Pathology - Research and Practice 215 (2019) 152405

Contents lists available at ScienceDirect

Pathology - Research and Practice

journal homepage: www.elsevier.com/locate/prp

Transcriptome sequencing profiles reveal lncRNAs may involve in breast cancer (ER/PR positive type) by interaction with RAS associated genes

Yongfeng Jia[a,b], Lin Shi[b], Fen Yun[b], Xia Liu[b], Yongxia Chen[c], Minjie Wang[a], Chen Chen[b], Yanni Ren[b], Yulong Bao[a,*], Li Wang[a,*]

[a] College of Basic Medicine, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China
[b] Department of Pathology, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China
[c] Tumor Molecular Diagnostic Laboratory, The Inner Mongolia Cancer Hospital, Hohhot, Inner Mongolia, China

ARTICLE INFO

Keywords:
Breast cancer
Transcriptome analysis
RAS genes
Differential expression long noncoding RNAs (lncRNAs)

ABSTRACT

To reveal novel lncRNAs and explore how could lncRNA affect the ER/PR positive type breast cancer, 16 different lncRNA transcriptomes (8 breast cancer tissues and 8 normal breast tissues) were successfully sequenced. In total, 8,954 high quality lncRNAs, including 5,516 lncRNAs reported in the previous studies and 3,438 novel lncRNAs, were annotated. The highest expressed lncRNAs were MALAT1, SCARNA10, RP11-206M11.7 and NEAT1, and the highest expressing mRNAs were RPL19, SCGB2A2, FTL and TMSB4 × . Of the 615 differentially expressed lncRNAs, 323 showed up regulated ($P < 0.05$) expression patterns in breast cancer, and 292 showed down regulated expression patterns. Of the 8,954 genes, 5,516 genes were upregulated in breast cancer, and 3,438 were downregulated. In total, the targets of 238 lncRNAs were confirmed by two lncRNA target prediction programs. Within these genes, Ras responsive element binding protein 1, Ras association domain family member 6, Ras association domain family member 8, Ras protein specific guanine nucleotide releasing factor 1and other 10 different Ras associated different expressed genes were predicted as targets of lncRNAs. These different expressed lncRNAs which could regulate the Ras gene families and ECM pathway may be another mechanism why the expression pattern of Ras genes changed in breast cancer. All these cancer-related genes (Ras genes) were annotated as targets of lncRNAs in the breast cancer transcriptome may provide us with a new way to understand the occurrence and development of breast cancer.

1. Introduction

Breast cancer (BC) is considered as the second (only after lung cancer) fatal malignant cancer among women [1]. The fatality rate of breast cancer patients is almost 1 per 10,000 individuals and increases with the progression of cancer (almost double or triple mortality in the fifth year compared to the first year after diagnosis. Every year, almost 60,000 breast cancer patients died. In general, breast cancer (at least hereditary breast cancers) is a disease related to genetic changes [2]. The diagnosis and classification of the breast cancer were based on the expression pattern of many breast cancer-related genes. The BC could be classified into different subtypes based on the status of ER, PR and Her2. The triple negative (ER-, PR- and Her2 -) breast cancer is the deadliest subtype [3]. Meanwhile, the ER positive type and PR positive type could be treated effectively by the same endocrine therapy. The 5-year survival rate of ER positive subtype breast cancer were also higher than the ER negative subtype, the ER/PR positive subtype was even more than 80% [4]. Changes of genome [5] or the expression pattern of many cancer-related genes that reflect the cellular phenotype and physiological function. The expression profile of transcriptome was discrepant in different kinds of tumor [6,7] or even the different subtype of the single cancer [8]. It is necessary to clarify the transcriptome of all different kinds of subtype cancer. Given the high mortality of triple negative breast cancer, more attention was paid on this subtype cancer [9,10].The ER(+)/PR(+) subtype breast cancer were the other extreme, a kind of subtype with a good prognostic and high survival rate. The transcriptome of ER positive type and negative type were significant different instead of the only ER gene itself [11,12].It is very hard to elucidate the mechanism of breast cancer which contained several subtype cancer. A detailed survey of transcriptome of all kinds of subtype breast cancer is necessary. Long non-coding RNA (LncRNA), an important part of transcript, may provide a new expound for us to clarify different clinical features and outcome of the subtype cancers [13].

* Corresponding authors at: Jinshan Development Zone, Hohhot, Inner Mongolia, China.
E-mail addresses: yulongbao0471@163.com (Y. Bao), liwang0472@163.com (L. Wang).

https://doi.org/10.1016/j.prp.2019.03.033
Received 17 December 2018; Received in revised form 18 March 2019; Accepted 31 March 2019
0344-0338/

The human genome project indicated that only 2% of the genome was protein-coding sequence (mRNA) [14,15] and the other part were non-coding dark sequences. Long noncoding RNAs (lncRNAs), one of the non-coding sequences, shared the same molecular structure with mRNA (coding genes, UTR and ORF region) but showed no protein coding capacity [14,16]. Previous studies suggest that lncRNAs may play an important role in the regulation of cell development, proliferation, differentiation and apoptosis [17]. LncRNAs had been studied in cancer for a long time and were considered as an important regulator in development of breast cancer [18]. Given all these features that are important in cancer, lncRNAs could be involved in the initiation and progression of many diseases in addition to cancer. Due to the cancer especially the breast cancer was a disease with a significant different gene profile with other tumor types or even the subtype of breast cancer [19].To explore the different subtype of breast cancer is meaningful.LncRNA sequencing of lung cancer and breast cancer provide 2,812 lncRNAs. Meanwhile, the different lncRNA associated with different development stage of lung cancer were identified. *NKX2-1-AS1* and *DSCAM-AS1*, which up-regulated in the lung cancer and regulated the protein-coding genes which were associated with lung cancer. Meanwhile, the subtype breast cancer ER(+), ER(-) shared 20 different expressed lncRNAs including LINC00324, *PTPRGAS1* (protein tyrosine phosphatase, receptor type, G, antisense) and *SNHG17* (small nucleolar RNA host gene 17).Except the identified of subtype-related lncRNAs, function research revealed many lncRNAs which could affect the function of breast cancer cells (Long noncoding RNA UCA1 modulates breast cancer cell growth and apoptosis through decreasing tumor suppressive miR-143)or the treatment effects(Regulation of apoptosis by long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy). H19, the first lncRNA identified in cancer were detected and functioned in many tumor types, such as colon, kidney, liver as well as breast cancer. The function of H19 in multi-cancer suggest that lncRNAs may play a key role in the development of many different cancers. Meanwhile, lncRNAs could regulate the gene expression profile by regulating genome activity via histone modification and DNA methylation and remodeling chromatin at the nucleic acid or molecular level. Targeting lncRNAs may be an easy way to treat cancer by elucidating the expression pattern or function of lncRNAs at the molecular level.

Except the subtype-related lncRNAs, many lncRNAs have also been studied in all breast cancer type. The first was HOX antisense intergenic RNA (HOTAIR) [20,21], which is expressed at a high level in breast cancer tissues and targets the HOXC1 and HOXC12 genes via histone modification. The overlapping lncRNA NKILA acts as a transcription factor in breast cancer (LncRNA NKILA suppresses TGF-β-induced epithelial–mesenchymal transition by blocking NF-κB signaling in breast cancer). Studies on lncRNAs may provide us a new direction to confirm and treat breast cancer. Even before we gained the benefits of next generation sequencing, many lncRNAs were identified and confirmed by traditional methods. The use of the Human Genome Project and transcriptome sequencing may greatly accelerate our discovery of novel lncRNAs for breast cancer. Meanwhile, a study pointed out that 60% of lncRNAs showed specificity in only one type of cancer [22]. Therefore, it is necessary to screen the lncRNAs that are specific to breast cancer tissues and systematically study the role of lncRNAs in breast cancer. ER (+) /PR (+) subtype breast cancer is a cancer with a low vulnerability and high cure rate. However, 2/3 of breast cancer patients were ER(+) type [11].In view of the important role of ER/PR (+) subtype cancer, in this study, the transcriptome expression profiles of breast cancer (ER/PR (+) type)tissues and normal breast tissues were detected to obtain an overview of the lncRNAs involved in this subtype breast cancer. The differentially expressed lncRNAs and co-expressed coding genes in breast cancer (ER(+)/PR (+) type) were filtered out. The functions of the lncRNAs and co-expressed mRNAs were identified by KEGG pathway and GO analyses. Our study may provide a foundation for the clinical research of breast cancer (ER(+)/PR (+) type) based on the research of lncRNA.

Table 1
The demography and clinical features.

NO	gender	age	ER	PR
1633649	female	35	3+	3+
1635819	female	53	3+	3+
1638292	female	51	3+	3+
1711197	female	34	2+	3+
1716677	female	53	2+	2+
1725159	female	43	2+	3+
1727975	female	46	3+	3+
1191141	female	41	2+	2+

Table 2
The details of the transcriptome assemble result.

Sample name	Raw reads	Clean reads	clean bases	Error rate (%)	GC content (%)
N8	107068208	101990210	15.3G	0.02	46.85
N2	152719916	145827132	21.87G	0.02	48.33
N3	101485610	97143966	14.57G	0.02	46.52
N4	103928234	98823036	14.82G	0.02	49.39
N5	126312546	120994424	18.15G	0.02	47.14
N7	101759050	97279986	14.59G	0.02	49.22
N1	107019882	103659356	15.55G	0.01	46.5
N6	108635026	105881964	15.88G	0.01	48.62
T1	115728976	112545512	16.88G	0.01	45.57
T8	115139232	110940972	16.64G	0.02	47.96
T6	117017282	113259066	16.99G	0.01	45.75
T2	94941988	91618018	13.74G	0.02	46.76
T3	123122094	119295204	17.89G	0.02	46
T4	147626842	142369812	21.36G	0.02	46.03
T5	127911670	124016248	18.6G	0.02	47.4
T7	103481692	98768054	14.82G	0.02	46.58

2. materials and methods

2.1. Ethics

Breast cancer tissue (invasive ductal carcinoma, tumor tissue) and adjacent normal breast tissue (precancerous tissues, normal tissue) samples from the same individual with breast cancer were obtained from 2016 to 2017 from the Department of Pathology, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China. All patients were informed and signed an agreement to be part of this study. All material and methods in this study were licensed and approved by the Ethics Committee of Inner Mongolia Medical University.

2.2. Study patients, tissue sample preparation and collection

Histopathological specimens from breast cancer patients were confirmed pathologically were selected from the Affiliated Hospital of Inner Mongolia Medical University (Inner Mongolia, China). All specimens (including BC tissues (tumor tissue) and paired adjacent normal tissues (normal tissues)) were cut into 3–8 mm^2 pieces after surgery. The sampled tissues used for transcriptome sequencing were submerged in an RNA protectant (RNA later, Sigma, USA) for 12 h at 4 °C, quickly placed into liquid nitrogen and transferred to −80 °C for RNA isolation. Meanwhile, the samples from the same patients used for histology testing were fixed with formalin. The clinical features of the patients such as the ER and PR status and Her2 status were all confirmed by histology.

