGR and CXCL10). Cartilage oligomeric matrix protein (COMP) is a soluble pentameric protein that is expressed in cartilage. It can lead to the severity of the disease by increasing the invasiveness and viability of tumor cells through metabolic conversion [35]. Studies have found that the imbalance of COMP expression is significantly related to the aggressiveness of breast cancer, which may affect the survival rate and recurrence rate of patients [36,37]. The polymeric immu-

noglobulin receptor (pIgR) is a key component of the mucosal immune system that mediates the transcytosis of immunoglobulin epithelial cells. In addition, up-regulation of pIgR expression was reported to relate to several human cancers [38,39], but there is currently little evidence in clinical trials. CXCL10 is a CXC chemokine [40,41], which mainly exerts its biological effects through CXCR3 [42], and is related to many human diseases. Bioinformatics and luciferase reporter gene analysis indicate that miR-34a can be used as a target of CXCL10, and over-expressed miR-34a can significantly inhibit breast cancer cell growth, invasion, migration and induce apoptosis [43,44]. Furthermore, Ln-cRNA UFC1 can regulate the biological activity of breast cancer through the miR-34a/CXCL10 axis, which provide new diagnostic biomarkers and potential therapeutic targets for breast cancer [44].

There are four kinds of drugs as molecular receptors: Tamoxifen, Lapatinib, Eribulin and Afimoxifene. Among them, Tamoxifen, Lapatinib and Eribulin have been approved by the FDA as clinically effective drugs for the treatment of BC, and Afimoxifene, the active metabolite of tamoxifen, is under clinical phase II study [45]. Tamoxifen, an estrogen receptor antagonist, is the most commonly used drug for the treatment of estrogen

receptor-positive breast cancer, or as an adjuvant therapy in the early stage of the disease [46-48]. Lapatinib is a dual tyrosine kinase inhibitor (TKI), which can block HER1 and HER2 tyrosine kinase activity by binding to the ATP-binding site of the receptor's intracellular domain [49]. Erebrin is highly effective for metastatic breast cancer (MBC), and combined treatment studies are currently undergoing combined therapy research [50].

Using molecular docking, we analyzed the interaction between the target protein and the drug and found no interaction between CXCL10 and Lapatinib. The docking effects of the other target proteins and drug were good, but the docking effects were inconsistent. According to the sequence of binding energy for COMP, Eribulin> Afimoxifene> Tamoxifen> Lapatinib, for PIGR, Eribulin> Lapatinib> Tamoxifen> Afimoxifene, for CXCL10, Afimoxifene>Eribulin> Tamoxifen. According to this ranking, it is found that the Leu residue of COMP and the GLN residue of PIGR can interact with Eribulin, and CXCL10 residues GLN interacts with Afimoxifene. Based on this finding, we suspect that these related amino acid residues of COMP, PIGR and CXCL10 are related with the efficiency of certain drugs for the treatment of breast cancer.

In addition, in the current study we constructed a risk model based on the constructed ceRNA network, and the model achieved AUC values of 0.60, 0.63, and 0.64 at 1, 3, and 5 years, and the model well divided the patients into two groups of high and low risk. 6 immune cells in both groups: T cells, CD8+ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, myeloid dendritic cells, and two stromal cells, endothelial cells and fibroblasts, were significantly different in abundance, suggesting that dysregulation of proliferation of these eight cell types may be associated with breast cancer patients.

Along with risk modeling we also identified three risk genes with risk ratios greater than 1, LRP1, ZIC1, and AXIN2. LRP1 (low-density lipoprotein receptor-related protein 1) is a large multifunctional endothelial cell surface receptor that recognizes numerous ligands and is associated with the regulation of multiple LRP1 has been shown to regulate tumor growth and progression in a variety of tumors and has been identified as a hub in a network of biomarkers used to predict clinical outcomes in multiple cancers [51], a conclusion supported by our study. ZIC1 (zinc finger of the cerebellum 1) is a multi-tumor It inhibits cancer cell growth by reducing Akt and Erk phosphorylation