2.3. Isolation of total RNA and construction of sequencing library

Only the ER/PR positive type breast cancer (the criteria was

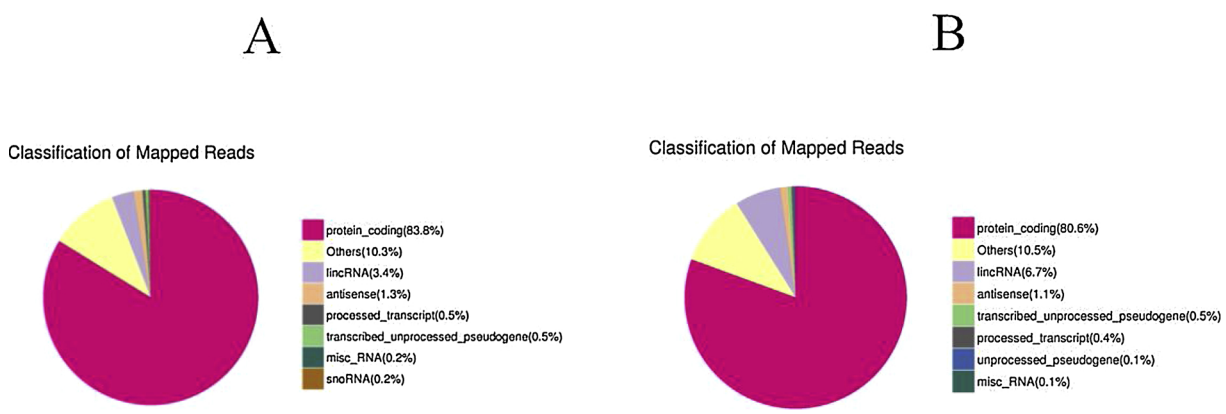

Fig. 1. classification of mapped reads from normal tissues(A) and cancer tissues(B). The different types of lncRNAs based on molecular structure.

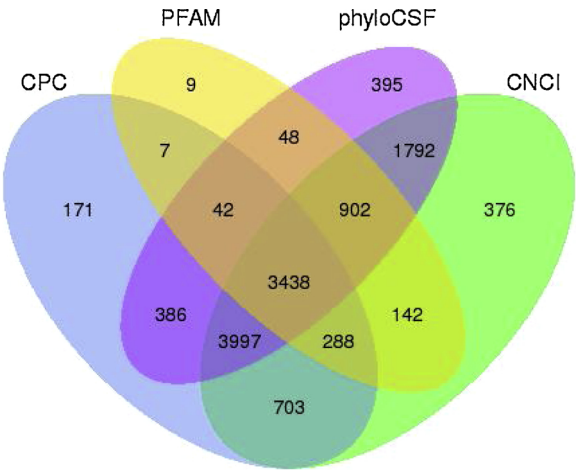

Fig. 2. The number of novel lncRNAs annotated based on different databases. The overlapped areas represent the same results from both databases.

Table 3
The different lncRNA type of annotated lncRNA and novel lncRNA.

lncRNA type	annotated lncRNA number	novel lncRNA number
3prime overlapping ncrna	13	0
antisense	1908	445
lincRNA	2051	2993
Macro lncRNA	1	0
Processed transcript	571	0
Sense intronic	596	0
Sense overlapping	106	0
TEC	270	0
in total	5516	3438

ER > 2+ and PR > 2+ at the same time) patients were selected as the sampled volunteers. Total RNA obtained from 8 normal breast tissue and 8 breast cancer (ER/PR positive type) tissue samples were extracted with TRIzol (Life Technologies, USA). The obtained RNA was analyzed by agarose gel electrophoresis (1%, 110 V). The amount of RNA used per sample was no less than 3 μg. The ribosomal RNA was eliminated. Subsequently, cDNA libraries were produced following the manufacturer's recommendations. The AMPure XP system was employed to purify the PCR products, and the libraries were analyzed on an Agilent Bioanalyzer 2100 system. All index-coded samples were clustered on a cBot Cluster Generation System with the assistant of TruSeq PE Cluster Kit v3-cBot-HS (Illumina) following a manual. After clustering, the Illumina HiSeq 4000 platform was employed to sequence all cDNA libraries to generate 150 bp paired-end raw reads.

2.4. Quality control, assembly and annotation

All primordial raw data (FASTQ format) generated from the sequencing platform were first processed with several homemade Perl scripts. All dirty or with low useful information raw data (such as the adapter, ploy-N sequence and low-quality reads) were removed. All clean reads for 16 different sequencing libraries were aligned to the reference human genome and gene model annotation databases using bowtie2 v2.2.8 and HISAT2 v2.0.4 [23], respectively. The aligned reads from all samples were assembled by StringTie (v1.3.1) [24]. All reads were assembled into transcripts based on the genome sequence and gene model annotation files.

2.5. Discovery of known and novel lncRNAs

The known lncRNAs were annotated based on the genome annotation result. The transcripts that failed to be annotated were subjected to coding potential analysis based on four databases: CNCI (Coding-Non-Coding-Index, with default parameters) (v2), CPC (Coding Potential Calculator (0.9-r2), the e-value were '1e-10', Pfam-sca (release 27; both Pfam A and Pfam B, with default parameters of 1e-10) and PhyloCSF (phylogenetic codon substitution frequency (v20121028), with default parameters). All transcripts that were successfully calculated by either/all of the four databases were identified as coding gene (mRNA). Only those transcripts that showed no coding potential were defined as candidate or novel lncRNAs. All novel lncRNAs and mRNAs were imputed into Phast (v1.3) [25] and phastCons to generate the conservation scores for all novel lncRNAs and coding genes (mRNA). When all the annotation progress were finished, the HISAT2 [23] were employed to map all annotation lncRNAs to the genome map and located all lncRNAs on different chromosomes with the help of R package.

2.6. Analysis of the differential expression of lncRNAs and genes

The software edgeR was used to calculate the Fragments Per Kilobase Million (FPKMs) of both lncRNAs and mRNAs for each sample [26]. The FPKMs for each gene (in the 16 different transcriptome datasets) were calculated by summing the FPKMs of the transcripts in each gene group. The software contains statistical routines for determining the differential expression of the digital transcript or gene expression data based on the negative binomial distribution [26]. Transcripts that showed a different expression level between breast cancer and normal tissues with a P-adjust < 0.05 and fold change > 2 were deemed differentially expressed.

All lncRNA expression data (HTseq-counts data) and the corresponding clinical data (clinical.car.tar.gz) were downloaded from TCGA

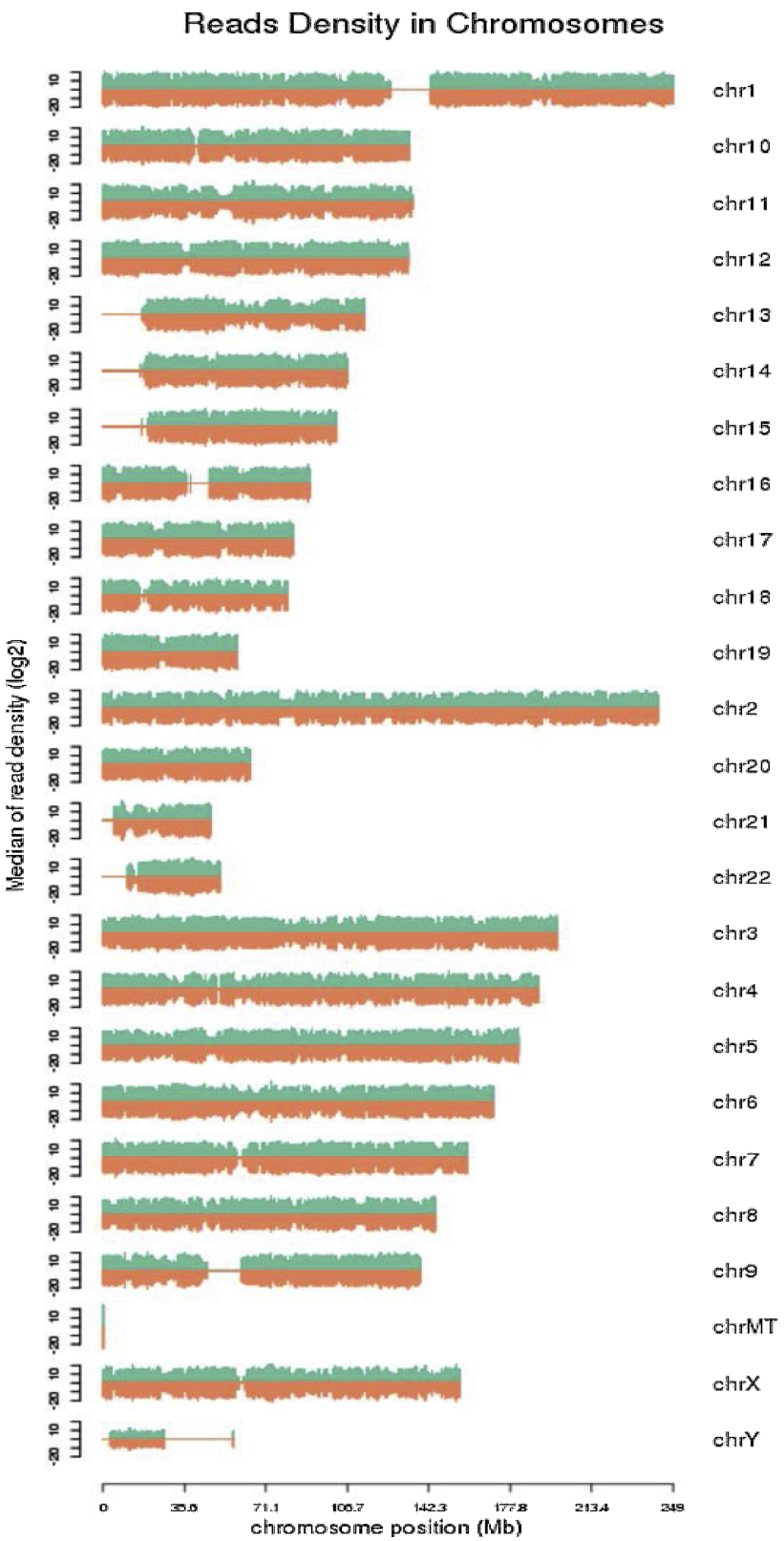

Fig. 3. the location of lncRNA on different chromosomes.
The horizontal coordinate is the length information of chromosomes (in millions of bases), the vertical coordinate is log2(the median density of reads), and green is the sense strand and red is antisense strand.

dataset (https://tcga-data.nci.nih.gov/tcga/)with the help of GDC client software. Up to December 2018, 1,109 breast cancer samples and 113 normal samples results could be downloaded from the TCGA database. The expression level of lncRNAs downloaded from TCGA dataset were calculated with edgeR and normalized to Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values. LncRNAs that showed different expression pattern with a P-adjust < 0.05 and fold change > 2 were deemed as differentially expressed

Table 4
Top 20 aberrantly expressed lncRNAs in tumor tissues.

Gene id	Official Symbol	Gene type	Normal FPKM	Tumor FPKM
ENSG00000251562.7	MALAT1	lincRNA	3376.575836	5579.980072
ENSG00000276232.1	SCARNA10	Sense intronic	1222.976543	1541.645969
ENSG00000244468.1	RP11-206M11.7	antisense	51.61778213	118.3694251
ENSG00000245532.5	NEAT1	lincRNA	377.8184568	690.604435
ENSG00000235687.8	LINC00993	lincRNA	72.94275638	127.7821443
ENSG00000234918.1	RP11-20F24.2	lincRNA	64.0717165	122.4735783
ENSG00000229807.9	XIST	lincRNA	112.8230591	111.8113195
ENSG00000260032.1	LINC00657	lincRNA	113.5950758	114.7980576
ENSG00000261294.1	RP11-616M22.3	antisense	0.48814725	13.57538138
ENSG00000255026.1	RP11-326C3.2	antisense	15.71662138	44.99087438
ENSG00000281508.1	CDR1-AS	antisense	341.9651606	94.10079388
ENSG00000259279.1	CTD-2033D15.1	antisense	17.53868488	31.54011925
ENSG00000254911.3	SCARNA9	antisense	145.2864723	92.12042913
ENSG00000234741.7	GAS5	Processed transcript	73.13883988	65.032235
ENSG00000230937.9	MIR205HG	Processed transcript	7.54118075	6.492219375
ENSG00000253361.1	RP11-675F6.3	lincRNA	0.68232525	5.176682625
ENSG00000276107.1	CTD-2033D15.2	Sense intronic	14.01177338	25.49593288
ENSG00000242588.5	RP11-274B21.14	Processed transcript	33.75894988	27.20923025
ENSG00000267519.3	CTD-3252C9.4	lincRNA	13.13186338	12.923282

Table 5
Top 20 aberrantly expressed mRNA in tumor tissues.

Gene id	Official Symbol	Gene type	Normal FPKM	Tumor FPKM
ENSG00000108298	RPL19	protein coding	498.999	2211.657
ENSG00000110484	SCGB2A2	protein coding	55.95528	1465.697
ENSG00000087086	FTL	protein coding	732.6121	1072.227
ENSG00000205542	TMSB4X	protein coding	395.6476	1017.151
ENSG00000108821	COL1A1	protein coding	98.67618	933.3021
ENSG00000034510	TMSB10	protein coding	399.2867	763.938
ENSG00000168542	COL3A1	protein coding	133.7993	676.1755
ENSG00000229117	RPL41	protein coding	779.847	670.8879
ENSG00000156482	RPL30	protein coding	326.5839	603.7048
ENSG00000112306	RPS12	protein coding	253.6478	550.9173
ENSG00000115414	FN1	protein coding	25.01122	538.5614
ENSG00000171345	KRT19	protein coding	194.3933	534.6924
ENSG00000147604	RPL7	protein coding	392.8142	528.301
ENSG00000164692	COL1A2	protein coding	107.0713	513.8003
ENSG00000231500	RPS18	protein coding	329.1737	512.9944
ENSG00000142534	RPS11	protein coding	721.0031	509.5972
ENSG00000177954	RPS27	protein coding	762.428	501.7894
ENSG00000147676	MAL2	protein coding	14.04005	484.7801
ENSG00000096384	HSP90AB1	protein coding	249.9771	482.9676
ENSG00000198918	RPL39	protein_coding	361.3886	460.7996

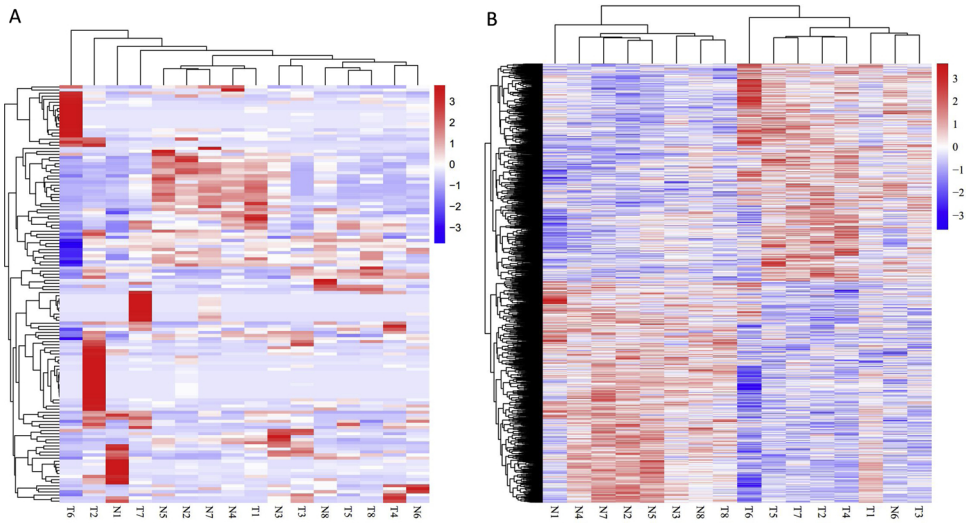

Fig. 4. The heat map and hierarchical clustering of the differentially expressed lncRNAs (a) and mRNAs (b) between normal and cancer samples.

Table 6
The top 10 up-regulated and 10 down-regulated lncRNA in tumor tissues.

Gene id	Gene name	Normal FPKM	Tumor FPKM	log2(fold change)	P value
ENSG00000254548.1	RP11-429J17.5	0.36169175	3.448151875	−3.252990723	0.018583
ENSG00000225768.1	RP11-127O4.3	0.114856875	0.950341375	−3.048608627	0.04399
ENSG00000224271.5	RP11-191L9.4	0.369097125	2.851944375	−2.949873438	0.000439
ENSG00000218357.3	LL22NC03-75H12.2	0.315174875	1.997651	−2.664080123	0.029554
XLOC_324131	–	0.163721875	0.94818225	−2.533917292	0.041175
XLOC_163714	–	0.48501325	2.545206875	−2.391686858	0.035804
XLOC_163712	–	0.4734865	2.320140375	−2.292816898	0.044643
ENSG00000228613.1	AC144450.1	1.65078125	7.0327315	−2.090936181	0.023233
ENSG00000261488.1	RP11-757F18.5	2.044084	6.90729	−1.756665312	0.043707
ENSG00000273576.1	RP11-390P24.1	4.421884	14.497324	−1.71305354	0.038067
ENSG00000254531.1	FLJ20021	2.476158625	1.493685	0.729227805	0.025339
ENSG00000255864.5	RP11-444D3.1	4.70513675	2.690764125	0.806220726	0.005876
ENSG00000214548.14	MEG3	2.142266875	1.175998125	0.865252456	0.033565
ENSG00000179818.13	PCBP1-AS1	2.006431375	1.094516625	0.874337945	0.006271
ENSG00000237438.6	CECR7	3.5568945	1.076878	1.723763368	0.048974
ENSG00000267532.3	MIR497HG	7.011348125	1.960515375	1.838458913	0.029331
XLOC_000333	–	136.3695363	36.37002088	1.906699736	0.007857
ENSG00000249669.7	MIR143HG	9.730973125	2.49426625	1.963968612	0.01,798
ENSG00000276649.1	RP5-858L17.1	2.78508725	0.541609625	2.362397242	0.03324
ENSG00000242908.6	AADACL2-AS1	3.568161375	0.549142875	2.699927404	0.029523
ENSG00000241684.5	ADAMTS9-AS2	4.560805	0.2512185	4.182273873	0.038657

lncRNAs. The differentially expressed lncRNAs from TCGA dataset and our sequencing data were compared.

2.7. Target gene prediction

All lncRNAs were subjectted to analysis of their Cis role or Trans role. For Cis role lncRNAs that are active on neighboring target genes, those coding genes located 10k/100k upstream and downstream of one lncRNA were searched, analyzed for their function and identified as potential target genes. For Trans role lncRNAs, the regulatory relationship (Transcriptome network analysis) was searched and analyzed based on the expressed correlation between lncRNAs and coding genes with Pearson correlation and WGCNA software [27]. Only the coding genes and lncRNAs that shared a common expression module (Pearson correlation, $R^2 > 0.95$) were identified as the co-expressed lncRNA and mRNA, the mRNA may be the potential.

2.8. GO and KEGG enrichment analyses

The GOseq R package [28] was employed to conduct the Gene Ontology (GO) enrichment analysis of differentially expressed lncRNA target genes. GO terms with corrected P values less than 0.05 were considered significantly differentially enriched lncRNA targets. To better understand the functions and impact of the biological system on the cell, the organism or the ecosystem level (http://www.genome.jp/kegg/,Kanehisa), KOBAS software was used to test the statistical enrichment of differential expression genes or lncRNA target genes in KEGG pathways [29].

2.9. Quantitative real-time PCR

At least three biological replicates were used for qPCR using Powers Green (Invitrogen). The comparative CT ($^{\Delta\Delta CT}$) method was used with values first normalized to the housekeeping RNA 18S. All primers were obtained from Integrated DNA Technologies (Coralville, IA, USA). Primer efficiency between 95 and 105% was determined for each primer candidate. The clinical experiment was prepared with total RNA from the 8 pairs of BC tissues and adjacent normal tissues, and first-strand cDNA was synthesized using a PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time). The expression levels of upregulated genes and downregulated genes selected as novel breast cancer-related genes were verified by qRT-PCR. The designed primers (listed in the Supplemental Table primers), PCR system and PCR amplification conditions were the same as those in the validation experiments. The data were analyzed using ABI 7500 H T SDS software 4.1 (Applied Biosystems, USA). DEG expression levels were analyzed using the $2^{-\Delta\Delta CT}$ analysis method.

3. Results

3.1. Patient demography and clinical features result

All volunteers in this study were the local people from the Hohhot, Inner Mongolia province, China (Table 1). To study the expression profile of lncRNAs and its target genes in the ER/PR positive type breast cancer, the status of ER and PR of all patients should be positive and > 2+ at the same time. At last, 8 female breast cancer patients aged from 34 to 53 years were selected as volunteers. The histology results of breast tissues from these volunteers were at least 2+ for both ER and PR. Four volunteers were ER/PR 3+ at the same time.

3.2. The results of sequencing and assembly of transcriptome data

In total, 16 different sequencing libraries (8 for breast cancer tissues and 8 for normal breast tissue) were constructed and sequenced successfully. The transcriptome results were named T1 to T8 for cancer tissues and N1 to N8 for normal tissues. To avoid the interference of genetic background and obtain the expression pattern of lncRNAs as well as its co-expressed gene during the development of breast cancer (genes in the ER/PR positive type breast cancer ER/PR(+)BC), ER(+)/PR(+) BC and normal breast tissues were collected from the same individual breast cancer patient (for example, T1 and N1 were obtained from the same patient volunteer with ER/PR + type breast cancer). After sequencing, a total of 1,853,898,248 paired-end raw reads with an error rate less than 0.02% were generated (Table 2). After removing the low-quality reads, redundant raw reads, adaptor sequence and joint base-containing sequences, 1,784,412,960 (96.25% of raw reads) clean reads were kept and used for transcriptome annotation and analysis a summary table for assembly was deposited in Table 2. The low error rate and high proportion of clean data suggest that the sequencing results were acceptable. To obtain an optimized reference transcriptome database and to better understand the transcriptomic lncRNA profile of ER/PR + type BC, all reads from the 16 different transcriptomes were mapped to the human genome data.