[52], and the high expression of ZIC1 has become a potential biomarker for good prognosis [53], and it has been demonstrated that ZIC1 expression is decreased in breast cancer samples as well as in breast cancer cell lines, while elevated ZIC1 expression inhibited the growth of breast cancer cells and xenograft tumors, while significantly downregulating survivin expression through the Akt/mTOR/P70S6K pathway [54], which presents opposite results to our results, and breast carcinogenesis may be associated with ZIC1 dysregulation, and the specific mechanism of ZIC1 in tumorigenesis still needs further study.AXIN2 (axis formation inhibitor 2) is an important regulator of the Wnt/β-catenin signaling pathway and is involved in cellular functions such as cell proliferation, cell chemotaxis, migration, and apoptosis [55]. Although AXIN2 is known as an oncogene, recent studies have reported that AXIN2 plays an oncogene role in cancers such as colorectal, hepatocellular and gastric cancers [56], and in our study, AXIN2 presented the same situation in breast cancer.

Similarly, when the joint model and clinical features were analyzed, it was found that the risk model was independent of the traditional clinical features and the clinical features with prognostic ability, lncRNA and risk model were used to construct the nomogram, and it was found that the model had the highest percentage score, but the nomogram had better predictive ability, which could provide a theoretical basis for the later multi-target and multi-mechanism combination of breast cancer.

Previous research shows that the patients with estrogen receptor (ER)-positive tumors, tumors with high CXCL10 level showed improved effect of tamoxifen treatment in terms of local recurrence-free survival and CXCL10 can be used as a prognosis biomarker for tamoxifen [57].

Until now, the pathogenesis of BC is still unclear, and there are few effective drugs approved by the FDA for clinical treatment, which poses a great challenge for future research. On the whole, the current analysis of drug targets is mostly based on existing databases, and the clinical experimental data is lacking. To further clarify the pathogenesis of BC and effective targets for treatment, more advanced biological technologies need to be used in basic and clinical research on BC and other tumor-related drug targets. This article mainly studies the related targets of BC from the theoretical level. Therefore, experiment is necessary to further verify the accuracy of prediction results.

Acknowledgments

References

1. Siegel RL, Miller KD, Jemal A (2018) Cancer Statistics, 2018. Ca-a Cancer Journal for Clinicians 68: 7-30.

2. Cho N (2016) Molecular subtypes and imaging phenotypes of breast cancer. Ultrasonography (Seoul, Korea) 35: 281-8.

3. Dai XF, Li T, Bai ZH, Yang YK, Liu XX et al. (2015) Breast cancer intrinsic subtype classification, clinical use and future trends. American Journal of Cancer Research 5: 2929-43.

4. Ontario H (2020) (Gene Expression Profiling Tests for Early-Stage Invasive Breast Cancer: A Health Technology Assessment. Ontario health technology assessment series 20: 1-234.

5. Liang YR, Zhang HW, Song XJ, Yang QF (2020) Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. Seminars in Cancer Biology 60: 14-27.

6. Tulotta C, Ottewell P (2018) The role of IL-1B in breast cancer bone metastasis. Endocrine-Related Cancer 25: R421-34.

7. Xiong ZC, Deng GZ, Huang XJ, Li X, Xie XH et al. (2018) Bone metastasis pattern in initial metastatic breast cancer: a population-based study. Cancer Management and Research 10: 287-95.

8. Smid M, Wang YX, Zhang Y, Sieuwerts AM, Yu J et al. (2008) Subtypes of breast cancer show preferential site of relapse. Cancer Research 68: 3108-14.

9. Tham YL, Sexton K, Kramer R, Hilsenbeck S, Elledge R (2006) Primary breast cancer phenotypes associated with propensity for central nervous system metastases 107: 696-704.

10. Mann RM, Kuhl CK, Kinkel K, Boetes C (2008) Breast MRI: guidelines from the European Society of Breast Imaging. European Radiology 18: 1307-18.

11. Jafari SH, Saadatpour Z, Salmaninejad A, Momeni F, Mokhtari M et al. (2018) Breast cancer diagnosis: Imaging techniques and biochemical markers. Journal of Cellular Physiology 233: 5200-13.