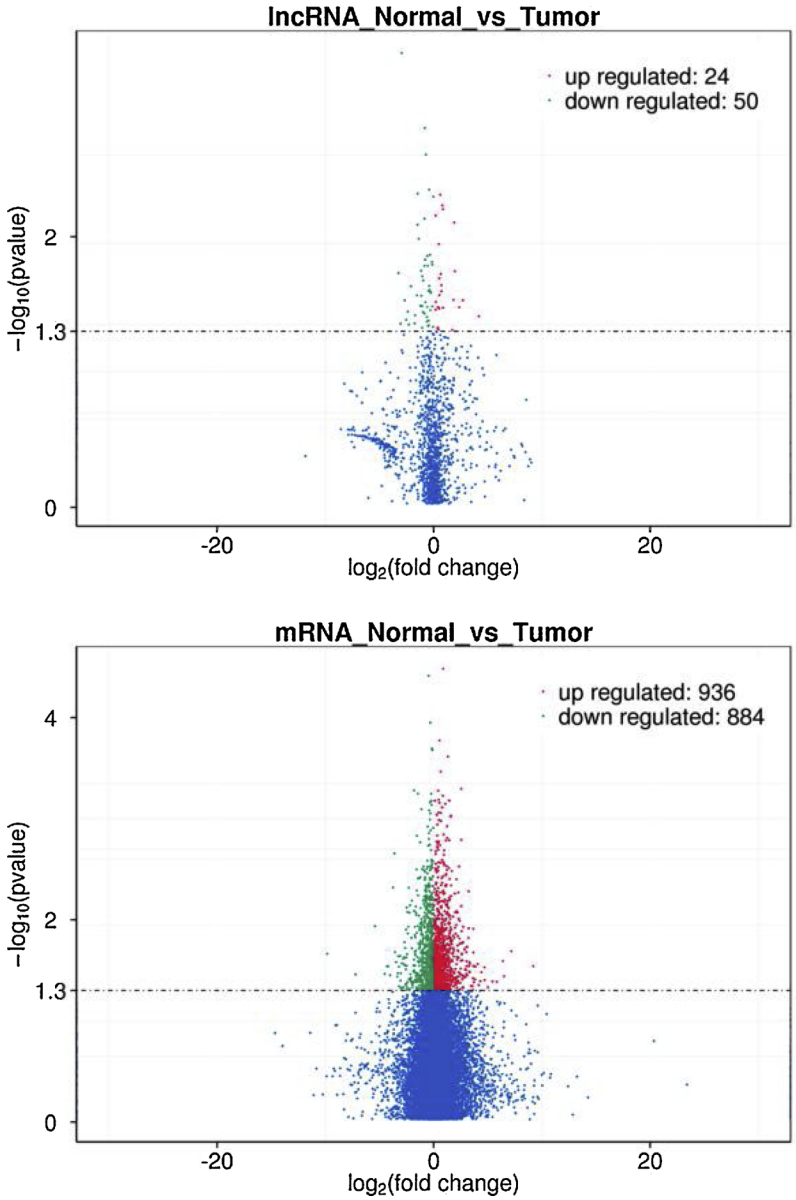

Fig. 5. The scatter plot of lncRNA (a) and mRNA (b) expression variation between the normal and cancer samples.
The red dot were up-regulated gene or lncRNA, the green dot were down-regulated gene or lncRNA and blue dot were genes or lncRNA without any change.

3.3. Identification of lncRNA and coding genes

For each sample (tumor tissues and normal tissues), the 1,784,412,960 sequence reads, and reference genome sequence were aligned using TopHat software. In total, 161,401 different transcripts were obtained and annotated. All transcripts were divided into 8 different types (Fig. 1 A and B). The results showed that approximately 80% of the transcripts were protein coding sequence(mRNA), and the lncRNAs comprised only approximately 3.4–6.7% of the transcripts. Meanwhile, the number of lincRNAs in the breast cancer tissues (Fig. 1B) is almost twice that in the normal breast tissues (Fig. 1A). In total, more than 3% of the protein coding transcripts in the breast tumor tissues were replaced by lncRNAs (Fig. 1A and B). Based on the referenced genome database and gene model file, most transcripts (146,252 transcripts) were annotated as coding genes or mRNAs, 5,516 transcripts were annotated as lncRNAs (Supplemental Table 1), and 9,632 were annotated as TUCP (transcripts of uncertain coding potential). Further analysis suggests that the 146,252 coding gene sequences belonged to 19,278 different genes (Supplemental Table 2).

During the identification progress of novel lncRNAs, transcripts from all samples failed to be annotated was performed another annotation progress that was referenced as the CPC, PFAM, phyloCSF and CNCI databases (Fig. 2). The numbers of transcripts that were annotated in the four databases were 9,032 (CPC), 4,876 (PFAM), 11,000 (phyloCSF) and 11,638 (CNCI). To ensure that the annotated transcripts were not false positive results (potential coding proteins), any transcripts successfully annotated by one database were filtered. Finally, only these transcripts confirmed as noncoding genes by all four protein databases were identified as novel lncRNAs. The number of novel lncRNAs from both breast cancer and normal tissue transcriptomes was 3,438 (Fig. 2, Supplemental Table 3). In total, 8,954 lncRNAs, including 5,516 lncRNA reported in the previous studies and 3,438 novel lncRNA confirmed by coding potential analysis, were annotated in this study. All these lncRNAs could be divided into 8 different lncRNA subtypes. Of the 3,438 novel lncRNAs, 2,993 lncRNAs were lncRNAs, and 445 were antisense lncRNAs (Table 3). The annotated 5,516 known lncRNAs belonged to 8 subtypes, including 2,051 lncRNAs, 1,980 antisense lncRNAs and only 1 macro lncRNA (Table 3). All 8,954 lncRNA covered

Table 7
The top 10 up-regulated and 10 down-regulated mRNA in tumor tissues.

Gene id	Gene name	Normal FPKM	Tumor FPKM	log2(foldchange)	P value
ENSG00000128918	ALDH1A2	1.928955	0.003254	9.211445	0.028752
ENSG00000094963	FMO2	3.509232	0.024225	7.178515	0.020483
ENSG00000077782	FGFR1	8.759106	0.097027	6.496254	0.036156
ENSG00000088992	TESC	1.516981	0.018053	6.39283	0.025533
ENSG00000165795	NDRG2	34.37104	0.601665	5.83609	0.041117
ENSG00000094755	GABRP	6.618527	0.153148	5.433512	0.040795
ENSG00000100146	SOX10	19.16088	0.587788	5.026725	0.046444
ENSG00000133392	MYH11	29.38619	1.125787	4.706133	0.03944
ENSG00000094755	GABRP	19.53561	0.756732	4.690181	0.029591
ENSG00000163017	ACTG2	5.912065	0.277506	4.413069	0.032957
ENSG00000050438	SLC4A8	0.263188	3.119803	−3.56729	0.020867
ENSG00000135272	MDFIC	0.295957	3.632073	−3.61733	0.002212
ENSG00000002834	LASP1	1.120908	15.13504	−3.75515	0.004802
ENSG00000170571	EMB	0.331906	5.386099	−4.0204	0.035283
ENSG00000007541	PIGQ	0.070927	1.301169	−4.19733	0.028961
ENSG00000080503	SMARCA2	0.095967	1.908851	−4.31403	0.028778
ENSG00000170442	KRT86	0.056961	1.303841	−4.51666	0.034335
ENSG00000100883	SRP54	0.14119	6.054887	−5.42239	0.011586
ENSG00000160221	C21orf33	0.009694	1.438299	−7.21313	0.034695
ENSG00000102158	MAGT1	0.003593	3.285402	−9.83677	0.021676

Table 8
The important Ras family gene and its associated lncRNAs in breast cancer.

lncRNA Gene ID	Gene Symbol	Gene Description	Pearson correlation	P value
ENSG00000225092.2	RREB1	ras responsive element binding protein 1	0.96	2.06E-09
ENSG00000231133.6	RASSF6	Ras association domain family member 6	0.95	1.75E-08
ENSG00000258520.1	RASSF8	Ras association domain family member 8	0.97	4.67E-10
XLOC-031380	RHOT1	ras homolog family member T1	0.95	1.79E-08
XLOC-124864	RAP1B	RAP1B, member of RAS oncogene family	0.95	8.53E-09
XLOC-262804	RASGEF1B	RasGEF domain family member 1B	0.95	1.67E-08
XLOC-358359	RASGRF1	Ras protein specific guanine nucleotide releasing factor 1	0.95	1.81E-08
ENSG00000254548.1	ARHGAP39	Rho GTPase activating protein 39	0.950378571	1.74E-08
ENSG00000262370.5	KALRN	kalirin, RhoGEF kinase	0.950973204	1.6E-08
ENSG00000236008.1	RAB11FIP1	RAB11 family interacting protein 1	0.95000194	1.83E-08

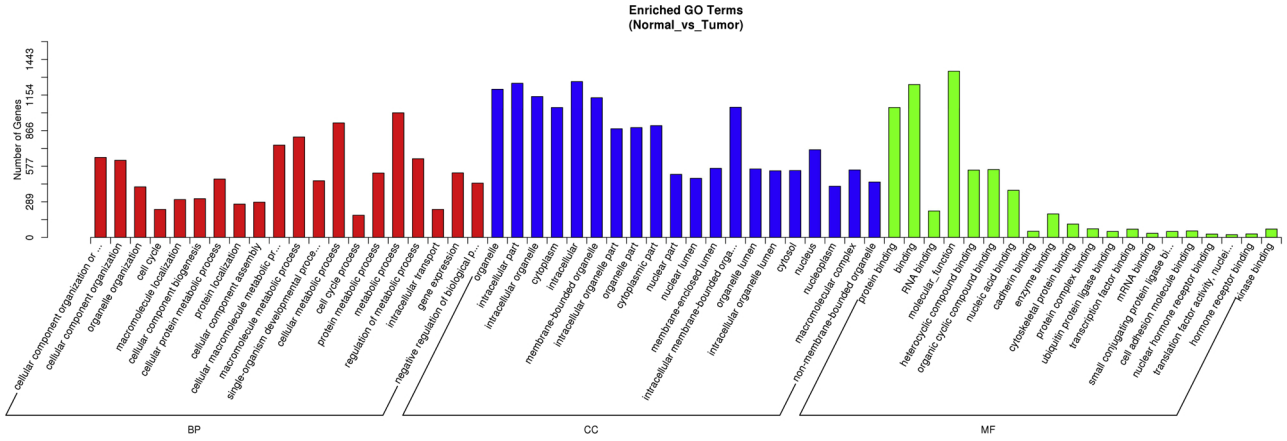

Fig. 6. GO terms which enriched the different expressed lncRNA.
BP is short for biological; CC is short for cellular; MF is short for molecular function.
The horizontal coordinate is the different classification, the vertical coordinate is the number of genes, and green is the sense strand and red is antisense strand.

the entire genome sequences. As for the location of lncRNA in different chromosomes (Fig. 3), lncRNAs could be mapped on all chromosomes except the front end of chr13, chr 14, chr15 and the second half of chr Y.

3.4. Expression profile of lncRNAs and coding genes in the breast transcriptome

All 1,222 different breast tissues samples from TCGA database generated 60,483 transcripts. Of these transcripts, 6,603 were annotated as lncRNAs. All annotated 5,516 known lncRNAs in our transcriptome result could be found in the lncRNAs from TCGA database. As for the novel lncRNAs, only 917 lncRNAs were the same with the TCGA lncRNAs.

The expression levels of lncRNAs and coding genes in breast cancer tissues relative to those in adjacent normal breast tissues were calculated by the FPKM method. The mean FPKM of lncRNAs in cancer and normal tissues is 3.8 and 2.7, respectively. The relative expression level

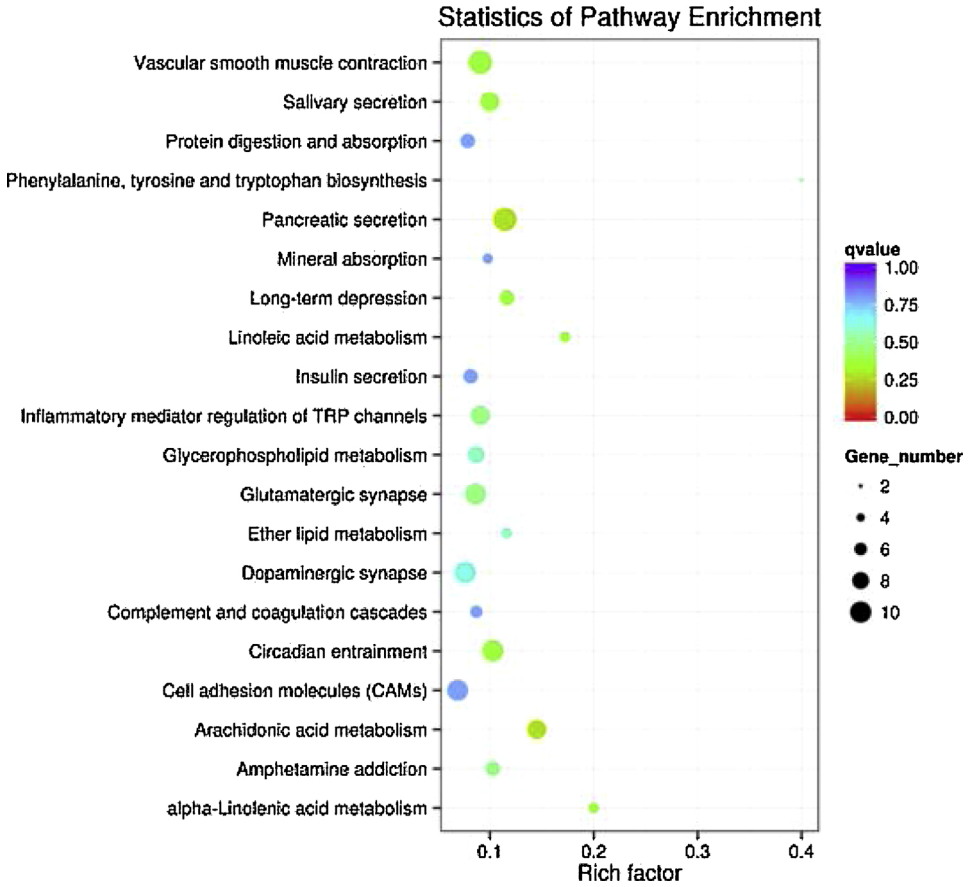

Fig. 7. KEGG pathway which enriched the different expressed lncRNA.

of lncRNA in tumor tissues is higher than that in normal tissues. The top 20 highest expressing lncRNAs included MALAT1, SCARNA10, RP11-206M11.7 and NEAT1 (Table 4), followed by LINC00993 and RP11-20F24.2. Of the top 20 highest lncRNAs in cancer tissues, 11 lncRNAs showed a higher expression level than that in normal tissues. Only two lncRNAs (CDR1-AS and SCARNA9) were expressed at higher levels in normal tissues. In particular, RP11-616M22.3 and RP11-675F6.3 were both the lncRNA type, and almost no expression was detected in normal tissues. The mean FPKM of mRNAs in both cancer and normal tissues is approximately 2.16. The top 20 highest expressing mRNAs included RPL19, SCGB2A2, FTL and TMSB4X (Table 5), followed by COL1A1 and TMSB10. Almost all the top 20 genes were upregulated in tumor tissues, except RPS27 (1.5-fold higher in normal tissue than in tumor tissue) and RPL41 (the same expression level in both tumor and normal tissues).

Of 6,603 lncRNAs from TCGA database, 2,688 lncRNAs were different expressed lncRNAs (1743 up-regulated and 945 down-regulated with a fold change > 2, $P < 0.05$). In total, 615 lncRNAs in our data showed different expression pattern ($P < 0.05$) in breast cancer compared to normal tissues (Supplemental Table 4). Of these differentially expressed lncRNAs, 323 were upregulated ((Fig. 4 A). The expression levels of the top 10 upregulated and 10 downregulated lncRNAs are listed in Table 6. All these lncRNAs were significantly changed ($P < 0.05$) in breast cancer compared with normal tissue. The highest fold-change (downregulated lncRNA) was ADAMTS9-AS2, which was 16 fold-change less than that in normal tissues, followed by AADACL2-AS1, RP5-858L17.1, and MIR143HG. The highest fold-change (upregulated lncRNA) was observed with RP11-429J17.5, RP11-127O4.3 and RP11-191L9.4, which all had a fold-change of more than 8. Considering all differentially expressed lncRNAs in all tested samples, only 74 lncRNAs (24 upregulated and 50 downregulated lncRNAs) showed the same expression pattern in all 8 transcriptome pairs (Fig. 5A). In total, 54different expressed lncRNAs including 22 lncRNA down-regulated and 32 up-regulated could compared to the TCGA data.

As for coding genes or mRNAs, 8,956 genes were differentially expressed. Of all 8,956 genes, 5,517 genes were upregulated in breast cancer, and 3,439 were downregulated. Meanwhile, the most differentially expressed mRNAs (upregulated in normal tissues) were ALDH1A2, FMO2, TESC, NDRG2 and FGFR1 (Table 7), and the mRNAs with the highest fold-change (downregulated mRNA) were MAGT1, C21 and SRP54, with fold-changes ranging from 64 to 256. Considering all differentially expressed mRNAs in all the tested samples, only 1,820 mRNAs (936 upregulated and 884 downregulated) showed the same expression pattern in all 8 transcriptome pairs (Figs. 4B and 5 B and Supplemental Table 5).

3.5. lncRNA target prediction in breast cancer

Cis lncRNAs are lncRNAs that are active on neighboring target genes. All coding genes located on the 10k/100k regions upstream and downstream of the lncRNA were searched. In total, of the 8,955 high quality lncRNAs, the target genes of 5,543 lncRNAs were predicted. Furthermore, 189,697 coding genes could be target genes for these lncRNAs. Moreover, our data showed that one lncRNA may correlate with 1679 mRNAs and that one mRNA may correlate with 1418 lncRNAs. Transcriptome (lncRNA and mRNA co-expression) network analysis was used to assess the target genes of lncRNAs (trans role lncRNAs) in breast cancer. In total, 3,217 lncRNAs were co-expressed with 10,500 coding genes ($R^2 > 0.95$). Considering all cis type and trans type lncRNAs, 2,079 lncRNAs could be annotated as both cis and trans. Of these, 238 lncRNAs shared the same target by the two different lncRNA target prediction results (Supplemental Table 6 and

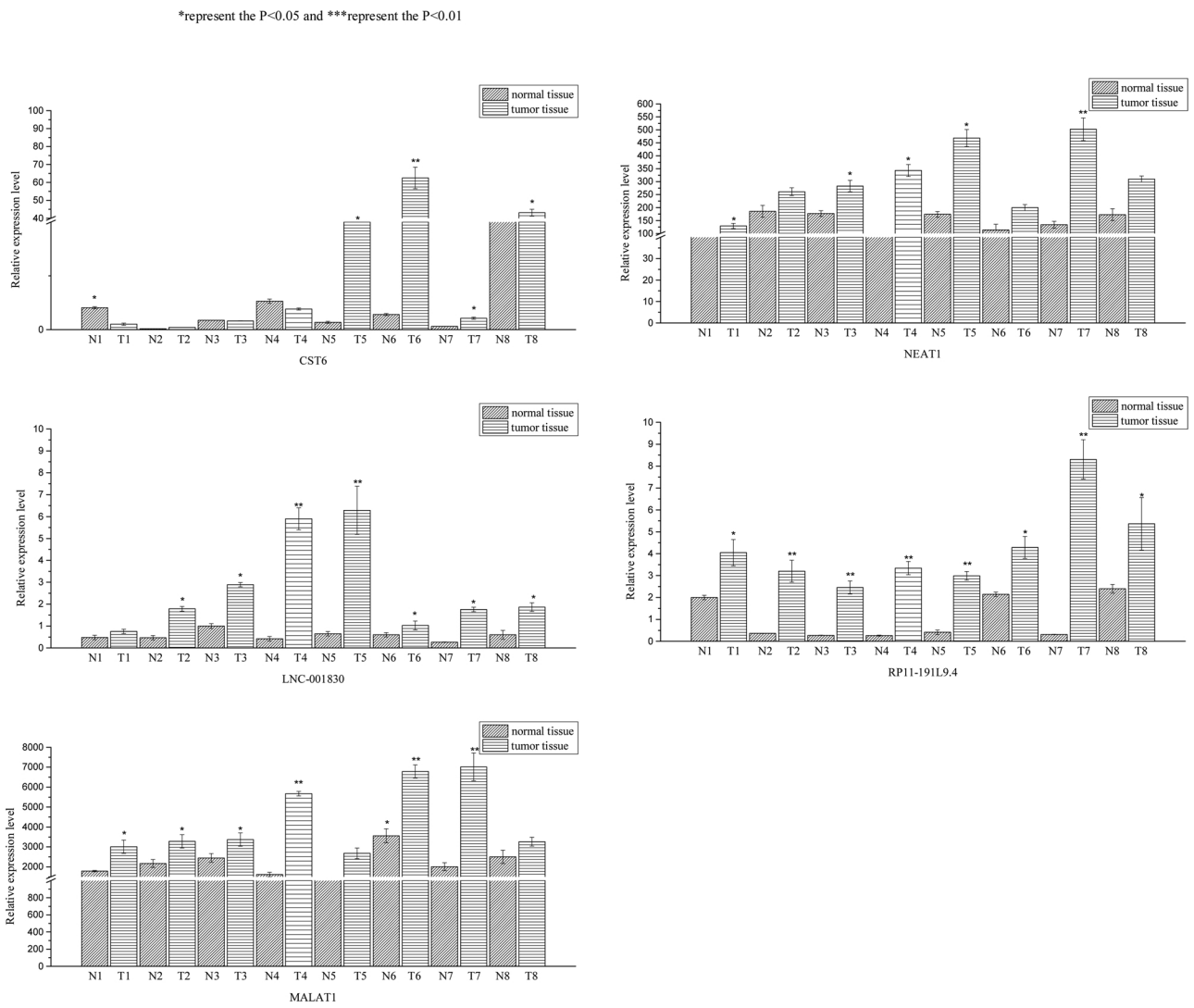

Fig. 8. The relative expression of up-regulated lncRNA in normal tissues and tumor tissues via qRT-PCR. The x-axis represents the samples, and the y-axis represents the relative gene expression levels.
*represent the $P < 0.05$ and ***represent the $P < 0.01$.

Table 8). Within these genes, domain family 2-member D, Ras responsive element binding protein 1, Ras association domain family member 6, Ras association domain family member 8, Ras protein specific guanine nucleotide releasing factor 1, Wilms tumor 1, fibroblast growth factor 1 and estrogen related receptor gamma were predicted as targets of lncRNAs (Table 8). These lncRNAs and their target genes may be important molecules that are associated with the development of breast cancer. In total, 10 different Ras family genes and its associated lncRNAs in breast cancer were annotated in breast cancer. All these Ras family genes or Ras association genes RREB1, RASSF6, RASSF8, RHOT1, RAP1B, RASGEF1B, RASGRF1 (Table 8) were different expressed genes and the Pearson correlation with its associated lncRNAs were higher than 0.95. On the other hand, the location of these genes and lncRNA are nearby which were not more than 100 K.These genes and lncRNAs may be important biomarkers in breast cancer.