12. Keshavarzi M, Sorayayi S, Rezaei MJ, Mohammadi M, Ghaderi A et al. (2017) MicroRNAs-Based Imaging Techniques in Cancer Diagnosis and Therapy. Journal of Cellular Biochemistry 118: 4121-8.

13. Shah K, Jacobs A, Breakefield XO, Weissleder R (2004) Molecular imaging of gene therapy for cancer. Gene Therapy 11: 1175-87.

14. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS et al. (2016) Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society. Obstetrical & Gynecological Survey 71: 153-5.

15. Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS et al. (2016) Supplemental Screening for Breast Cancer in Women with Dense Breasts: A Systematic Review for the US Preventive Services Task Force. Annals of Internal Medicine 164: 268.

16. Comstock CE, Gatsonis C, Newstead GM, Snyder BS, Gareen IF et al. (2020) Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women with Dense Breasts Undergoing Screening. Jama 323: 746-56.

17. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? Cell 146: 353-8.

18. Amaral PP, Dinger ME, Mercer TR, Mattick JS (2008) The eukaryotic genome as an RNA machine. Science 319: 1787-9.

19. Olgun G, Sahin O, Tastan O (2018) Discovering lncRNA mediated sponge interactions in breast cancer molecular subtypes. Bmc Genomics 19.

20. Xiao B, Zhang WY, Chen LD, Hang JF, Wang LZ et al. (2018) Analysis of the miRIVA miRNA-mRNA-lncRNA network in human estrogen receptor-positive and estrogen receptor-negative breast cancer based on TCGA data. Gene 658: 28-35.

21. Shen XK, Xie BJ, Ma ZS, Yu WJ, Wang WM et al. (2015) Identification of novel long non-coding RNAs in triple-negative breast cancer. Oncotarget 6: 21730-9.

22.   Lv MM, Xu PF, Wu Y, Huang L, Li WQ et al. (2016) LncRNAs as new biomarkers to differentiate triple negative breast cancer from non-triple negative breast cancer. Oncotarget 7: 13047-59.

23.   Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D et al. (2007) NCBI GEO: mining tens of millions of expression profiles - database and tools update. Nucleic Acids Research 35: D760-5.

24.   The Cancer Genome Atlas Program. In. Edited by Services USDoHaH, Health NIo, Institute NC.

25.   Sean D, Meltzer PS (2007) GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics 23: 1846-7.

26.   Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CYJ et al. (2015) FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15: 2597-601.

27.   Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4: 44-57.

28.   Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S et al. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Research 45: D362-8.

29.   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT et al. (2003) A software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498-504.

30.   Yang JH, Li JH, Shao P, Zhou H, Chen YQ et al. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Research 39: D202-9.

31.   Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC et al. (2018) mirDIP 4.1-integrative database of human microRNA target predictions. Nucleic Acids Research 46: D360-70.

32.   Rose PW, Prlic A, Bi CX, Bluhm WF, Christie CH et al. (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Research 43: D345-56.

33.   Seeliger D, de Groot BL (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. Journal of Computer-Aided Molecular Design 24: 417-22.

34.   Lill MA, Danielson ML (2011) Computer-aided drug design platform using PyMOL. Journal of Computer-Aided Molecular Design 25: 13-9.

35.   Englund E, Bartoschek M, Reitsma B, Jacobsson L, Escudero-Esparza A et al. (2016) Cartilage oligomeric matrix protein contributes to the development and metastasis of breast cancer. Oncogene 35: 5585-96.

36.   Papadakos KS, Bartoschek M, Rodriguez C, Gialeli C, Jin SB et al. (2019) Cartilage Oligomeric Matrix Protein initiates cancer stem cells through activation of Jagged1-Notch3 signaling. Matrix Biology 81: 107-21.

37.   Li Q, Wang C, Wang YF, Sun LK, Liu ZK et al. (2018) HSCs-derived COMP drives hepatocellular carcinoma progression by activating MEK/ERK and PI3K/AKT signaling pathways. Journal of Experimental & Clinical Cancer Research 37.

38.   Qi XC, Li XC, Sun XX (2016) Reduced expression of polymeric immunoglobulin receptor (pIgR) in nasopharyngeal carcinoma and its correlation with prognosis. Tumor Biology 37: 11099-104.