3.6. GO analysis and KEGG enrichment analysis

To predict the potential functions of the lncRNAs and their target genes, the GO terms of the target genes for all differentially expressed lncRNAs were enriched and are shown in Fig. 6. The GO enriched analysis indicated that the most enriched GO terms targeted by the mRNAs coexpressed with lncRNAs were molecular function and binding, which belongs to the molecular function term, followed by intracellular function. Furthermore, the KEGG pathway enrichment analysis suggested that the mRNAs co-expressed with lncRNAs were involved in alph A-L inolenic acid metabolism (Rich factor 0.2) and linoleic acid metabolism (Rich factor 0.17). The 20 enriched KEGG pathways are listed in Fig. 7. The cell adhesion molecules (CAMs), complement and coagulation cascades and insulin secretion were identified in the KEGG pathway analysis as containing the highest enrichment of DELs (differentially expressed lncRNAs).

3.7. Clinical experiment with qRT-PCR

To determine the clinical effects of all selected genes, including high-level and low-level lncRNAs and their target genes, all primers were designed based on the lncRNAs and coding mRNA sequences reported in the NCBI database (Supplemental Table primers). We screened the 5 highest expressing and 5 lowest expressing lncRNAs to detect the expression patterns of these lncRNAs in breast cancer tissues and adjacent tissues from 16 different clinical individuals by qRT-PCR. The results showed that the upregulated lncRNAs NEAT1, MALAT1, CST6, RP11-191L9.4, and LNC-001830 (Fig. 8) as well as the down-regulated lncRNAs ADAMTS9-AS2, MIR497HG, CECR7, AADACL2-AS1, and RP5-858L17.1 (Fig. 9) showed the same result in different patients,

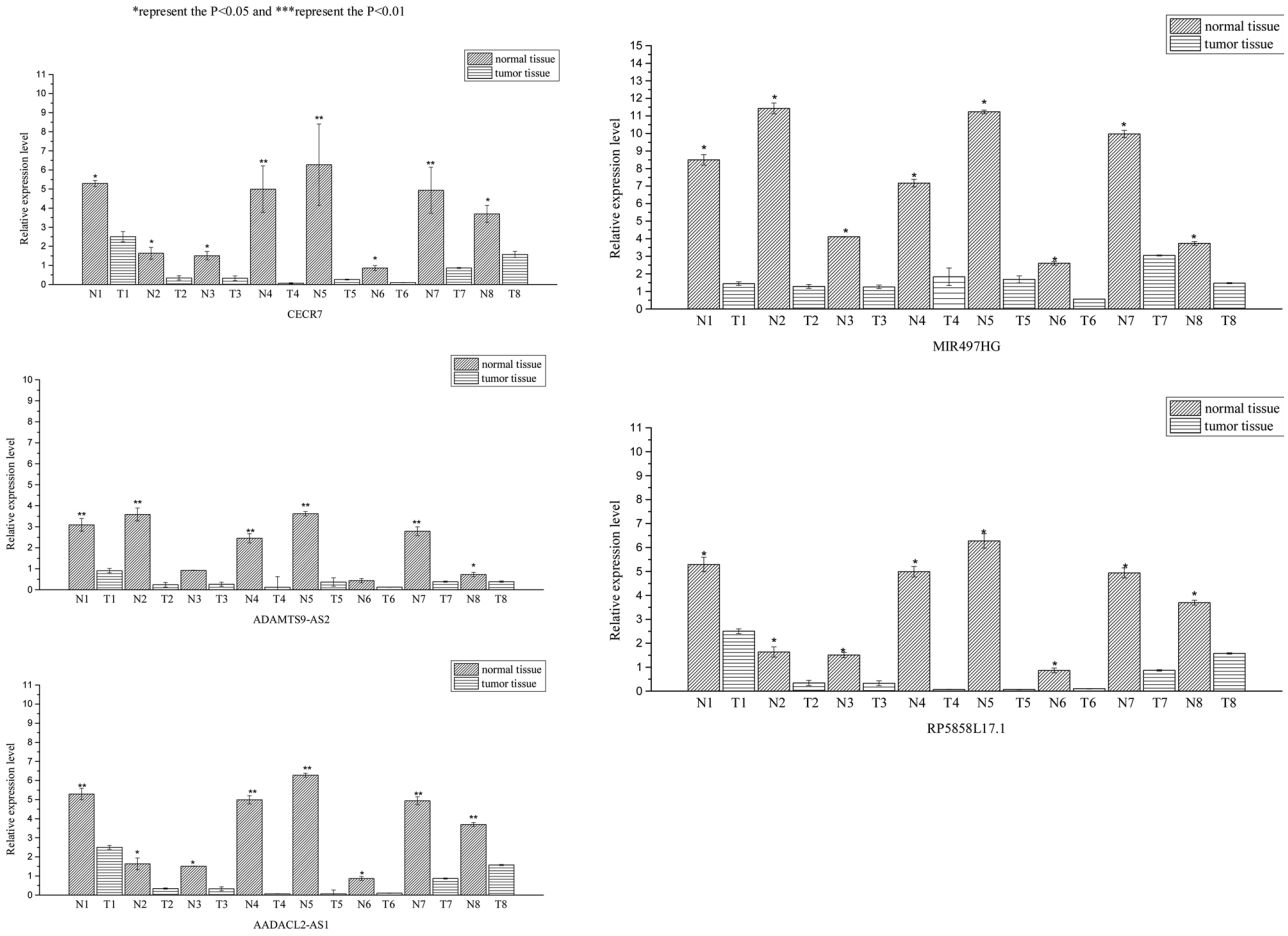

Fig. 9. The relative expression of down-regulated lncRNA in normal tissues and tumor tissues via qRT-PCR. The x-axis represents the samples, and the y-axis represents the relative gene expression levels.
*represent the P < 0.05 and ***represent the P < 0.01.

which verified the RNA-Seq results. Meanwhile, the expression pattern of three important coding mRNAs (HSP90AB1, TMSB10, and COL1A2) (Fig. 10) were also verified by qRT-PCR.

4. Discussion

Different type or subtype of cancers are diseases caused by genetic changes. The expression patterns on transcriptome level for mRNA and non-coding RNA are a molecular basis for development of tumors, the different profile represent the biological and histological diversity of tumor type or subtype [12,30]. For these reasons, the study of different kinds of subtype cancers was meaningful to better understand the development of breast cancer. LncRNAs are a set of RNAs that do not code proteins but could regulate the expression profile of target genes at the molecular level [31]. Analysis of the expression pattern of lncRNAs and co-expressed mRNAs may provide us with new insights to explain the mechanism of the occurrence and progression of cancer. In previous studies, lncRNAs in many subtype breast cancers, such as ER (+), ER (-), Her2 (-) and triple negative (ER-, PR- and Her2 -) breast cancer were sequenced. The whole specific lncRNA transcriptome expression profile which could reflect the progression stages of different subtype breast cancer. The transcriptome profile of double positive ER/PR (+) type breast cancer was still undercover. In this study, the ER (2+)/PR (2+) type breast cancer were sampled and sequenced to identify its specific lncRNAs expression profile may provide us a new sight to understand the reason for the high survival rate. The confirmation of a breast cancer diagnosis is complex because the expression pattern or profile of breast cancer changes with age, developmental stages and breast cancer subtype [32,33]. To avoid differences between individuals, breast cancer tissues and normal tissues in this study were collected from the same individual. We focused on the genetic information and removed the influence of environment, food and other factors.

Due to the important role of non-coding RNAs in cancer, many studies have focused on the discovery of non-coding RNA as well as the correlation between the ncRNAs and coding-RNA which were involved in cancer [34]. The discovery of differentially expressed miRNAs [35] and novel lncRNAs [36] and functional analysis of lncRNAs [37] were performed successfully. Meanwhile, as a useful experimental method, high-throughput sequencing provides us with an easy way to undercover vast amounts of lncRNA data at a lower cost. Transcriptome sequencing has been successfully employed to reveal lncRNAs in many types of cancers, such as renal cells [38], colon cancer cells [39], gastric cancer [40], lung cancer [41] and breast cancer [42]. In the present study, 5,516 lncRNAs reported in the previous studies and 3,439 novel lncRNAs were obtained. When compared with the lncRNA annotated from TCGA database, all known lncRNAs were found in the TCGA database. Meanwhile, almost 2500 novel lncRNAs were faild to mapped to the TCGA database. The number of lncRNAs identified in double positive ER/PR (+) type breast cancer is greater than that in the lung transcriptome [41] and the breast cancer transcriptome [42], which may be because of the improvement of the reference genome database and the development of software. The assembly result reveals that breast cancer tissues express more lncRNAs than normal tissues, the extra lncRNAs may be important regulator in breast cancer tissues. Meanwhile, many novel lncRNAs may be false positive, these lncRNAs may be part of untransltion region of mRNA [43].

*represent the P<0.05 and ***represent the P<0.01

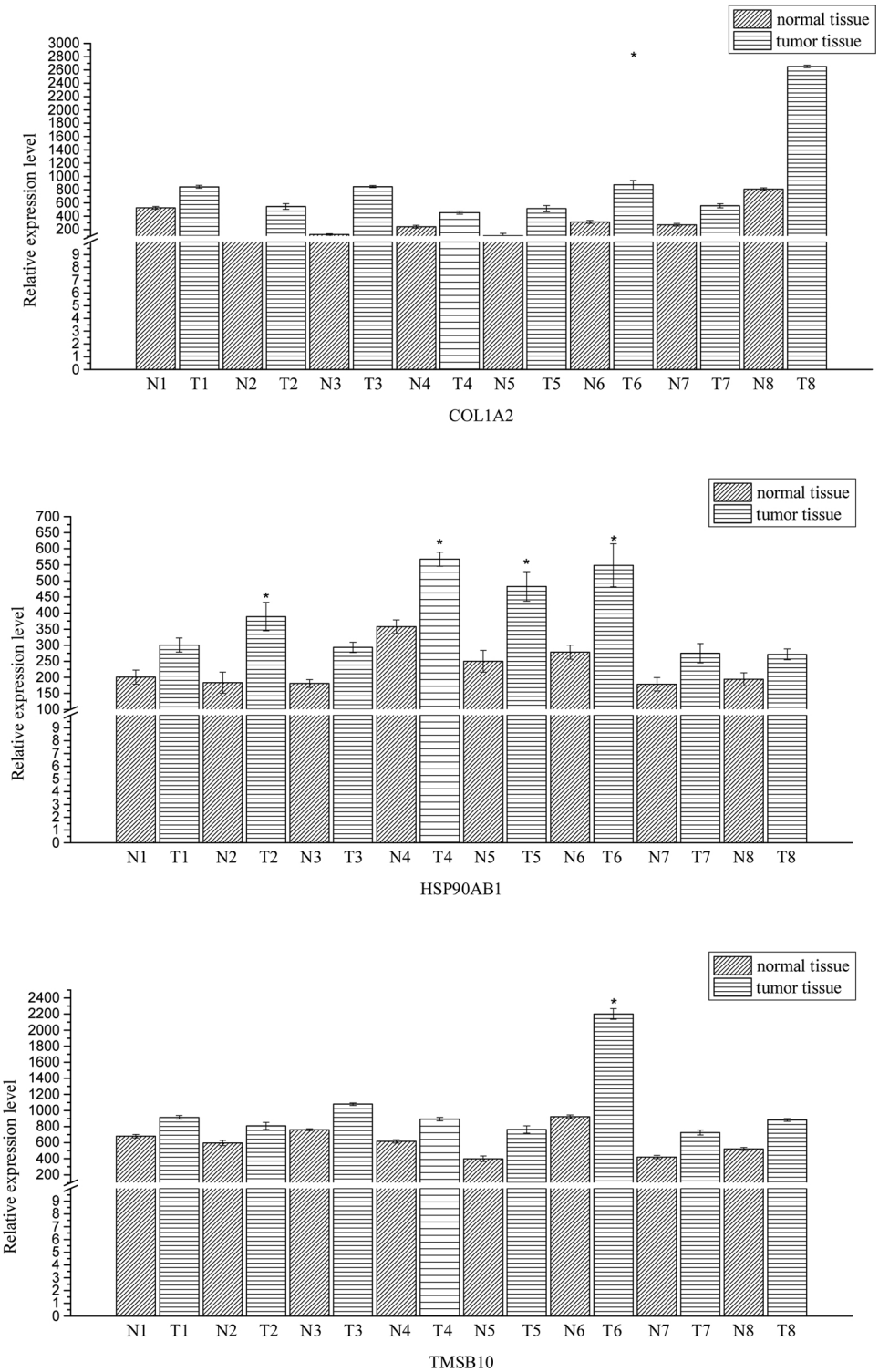

Fig. 10. The relative expression of coding genes in normal tissues and tumor tissues via qRT-PCR. The x-axis represents the samples, and the y-axis represents the relative gene expression levels.
*represent the P < 0.05 and ***represent the P < 0.01.