39.   Li D, Wang FJ, Yu L, Yao WR, Cui YF et al. (2017) Expression of pIgR in the tracheal mucosa of SHIV/SIV-infected rhesus macaques. Zoological Research 38: 44-8.

40.   Zlotnik A, Yoshie O (2012) The Chemokine Superfamily Revisited. Immunity 36 :705-16.

41.   Karin N, Razon H (2018) Chemokines beyond chemo-attraction: CXCL10 and its significant role in cancer and autoimmunity. Cytokine 109: 24-8.

42.   Loetscher M, Gerber B, Loetscher P, Jones SA, Piali L et al. (1996) Chemokine receptor specific for IP10 and Mig: Structure, function, and expression in activated T-lymphocytes. Journal of Experimental Medicine 184: 963-9.

43.   Xu M, Li D, Yang C, Ji JS (2018) MicroRNA-34a Inhibition of the TLR Signaling Pathway Via CXCL10 Suppresses Breast Cancer Cell Invasion and Migration. Cellular Physiology and Biochemistry 46: 1286-304.

44. Xie RL, Wang MY, Zhou WT, Wang D, Yuan Y et al. (2019) Long Non-Coding RNA (LncRNA) UFC1/miR-34a Contributes to Proliferation and Migration in Breast Cancer. Medical Science Monitor 25: 7149-57.

45. Mansel R, Goyal A, Le Nestour E, Masini-Eteve V, O'Connell K et al. (2007) A phase II trial of Afimoxifene (4-hydroxytamoxifen gel) for cyclical mastalgia in premenopausal women. Breast Cancer Research and Treatment 106: 389-97.

46. Shagufta, Ahmad I (2018) Tamoxifen a pioneering drug: An update on the therapeutic potential of tamoxifen derivatives. European Journal of Medicinal Chemistry 143: 515-31.

47. Radin DP, Patel P (2016) Delineating the molecular mechanisms of tamoxifen's oncolytic actions in estrogen receptor-negative cancers. European Journal of Pharmacology 781: 173-80.

48. Briest S, Stearns V (2009) Tamoxifen metabolism and its effect on endocrine treatment of breast cancer. Clinical advances in hematology & oncology: H&O 7: 185-92.

49. Nolting M, Schneider-Merck T, Trepel M (2014) Lapatinib. Recent results in cancer research Fortschritte der Krebsforschung Progres dans les recherches sur le cancer 201: 125-43.

50. Lillis AP, Van Duyn LB, Murphy-Ullrich JE, Strickland DK (2008) LDL receptor-related protein 1: Unique tissue-specific functions revealed by selective gene knockout studies. Physiological Reviews 88: 887-918.

51. Martinez-Ledesma E, Verhaak RGW, Trevino V (2015) Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. Scientific Reports 5.

52. Gan LH, Chen SJ, Zhong J, Wang X, Lam EKY et al. (2011) ZIC1 Is Downregulated through Promoter Hypermethylation, and Functions as a Tumor Suppressor Gene in Colorectal Cancer. Plos One 6.

53. Ma G, Dai WJ, Sang AY, Yang XZ, Li QJ (2016) Roles of ZIC family genes in human gastric cancer. International Journal of Molecular Medicine 38: 259-66.

54. Han W, Cao F, Gao XJ, Wang HB, Chen F et al. (2018) ZIC1 acts a tumor suppressor in breast cancer by targeting survivin. International Journal of Oncology 53: 937-48.

55. Dong X, Seelan RS, Qian C, Mai M, Liu W (2001) Genomic structure, chromosome mapping and expression analysis of the human AXIN2 gene. Cytogenetics and Cell Genetics 93: 26-8.

56. Ying Y, Tao Q (2009) Epigenetic disruption of the WNT/beta-catenin signaling pathway in human cancers. Epigenetics 4: 307-12.

57. Hilborn E, Sivik T, Fornander T, Stal O, Nordenskjold B et al. (2014) C-X-C ligand 10 and C-X-C receptor 3 status can predict tamoxifen treatment response in breast cancer patients. Breast Cancer Research and Treatment 145: 73-82.