In the present study, MALAT1 and NEAT1 were annotated as lncRNAs with the highest expression level in breast cancer. MALAT1 (targets the SCVL1 gene) has been found to function in the post-transcriptional stage and in lung cancer [44,45]. MALAT1 showed different expression pattern in the ER + type breast cancer before and after the treatment of Tamoxifen. [46].NEAT1 is another important lncRNA that targets FRMD8 during progesterone production [47,48], suppresses MHC class II ag and involved in the formation of paraspeckle [49,50] COL1A1 is a lncRNA that binds to the estrogen receptor gene and functions in the development of breast cancer [51]. NEAT1 and

COL1A1 related to progesterone and estrogen respectively, the tissues were sampled from double positive ER/PR (+) type breast cancer patients. The higher expression level of NEAT1 and COL1A1 suggest that the level of progesterone receptor and estrogen receptor were up-regulated. These lncRNAs NEAT1 and mRNA COL1A1 may be new break out for the treatment of ER/PR (+) type breast cancer.

Unlike the protein-coding mRNAs, lncRNAs have very complex functions that are difficult to assess [16]. To date, most of the lncRNAs have been annotated by bioinformation analysis and have not yet been confirmed by experimental methods. The reference database of lncRNA function is very limited. For these reasons, the functional research of lncRNAs must be performed along with the study of their target genes, which means that target gene prediction for lncRNAs is very important and urgent. For Cis lncRNAs that are active on neighboring target genes, colocalization is the first and most important parameter to assess. In this study, combined with all the Cis and trans results, only 238 lncRNAs shared the same target based on the two different lncRNA target prediction results. Within these genes, three complement genes, C3, C6 and C-type lecti from the complement pathway were annotated. Several studies have noted that the complement system may affect the form of the tumor. Persistent complement activation in breast cancer and C3 were present only in carcinoma samples and showed pathogenetic effects in breast cancer [52,53]. Wilms' tumor 1 is another important cancer-related gene, and downregulation of WT1 protein could inhibit the proliferation of breast cancer [54,55]. Estrogen-related receptor gamma is a therapeutic target in cancer [56] and can promote mesenchymal-to-epithelial transition and suppress the growth of breast tumors [57], the discovery of the lncRNA for estrogen-related receptor gamma which may be a new regulation factor. Our samples were ER/PR (+) type breast cancer, this novel lncRNA may be a new therapeutic target.

Co-expression analysis reveals that Ras responsive element binding protein 1, Ras association domain family member 6 and other 10 Ras gene family were found a co-expressed with 10 different lncRNAs. Ras gene family played important role in the cell growth and differentiation, cellular transformation [58], integrin-mediated cell adhesion [59,60] and processes including tumor angiogenesis. In prostate cancer cells, Ras responsive element binding protein-1 (RREB-1) could down-regulate the expression level of hZIP1 [61,62]. Many researches reveal that there are relations between the occurrence and progression of cancer and Ras gene family. Ras genes (HRAS, NRAS and KRAS) are human oncogenes and found to be mutated in many kinds of human cancers. More and more evidence suggested that Ras functioned as tumor-induce role. A study on breast cancer suggest that the R-Ras 38 V could affect the cell migration, cell invasion and stimulate the estrogen-independent proliferation of breast cancer cells [63]. Meanwhile, the Ras has been implicated in breast tumor inducing [64]. Many mechanisms could affect the oncogenic of Ras gene family. These different expressed lncRNAs target on Ras genes could be a new method which performed the regulation role in the development of ER/PR (+) type breast tumor. In the KEGG annotation result, the ECM-receptor interaction pathways (ECM) and cell adhesion molecules (CAMs) enriched the genes which regulated by different expressed lncRNAs. Ras could upregulate the integrins which could active by extracellular matrices (ECM) [65,66]. These different expressed lncRNAs which could regulate the Ras gene family and ECM pathway may be another mechanism why the expression pattern of Ras genes changed in breast cancer. All these cancer-related genes (Ras genes) were annotated as targets of lncRNAs in the breast cancer transcriptome and may provide us with a new way to understand the occurrence and development of breast cancer.

5. Conclusion

In this study, 16 different transcriptomes (8 ER/PR (+) type breast cancer tissues and 8 normal breast tissues) were successfully sequenced. In total, 8,954 high quality lncRNAs, including 5,516 lncRNAs reported in the previous study and 3,438 novel lncRNAs confirmed by coding potential analysis, were annotated. The highest expressing lncRNAs were MALAT1, SCARNA10, RP11-206M11.7 and NEAT1, and the highest expressing mRNAs were RPL19, SCGB2A2, FTL and TMSB4 × . Of these 615 differentially expressed lncRNAs, 323 showed upregulated ($P < 0.05$) expression patterns in breast cancer, and 292 showed downregulated expression patterns. Of all 8,954 coding genes, 5,516 genes were up-regulated and 3,438 were down-regulated in breast cancer. In total, the targets of 238 lncRNAs were confirmed by two lncRNA target prediction tools. Among these genes, Ras responsive element binding protein 1, Ras association domain family member 6, Ras association domain family member 8, Ras protein specific guanine nucleotide releasing factor 1, and other 10 Ras association genes were predicted as the targets of lncRNAs. Ras gene family played important role in the cell growth and differentiation, cellular transformation, integrin-mediated cell adhesion and processes including tumor angiogenesis. The Ras has been implicated in breast tumor suppression. These new lncRNAs which could regulate the RAS genes could be a new mechanism to explain the important role of Ras genes in breast cancer. All these lncRNAs targeting on cancer-related genes (Ras genes) in the breast cancer transcriptome may provide us a new way to understand the occurrence and development of breast cancer.

Declarations of interest

None.

Acknowledgments

This work has received funding from the Inner Mongolia Autonomous Region Tumor Biotherapy Collaborative Innovation Center (grant No 2016ZLXT006). In addition, this work was supported by grants from the Inner Mongolia Natural Science Fund (grant No 2018MS08088), and from the Inner Mongolia Autonomous Region Universities "Innovation Research Team project"(grant No NMGIRT-A1604).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prp.2019.03.033.

References

[1] W. Chen, R. Zheng, P.D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X.Q. Yu, J. He, Cancer statistics in china, 2015, CA Cancer J. Clin. 66 (2) (2016) 115–132.
[2] A.M. Aravanis, M. Lee, R.D. Klausner, Next-generation sequencing of circulating tumor DNA for early cancer detection, Cell 168 (4) (2017) 571–574.
[3] W.D. Foulkes, I.E. Smith, J.S. Reis-Filho, Triple-negative breast cancer, N. Engl. J. Med. 363 (20) (2010) 1938–1948.
[4] K.D. Voduc, M.C. Cheang, S. Tyldesley, K. Gelmon, T.O. Nielsen, H. Kennecke, Breast cancer subtypes and the risk of local and regional relapse, J. Clin. Oncol. 28 (10) (2010) 1684–1691.
[5] T. Walsh, King M-C: Ten genes for inherited breast cancer, Cancer Cell 11 (2) (2007) 103–105.
[6] D.R. Rhodes, A.M. Chinnaiyan, Integrative analysis of the cancer transcriptome, Nat. Genet. 37 (6s) (2005) S31.
[7] M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, A pathology atlas of the human cancer transcriptome, Science 357 (6352) (2017) eaan2507.
[8] W. Zhao, J. Luo, S. Jiao, Comprehensive characterization of cancer subtype associated long non-coding rnas and their clinical implications, Sci. Rep. 4 (6591) (2014).
[9] K.R. Bauer, M. Brown, R.D. Cress, C.A. Parise, V. Caggiano, Descriptive analysis of estrogen receptor (er)-negative, progesterone receptor (pr)-negative, and her2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the california cancer registry, Cancer 109 (9) (2007) 1721–1728.
[10] R. Dent, M. Trudeau, K.I. Pritchard, W.M. Hanna, H.K. Kahn, C.A. Sawka, L.A. Lickley, E. Rawlinson, P. Sun, S.A. Narod, Triple-negative breast cancer: clinical features and patterns of recurrence, Clin. Cancer Res. 13 (15) (2007) 4429–4434.

[11] S. Gruvberger, M. Ringnér, Y. Chen, S. Panavally, L.H. Saal, Å Borg, M. Fernö, C. Peterson, P.S. Meltzer, Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns, Cancer Res. 61 (16) (2001) 5979–5984.

[12] T. Sørlie, C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, M. Van De Rijn, S.S. Jeffrey, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, Proc. Natl. Acad. Sci. 98 (19) (2001) 10869–10874.

[13] X. Su, G.G. Malouf, Y. Chen, J. Zhang, H. Yao, V. Valero, J.N. Weinstein, J.-P. Spano, F. Meric-Bernstam, D. Khayat, Comprehensive analysis of long non-coding rnas in human breast cancer clinical subtypes, Oncotarget 5 (20) (2014) 9864.

[14] M.K. Iyer, Y.S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T.R. Barrette, J.R. Prensner, J.R. Evans, S. Zhao, The landscape of long noncoding rnas in the human transcriptome, Nat. Genet. 47 (3) (2015) 199.

[15] G.S. Laurent, C. Wahlestedt, P. Kapranov, The landscape of long noncoding rna classification, Trends Genet. 31 (5) (2015) 239–251.

[16] M. Huarte, The emerging role of lncrnas in cancer, Nat. Med. 21 (11) (2015) 1253.

[17] A.C. Mallory, A. Shkumatava, Lncrnas in vertebrates: advances and challenges, Biochimie 117 (2015) 3–14.

[18] Piao H-l, L. Ma, Non-coding rnas as regulators of mammary development and breast cancer, J. Mammary Gland Biol. Neoplasia 17 (1) (2012) 33–42.

[19] K. Polyak, Heterogeneity in breast cancer, J. Clin. Invest. 121 (10) (2011) 3786–3788.

[20] J.L. Rinn, M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, Functional demarcation of active and silent chromatin domains in human hox loci by noncoding rnas, Cell 129 (7) (2007) 1311–1323.

[21] P. Schorderet, D. Duboule, Structural and functional differences in the long non-coding rna hotair in mouse and human, PLoS Genet. 7 (5) (2011) e1002071.

[22] X. Yan, Z. Hu, Y. Feng, X. Hu, J. Yuan, S.D. Zhao, Y. Zhang, L. Yang, W. Shan, Q. He, Comprehensive genomic characterization of long non-coding rnas across human cancers, Cancer Cell 28 (4) (2015) 529–540.

[23] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (4) (2012) 357.

[24] M. Pertea, D. Kim, G.M. Pertea, J.T. Leek, S.L. Salzberg, Transcript-level expression analysis of rna-seq experiments with hisat, stringtie and ballgown, Nat. Protoc. 11 (9) (2016) 1650.

[25] A. Siepel, G. Bejerano, J.S. Pedersen, A.S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L.W. Hillier, S. Richards, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes, Genome Res. 15 (8) (2005) 1034–1050.

[26] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (5) (2010) 511.

[27] P. Langfelder, S. Horvath, Wgcna: An r package for weighted correlation network analysis, BMC Bioinformatics 9 (1) (2008) 559.

[28] M.D. Young, M.J. Wakefield, G.K. Smyth, A. Oshlack, Gene ontology analysis for rna-seq: accounting for selection bias, Genome Biol. 11 (2) (2010) R14.

[29] X. Mao, T. Cai, J.G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the kegg orthology (ko) as a controlled vocabulary, Bioinformatics 21 (19) (2005) 3787–3793.

[30] T. Sørlie, R. Tibshirani, J. Parker, T. Hastie, J.S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, Repeated observation of breast tumor subtypes in independent gene expression data sets, Proc. Natl. Acad. Sci. 100 (14) (2003) 8418–8423.

[31] G. Yang, X. Lu, L. Yuan, Lncrna: a link between rna and cancer, Biochim. et Biophys. Acta (BBA)-Gene Regul. Mech. 1839 (11) (2014) 1097–1109.

[32] J. Meng, P. Li, Q. Zhang, Z. Yang, S. Fu, A four-long non-coding rna signature in predicting breast cancer survival, J. Exp. Clin. Cancer Res. 33 (1) (2014) 84.

[33] G. Ciriello, M.L. Gatza, A.H. Beck, M.D. Wilkerson, S.K. Rhie, A. Pastore, H. Zhang, M. McLellan, C. Yau, C. Kandoth, Comprehensive molecular portraits of invasive lobular breast cancer, Cell 163 (2) (2015) 506–519.

[34] J. Li, A.E. Lenferink, Y. Deng, C. Collins, Q. Cui, E.O. Purisima, M.D. O'Connor-McCourt, E. Wang, Identification of high-quality cancer prognostic markers and metastasis network modules, Nat. Commun. 1 (34) (2010).

[35] K.K. Sahlberg, G. Bottai, B. Naume, B. Burwinkel, G.A. Calin, A.-L. Børresen-Dale, L. Santarpia, A serum microrna signature predicts tumor relapse and survival in triple-negative breast cancer patients, Clin. Cancer Res. 21 (5) (2015) 1207–1214.

[36] A.E. Zou, J. Ku, T.K. Honda, V. Yu, S.Z. Kuo, H. Zheng, Y. Xuan, M.A. Saad, A. Hinton, K.T. Brumund, Transcriptome sequencing uncovers novel long non-coding and small nucleolar rnas dysregulated in head and neck squamous cell carcinoma, Rna (2015).

[37] Y. Wang, A.J. Bernhardy, C. Cruz, J.J. Krais, J. Nacson, E. Nicolas, S. Peri, H. Van Der Gulden, I. Van Der Heijden, S.W. O'Brien, The brca1-δ11q alternative splice isoform bypasses germline mutations and promotes therapeutic resistance to parp inhibition and cisplatin, Cancer Res. 76 (9) (2016) 2778–2790.

[38] G.G. Malouf, J. Zhang, Y. Yuan, E. Compérat, M. Rouprêt, O. Cussenot, Y. Chen, E.J. Thompson, N.M. Tannir, J.N. Weinstein, Characterization of long non-coding rna transcriptome in clear-cell renal cell carcinoma by next-generation deep sequencing, Mol. Oncol. 9 (1) (2015) 32–43.

[39] Q.-L. Sun, C.-P. Zhao, T.-Y. Wang, X.-B. Hao, X.-Y. Wang, X. Zhang, Y.-C. Li, Expression profile analysis of long non-coding rna associated with vincristine resistance in colon cancer cells by next-generation sequencing, Gene 572 (1) (2015) 79–86.

[40] T. Li, X. Mo, L. Fu, B. Xiao, J. Guo, Molecular mechanisms of long noncoding rnas on gastric cancer, Oncotarget 7 (8) (2016) 8601.

[41] M.L. Morton, X. Bai, C.R. Merry, P.A. Linden, A.M. Khalil, R.S. Leidner, C.L. Thompson, Identification of mrnas and lincrnas associated with lung cancer progression using next-generation rna sequencing from laser micro-dissected archival ffpe tissue specimens, Lung cancer 85 (1) (2014) 31–39.

[42] W. Guo, Q. Wang, Y. Zhan, X. Chen, Q. Yu, J. Zhang, Y. Wang, Xu X-j, L. Zhu, Transcriptome sequencing uncovers a three–long noncoding rna signature in predicting breast cancer survival, Sci. Rep. 6 (27931) (2016).

[43] C.-C. Hu, P. Gan, R.-Y. Zhang, J.-X. Xue, L.-K. Ran, Identification of prostate cancer lncrnas by rna-seq, Asian Pac. J. Cancer Prev. 15 (2014) 9439–9444.

[44] D. Bernard, K.V. Prasanth, V. Tripathi, S. Colasse, T. Nakamura, Z. Xuan, M.Q. Zhang, F. Sedel, L. Jourdren, F. Coulpier, A long nuclear-retained non-coding rna regulates synaptogenesis by modulating gene expression, EMBO J. 29 (18) (2010) 3082–3093.

[45] V. Tripathi, J.D. Ellis, Z. Shen, D.Y. Song, Q. Pan, A.T. Watt, S.M. Freier, C.F. Bennett, A. Sharma, P.A. Bubulya, The nuclear-retained noncoding rna malat1 regulates alternative splicing by modulating sr splicing factor phosphorylation, Mol. Cell 39 (6) (2010) 925–938.

[46] S. Loi, B. Haibe-Kains, C. Desmedt, F. Lallemand, A.M. Tutt, C. Gillet, P. Ellis, A. Harris, J. Bergh, J.A. Foekens, Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade, J. Clin. Oncol. 25 (10) (2007) 1239.

[47] S. Nakagawa, T. Naganuma, G. Shioi, T. Hirose, Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice, J. Cell Biol. 193 (1) (2011) 31–39.

[48] S. Nakagawa, M. Shimada, K. Yanaka, M. Mito, T. Arai, E. Takahashi, Y. Fujita, T. Fujimori, L. Standaert, J.-C. Marine, The lncrna neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice, Development 110544 (2014).

[49] A. Geirsson, R.J. Lynch, I. Paliwal, A.L. Bothwell, G.L. Hammond, Human trophoblast noncoding rna suppresses ciita promoter iii activity in murine b-lymphocytes, Biochem. Biophys. Res. Commun. 301 (3) (2003) 718–724.

[50] A. Geirsson, A.L. Bothwell, G.L. Hammond, Inhibition of alloresponse by a human trophoblast non-coding rna suppressing class ii transactivator promoter iii and major histocompatibility class ii expression in murine b-lymphocytes, J. Heart Lung Transplant. 23 (9) (2004) 1077–1081.

[51] J. Loughlin, J.S. Sinsheimer, Z. Mustafa, A.J. Carr, K. Clipsham, V.A. Bloomfield, J. Chitnavis, A. Bailey, B. Sykes, K. Chapman, Association analysis of the vitamin d receptor gene, the type i collagen gene col1a1, and the estrogen receptor gene in idiopathic osteoarthritis, J. Rheumatol. 27 (3) (2000) 779–784.

[52] F. Niculescu, H. Rus, M. Retegan, R. Vlaicu, Persistent complement activation on tumor cells in breast cancer, Am. J. Pathol. 140 (5) (1992) 1039.

[53] M.M. Markiewski, J.D. Lambris, Is complement good or bad for cancer patients? A new perspective on an old dilemma, Trends Immunol. 30 (6) (2009) 286–292.

[54] P. Zapata-Benavides, M. Tuna, G. Lopez-Berestein, A.M. Tari, Downregulation of wilms' tumor 1 protein inhibits breast cancer proliferation, Biochem. Biophys. Res. Commun. 295 (4) (2002) 784–790.

[55] H. Sugiyama, Wt1 (wilms' tumor gene 1): Biology and cancer immunotherapy, Jpn. J. Clin. Oncol. 40 (5) (2010) 377–387.

[56] R. Stein, D. McDonnell, Estrogen-related receptor α as a therapeutic target in cancer, Endocr. Relat. Cancer 13 (Supplement 1) (2006) S25–S32.

[57] C. Tiraby, B.C. Hazen, M.L. Gantner, A. Kralli, Estrogen-related receptor gamma promotes mesenchymal-to-epithelial transition and suppresses breast tumor growth, Cancer Res. 1315 (2010) 2011.

[58] G.W. Reuther, C.J. Der, The ras branch of small gtpases: ras family members don't fall far from the tree, Curr. Opin. Cell Biol. 12 (2) (2000) 157–165.

[59] B. Boettner, L. Van Aelst, Control of cell adhesion dynamics by rap1 signaling, Curr. Opin. Cell Biol. 21 (5) (2009) 684–693.

[60] A. Arai, Y. Nosaka, H. Kohsaka, N. Miyasaka, O. Miura, Crkl activates integrin-mediated hematopoietic cell adhesion through the guanine nucleotide exchange factor c3g, Blood 93 (11) (1999) 3713–3722.

[61] B.C. Milon, A. Agyapong, R. Bautista, L.C. Costello, R.B. Franklin, Ras responsive element binding protein-1 (rreb-1) down-regulates hzip1 expression in prostate cancer cells, Prostate 70 (3) (2010) 288–296.

[62] J. Imler, C. Schatz, C. Wasylyk, B. Chatton, B. Wasylyk, A harvey-ras responsive transcription element is also responsive to a tumour-promoter and to serum, Nature 332 (6161) (1988) 275.

[63] P.J. Keely, E.V. Rusyn, A.D. Cox, L.V. Parise, R-ras signals through specific integrin α cytoplasmic domains to promote migration and invasion of breast epithelial cells, J. Cell Biol. 145 (5) (1999) 1077–1088.

[64] J. Song, B. Zheng, X. Bu, Y. Fei, S. Shi, Negative association of r-ras activation and breast cancer development, Oncol. Rep. 31 (6) (2014) 2776–2784.

[65] J.D. Humphries, A. Byron, M.J. Humphries, Integrin ligands at a glance, J. Cell. Sci. 119 (19) (2006) 3901–3903.

[66] R.O. Hynes, Integrins: bidirectional, allosteric signaling machines, Cell 110 (6) (2002) 673–687.